

# ОГЛАВЛЕНИЕ

<b>Предисловие</b> .....	9
О содержании книги .....	9
О названии книги .....	12
Благодарности .....	13
<b>Введение</b> .....	15
<i>Цитируемая литература</i> .....	17
<b>Глава 1. Секвенирование ДНК методом химической деградации по Максиму-Гилберту</b> .....	19
1.1. Принцип секвенирования ДНК методом химической деградации по Максиму-Гилберту .....	20
1.2. Концевое мечение фрагментов ДНК .....	22
1.3. Подготовка фрагментов ДНК, пригодных для секвенирования .....	24
1.4. Проведение реакций модификации азотистых оснований и расщепления цепи ДНК по модифицированным основаниям .....	33
1.5. Электрофорез и радиоавтография .....	40
<i>Цитируемая литература</i> .....	41
<i>Дополнительная литература</i> .....	43
<b>Глава 2. Секвенирование ДНК ферментативным методом по Сэнгеру</b> .....	44
2.1. Принцип секвенирования ДНК ферментативным методом по Сэнгеру .....	44
2.2. Матрицы для секвенирования .....	48
2.3. Затравочные молекулы .....	58
2.4. ДНК-полимеразы .....	78
2.5. Мечение новосинтезируемой цепи ДНК .....	98
2.6. Терминирующие реакции .....	109
2.7. Другие подходы к ферментативному определению нуклеотидной последовательности ДНК .....	118
<i>Цитируемая литература</i> .....	121
<i>Дополнительная литература</i> .....	133
<b>Глава 3. ПЦР-секвенирование ДНК</b> .....	134
3.1. Секвенирование клонированных фрагментов ДНК, полученных с помощью ПЦР .....	138
3.2. Прямое секвенирование фрагментов ДНК, полученных с помощью ПЦР .....	139
3.3. Твердофазное секвенирование ДНК .....	149
3.4. Секвенирование фрагментов ДНК с помощью линейной ПЦР или циклическое секвенирование ДНК .....	152
<i>Цитируемая литература</i> .....	161
<i>Дополнительная литература</i> .....	167
<b>Глава 4. Секвенирующий гель-электрофорез</b> .....	168
4.1. Секвенирующий гель .....	168
4.2. Гель-электрофорез .....	181
4.3. Перенос фрагментов ДНК из секвенирующего геля на мембранные фильтры .....	186
<i>Цитируемая литература</i> .....	189
<i>Дополнительная литература</i> .....	191

<b>Глава 5. Радиоавтография</b> .....	192
5.1. Радиоавтография секвенирующего геля .....	192
5.2. Рентгеновские пленки .....	195
5.3. Прочие способы регистрации радиоактивности в секвенирующих гелях .....	196
<i>Цитируемая литература</i> .....	197
 <b>Глава 6. Чтение нуклеотидной последовательности с радиоавтографа геля</b> .....	198
6.1. Ручное чтение .....	199
6.2. Полуавтоматическое и автоматическое чтение .....	204
<i>Цитируемая литература</i> .....	207
<i>Дополнительная литература</i> .....	208
 <b>Глава 7. Векторы для молекулярного клонирования и секвенирования</b> .....	209
7.1. Первое поколение плазмидных векторов .....	209
7.2. Второе поколение плазмидных векторов .....	210
7.3. Векторы на основе нитевидного фага M13 .....	213
7.4. Третье поколение плазмидных векторов .....	216
7.5. Четвертое поколение плазмидных векторов .....	217
7.6. Пятое поколение плазмидных векторов .....	218
7.7. Шестое поколение плазмидных векторов .....	221
7.8. Развитие векторных систем для клонирования и секвенирования ДНК .....	224
7.9. Другие специализированные векторы для секвенирования ДНК .....	227
7.10. Векторы для клонирования ПЦР-фрагментов ДНК .....	229
7.11. О наименованиях векторов .....	230
<i>Цитируемая литература</i> .....	231
<i>Дополнительная литература</i> .....	233
 <b>Глава 8. Стратегии секвенирования протяженных фрагментов ДНК</b> .....	234
8.1. Стратегии секвенирования протяженных фрагментов ДНК методом химической дегградации по Максаму–Гилберту .....	234
8.2. Стратегии секвенирования протяженных фрагментов ДНК ферментативным методом по Сэнгеру с помощью случайного подхода .....	240
8.3. Стратегии секвенирования протяженных фрагментов ДНК ферментативным методом по Сэнгеру с помощью направленного подхода .....	248
8.4. Стратегии секвенирования протяженных фрагментов ДНК ферментативным методом по Сэнгеру праймерной “прогулкой” .....	281
8.5. Стратегии секвенирования протяженных фрагментов ДНК путем получения делеционных субклонов <i>in vivo</i> .....	295
8.6. Мультиплексное секвенирование ДНК .....	298
<i>Цитируемая литература</i> .....	302
 <b>Глава 9. Автоматическое секвенирование ДНК</b> .....	309
9.1. Матрицы для секвенирования ДНК автоматическим методом .....	309
9.2. Проведение терминирующих реакций при автоматическом секвенировании ДНК ферментативным методом .....	311
9.3. Разделение продуктов терминирующих реакций в автоматическом секвенировании ДНК .....	320
9.4. Автоматизация секвенирования ДНК методом химической дегградации по Максаму–Гилберту .....	329
<i>Цитируемая литература</i> .....	331
<i>Дополнительная литература</i> .....	338

<b>Глава 10. Компьютерный анализ и базы данных</b> .....	339
10.1. Пакеты прикладных программ для полного анализа нуклеотидных последовательностей секвенированных фрагментов ДНК .....	340
10.2. Базы данных по нуклеотидным и аминокислотным последовательностям ....	345
<i>Цитируемая литература</i> .....	354
<i>Дополнительная литература</i> .....	356
<b>Глава 11. Секвенирование целых геномов</b> .....	357
11.1. Геномы прокариотических организмов .....	358
11.2. Геномы эукариотических организмов .....	367
11.3. Геном человека .....	370
11.4. Стратегии секвенирования полных геномов .....	372
11.5. Достоверность данных при секвенировании ДНК .....	376
11.6. Стратегии анализа полных геномов .....	378
11.7. Будущее секвенирования полных геномов .....	381
<i>Цитируемая литература</i> .....	382
<i>Дополнительная литература</i> .....	387
<b>Глава 12. Секвенирование ДНК посредством синтеза</b> .....	388
12.1. Пиросеквенирование ДНК .....	389
12.2. Секвенирование ДНК путем детекции одиночных флуоресцентно меченных дНМФ, последовательно реализуемых с 3'-конца фрагмента ДНК под действием экзонуклеазы .....	393
12.3. Секвенирование ДНК путем возобновления синтеза комплементарной цепи ДНК после удаления терминаторной группы у концевой нуклеотида .....	395
12.4. Прочие методы секвенирования ДНК без использования электрофоретического разделения .....	397
<i>Цитируемая литература</i> .....	398
<b>Глава 13. Секвенирование ДНК посредством гибридизации</b> .....	400
13.1. Гибридизация меченых олигонуклеотидов с фиксированной секвенируемой ДНК .....	401
13.2. Гибридизация меченой ДНК с фиксированными олигонуклеотидами – “ДНК-чип” технология .....	402
<i>Цитируемая литература</i> .....	412
<i>Дополнительная литература</i> .....	415
<b>Глава 14. Эволюционное развитие методов секвенирования ДНК</b> .....	416
<b>Приложение</b> .....	422
<b>Отечественные поставщики материалов и оборудования для секвенирования ДНК</b> .....	428

# CONTENTS

<b>Preface</b> .....	9
About the book .....	9
About the name of the book .....	12
Acknowledgements .....	13
<b>Introduction</b> .....	15
<i>References</i> .....	17
<b>Charter 1. DNA sequencing by Maxam-Gilbert chemical degradation method</b> .....	19
1.1. The principle of DNA sequencing by Maxam-Gilbert chemical degradation method ..	20
1.2. End-labeling DNA fragments .....	22
1.3. Preparation of DNA fragments suitable for sequencing .....	24
1.4. Chemical degradation reactions .....	33
1.5. Electrophoresis and autoradiography .....	40
<i>References</i> .....	41
<i>Further reading</i> .....	43
<b>Charter 2. DNA sequencing by Sanger chain termination method</b> .....	44
2.1. The principle of DNA sequencing by Sanger chain termination method .....	44
2.2. Templates for DNA sequencing .....	48
2.3. Primers .....	58
2.4. DNA polymerases .....	78
2.5. Labeling of a newly synthesized DNA strand .....	98
2.6. The termination procedure .....	109
2.7. Other approaches to enzymatic determination of DNA sequences .....	118
<i>References</i> .....	121
<i>Further reading</i> .....	133
<b>Charter 3. PCR sequencing</b> .....	134
3.1. Sequencing of cloned DNA fragments generated by PCR .....	138
3.2. Direct DNA sequencing of PCR products .....	139
3.3. Solid-phase DNA sequencing .....	149
3.4. DNA sequencing via linear PCR or cycle DNA sequencing .....	152
<i>References</i> .....	161
<i>Further reading</i> .....	167
<b>Charter 4. Sequencing gel-electrophoresis</b> .....	168
4.1. Sequencing gels .....	168
4.2. Gel-electrophoresis .....	181
4.3. Transfer of DNA fragments from sequencing gel on membrane filters .....	186
<i>References</i> .....	189
<i>Further reading</i> .....	191
<b>Charter 5. Autoradiography</b> .....	192
5.1. Autoradiography of gels .....	192
5.2. X-ray films .....	195
5.3. Other possibilities of radioactivity registration in sequencing gels .....	196
<i>References</i> .....	197

<b>Charter 6. Reading the sequence of nucleotides from autoradiograph .....</b>	<b>198</b>
6.1. Manual reading .....	199
6.2. Semiautomatic and automatic reading .....	204
<i>References</i> .....	207
<i>Further reading</i> .....	208
<b>Charter 7. Vectors for molecular cloning and DNA sequencing .....</b>	<b>209</b>
7.1. The first generation of plasmid vectors .....	209
7.2. The second generation of plasmid vectors .....	210
7.3. Vectors on the basis of bacteriophage M13 .....	213
7.4. The third generation of plasmid vectors .....	216
7.5. The fourth generation of plasmid vectors .....	217
7.6. The fifth generation of plasmid vectors .....	218
7.7. The sixth generation of plasmid vectors .....	221
7.8. Development of vector systems for cloning and DNA sequencing .....	224
7.9. Other specialized vectors for DNA sequencing .....	227
7.10. Vectors for cloning PCR-products .....	229
7.11. About vector names .....	230
<i>References</i> .....	231
<i>Further reading</i> .....	233
<b>Charter 8. DNA sequencing strategies .....</b>	<b>234</b>
8.1. Strategies for DNA sequencing of long DNA fragments by Maxam-Gilbert chemical degradation method .....	234
8.2. Strategies for DNA sequencing of long DNA fragments by Sanger chain termination method via random approach .....	240
8.3. Strategies for DNA sequencing of long DNA fragments by Sanger chain termination method via directed approach .....	248
8.4. Strategies for DNA sequencing of long DNA fragments by Sanger chain termination method via primer walking .....	281
8.5. Strategies for DNA sequencing of long DNA fragments via generating subclones <i>in vivo</i> .....	295
8.6. Multiplex DNA sequencing .....	298
<i>References</i> .....	302
<b>Charter 9. Automated DNA sequencing .....</b>	<b>309</b>
9.1. Templates for DNA sequencing by automated method .....	309
9.2. Chain termination reactions in automated DNA sequencing by enzymatic method ...	311
9.3. Separation of the products of chain termination reactions in automated DNA sequencing .....	320
9.4. Automation of DNA sequencing by Maxam-Gilbert chemical degradation method ...	329
<i>References</i> .....	331
<i>Further reading</i> .....	338
<b>Charter 10. Computer analysis and databases .....</b>	<b>339</b>
10.1. Packets of applied programs .....	340
10.2. Databases of nucleotide and amino acid sequences .....	345
<i>References</i> .....	354
<i>Further reading</i> .....	356
<b>Charter 11. Sequencing of the whole genomes .....</b>	<b>357</b>
11.1. Genomes of prokaryotic organisms .....	358
11.2. Genomes of eukaryotic organisms .....	367
11.3. The Human genome .....	370
11.4. Strategies for sequencing of the whole genomes .....	372

11.5. Reliability of data in DNA sequencing .....	376
11.6. Strategies for the whole genome analysis .....	378
11.7. Perspectives for sequencing of the whole genomes .....	381
<i>References</i> .....	382
<i>Further reading</i> .....	387
<b>Charter 12. DNA sequencing by synthesis .....</b>	<b>388</b>
12.1. Pyrosequencing of DNA .....	389
12.2. DNA sequencing via detection of a single molecule realized from 3' end of DNA fragment after exonuclease attack .....	393
12.3. DNA sequencing via resumption of synthesis complement strand of DNA after removing terminating reversible 3'-tags .....	395
12.4. Other methods of DNA sequencing by synthesis without use electrophoretic separation of reaction products .....	397
<i>References</i> .....	398
<b>Charter 13. DNA sequencing by hybridization .....</b>	<b>400</b>
13.1. Sequencing fixed DNA by hybridization with labeled oligonucleotides .....	401
13.2. Hybridization of labeled DNA with fixed oligonucleotides – “DNA chip” technology .....	402
<i>References</i> .....	412
<i>Further reading</i> .....	415
<b>Charter 14. Evolution of DNA sequencing methods .....</b>	<b>416</b>
<b>Appendixes .....</b>	<b>422</b>
<b>Suppliers of home-produced materials and equipment for DNA sequencing .....</b>	<b>428</b>

# ПРЕДИСЛОВИЕ

## О СОДЕРЖАНИИ КНИГИ

Написание этой книги вызвано отсутствием соответствующей литературы на русском языке и ограниченной доступностью ряда обзоров и специальных изданий зарубежных авторов для большинства российских читателей, главным образом в связи с высокой стоимостью (поскольку в настоящее время приобретение иностранных книг как библиотечками, так и частными лицами не кажется столь невозможным, как раньше), а также и из-за некоторого языкового барьера. Немаловажной причиной является и тот факт, что роль секвенирования ДНК в различных научных исследованиях будет, несомненно, возрастать и количества определяемых нуклеотидных последовательностей будут год от года увеличиваться и, причем не только за счет существующих грандиозных проектов определения нуклеотидных последовательностей целых геномов прокариотических и эукариотических организмов.

Особенностью данной книги можно считать то, что в ней рассмотрены практически все методы и подходы к секвенированию ДНК, включая как еще не получившие, также так и не получившие широкого развития. При этом наш ответ на вполне резонный вопрос – “зачем рассматривать (даже не очень подробно) малоэффективные подходы к модификации отдельных методов?” – заключается в следующем. Для того чтобы с высокой производительностью использовать существующие методы, а тем более разрабатывать их новые версии и модификации, необходимо знать и относительно неудачные попытки во избежание их повторения, ибо, слегка перефразируя известную поговорку, можно сказать, что кое-что из нового – это хорошо забытое старое. Нельзя также сбрасывать со счетов и то, что некоторые ранее разработанные и неиспользуемые сейчас подходы могут обрести новую жизнь и мощное продолжение в современных условиях, поскольку молекулярная биология столь бурно развивается, что возможности сегодняшнего дня и дня вчерашнего несопоставимы. Так, в качестве простого примера, оказавшего заметное влияние на производительность довольно старых подходов к секвенированию ДНК, можно привести обнаружение рестрикционных эндонуклеаз с новыми специфичностями, такими как *CviI* (RGCY) и *NotI* (GCGGCCGC), или аналогичных, о существовании которых когда-то даже и не предполагалось.

Необходимо признать, что наше желание максимально охватить все то, что называется секвенированием ДНК, привело к описанию в этой книге отдельных методов и подходов, знакомство с которыми было для нас ранее довольно поверхностным. Поэтому кому-то некоторые места этой книги могут показаться неинтересными и слишком простыми (поскольку он все это хорошо знает), тогда как другие читатели, возможно, сочтут отдельные главы чрезмерно сложными. Вероятно,

такое неизбежно, и авторы заранее просят и тех, и других читателей быть снисходительными.

В отличие от большинства издаваемых книг, Оглавление предваряет основной материал, поскольку авторский коллектив считает, что знакомство с книгой начинается все же с ее Оглавления и для этого совсем не обязательно начинать листать ее с конца. Построено достаточно удобно, оно позволит читателю довольно легко найти интересующий его раздел. Следует отметить, что началу каждой главы предшествует специальное вступление, вводящее читателя в рассматриваемую проблему.

Во многих современных книгах, посвященных секвенированию ДНК, методу секвенирования путем химической дегградации по Максиму–Гилберту обычно отводят последние главы. Однако на протяжении продолжительного времени он все же был для большинства лабораторий основным методом и уступил свои позиции лишь после разработки серии удобных одноцепочечных фаговых векторов, а также после применения в качестве ДНК-полимеразы модифицированной T7 ДНК-полимеразы или секвеназы. При этом данный метод не изменился столь существенно по сравнению с ферментативным и его описание требует значительно меньше места. В связи с этим, на наш взгляд, более логично представить его в главе 1 – перед главами, посвященными ферментативному методу секвенирования ДНК, претерпевшему целый ряд всевозможных усовершенствований и улучшений и являющемуся в настоящее время основным.

Хотя сам принцип специфической терминации, положенный в основу первоначального метода секвенирования ДНК с помощью дидезокситерминаторов, остался неизменным, само секвенирование ДНК по Сэнгеру все же стало в значительной степени другим. Причем для описания всех этих последовательных усовершенствований одной главы оказалось недостаточно и, кроме главы 2, целиком посвященной этому методу, еще ряд глав (глава 3. ПЦР-секвенирование ДНК; глава 9. Автоматическое секвенирование ДНК) имеют к ферментативному методу секвенирования ДНК самое непосредственное отношение. Разработки векторов для молекулярного клонирования и стратегий секвенирования (главы 7 и 8 соответственно) также в большей степени ориентированы на ферментативное секвенирование. В связи с разделением нами в главе 7 векторов для молекулярного клонирования на несколько поколений следует отметить, что оно, конечно же, условно хотя бы по той причине, что огромная группа специализированных векторов оказалась не включенной в приведенную схему. Подробное рассмотрение стратегий секвенирования объясняется тем, что именно от их выбора часто зависит общая производительность метода секвенирования ДНК.

Поскольку оба классических метода секвенирования ДНК – путем химической дегградации и ферментативным построением комплементарной цепи – в условиях специфической терминации требуют разделения меченых фрагментов ДНК, различающихся по длине на один нуклеотид, в полиакриламидном гель-электрофорезе высокого разрешения, то этому процессу, а также и радиоавтографии, и “чтению” нуклео-



тидной последовательности с радиоавтографа геля посвящены главы 4–6, которые являются общими для обоих методов. (Что касается содержащейся в главе 1 краткой информации о секвенирующем геле-электрофорезе, то она лишь вводит читателя в круг проблем.)

Ввиду большого объема информации анализ нуклеотидной последовательности ДНК уже давно немислим без компьютера и без банков данных. Современный молекулярный биолог должен хорошо разбираться в возможностях, предоставляемых ему той или иной программой, а также уметь пользоваться для этих целей сетью Интернет, что кратко изложено в главе 10. Глава 11 посвящена грандиозным успехам всей молекулярной биологии в виде расшифровки первичных последовательностей геномов разных свободноживущих организмов. В ней приводятся сведения в этой области по состоянию на конец ноября 1998 г. Несколько отличающиеся от классического метода ферментативного секвенирования подходы описаны в главе 12. Некоторые из них обладают значительным потенциалом. Глава 13 посвящена еще одному способу секвенирования ДНК с помощью гибридизации, где вклад отечественных ученых во главе с академиком А.Д. Мирзабековым особенно значителен. В заключительной, 14-й, главе нами осуществлена попытка отобразить изменения методов секвенирования ДНК, произошедшие с ними за эти годы и описанные в данной книге.

В подавляющем большинстве книг, посвященных секвенированию ДНК и изданных в начале 1980-х годов, имеющиеся там приложения содержат нуклеотидные последовательности некоторых векторных молекул, например плазмиды pBR322 и др. Однако в настоящее время это не так актуально, поскольку все эти последовательности доступны посредством сети Интернет, да и молекулярная биология преодолела уже новые рубежи в виде полных геномов свободноживущих организмов, однако, если отобразить на бумаге последовательность нуклеотидов самого маленького из них, принадлежащего микроорганизму *Mycoplasma genitalum* (580 тыс 70 пн), то это заняло бы около половины данной книги. Поэтому в Приложении мы ограничились лишь рядом сведений, главным образом справочного характера. Даже довольно скудная информация об отечественных поставщиках материалов и оборудования для секвенирования ДНК, приведенная в специальном разделе Приложения, возможно, окажется кому-то полезной.

Большое количество иллюстративного материала (особенно в главе 8) призвано облегчить восприятие изложенного, иногда довольно сложного материала. При этом следует отметить, что подготовленные иллюстрации не являются простой перерисовкой рисунков из оригинальных статей и обзоров, а есть продукт их тщательного анализа, имеющего целью максимально наглядно отобразить последовательность стадий того или иного процесса. Авторы цитированных статей, отчасти по причине экономии журнального пространства, во многих случаях приводили довольно краткие схемы, отображающие лишь наиболее значимые черты какого-либо метода или подхода. Так, в качестве примера кардинального изменения представленного в оригинальной статье

иллюстративного описания можно привести рис. 12.1, на котором довольно детально показан метод пиросеквенирования ДНК, развивающийся по спирали.

В процессе написания данной книги было проанализировано большое количество литературных источников, из которых приблизительно пятая часть по разным причинам не была включена в окончательные списки цитируемой литературы, состоящей из более чем 1000 источников и охватывающей период с апреля 1953 г. по декабрь 1998 г. Многие из так и не процитированных статей являются краткими обзорами текущего состояния дел в секвенировании ДНК на момент их написания и за счет этого безнадежно устаревшими. Такая же участь постигла и ряд статей, посвященных рассмотрению перспектив развития методов секвенирования ДНК. Однако самая большая доля процитированных работ принадлежит статьям, посвященным компьютерному анализу нуклеотидных последовательностей, потерявшим свою актуальность благодаря чрезвычайно быстрому совершенствованию компьютерной техники. Что касается цитируемой в данной книге литературы, то она, за редкими исключениями, представлена методическими статьями, посвященными описанию новых методов или их модификации. К исключениям же относятся статьи, являющиеся настоящими вехами в молекулярной биологии, как, например, статьи Дж. Уотсона и Ф. Крика, опубликованные в журнале "Nature" в 1953 г. или статьи, описывающие секвенирование и анализ полных геномов каких-либо организмов.

Приведенный в конце большинства глав список литературы для дополнительного чтения позволит заинтересованному читателю осуществлять целенаправленный поиск конкретного издания. При его составлении, однако, не ставилась цель перечислить всю имеющуюся в мире литературу по данной теме, поскольку в таком случае списки могли оказаться довольно большими, а их пригодность для российского читателя весьма ограниченной. В этой связи выбирались, во-первых, те издания, с которыми сами авторы знакомы не понаслышке, и, во-вторых, издания, имеющиеся (по крайней мере, большинство) в российских библиотеках.

## О НАЗВАНИИ КНИГИ

Возможно, более правильным и точнее отражающим ее суть, было бы назвать данную книгу "Эволюционное развитие методов секвенирования ДНК", однако мы решили ограничиться названием "Секвенирование ДНК". И тому есть целый ряд причин. Во-первых, в таком случае нам пришлось бы значительно углубиться в 1960-е годы и самое начало 1970-х, что уже было сделано до нас путем перевода книги С. Манделеса "Установление первичной структуры нуклеиновых кислот" (М.: Мир, 1975). Во-вторых, ввиду полного отсутствия подобных изданий на русском языке, включая и переводные (которых нет, кроме уже упомянутой выше книги С. Манделеса, весьма устарев-

шей даже к моменту издания ее перевода, поскольку сам оригинал датируется 1972 г., а за это время в молекулярной биологии произошли весьма важные события), на наш взгляд, было необходимо написать книгу, которая бы максимально восполнила этот досадный пробел. Поэтому наш труд “Секвенирование ДНК” представляет собой монографическое издание с некоторыми элементами лабораторного руководства. Имеющиеся же на русском языке отдельные обзорные статьи в журналах и сборниках ВИНТИ, причем вышедшие далеко не в последние годы, конечно, не в состоянии осветить многочисленные аспекты секвенирования ДНК. Еще одной причиной такого краткого названия послужило то, что для такого “громкого” названия со словом “эволюционное” должны быть веские основания и полная уверенность авторов, что ими поднят весь пласт литературы и никто не пропущен и ничто не забыто. Такой уверенности у нас нет, поскольку то, что называется сегодня СЕКВЕНИРОВАНИЕМ ДНК, стало практически необъятным и наша попытка все это объять, конечно же, далека от совершенства. Однако, так как нам самим стало интересно хотя бы на плоской схеме в виде филогенетического древа (да не одного) взглянуть, как развивались те или иные методы и подходы к секвенированию ДНК, мы все же не удержались и ввели специальную главу с названием, одноименным несостоявшемуся названию всей книги (глава 14). Насколько нам известно – это первый опыт такого рода, поскольку нигде ничего подобного нам не встречалось, и может быть несколько нарушена последовательность событий или что-то достаточно важное вообще было упущено. Поэтому авторы будут признательны всем приславшим свои отзывы и замечания, касающиеся, впрочем, не только этой главы, но и всей книги. Это поможет сделать книгу более точной, содержательной и полезной при ее возможных будущих переизданиях. С этой целью создан специальный электронный почтовый ящик [gen\\_seq@anrb.ru](mailto:gen_seq@anrb.ru), куда все желающие могут направлять свою корреспонденцию (некоторое звуковое сходство данного адреса с не так давно широко употребляемой аббревиатурой является случайным совпадением и, на наш взгляд, не должно напоминать о чем-то особенном, потому что имелось в виду совсем иное). С не меньшей благодарностью будут приняты и обычные письма, направляемые по адресу: Россия, 450054, Уфа, проспект Октября, 69, Отдел биохимии и цитохимии Уфимского научного центра РАН (ОБЦ УНЦ РАН), любому из авторов данной книги.

## БЛАГОДАРНОСТИ

В завершение Предисловия авторы хотели бы выразить свою признательность Российскому фонду фундаментальных исследований, сделавшему возможным издание данной книги, и Академии наук Республики Башкортостан, также внесшей свой вклад в ее издание.

Сама работа над книгой стала возможной отчасти благодаря Программе поддержке ведущих научных школ Российской Федерации, тех-

ническая поддержка которой осуществлялась также через РФФИ (грант № 96-15-98061 – Вахитов В.А.)

Особая благодарность должна быть высказана своим домашним за их долготерпение и стойко переносимые неудобства, связанные с горами литературы (книгами, журналами, оттисками статей, каталогами, рекламными буклетами и пр.).

Отдельную благодарность авторы выражают своему коллеге Ф.Р. Гималову, явившемуся первым читателем данной книги в ее еще нетипографском варианте и сделавшему ряд ценных замечаний.

## ВВЕДЕНИЕ

Важность определения нуклеотидных последовательностей ДНК уже давно не вызывала никаких сомнений. Однако отсутствие производительных методов секвенирования ДНК делало решение этой проблемы малореальным, в связи с чем оно многими учеными откладывалось на следующее столетие или, вернее, тысячелетие. Так, в 1968 г. известный ученый Э. Чаргафф писал, что "... чтение последовательности ДНК может стать задачей XXI века..." [Chargaff, 1968]. Тем сильнее оказался эффект от сообщений о разработанных быстрых и эффективных методах секвенирования ДНК [Maxam, Gilbert, 1977; Sanger et al., 1977].

В 1953 г. в журнале "Nature" появились сообщения об открытии двойной спирали ДНК [Watson, Crick, 1953a] и о важности такой структуры для генетики [Watson, Crick, 1953b], но потребовалась почти четверть столетия, чтобы первичная структура в виде последовательности нуклеотидов этих молекул стала относительно легкодоступной. Главными препятствиями для определения нуклеотидных последовательностей молекул ДНК были их огромный размер и невозможность получения небольших дискретных фрагментов ДНК, пригодных для секвенирования. Значительно меньший размер молекул РНК и наличие определенных классов этих молекул, а также выявление специфических рибонуклеаз, расщепляющих молекулы РНК в строго определенных местах, обусловили успехи в определении нуклеотидных последовательностей именно этих нуклеиновых кислот. Однако стремительное развитие молекулярной биологии привело во второй половине 1970-х годов к долгожданной возможности определения нуклеотидных последовательностей достаточно длинных фрагментов ДНК. Этим событиям предшествовал целый ряд важных открытий. Очень большое значение для всей молекулярной биологии несло обнаружение ферментов рестрикции, позволяющих расщеплять молекулы ДНК в строго определенных местах. С открытием ферментов ДНК-лигаз появилась возможность "сшивать" расщепленные рестрикционными эндонуклеазами фрагменты ДНК с векторными последовательностями бактериальных плазмид и получать таким образом на их основе рекомбинантные клоны. Так, технология рекомбинантных ДНК позволила нарабатывать необходимые небольшие фрагменты ДНК в количествах, достаточных для определения их нуклеотидной последовательности.

Возможность получения благодаря разработанным методам секвенирования ДНК принципиально новых сведений в виде генетических текстов, ранее недоступных, оказала значительное влияние на дальнейшее развитие не только самой молекулярной биологии, но и общей биологии и других смежных областей знаний. Заставила по-новому взгля-

нуть на планирование и проведение экспериментов, на адекватность получаемых результатов поставленным задачам, вызвала к жизни разработку соответствующего компьютерного обеспечения и пр. Словом, вся биологическая наука и связанные с ней смежные дисциплины получили мощный импульс. Поэтому вполне оправданным выглядит решение Нобелевского комитета о присуждении в 1980 г. Нобелевской премии по химии У. Гилберту и Ф. Сэнгеру<sup>1</sup> за разработку методов секвенирования ДНК путем химической дегградации и ферментативного построения соответственно. В том же году эту премию с ними разделил П. Берг, получивший ее за метод создания рекомбинантной ДНК, без которой тогдашнее секвенирование было просто невозможно. К другим Нобелевским премиям, присужденным за открытия, оказавшие непосредственное влияние на развитие молекулярной биологии и приведшие таким образом, к созданию методов секвенирования ДНК, можно отнести, во-первых, премию по физиологии 1962 г., доставшуюся Ф. Крику, Дж. Уотсону, а также М. Уилкинсу за открытие двойной спирали ДНК. И, как можно видеть из списка нобелевских лауреатов, доступного посредством сети Интернет из web-сайта Нобелевского комитета ([www.nobel.se/prize/index.html](http://www.nobel.se/prize/index.html)), есть еще более ранние премии, связанные с разработанными методами секвенирования ДНК. Так, в 1959 г. премия по физиологии досталась С. Очоа и А. Корнбергу за открытие механизма биосинтеза нуклеиновых кислот. Открытие специфических ферментов – рестрикционных эндонуклеаз, без которых не могли быть разработаны данные производительные методы секвенирования ДНК, да и нынешняя молекулярная биология была бы не просто немыслима, было отмечено Нобелевской премией в 1978 г., которая была поделена между В. Арбером, Д. Натансом и Г. Смитом. Премия 1993 г. по химии была присуждена за метод амплификации ДНК с помощью полимеразной цепной реакции (К. Муллис) и сайт-направленный мутагенез (М. Смит). И хотя последний метод не имеет прямого отношения к секвенированию ДНК, его появление способствовало заметному прогрессу в создании новых улучшенных векторов, ферментов (ДНК-полимераз) и через это оказало серьезное влияние на секвенирование ДНК. Полимеразная же цепная реакция сделала секвенирование ДНК более доступным, и этот метод стали осваивать даже немолекулярно-биологические лаборатории, ранее о нем и не помышлявшие.

Упоминание нобелевских лауреатов в связи с секвенированием ДНК здесь не случайно. Оно показывает оценку мировым научным сообществом<sup>2</sup> и самих методов определения нуклеотидных последовательностей ДНК, и тех открытий, которые послужили основой этих методов или оказали заметное влияние на их дальнейшее развитие.

---

<sup>1</sup> Ф. Сэнгер является одним из немногих ученых – дважды лауреатов Нобелевской премии. Первая премия была присуждена ему в 1958 г. за вклад в изучение вторичной структуры белков и, в частности, инсулина.

<sup>2</sup> В то же время необходимо отметить, что некоторые из последних премий Нобелевского комитета вызвали жаркие споры в ученом мире.

Таким образом, успехи в определении нуклеотидной последовательности ДНК были подготовлены как всем развитием молекулярной биологии (обнаружение рестрикционных эндонуклеаз, других ферментов обмена нуклеиновых кислот, плазмидных векторов, разработка методов молекулярного клонирования, методов электрофоретического разделения нуклеиновых кислот), так и накопленным знанием в области химии нуклеозидов, нуклеотидов и нуклеиновых кислот, где вклад отечественных ученых весьма значителен. Так, уже в 1960-е годы и в начале 1970-х в работах С.К. Василенко и соавт. [Василенко и др., 1965], а также Е.Д. Свердлова и соавт. [Свердлов, 1975; Sverdlov et al., 1972; Sverdlov et al., 1973] высказывалась мысль о необходимости определения последовательности нуклеотидов в олигонуклеотидах и ДНК путем их частичного расщепления, начиная с одного из концов. Нельзя не упомянуть о важной роли академика А.Д. Мирзабекова, совместная работа с которым, по признанию самого У. Гилберта, послужила толчком к разработке метода секвенирования ДНК путем химической деградации [Gilbert, 1981]. Впоследствии уже сам А.Д. Мирзабеков был одним из тех, кто в конце 1980-х годов предложил новый оригинальный способ секвенирования ДНК с помощью гибридизации [Лысов и др., 1988; Khrapko et al., 1989].

Следует отметить, что классические методы секвенирования ДНК химической деградацией и ферментативным построением комплементарной цепи в условиях специфической терминации, являющиеся главным предметом рассмотрения в этой книге, кардинально различаясь в своих подходах, требуют разделения меченых фрагментов ДНК, отличающихся по длине на один нуклеотид, в полиакриламидном гель-электрофорезе высокого разрешения. Все составляющие процесса секвенирования ДНК как тем, так и другим методами за годы прошедшие с их разработки, подверглись сильной модификации, и производительность сегодняшнего секвенирования ДНК просто поражает. Однако отражение только нынешнего состояния дел в этой области молекулярной биологии в отрыве от некоей цепочки последовательных улучшений той или иной составляющей этих методов не способно показать тот огромный прогресс, который был достигнут усилиями очень многих ученых.

### *Цитируемая литература*

*Василенко С.К., Демушкин В.П., Будовский Э.И., Кнорре Д.Г.* Определение нуклеотидной последовательности в олигонуклеотидах // Докл. АН СССР. 1965. Т. 162. С. 694–697.

*Лысов Ю.П., Флорентьев В.Л., Хорлин А.А., Храпко К.Р., Шик В.В., Мирзабеков А.Д.* Определение нуклеотидной последовательности ДНК гибридизацией с олигонуклеотидами: Новый метод // Там же. 1988. Т. 303. С. 1508–1511.

*Свердлов Е.Д.* Подходы к исследованию первичной структуры ДНК // Итоги науки и техники. М.: ВИНТИ, Молекулярная биология. 1975. Т. 4. С. 7–88.

*Chargaff E.* What really is DNA? Remarks on the changing aspects of a scientific concept // Progr. Nucl. Acids Res. and Mol. Biol. 1968. Vol. 8. P. 297–333.

*Khrapko K.R., Lysov Yu.P., Khorlin A.A., Shick V.V., Florentiev V.L., Mirzabekov A.D.* An oligonucleotide hybridization approach to DNA sequencing // FEBS Lett. 1989. Vol. 256. P. 118–122.

*Maxam A.M., Gilbert W.* A new method for sequencing DNA // Proc. Nat. Acad. Sci. USA. 1977. Vol. 74. P. 560–564.

*Sanger F., Nicklen S., Coulson A.R.* DNA sequencing with chain-terminating inhibitors / Ibid. 1977. Vol. 74. P. 5463–5467.

*Sverdlov E.D., Monastyrskaya C.S., Budowsky E.I.* Determination of the number of cytidine residues in oligonucleotide // FEBS Lett. 1972. Vol. 28. P. 236–238.

*Sverdlov E.D., Monastyrskaya C.S., Chestukhina A.V., Budowsky E.I.* The primary structure of oligonucleotides: Partial apurinization as a method to determine the positions of purine and pyrimidine residues // Ibid. 1973. Vol. 33. P. 15–17.

*Watson J.D., Crick F.H.C.* A structure for deoxyribose nucleic acid // Nature. 1953a. Vol. 171. P. 737–738.

*Watson J.D., Crick F.H.C.* Genetic implication of the structure of deoxyribonucleic acid // Ibid. 1953b. Vol. 171. P. 964–967.



## Глава 1

# СЕКВЕНИРОВАНИЕ ДНК МЕТОДОМ ХИМИЧЕСКОЙ ДЕГРАДАЦИИ ПО МАКСАМУ-ГИЛБЕРТУ

За 20 с лишним лет, прошедших с момента разработки метода определения нуклеотидной последовательности ДНК с помощью ограниченной химической модификации азотистых оснований, происходящей под действием различных реагентов и дальнейшего расщепления цепи ДНК по местам таких модифицированных оснований, заметно изменились многие составляющие всего этого процесса. Так, для получения фрагментов ДНК, несущих метку на одном из своих концов и пригодных для их последующего секвенирования, стали использовать несколько иные подходы, включая и ПЦР. Расширился выбор меченых молекул, повысилась разрешающая способность электрофоретического разделения, “упростились” радиоавтография, усовершенствовался этап “чтения” нуклеотидной последовательности с радиоавтографа геля. Значительный прогресс достигнут в области компьютеризации секвенирования ДНК. Однако центральный элемент метода секвенирования ДНК по Максому–Гилберту – химическая деградация меченой цепи ДНК – изменился за эти годы не так сильно. Хотя многими исследователями и предлагались различные варианты модификаций тех или иных азотистых оснований, все же в основном арсенале этого метода находятся 4–6 химических реакций, показывающих наиболее стабильные результаты. К числу важных результатов, полученных с помощью данного метода, следует отнести нуклеотидную последовательность плазмидного вектора pBR322 [Sutcliffe, 1978], впоследствии несколько уточненную (см. разд. 7.2).

Несмотря на относительно низкую производительность метода секвенирования ДНК путем химической деградации по Максому–Гилберту в сравнении с ферментативным методом секвенирования ДНК по Сэнгеру, этот метод в настоящее время все же продолжает использоваться и в отдельных случаях почти незаменим. Так, метод химической деградации применяется для секвенирования синтетических олигонуклеотидов в тех случаях, когда это необходимо. Особо “трудные” участки с сильной вторичной структурой не всегда бывает возможно секвенировать с помощью ферментативного построения новой цепи ДНК и тогда используют данный метод. Кроме этого, с помощью секвенирующего гель-электрофореза возможно выявление ДНК/белковых взаимодействий после того, как исследуемая ДНК в комплексе с белком была подвергнута химической модификации по Максому–Гилберту. К некоторому преимуществу метода секвенирования ДНК химической деградацией можно отнести то, что здесь определяется последовательность фрагмента ДНК, или геномного, или клонированного, в каком-либо подходящем векторе (т.е. реплицировавшегося *in vivo*), а не ново-

синтезированная *in vitro* копия, как в ферментативном методе с дидезокси-терминаторами. Еще одно отличие метода секвенирования ДНК по Максаму–Гилберту от метода Сэнгера заключается в том, что его осуществление может начаться практически с любого сайта узнавания какой-нибудь рестрикционной эндонуклеазы, присутствующего во вставке и поэтому не требуется предварительного знания даже небольшого участка нуклеотидной последовательности, окружающего данное место. В этой связи метод секвенирования ДНК путем химической деградации иногда выступает в качестве стартового при выполнении крупномасштабных проектов по определению нуклеотидных последовательностей протяженных фрагментов ДНК. В то же время нельзя не отметить серьезный недостаток метода секвенирования ДНК по Максаму–Гилберту, заключающийся в высокой токсичности большинства используемых реагентов, обращение с которыми и их дальнейшая утилизация требуют соблюдения определенных правил.

### 1.1. ПРИНЦИП СЕКВЕНИРОВАНИЯ ДНК МЕТОДОМ ХИМИЧЕСКОЙ ДЕГРАДАЦИИ ПО МАКСАМУ–ГИЛБЕРТУ

В основе метода секвенирования ДНК путем химической деградации лежит ограниченное расщепление меченого фрагмента ДНК под действием специфических реагентов. Непременным условием проведения секвенирования этим методом является наличие фрагмента ДНК, меченного только по одному концу. Разделение продуктов деградации по размеру с помощью высоковольтного электрофореза в полиакриламидном геле высокого разрешения, способного разделять фрагменты ДНК, различающиеся между собой по длине всего на один нуклеотид в достаточно широком диапазоне, и последующая радиоавтография геля позволяет определить нуклеотидную последовательность секвенированного участка ДНК.

Первым этапом при проведении реакции химической деградации является ограниченная модификация определенных нуклеотидов под действием различных химических агентов. Концентрация агента и продолжительность его воздействия на молекулы ДНК подбирается с таким расчетом, чтобы в каждой молекуле произошла модификация только одного нуклеотида, а поскольку в реакционной смеси присутствует огромное количество таких молекул, то, согласно теории вероятности, все основания данного типа в секвенируемом фрагменте ДНК окажутся модифицированными. Следующие этапы удаления модифицированных оснований и  $\beta$ -элиминации обоих фосфатов, окружающих дезоксирибозу, и разрыва цепи должны проходить уже количественно. Для каждого типа нуклеотидов или их комбинации проводят отдельные реакции ограниченной модификации и количественного расщепления. Таким образом, в результате четырех (или иногда трех, пяти или даже шести) типов реакций образуется смесь олигонуклеотидных молекул, различающихся по размеру на один нуклеотид и несущих на одном из концов метку, обычно радиоактивную. Следует отметить, что кроме

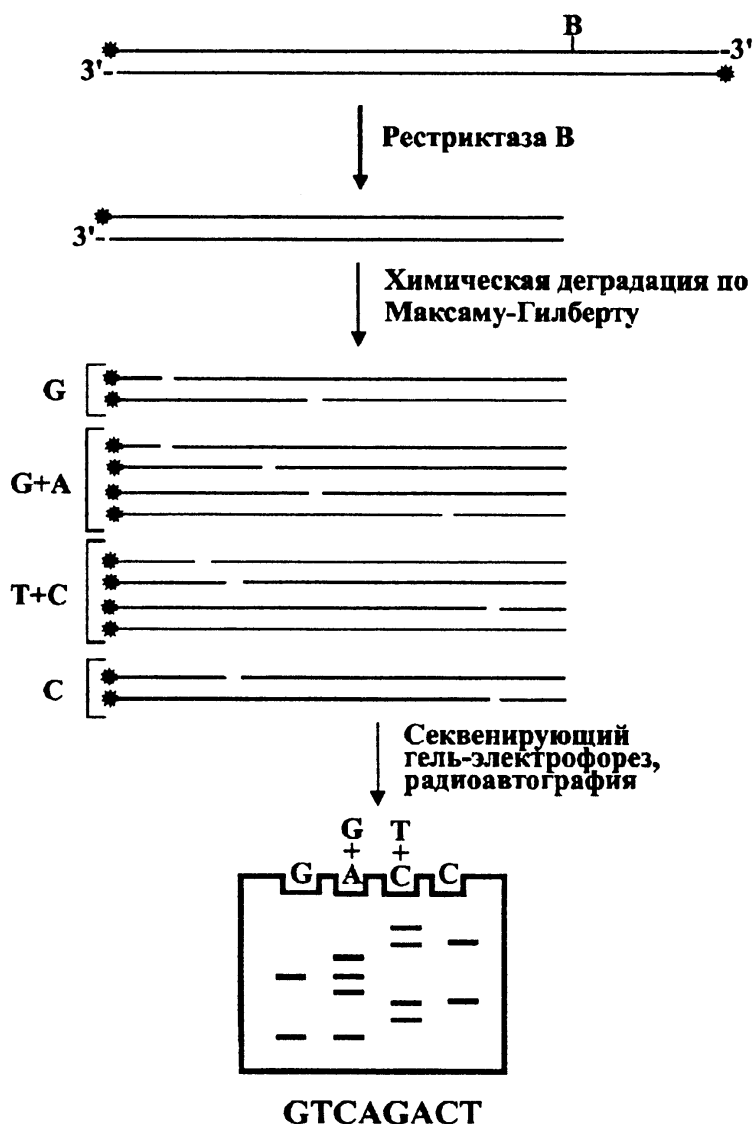


Рис. 1.1. Схема определения нуклеотидной последовательности ДНК методом химической деградации

меченых молекул в реакционной смеси будут представлены, и олигонуклеотидные фрагменты, не несущие метки, но на этапе радиоавтографии они окажутся невидимыми и поэтому для данного метода они как бы не существуют. После разделения продуктов реакции в соседних дорожках секвенирующего денатурирующего полиакриламидного геля и этапа радиоавтографии на рентгеновской пленке будет видна лестница из полос ДНК на соседних дорожках, “чтение” которой позволяет восстановить последовательность нуклеотидов секвенируемого фрагмента ДНК.

На рис. 1.1 в виде упрощенной схемы показан весь процесс определения последовательности нуклеотидов ДНК данным методом, из которого у читателя уже должно сложиться общее представление о секвенировании ДНК методом химической деградации, подробному рассмотрению которого и посвящены следующие разделы этой главы.

## 1.2. КОНЦЕВОЕ МЕЧЕНИЕ ФРАГМЕНТОВ ДНК

Как уже отмечалось выше, для определения нуклеотидных последовательностей ДНК методом Максама–Гилберта необходимо наличие фрагментов ДНК, несущих на одном из концов какую-либо метку. Отдельные частные случаи секвенирования ДНК методом химической дегградации (геномное и мультиплексное секвенирование, рассматриваемые в других главах) могут проводиться даже с немечеными препаратами ДНК. Концевое мечение секвенируемых фрагментов ДНК можно осуществить при помощи различных ферментов. Так, с помощью полинуклеотидкиназы фага Т4 и молекулы АТФ, несущей радиоактивный фосфор в  $\gamma$ -положении, проводится мечение 5'-концов одно- или двуцепочечного фрагмента ДНК. Эффективность мечения тупых или 5'-углубленных концов двуцепочечного фрагмента ДНК значительно ниже, чем 5'-выступающих. Для более эффективного мечения необходим предварительный этап дефосфорилирования фрагмента ДНК с помощью щелочной фосфатазы. Особая забота должна быть проявлена при последующем удалении щелочной фосфатазы, что достигается обычно фенольной депротенинизацией, так как остаточная фосфатазная активность может резко снизить эффективность этапа мечения. Мечение путем обменной реакции позволяет обойтись без этапа дефосфорилирования и требует присутствия в реакционной смеси в качестве акцептора молекул АДФ, однако эффективность мечения при этом будет значительно ниже.

Мечение 3'-концов двуцепочечного фрагмента ДНК можно провести с помощью Кленовского фрагмента ДНК-полимеразы I *E. coli* или ДНК-полимеразы фага Т4 и соответствующего меченного по  $\alpha$ -положению дНТФ. Мечение 3'-концов как одно-, так и двуцепочечного фрагмента ДНК можно осуществить с использованием терминальной трансферазы и меченного по  $\alpha$ -положению подходящего нуклеотидного предшественника. В случае с Кленовским фрагментом ДНК-полимеразы I для эффективного мечения необходимо наличие фрагментов ДНК с выступающими 5'-концами, образованными при расщеплении ДНК определенными рестрикционными эндонуклеазами. Для этого необходимо добавление в реакционную смесь конкретного дНТФ, комплементарного первому нуклеотиду в выступающем участке молекулы (считая от двуцепочечного участка). В тех случаях, когда по какой-либо причине это невозможно, и имеющийся меченый дНТФ комплементарен следующим нуклеотидам, расположенным далее, в реакционную смесь необходимо добавить “холодные” (немеченные) соответствующие дНТФ, комплементарные первым нуклеотидам выступающего участка фрагмента ДНК. При этом следует учитывать, что чем дальше расположен нуклеотид, комплементарный меченому, тем более низкую эффективность мечения можно ожидать. Поскольку Кленовский фрагмент ДНК-полимеразы I имеет довольно низкую 3'  $\rightarrow$  5'-экзонуклеазную активность, то возможное мечение тупых и 3'-выступающих концов (за счет данной экзонуклеазной активности), которые могут находиться на противоположном конце фрагмента ДНК, можно не прини-

мать во внимание ввиду неэффективности этого процесса. Однако добавление в реакционную смесь всех четырех дНТФ (один из которых меченый) позволяет полностью исключить эту вероятность, кроме случаев, когда для обоих концов будут требоваться одинаковые меченые дНТФ. Следует отметить, что добавление в реакционную смесь всех четырех дНТФ, включая меченый, в отдельных случаях может привести к образованию фрагментов ДНК с гетерогенными концами, непригодными для секвенирования, ввиду потенциально возможного включения в различные молекулы разного числа нуклеотидов.

Более высокая эффективность мечения может быть достигнута, если в результате расщепления фрагмента ДНК какой-либо рестрикционной эндонуклеазой образуются 5'-выступающие концы с одинаковыми нуклеотидами. Например, рестрикционная эндонуклеаза *EcoRI* генерирует выступающие концы со следующей последовательностью: 5'-AATT-3'. Таким образом, в результате мечения могут образоваться более высоко меченные фрагменты ДНК с двумя нуклеотидами, несущими метку (при добавлении в реакционную смесь "горячего" дАТФ). Однако при этом существует упоминавшаяся выше опасность неполного заполнения обоими этими нуклеотидами концов во всех молекулах, что приведет к наличию в образце ДНК, предназначенном для секвенирования гетерогенных фрагментов, отличающихся на один нуклеотид. Результатом этого может стать или очень высокий фон, затрудняющий достоверное "чтение" нуклеотидов с радиоавтографа геля, или полная невозможность его прочтения в случаях, когда в секвенируемом образце ДНК один из фрагментов или с одним, или двумя мечеными нуклеотидами будет представлен в достаточно высокой пропорции. Обойти это неудобство можно дополнительной инкубацией с добавлением в реакционную смесь (после завершения основной реакции мечения) этого же "холодного" дНТФ в высокой концентрации с целью гарантированного включения обоих нуклеотидов. Что касается фрагментов ДНК с такими концами, что в них теоретически может встраиваться только один нуклеотид, то и в этом случае ввиду не 100%-й эффективности мечения образуются гетерогенные фрагменты ДНК, однако поскольку один из них остается немеченным, то никакого вклада в заключительный радиоавтограф это не внесет.

Терминальная трансфераза осуществляет нематричный синтез полинуклеотидной цепи ДНК и способна включать произвольное количество нуклеотидов на 3'-конце. Однако в связи с тем, что для секвенирования ДНК требуется наличие фрагментов с гомогенными мечеными концами, необходим этап щелочного гидролиза, обычно осуществляемого пиперидином, с целью удаления достроенных нуклеотидов, кроме первого, и остатка фосфорной кислоты от второго. Таким образом, фрагмент ДНК, меченый подобным образом, будет нести в своем 3'-конце один дополнительный нуклеотид и два меченых остатка фосфорной кислоты. Более простой путь, исключаяющий этап гидролиза, для получения фрагментов ДНК с гомогенными мечеными 3'-концами с помощью терминальной трансферы достигается при использовании в качестве меченого предшественника или дидезоксинуклеотид трифосфа-

та, или кордицепин трифосфата, характеризующихся отсутствием гидроксильной группы в 3'-положении, делающем невозможным присоединение дальнейших нуклеотидов.

Интересный способ мечения 3'-выступающих концов, образуемых при расщеплении ДНК рестрикционной эндонуклеазой *Pst*I с помощью Кленовского фрагмента ДНК-полимеразы I, заключается в предварительном отжиге пятичленного олигонуклеотида 5'-GTGCA-3', комплементарного выступающему участку данного рестрикционного фрагмента ДНК, в результате чего образуется конец нового типа с одним 5'-выступающим нуклеотидом G [Chirala, Wakil, 1986]. Авторы уверяют, что при добавлении в реакционную смесь необходимого радиоактивного дЦТФ эффективность мечения таких вновь полученных концов достигает 80%.

Следует отметить, что, несмотря на довольно большой выбор ферментов и соответствующих меченых нуклеотидов для мечения фрагментов ДНК по 3'- или 5'-концам, наиболее часто используемыми являются варианты с применением полинуклеотидкиназы фаза Т4 для мечения 5'-выступающих концов и Кленовского фрагмента ДНК-полимеразы I для мечения 3'-углубленных концов ввиду более высокой эффективности и относительной простоты данных процессов. Дополнительным удобством является то, что Кленовский фрагмент ДНК-полимеразы I хорошо работает в большинстве рестрикционных буферов, и, таким образом, реакции расщепления и последующего мечения можно проводить в одной пробирке и даже совместно, что впрочем нежелательно из-за все-таки имеющейся 3' → 5'-экзонуклеазной активности у этой ДНК-полимеразы.

При написании этого раздела намеренно не упоминался тип используемой метки, поскольку он может быть разным, но принципиальной разницы на этапе мечения не существует. Обычно при секвенировании фрагментов ДНК методом химической дегградации в качестве радионуклида используется <sup>32</sup>P ввиду его более высокой энергии. Использование имеющих меньшую энергию радионуклидов <sup>33</sup>P и <sup>35</sup>S также возможно, но при этом время экспозиции сильно увеличивается. При секвенировании фрагментов ДНК методом химической дегградации применяются также и хемилюминесцентные, и флуоресцентные метки, но при этом требуется проведение несколько других (более сложных) процедур для выявления полос ДНК, чем простая радиоавтография на рентгеновскую пленку [Voss et al., 1989].

### 1.3. ПОДГОТОВКА ФРАГМЕНТОВ ДНК, ПРИГОДНЫХ ДЛЯ СЕКВЕНИРОВАНИЯ

Метод секвенирования ДНК путем химической дегградации применим как для одно-, так и для двуцепочечных фрагментов, но, как уже отмечалось выше, непременным условием является наличие в секвенируемом фрагменте только одного гомогенного меченого конца. Как правило, в результате мечения двуцепочечного фрагмента ДНК обра-

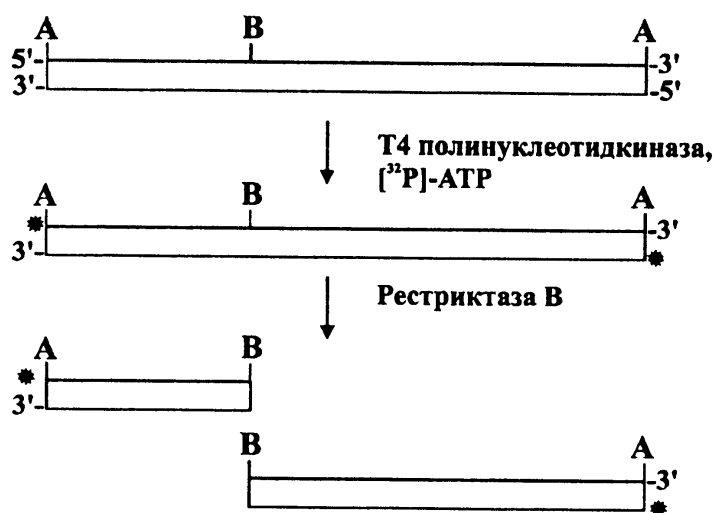


Рис. 1.2. Получение двуцепочечных фрагментов ДНК, меченных только по одному из четырех концов

зуются два меченых 3'- или 5'-конца. Стратегия получения одноцепочечных фрагментов ДНК, меченных по одному из этих концов, предполагающая денатурацию двуцепочечной ДНК, весьма проста в отличие от самого процесса. Так, для достижения этой цели используется метод деления цепей ДНК в денатурирующем полиакриламидном геле. Различия в подвижности разных цепей одного и того же фрагмента ДНК обычно не столь значительны, и получение одной из цепей в чистом виде без примесей другой часто бывает проблематичным. Другими отрицательными моментами являются и трудности детекции в геле нужного фрагмента ДНК, и относительно низкий выход искомого продукта при его элюции из геля. Ввиду этого для получения меченных по одному из концов фрагментов ДНК данный подход с использованием денатурации и последующим делением цепей в настоящее время практически не применяется и упомянут здесь лишь из-за принципиальной возможности его осуществления.

Прежде чем перейти к описанию получения двуцепочечных фрагментов ДНК, несущих единственную метку на одном из концов, все же необходимо упомянуть о других одноцепочечных фрагментах ДНК, синтезируемых химическим путем, – олигонуклеотидах, точность синтеза которых иногда проверяют секвенированием. Следует отметить, что после мечения 5'-конца синтезированного одноцепочечного олигонуклеотида с помощью полинуклеотидкиназы (причем предварительное дефосфорилирование в данном случае не требуется, так как в результате синтеза 5'-конец олигонуклеотида не несет остатка фосфорной кислоты) образуется сразу пригодный для секвенирования фрагмент ДНК с меткой на одном конце.

Получение двуцепочечных фрагментов ДНК, меченных только по одному из четырех концов, возможно различными способами. Фрагменты ДНК, несущие метку или на 3'- или 5'-концах, подвергаются дополнительному расщеплению подходящей рестрикционной эндонуклеазой с таким расчетом, чтобы получались рестриктазные фрагменты,

заметно различающиеся по размеру (рис. 1.2). Далее следует этап препаративного гель-электрофореза в полиакриламидном или агарозном геле и элюция необходимого фрагмента(ов). Данный подход требует или этапа предварительного рестриктазного картирования исходного меченого (лучше немеченого еще) фрагмента, или знания всей так называемой рестриктазной карты клонированного фрагмента ДНК и разработанной на ее основе стратегии секвенирования.

Несмотря на кажущуюся простоту, при получении фрагмента ДНК с меткой только на одном конце этим способом, неприятным моментом является необходимость дополнительных этапов обращения с радиоактивно меченым препаратом, заставляющих проявлять особую осторожность ввиду недопущения радиоактивного загрязнения во время осуществления процедур гель-электрофореза и элюции фрагментов ДНК из геля. Однако этот подход является неизбежным при мечении 5'-концов полинуклеотидкиназой. Если мечение 3'-конца еще позволяет получить секвенируемый фрагмент ДНК с различными концами, один из которых практически не будет метиться в одной реакции, то мечение 5'-конца не допускает такой возможности, поскольку, несмотря на существенные различия в эффективности мечения 5'-выступающих и тупых или углубленных концов фрагмента ДНК, все они будут нести метку, что приведет к невозможности прочтения нуклеотидной последовательности с радиоавтографа секвенирующего геля. Уменьшить неизбежный фон и как-то прочесть нуклеотидную последовательность теоретически можно, проведя вначале этап обязательного дефосфорилирования концов фрагмента ДНК, мечение которых предполагается осуществить, затем – этап повторного расщепления другой рестрикционной эндонуклеазой, генерирующей концы, менее удобные для мечения полинуклеотидкиназой, этап элюции из геля и только потом – мечения. Уже простое перечисление всех этих этапов заставляет отказаться от такого варианта и выбрать способ обращения с фрагментом ДНК, несущим оба радиоактивно меченых 5'-конца, предполагающий расщепление еще одной рестрикционной эндонуклеазой, как описано выше.

Как уже отмечалось, получение фрагментов ДНК, несущих только один меченый 3'-конец, возможно также путем асимметричного (здесь – генерирующего фрагменты разной длины) расщепления соответствующей рестрикционной эндонуклеазой и последующего препаративного электрофореза, как и в случае фрагментов ДНК с мечеными 5'-концами. Однако более привлекательным является подход, основанный на знании рестриктазной карты всего клонированного фрагмента ДНК и разработанной стратегии секвенирования, позволяющей в результате препаративного расщепления разными рестрикционными эндонуклеазами и препаративного электрофореза получить целый набор фрагментов ДНК, несущих разные концы, как 5'-, так и 3'-выступающие, причем в случае обоих 5'-выступающих концов у одного фрагмента ДНК они могут иметь отличающиеся последовательности. Например, фрагменты ДНК, образованные совместным расщеплением двух рестрикционных эндонуклеаз *Hind*III и *Sal*I, будут иметь 5'-AGCT-3' и 5'-TCGA-3' выступающие концы соответственно. Так, для их мечения Кленовским



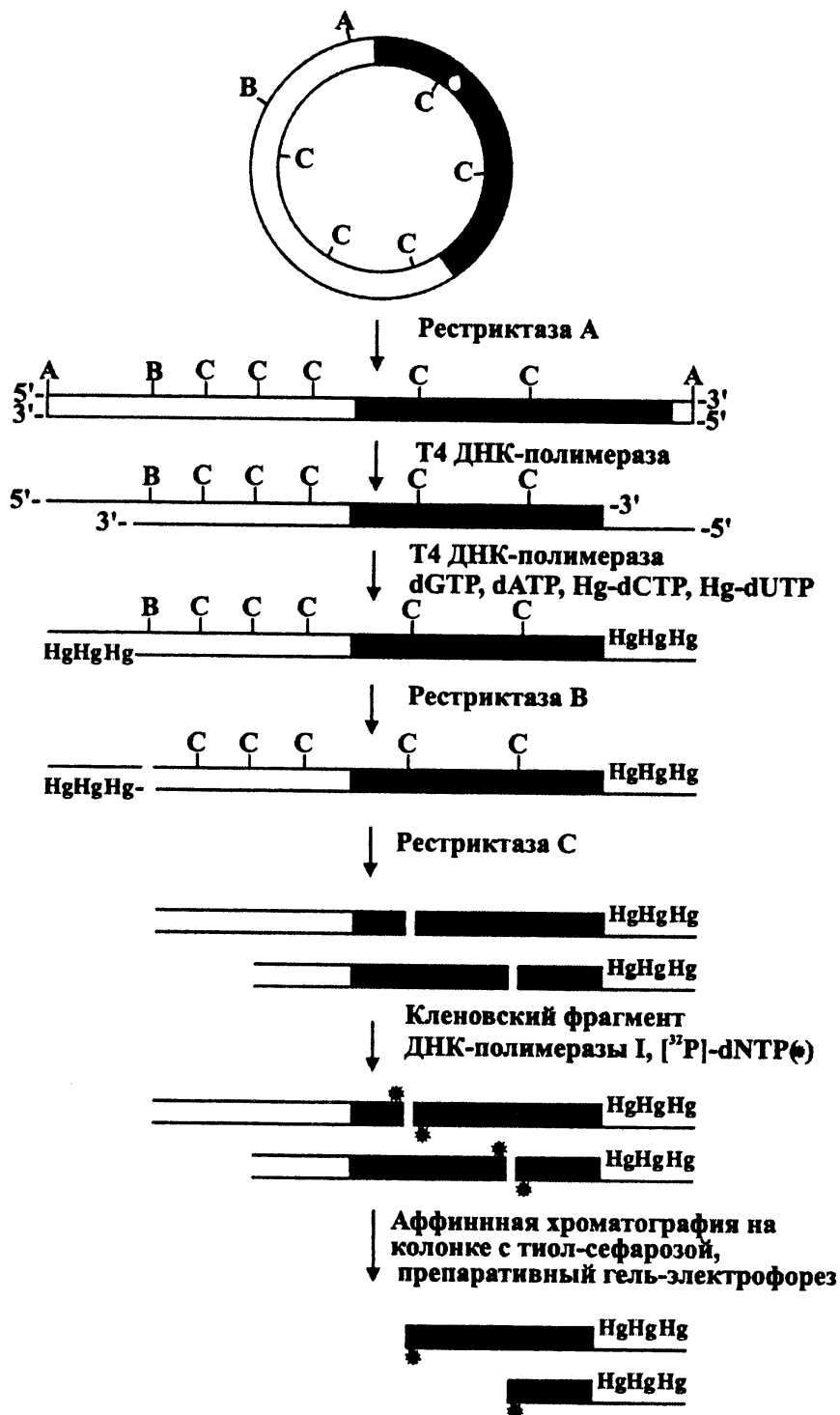


Рис. 1.3. Получение двуцепочечных фрагментов ДНК, меченных по одному из концов с помощью фракционирования ртутьсодержащих цепей на колонках с тиол-сефарозой

фрагментом ДНК-полимеразы I будут необходимы комплементарные первым нуклеотидам в выступающем одноцепочечном участке предшественники дАТФ и ТТФ соответственно. Добавляя в реакционную смесь тот или иной дНТФ, можно избирательно метить требуемые концы и получать меченые фрагменты, непосредственно пригодные для секвенирования разных цепей одного и того же фрагмента ДНК и не

требующие дополнительных этапов рестриктазного расщепления и электрофореза. Вся подготовка таких фрагментов к проведению реакций химической модификации будет заключаться в осаждении ДНК этанолом с целью удаления как невключившейся радиоактивной метки, так и солей, о чем более подробно речь пойдет в следующем разделе.

Был предложен интересный, хотя и весьма трудоемкий, способ получения меченных по одному концу фрагментов ДНК, заключающийся в их разделении аффинной хроматографией на колонках с тиол-сефарозой [Hartley et al., 1982]. Схема получения таких фрагментов, приведенная на рис. 1.3, состоит в следующем. Первым этапом является линеризация вектора со вставкой путем расщепления одной из рестрикционных эндонуклеаз, сайт узнавания которой расположен в последовательности вектора. Вторым этапом, состоящий как бы из двух стадий, необходимо сначала для удаления 100–200 нуклеотидов с обоих 3'-концов противоположных цепей ДНК, осуществляемого при помощи 3' → 5'-экзонуклеазной активности ДНК-полимеразы фаза Т4 в отсутствие в реакционной смеси дНТФ и затем достраивания удаленных нуклеотидов за счет полимеразной активности этого же фермента после добавления в реакционную смесь дНТФ, часть из которых несла в 5'-положении пиримидиновых оснований атом ртути. На третьем этапе проводилось еще одно расщепление вектора подходящим ферментом, в результате чего образовывались два фрагмента, каждый из которых нес на своем 3'-конце ртутную метку. Четвертый этап заключался в частичном расщеплении в условиях недорестрикции уже самой вставки частотщепляющей рестрикционной эндонуклеазой. На пятом этапе осуществлялось мечение 3'-концов, генерируемых этой частотщепляющей рестрикционной эндонуклеазой с помощью Кленовского фрагмента ДНК-полимеразы I и соответствующего меченого (<sup>32</sup>P) дНТФ. После разделения на следующем этапе продуктов всех этих реакций с помощью аффинной хроматографии на колонке с тиол-сефарозой с последней элюировалась фракция ртутьсодержащих фрагментов ДНК, меченных радионуклидом по одному концу. Далее следовал этап электрофоретического разделения полученных меченых фрагментов ДНК уже по размеру для проведения затем с ними реакций химической модификации. Даже простое описание всего этого процесса позволяет оценить всю его трудоемкость, поэтому создание специализированных векторов для секвенирования ДНК позволило исключить трудоемкий этап препаративного электрофореза и получать меченые по одному концу фрагменты ДНК сразу после расщепления двумя рестрикционными эндонуклеазами и мечения Кленовским фрагментом ДНК полимеразы I.

Третье поколение плазмидных векторов общего назначения (см. главу 7) в числе новых черт несло синтетический участок ДНК, называемый полилинкером. Создание полилинкера оказало значительное влияние на развитие стратегии секвенирования ДНК методом химической дегградации в целом [Ruther et al., 1981; Ruther, 1982]. Не менее важна его роль в получении фрагментов ДНК, меченных по одному из 3'-концов. Рис. 1.4 наглядно иллюстрирует возможность получения меченых фрагментов ДНК, пригодных для секвенирования, не прибегая к этапу

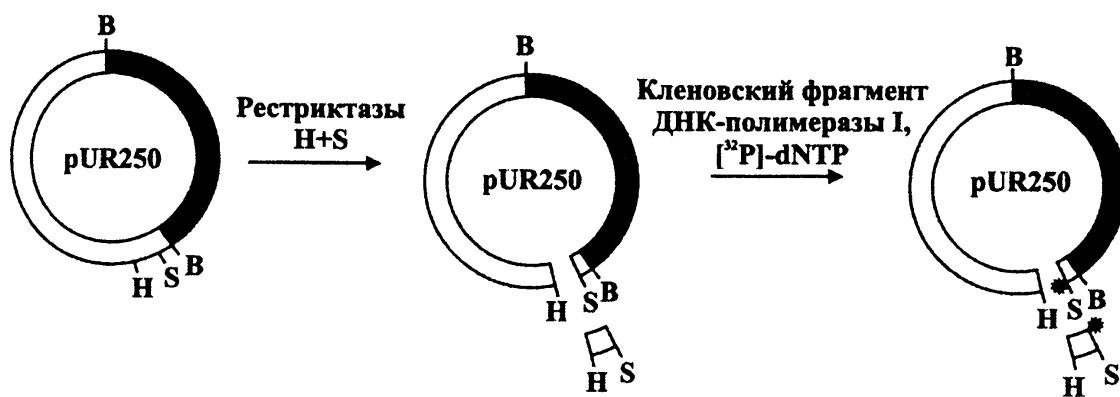


Рис. 1.4. Получение двуцепочечных фрагментов ДНК, меченных по одному из концов, после их клонирования в полилинкере плазмидного вектора с помощью двух рестрикционных эндонуклеаз

препаративного гель-электрофореза, обычно трудоемкого и приводящего к значительным потерям меченого препарата ДНК.

В результате расщепления рекомбинантной плазмиды, несущей вставку, например по сайту узнавания рестрикционной эндонуклеазы *Bam*HI, уже упомянутыми выше рестрикционными эндонуклеазами *Hind*III и *Sal*I, генерирующими выступающие концы 5'-AGCT-3' и 5'-TCGA-3' соответственно, образуется два фрагмента ДНК. Один из этих фрагментов будет содержать всю вставку и почти весь вектор, а второй – только небольшой участок полилинкера, ограниченный сайтами рестрикционных эндонуклеаз *Hind*III и *Sal*I. После этапа мечения с помощью <sup>32</sup>P дТТФ и Кленовского фрагмента ДНК-полимеразы I концов, образованных рестрикционной эндонуклеазой *Sal*I, в реакционной смеси будут присутствовать два меченых фрагмента ДНК, резко различающиеся по своей длине. Ввиду того, что *Hind*III–*Sal*I-участок полилинкера имеет протяженность всего 18 пн, то нет особой надобности в проведении препаративного электрофореза с целью элюции основного фрагмента ДНК, предназначенного для секвенирования. Некоторым недостатком данного подхода является то, что при “чтении” радиоавтографа секвенирующего геля из-за наложения полос двух секвенируемых фрагментов ДНК (вставки с основной частью вектора и участка полилинкера) первые нуклеотиды вставки, соответствующие длине меченого фрагмента полилинкера (в данном случае 18), или не будут читаться совсем, или придется принимать во внимание последовательность этого участка заодно секвенируемого полилинкера. С уверенностью можно сказать лишь то, что, по крайней мере, один нуклеотид вставки, соответствующий по размеру данному участку полилинкера, будет невозможно идентифицировать, поскольку, скорее всего, при проведении реакций химической деградации некоторое количество исходного меченого фрагмента полилинкера все же останется и даст на всех дорожках сигналы обычно более интенсивные, чем принадлежащие остальным расщепленным фрагментам ДНК. Справедливости ради следует отметить, что подобный подход был применен ранее еще с вектором pBR322, не имеющим полилинкера, но несущим сайты узна-

вания рестрикционных эндонуклеаз *Hind*III и *Cla*I, расположенных на расстоянии всего 29 и 23 пн от сайта рестрикционной эндонуклеазы *Eco*RI соответственно [Frischauf et al., 1980].

Однако такое наложение полос двух секвенируемых фрагментов не представляет серьезного неудобства. С развитием методов секвенирования протяженных фрагментов ДНК стало ясно, что в каждом эксперименте читать начальные нуклеотиды часто не обязательно и их выявление обычно происходит при осуществлении перекрытия с другими фрагментами ДНК. Более того, рестрикционные эндонуклеазы для расщепления и последующего мечения концов, образованных одной из них, стараются подбирать так, чтобы выщепляемый фрагмент полилинкера был бы минимально возможным, принимая во внимание возможность рестрикции ДНК отдельными ферментами, поскольку для различных рестрикционных эндонуклеаз существуют свои определенные требования к локализации места узнавания на определенном расстоянии от края фрагмента ДНК для эффективного расщепления. Некоторая опасность в виде неполного расщепления рекомбинантной плазмиды одной или сразу двумя рестрикционными эндонуклеазами все же подстерегает экспериментаторов. Поскольку разница в размерах линеаризованной вставки с полноразмерным вектором и с вектором без небольшого участка полилинкера крайне незначительна, то проверить полноту расщепления ДНК обоими ферментами с помощью аналитического гель-электрофореза очень трудно. В этом случае (недорасщепления второй рестрикционной эндонуклеазой в дистальной части полилинкера) наложение полос может быть на всем протяжении секвенируемого фрагмента ДНК, что сведет на нет все предыдущие усилия. Избежать этого можно, проводя последовательные расщепления двумя ферментами. При этом сначала надо использовать второй фермент (в нашем примере *Hind*III), расположенный дальше от места клонирования вставки, проверить полноту расщепления аналитическим гель-электрофорезом и только затем добавить первый фермент, по сайту которого и будет осуществляться мечение. Заранее зная, что этот фермент работает в таком буфере и данная фасовка содержит работающий фермент, можно не проводить дополнительную стадию гель-электрофореза (тем более, что в ней мало что можно будет увидеть нового, если не прилагать значительных усилий) и сразу осуществить этап мечения Кленовским фрагментом ДНК-полимеразы I. Конечно, можно попытаться увидеть появляющийся небольшой фрагмент ДНК, принадлежащий полилинкеру в высокопроцентном агарозном геле или даже полиакриламидном, но все это весьма трудоемко и, более того, надо обязательно обработать препарат ДНК РНКазой, поскольку в противном случае тРНК и обломки прочих РНК будут маскировать маленький фрагмент ДНК полилинкера. Таким образом, лучше не проводить столь сложных процедур отчасти из-за того, что если вторая рестрикционная эндонуклеаза в этом способе даже не полностью расщепила свой сайт, часть образованных концов будет все же доступна для мечения и этого количества может оказаться достаточно

для проведения секвенирования данного фрагмента ДНК. Единственное, что остается порекомендовать в подобной ситуации – это полагаться на свой собственный опыт.

Дальнейшее совершенствование плазмидных векторов для секвенирования ДНК позволило создать еще более удобные специальные полилинкеры, содержащие в своей последовательности два участка расщепления рестрикционной эндонуклеазы *Hinf*I, имеющей сайт узнавания GACN<sup>v</sup>NNGTC [Eckert, 1987; Volckaert, 1987]. Уникальность этой рестрикционной эндонуклеазы применительно к получению меченых фрагментов ДНК, пригодных для секвенирования, заключается в следующем. В результате расщепления ДНК этим ферментом образуются фрагменты ДНК с одним выступающим нуклеотидом на 5'-конце, пригодные для мечения Кленовским фрагментом ДНК-полимеразы I. Сайт узнавания этого фермента представляет собой прерванный гексануклеотидный палиндром, центральные нуклеотиды которого могут быть любыми. Поскольку в описываемых полилинкерах векторов pSP64CS, pSP65CS и pGV451 центральные нуклеотиды в одном случае были G и T, а в другом A и C, то при мечении Кленовским фрагментом ДНК-полимеразы I при добавлении какого-либо одного дНТФ метится только один 3'-конец из четырех возможных. Таким образом, отпадает необходимость расщепления полилинкера вектора вторым ферментом и, более того, секвенируемый фрагмент, теоретически, может “читаться” с первого нуклеотида.

В литературе описан интересный способ получения меченого по одному концу одноцепочечного фрагмента ДНК, клонированного в каком-либо из фаговых векторов серии M13mp [Bellemare et al., 1986]. Его суть заключается в отжиге специального олигонуклеотида на комплементарном ему участке одноцепочечной векторной молекулы, прилегающему к полилинкеру и захватывающему сайт подходящей полилинкерной рестрикционной эндонуклеазы (в рассматриваемом случае – *Eco*RI) и последующем расщеплении образовавшегося двуцепочечного участка данным ферментом (рис. 1.5). Что касается самой вставки, то даже в случае присутствия в ней последовательности, узнаваемой используемым ферментом, она не будет подвержена расщеплению, поскольку остается одноцепочечной. В результате линейаризации данного одно/двуцепочечного вектора, несущего вставку, образуется два 5'-выступающих конца, пригодных для их мечения полинуклеотидкиназой фага T4 и радиоактивным  $\gamma$ -АТФ после этапа дефосфорилирования. Таким образом, один радиоактивный 5'-конец будет принадлежать образовавшейся после расщепления части использованного олигонуклеотида, а другой – сайту использованной рестрикционной эндонуклеазы полилинкера, “продолжением” которого является клонированная вставка. Из-за большой разницы в размерах этих меченых фрагментов ДНК нет необходимости в проведении дополнительных этапов их разделения, что позволяет непосредственно после этапа мечения, предварительно удалив непрореагировавшую метку этанолом, проводить реакции химической модификации и пиперидинового гидролиза.

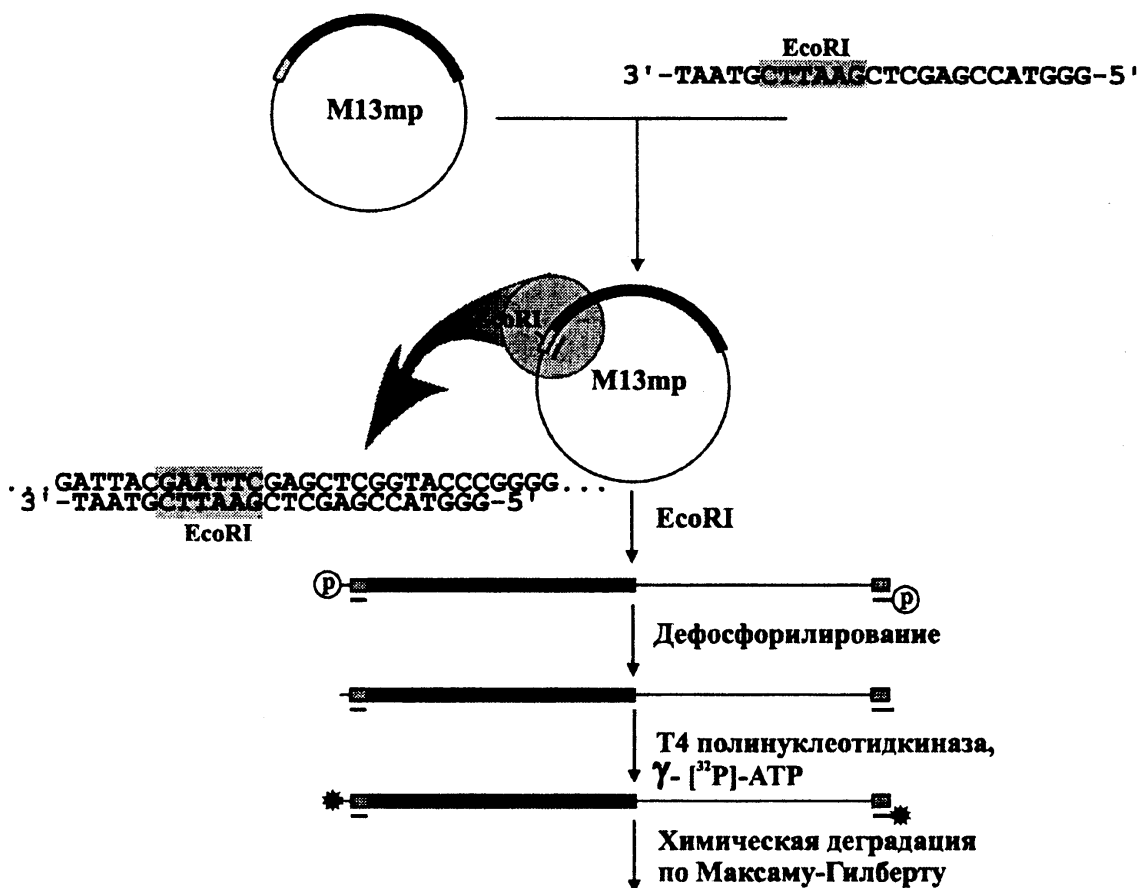


Рис. 1.5. Получение меченого одноцепочечного фрагмента ДНК с помощью вектора на основе фага M13 и специального олигонуклеотида

Другой способ получения меченого по одному концу фрагмента ДНК также на основе одноцепочечного фага M13 описан другими авторами [Rosenthal et al., 1990]. В данной работе осуществлялся отжиг меченого праймера на одноцепочечной матрице ДНК с таким расчетом, что построение комплементарной цепи с помощью высокопроцессивной ДНК-полимеразы приводило, в первую очередь, к копированию вставки, секвенирование которой и предполагалось.

С разработкой метода полимеразной цепной реакции (ПЦР) появилась еще одна возможность получения меченого по одному концу фрагмента ДНК. Так, если предварительно один из праймеров помечить с помощью полинуклеотидкиназы и меченого  $\gamma$ -АТФ, то при амплификации нарабатываемый фрагмент ДНК будет нести метку на одном из 5'-концов [Ohara et al., 1989]. Однако при наличии нескольких различающихся продуктов ПЦР необходим этап повторной амплификации с выделенным из геля фрагментом ДНК, предназначенным для секвенирования. Естественно, в этом случае первая амплификация проводится с немечеными праймерами и только во время повторной – один из праймеров должен быть меченым. В то же время можно проводить реакции химической дегградации с мечеными фрагментами ДНК, выделенными из геля, не прибегая к повторной амплификации [Tahara et al., 1990; Gal, 1996].

Существует и другая возможность получения с помощью ПЦР фрагмента ДНК, пригодного для мечения только по одному концу и ранее успешно реализованная [Nakamae et al., 1988]. В этом случае амплификация проводится с праймерами, один из которых предварительно фосфорилирован. После завершения амплификации у конечного продукта ДНК на одном из 5'-концов, принадлежащему изначально нефосфорилированному праймеру, будет отсутствовать фосфатная группа и, таким образом, он будет пригодным для мечения радиоактивным  $\gamma$ -АТФ с помощью полинуклеотидкиназы фага T4.

#### 1.4. ПРОВЕДЕНИЕ РЕАКЦИЙ МОДИФИКАЦИИ АЗОТИСТЫХ ОСНОВАНИЙ И РАСЩЕПЛЕНИЯ ЦЕПИ ДНК ПО МОДИФИЦИРОВАННЫМ ОСНОВАНИЯМ

Не желая перегружать данный раздел детальной информацией о протекании реакций химической модификации нуклеотидов и промежуточных продуктах этих реакций (заинтересованный читатель может почерпнуть эти сведения из оригинальных статей или обзоров – Максам, Гилберт, 1986; Maxam, Gilbert, 1977 и др.), ограничимся лишь указанием специфичности тех или иных реагентов к отдельным основаниям или их типам. Так, модификация части азотистых оснований подготовленного меченого фрагмента ДНК является первым этапом при проведении реакций химической дегградации. В оригинальной работе Максама и Гилберта [Maxam, Gilbert, 1977] для модификации пуринов была проведена предварительная стадия ограниченного метилирования остатков гуанина с помощью диметилсульфата. Затем инкубация этой частично метилированной ДНК в фосфатном буфере при рН 7,0 и последующий разрыв сахарофосфатной цепи с помощью гидроокиси натрия привели к расщеплению, имеющим характер  $G > A$ . Аналогичная процедура отщепления оснований, но с использованием вместо фосфатного буфера соляной кислоты, и дальнейший разрыв цепи дали реакцию  $A > G$ . Была также показана принципиальная возможность осуществления расщепления ДНК только по гуанинам, заключающаяся в дальнейшей обработке частично метилированной с помощью диметилсульфата ДНК пиперидином. Эта реакция была названа альтернативной и не вошла тогда в основной арсенал методов секвенирования ДНК путем химической дегградации. Однако впоследствии эта реакция модификации гуанинов стала обязательной ввиду ее относительной простоты и воспроизводимости. Еще одна альтернативная реакция  $A > C$ , где модификация оснований проводилась с помощью гидроокиси натрия с последующим разрывом сахарофосфатной цепи пиперидиновым гидролизом, оказалась менее удобной и в дальнейшем использовалась не так широко. Реакцию модификации пиримидинов  $C + T$  проводили с помощью гидразина и последующего гидролиза пиперидином. С целью модификации только цитозиновых остатков использовали раствор гидразина вместе с NaCl.

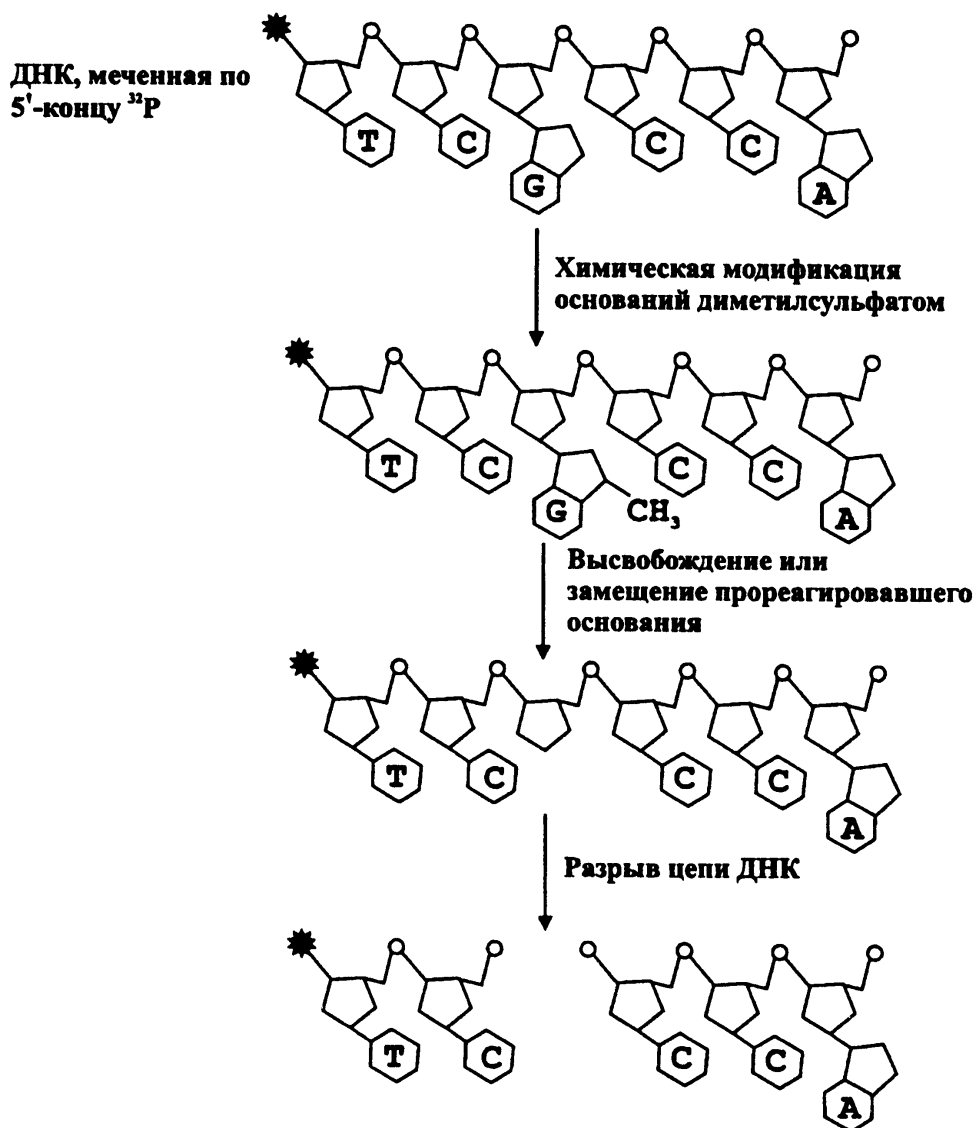


Рис. 1.6. Схема проведения секвенирующей реакции по гуаниновым остаткам

В дальнейшем этими авторами была опубликована большая статья с подробным изложением не только этапов модификации и гидролиза азотистых оснований, но и с многочисленными протоколами проведения различных реакций подготовительного характера [Maxam, Gilbert, 1980; русский перевод данной статьи Максам, Гилберт, 1986]. Приведенные в этой статье прописи реакций модификации G, G + A, C + T и C оказались весьма удобными как при их проведении, так и при чтении радиоавтографа геля. Однако, пожалуй, единственная реакция модификации G с помощью диметилсульфата в последующие годы практически не подверглась сколь-нибудь серьезным изменениям и использовалась многими исследователями при секвенировании ДНК (рис. 1.6).

Реакция G + A существует в двух вариантах, общим для которых является кислотный гидролиз. В первом варианте, предложенном Максамом и Гилбертом [Maxam, Gilbert, 1980], для осуществления кислотного гидролиза использовался формиат пиперидиния, представляющий собой 4%-ную муравьиную кислоту, доведенную пиперидином до pH 2,0.



Главным недостатком этого подхода (кроме значительной продолжительности протекания реакции) является необходимость трудоемкого и длительного этапа лиофилизации для удаления пиперидина. Кислотный гидролиз с помощью 66%-ной муравьиной кислоты, предложенный Скрыбиным и соавт. [Скрыбин и др., 1978], обеспечивает сравнимую специфичность расщепления ДНК по пуриновым основаниям (G + A) и не требует этапа лиофилизации.

Что касается пиримидинов, то как можно видеть из табл. 1.1., были предложены различные реагенты для их модификации. Некоторые из этих реакций не нашли широкого применения ввиду или нестрогой специфичности, или высокой сложности. К наиболее широко применяемым можно отнести классические реакции с использованием гидразина [Maxam, Gilbert, 1980], а также перекиси водорода [Свердлов, Калинина, 1983; Richterich et al., 1995], солянокислого гидроксилamina, перманганата калия или гидразин ацетата [Rubin, Schmid, 1980; McCarthy, 1989]. Однако эффективность модификации одно- или двуцепочечной ДНК некоторыми из этих реагентов весьма различается, что необходимо учитывать при выборе условий проведения реакций.

Ввиду скоротечности большинства реакций модификации азотистых оснований особая забота проявлялась по их одномоментной остановке с тем, чтобы произошла модификация лишь ограниченной части оснований. Для этой цели в классических реакциях (G – диметилсульфат; C + T – гидразин; C – гидразин + 1M NaCl) использовались специальные стоп-буферы, не позволяющие реакциям пройти слишком далеко и произвести модификацию излишнего количества оснований. Следующим этапом при проведении упомянутых реакций являлась стадия осаждения препаратов ДНК этанолом. Полнота осаждения достигалась выдерживанием при температуре  $-70^{\circ}\text{C}$  или даже при температуре жидкого азота. Затем следовало переосаждение ДНК и многократные промывки этанолом для удаления примеси солей. Весьма трудоемкой была процедура удаления пиперидина после этапа  $\beta$ -элиминации с помощью лиофилизации. Все это отнимало много времени, поэтому были предложены способы одновременной остановки реакций и осаждения препаратов ДНК бутанолом [Bencini et al., 1984] или ацетоном [Барам, Грачев; 1985; Данилюк и др., 1986]. В первой работе, кроме осаждения бутанолом, применялось также трехдорожечное секвенирование ДНК в виде реакций A > C (гидроокись натрия,  $90^{\circ}\text{C}$ ), T + C (гидразин), G + A (муравьиная кислота). Сопоставление полос на трех соседних дорожках все же позволяло “читать” последовательность ДНК. К сомнительному преимуществу такого подхода можно отнести экономию числа дорожек. В дальнейшем способ осаждения бутанолом был с успехом применен и для остальных классических реакций G – диметилсульфат; C – гидразин + 1M NaCl [Chirala, Wakil, 1986]. Однако при использовании бутанола имел место эффект не столько осаждения ДНК, а скорее полного удаления воды и в связи с этим с осадком ДНК могли оставаться следовые количества солей, которые необходимо было тщательно удалить перед этапом гидролиза пиперидином во избежание неспецифического расщепления.

Таблица 1.1

## Условия проведения химических реакций в методе секвенирования ДНК по Максаму-Гилберту

Реакция	Реагент для модификации оснований / Температура / Продолжительность инкубации	Реагент для замещения или удаления оснований / Температура / Продолжительность инкубации	Реагент для β-элиминации / Температура / Продолжительность инкубации	Литература
G > A	0,5%-ный диметилсульфат / Комн. / 15 мин	0,01 М натрий-фосфатный буфер, рН7,0 / 90°C / 15 мин	0,1 М NaOH / 90°C / 30 мин	Maxam, Gilbert, 1977
A > G	0,5%-ный диметилсульфат / Комн. / 15 мин	0,1 М HCl / 0°C	0,1 М NaOH / 90°C / 30 мин	Maxam, Gilbert, 1977
G	0,5%-ный диметилсульфат / Комн. / 15 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Maxam, Gilbert, 1977
A > C	1,2 М NaOH / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Maxam, Gilbert, 1977
C + T	18 М гидразин / Комн. / 15 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Maxam, Gilbert, 1977
C	18 М гидразин + 2 М NaCl / Комн. / 15 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Maxam, Gilbert, 1977
A + G	66%-ная муравьиная кислота / Комн. / 5 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Скрябин и др., 1978
T	$7 \cdot 10^{-3}$ М OsO <sub>4</sub> (рН10) / Комн.	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Friedmann, Brown, 1978

С	2M H <sub>2</sub> O <sub>2</sub> (pH10) / 37°C / 60 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Свердлов, Калинина, 1983
Т	1 М NaBH <sub>4</sub> (pH9,3) / 37°C / 10-60 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Свердлов, Калинина, 1984
Т	2 М H <sub>2</sub> O <sub>2</sub> (pH7,4) / 37°C / 60 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Свердлов, Калинина, 1983
С	0,25 М H <sub>2</sub> O <sub>2</sub> / Комн. / 20 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Richterich et al., 1995
С	4 М NH <sub>2</sub> OH·HCl / Комн. / 10 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Rubin, Schmid, 1980
С	18М N <sub>2</sub> H <sub>2</sub> ·CH <sub>3</sub> COOH / Комн. / 10 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Rubin, Schmid, 1980
Т	УФ-свет (254 нм) / 1 М спермин / Комн. / 10 мин	Не требуется	Не требуется	Saito et al., 1984
G > T	УФ-свет (254 нм) / 1 М метиламин / Комн. / 10 мин	"	"	Saito et al., 1984
Т	0,2 mM KMnO <sub>4</sub> / Комн. / 10 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	Rubin, Schmid, 1980
Т	25 μM KMnO <sub>4</sub> / 94°C / 2 мин	1 М пиперидин / 90°C / 30 мин	1 М пиперидин / 90°C / 30 мин	McCarthy, 1989
A > C = = G > T	0,5 М пиперидин + 0,3 М NaCl / 95°C	0,5 М пиперидин + 0,3 М NaCl / 95°C	0,5 М пиперидин + 0,3 М NaCl / 95°C	Ambrose, Pless, 1985

Применение для остановки течения реакций и осаждения препаратов ДНК ацетона с использованием в качестве соли имеющего высокую растворимость в органических растворителях перхлората лития в значительной степени решало эту проблему. Более того, ацетон количественно и практически мгновенно осаждал не только длинные фрагменты ДНК, но и короткие олигонуклеотиды. Этот вариант быстрого секвенирования, разработанный Данилюком и соавт. [Данилюк и др., 1986], позволял в короткий срок и с минимальными усилиями получать препараты ДНК, пригодные для нанесения на секвенирующий гель. Замена нерастворимого в ацетоне NaCl на LiCl и LiOH, обладающих лучшей растворимостью, в реакции модификации цитозиновых остатков позволила повысить специфичность расщепления и уменьшила фон, возникающий после пиперидинового гидролиза ввиду обычно неполного удаления NaCl.

Разработанные способы твердофазного секвенирования ДНК методом химической деградации также позволяли избежать большинства длительных и трудоемких процедур и значительно сократить время на приготовление препаратов ДНК, пригодных к нанесению на секвенирующий гель. Было показано, что специфичность реакций химической модификации и последующего гидролиза пиперидином меченых препаратов ДНК в растворе и иммобилизованных на DEAE-бумаге одинакова [Чувпило, Кравченко, 1983]. Относительными недостатками этого подхода были непрочность самого носителя – DEAE-бумаги и чувствительные потери препаратов ДНК в реакции модификации цитозина из-за присутствия солей на стадии пиперидинового гидролиза и при элюции образцов. Новый носитель для твердофазного секвенирования ДНК, CCS-бумага с анионообменными свойствами, обладающая высокой механической прочностью, позволила проводить одновременные реакции ограниченной модификации азотистых оснований на одной полоске бумаги различных образцов ДНК и, таким образом, даже автоматизировать процесс секвенирования [Rosenthal et al., 1985]. Важным свойством этого носителя является также относительно незначительная потеря образцов ДНК во время проведения реакций ограниченной модификации, а также осуществляемые в одну стадию пиперидиновый гидролиз и элюция ДНК. Данная CCS-бумага под торговым названием M & G Hybond в течение ряда лет поставляется фирмой “Amersham International plc” и пользуется заслуженной популярностью.

Были предложены и другие подходы, заменяющие этапы центрифугирования при проведении реакций химической модификации оснований в методе Максама–Гилберта [Jagadeeswaran, Kaul, 1986; Boland et al., 1994]. Так, в первой из этих работ использовалась обращенно-фазовая хроматография с носителем C<sub>18</sub>, что обеспечило быстрое удаление химических реагентов после завершения первого этапа модификации оснований. Далее элюированные с помощью ацетонитрила модифицированные препараты ДНК высушивали и подвергали стандартному гидролизу пиперидином, однако его удаление представляло уже определенную проблему из-за плохого связывания ДНК с носителем C<sub>18</sub> в присутствии пиперидина (даже разбавленного в 10 раз после завершения

реакции). Добавление в реакционную смесь 1М уксусной кислоты перед этапом хроматографии позволило разрешить эту проблему, и после элюции очищенные от следов пиперидина препараты ДНК разделялись секвенирующим гель-электрофорезом. Таким образом, при осуществлении данного подхода вместо этапов центрифугирования и лиофилизации использовалась колоночная хроматография. В другой своей работе эта же группа авторов [Boland et al., 1994] заметно усовершенствовала данный подход, заменив, в первую очередь, носитель  $C_{18}$ , часто необратимо сорбирующий обработанные пиперидином препараты ДНК, на стеклянный порошок, лучше подходящий для проведения секвенирования методом химической дегградации. Более того, оказалось возможным проводить все химические реакции этого метода секвенирования с препаратами ДНК, сорбированными на стеклянном порошке с завершающей элюцией уже готовой для электрофоретического разделения ДНК с помощью высокой концентрации NaI. Это позволило авторам с помощью лабораторного робота Biomek 1000 (Beckman, США) автоматизировать трудоемкий процесс проведения химических реакций при секвенировании ДНК по методу Максама–Гилберта.

Относительно недавно был предложен новый вариант твердофазного секвенирования ДНК с использованием магнитных частиц, покрытых стрептавидином [Ohara R., Ohara O., 1995]. Фрагменты ДНК, меченные биотином по 5'-концу с помощью ПЦР, сорбировались на данные магнитные частицы и подвергались реакциям химической модификации и после пиперидинового гидролиза наносились на секвенирующий гель. С помощью магнитных частиц достигалось быстрое и полное удаление реагентов, что не могло не привести к получению высококачественных результатов.

Интересный вариант метода секвенирования ДНК путем химической дегградации с помощью только одного пиперидина и 0,5М NaCl известен как однопорожечное секвенирование [Ambrose, Pless, 1985, 1987; Ambrose et al., 1988]. Несмотря на некоторые сложности при "чтении" радиоавтографа геля, заставляющие учитывать не только интенсивность полос, соответствующих определенным азотистым основаниям (А:С:G:Т – 8:4:4:1), но и вертикальное расстояние между ними, где порядок уже следующий: G > Т > А > С. Однако такой гидролиз не выявлял разницы между цитозиновыми и гуаниновыми остатками и их приходилось различать по подвижности, принимая во внимание, что полосам цитозина предшествовал больший промежуток от вышестоящей полосы по сравнению с гуаниновыми полосами. В дальнейшем была исследована возможность использования для гидролиза молекул ДНК различных аминов, дающих разницу между G- и С-остатками, и такие были найдены, но при этом, к сожалению, не выявлялась разница между Т и А [Testoff, Pless, 1991]. К преимуществам данного способа относится также и то, что отпадает необходимость обращения с такими вредными реагентами, как диметилсульфат и гидразин, хотя не менее опасный пиперидин все же остается. Справедливости ради надо заметить, что данный способ секвенирования ДНК, несмотря на некоторые незначительные преимущества, так и не стал массовым.

Следует отметить, что из-за того, что пиперидин является предшественником в синтезе фенилциклидина и его продажа и использование в США строго регулируется, были попытки заменить его более безопасным пирролидином, имеющим сходные характеристики и требующим всего 15 мин инкубации (вместо 30 мин для пиперидина) при 90°C для расщепления модифицированных цепей ДНК [Shi, Tyler, 1989], однако широкого распространения это не получило.

Сообщается о секвенировании ДНК путем химической деградации, завершаемого в одну стадию с помощью одного N-метилформамида в присутствии ионов марганца сразу всех четырех типов оснований [Negri et al., 1996]. Авторы продемонстрировали возможность однократного разделения продуктов секвенирующих реакций, причем как радиоактивно, так и флуоресцентно меченных по 5'-концу молекул ДНК. Эта же группа авторов исследовала механизм деградации формамидом 2'-деоксицитидина и пришла к выводу, что в случае замены последнего на 5-бромо-2'-деоксицитидин, производимой во время ферментативной амплификации фрагмента ДНК, заметно улучшается однократное секвенирование ДНК путем химической деградации за счет лучшего распознавания остатков цитидина и тимидина, являющегося, обычно, главным источником ошибок [Saladino et al., 1997]. Весьма просто проблема одинаковой высоты пиков, принадлежащих остаткам гуанина и цитозина, в однократном флуоресцентном секвенировании ДНК с помощью гидролиза пиперидином в присутствии NaCl была решена путем предварительной обработки ДНК диметилсульфатом, приводящим к метилированию гуанинов и формированию иного характера пиков, принадлежащих разным основаниям – G > A > C > T [Rosenthal et al., 1990].

Завершая данный раздел, следует отметить, что предложенное проведение реакций химической модификации и гидролиза в 96 лунках микротитратора [Church, Kieffer-Higgins, 1988] представляло собой некий элемент автоматизации и позволило значительно повысить производительность данного метода по сравнению с применявшимися ранее для этой цели полипропиленовыми пробирками [Maxam, Gilbert, 1980; Максам, Гилберт, 1986] и даже (для пиперидинового гидролиза) запаянными капиллярами [Maxam, Gilbert, 1977].

## 1.5. ЭЛЕКТРОФОРЕЗ И РАДИОАВТОГРАФИЯ

В первоначальных экспериментах меченные по одному концу продукты реакций химической деградации разделяли электрофорезом в 20%-ном полиакриламидном геле в денатурирующих условиях с 7М мочевиной в трис-боратном буфере. Для успешного разделения смеси олигонуклеотидных фрагментов, отличающихся на один нуклеотид, применяли плоские пластины гелей толщиной 1,5 мм и размером 30 на 40 см, что обеспечивало необходимое разрешение и соответствовало имеющимся рентгеновским пленкам с размерами 35,5 на 43 см (или 30 на 40 см для отечественных). Для эффективного разделения градиент напряжения составлял около 30 V на см длины геля. После завершения электрофореза стекла разъединяли и гель с одним стеклом, используемым

для поддержки, быстро замораживали, заворачивали в кухонную пленку Saran Wrap и экспонировали на рентгеновскую пленку, обеспечивая ее тесный контакт с гелем при помощи груза при температуре  $-20^{\circ}\text{C}$  и, таким образом, не допуская размораживания геля и диффузии полос ДНК (Maxam, Gilbert, 1977). После проявления экспонированной рентгеновской пленки на ней была видна лестница полос. Начиная “читать” данный радиоавтограф снизу восстанавливали последовательность секвенированного фрагмента ДНК.

Секвенирующий электрофорез, радиоавтография и “чтение” нуклеотидных последовательностей изложены гораздо подробнее в соответствующих главах, специально посвященных этим вопросам.

### Цитируемая литература

Барам Г.И., Грачев С.А. Использование перхлората лития при выделении и анализе олиго- и полинуклеотидов // *Биоорганическая химия*. 1985. Т.11. С. 1420–1422.

Данилюк Н.К., Ястребов С.И., Артамонова Т.П., Попов С.Г. Упрощенный вариант метода Максама–Гилберта для определения первичной структуры олигонуклеотидов и фрагментов ДНК // Там же. 1986. Т. 12. С. 1185–1188.

Максам А.М., Гилберт У. Метод определения последовательности ДНК (секвенирования) после введения метки в конец молекулы и расщепления ее по основаниям // *Молекулярная биология*. 1986. Т. 20. С. 581–637.

Свердлов Е.Д., Калинина Н.Ф. Взаимодействие ДНК с перекисью водорода. Метод локализации пиримидиновых оснований в ДНК // *Биоорганическая химия*. 1983. Т. 9. С. 1696–1698.

Свердлов Е.Д., Калинина Н.Ф. Расщепление одноцепочечных ДНК по остаткам Т, двухцепочечных – по последовательностям ТА, ТГ, ТС при действии  $\text{NaBH}_4$ . Увеличенная реакционная способность центрального Т в боксе Прибнова ряда промоторов // Там же. 1984. Т. 10. С. 418–421.

Скрябин К.Г., Захарьев В.М., Баев А.А. Структура внешнего транскрибируемого спейсера рибосомной РНК дрожжей // *Докл. АН СССР*. 1978. Т. 241. С. 488–490.

Чувпило С.А., Кравченко В.В. Твердофазный метод определения нуклеотидной последовательности ДНК // *Биоорганическая химия*. 1983. Т. 9. С. 1634–1637.

Ambrose B.J., Castro M.M., Pless R.C. One-lane sequence analysis of oligodeoxyribonucleotides // *Anal. Biochem.* 1986. Vol. 159. P. 24–28.

Ambrose B.J.B., Pless R.C. Analysis of DNA sequences using chemical cleavage procedure // *Biochemistry*. 1985. Vol. 24. P. 6194–6200.

Ambrose B.J., Pless R.C. DNA sequencing: Chemical methods // *Meth. Enzymol.* 1987. Vol. 152: Guide to molecular cloning techniques. A.R. P. 522–538.

Ambrose B.J., Pless R.C., Ayers M.E. Sequence analysis of end-labeled DNA fragments by solvolysis in hot aqueous piperidine solutions // *Anal. Biochem.* 1988. Vol. 169. P. 151–158.

Bellemare G., Potvin C., Simard C. Chemical DNA sequencing in the opposite direction to the dideoxy chain termination method, using a single preparation of M13 single-stranded DNA // *Biochem. Cell. Biol.* 1986. Vol. 64. P. 70–72.

Bencini D.A., O'Donovan G.A., Wild J.R. Rapid chemical degradation sequencing // *Biotechniques*. 1984. Vol. 2. P. 4–5.

Boland E.J., Pillai A., Odom M.W., Jagadeeswaran P. Automation of the Maxam–Gilbert chemical sequencing reactions // *Ibid.* 1994. Vol. 16. P. 1088–1095.

Chirala S.S., Wakil S.J. Radio-labelling of the *Pst*I restriction fragments and improvement in the sequencing procedure // *Gene*. 1986. Vol. 47. P. 297–300.

Church G.M., Kieffer-Higgins S. Multiplex DNA sequencing // *Science*. 1988. Vol. 240. P. 47–50.

Eckert R.L. New vectors for rapid sequencing of DNA fragments by chemical degradation // *Gene*. 1987. Vol. 51. P. 247–254.

Friedmann T., Brown D.M. Base-specific reactions useful for DNA sequencing: methylene

- blue – sensitized photooxidation of guanine and osmium tetroxide modification of thymine // *Nucl. Acids Res.* 1978. Vol. 5. P. 615–622.
- Frischauf A.M., Garoff H., Lehrach H.* A subcloning strategy for DNA sequence analysis // *Ibid.* 1980. Vol. 8. 5541–5549.
- Gal S.* Sequencing of double-stranded PCR products // *Mol. Biotechnol.* 1996. Vol. 5. P. 159–164.
- Hartley J.L., Chen K.K., Donelson J.E.* A mercury-thiol affinity system for rapid generation of overlapping labeled DNA fragments for DNA sequencing // *Nucl. Acids Res.* 1982. Vol. 10. P. 4009–4025.
- Jagadeeswaran P., Kaul R.K.* Use of reverse-phase chromatography in the Maxam–Gilbert method of DNA sequencing: A step toward automation // *Gene Anal. Techn.* 1986. Vol. 3. P. 79–85.
- Maxam A.M., Gilbert W.* A new method for sequencing DNA // *Proc. Nat. Acad. Sci. USA.* 1977. Vol. 74. P. 560–564.
- Maxam A.M., Gilbert W.* Sequencing end-labeled DNA with base-specific chemical cleavage // *Meth. Enzymol.* 1980. Vol. 65. Nucleic acids, pt. 1. P. 499–560.
- McCarthy J.G.* An improvement in thymine specific chemical DNA sequencing // *Nucl. Acids Res.* 1989. Vol. 17. P. 7541.
- Nakamaye K.L., Gish G., Eckstein F., Vosberg H.-P.* Direct sequencing of polymerase chain reaction amplified DNA fragments through the incorporation of deoxynucleoside  $\alpha$ -thiotriphosphates // *Ibid.* 1988. Vol. 16. P. 9947–9959.
- Negri R., Costanzo G., Saladino R., Di Mauro E.* One-step, one-lane chemical DNA sequencing by N-methylformamide in the presence of metal ions // *Biotechniques.* 1996. Vol. 21. P. 910–917.
- Ohara O., Dorit R.L., Gilbert W.* One-sided polymerase chain reaction: The amplification of cDNA // *Proc. Nat. Acad. Sci. USA.* 1989. Vol. 86. P. 5673–5677.
- Ohara R., Ohara O.* A new solid-phase chemical DNA sequencing method which uses streptavidin-coated magnetic beads // *DNA Res.* 1995. Vol. 2. P. 123–128.
- Richterich P., Lakey N.D., Lee H.-M., Mao J., Smith D., Church C.M.* Cytosine specific DNA sequencing with hydrogen peroxide // *Nucl. Acids Res.* 1995. Vol. 23. P. 4922–4923.
- Rosenthal A., Schewrtner S., Hahn V., Hunger H.-D.* Solid-phase methods for sequencing of nucleic acids. 1. Simultaneous sequencing of different oligodeoxyribonucleotides using a new, mechanically stable anion-exchange paper // *Ibid.* 1985. Vol. 13. P. 1173–1184.
- Rosenthal A., Sproat B., Voss H., Stegemann J., Schwager C., Erfle H., Zimmermann J., Coutelle C., Ansoerge W.* Automated sequencing of fluorescently labelled DNA by chemical degradation // *DNA Sequenc.* 1990. Vol. 1. P. 63–71.
- Rubin C.M., Schmid C.W.* Pyrimidine-specific chemical reactions useful for DNA sequencing // *Nucl. Acids Res.* 1980. Vol. 8. P. 4613–4619.
- Ruther U.* pUR250 allows rapid chemical sequencing of both DNA stands of its inserts // *Ibid.* 1982. Vol. 10. P. 5765–5772.
- Ruther U., Koenen M., Otto K., Muller-Hill B.* pUR222, a vector for cloning and rapid chemical sequencing of DNA // *Ibid.* 1981. Vol. 9. P. 4087–4098.
- Saito I., Sugiyama H., Matsuura T., Ueda K., Komano T.* A new procedure for determining thymine residues in DNA sequencing. Photoinduced cleavage of DNA fragments in the presence of spermine // *Ibid.* 1984. Vol. 12. P. 2879–2885.
- Saladino R., Crestini C., Mincione E., Costanzo G., Di Mauro E., Negri R.* Mechanism of degradation of 2'-deoxycytidine by formamide: Implication for chemical DNA sequencing procedures // *Bioorg. Med. Chem.* 1997. Vol. 5. P. 2041–2048.
- Shi Y., Tyler B.M.* Pyrrolidine, a non-controlled substance, can replace piperidine for the chemical sequencing of DNA // *Nucl. Acids Res.* 1989. Vol. 17. P. 3317.
- Sutcliffe J.G.* Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322 // *Cold Spring Harbor Symp. Quant. Biol.* 1978. Vol. 43. P. 77–90.
- Tahara T., Kraus J.P., Rosenberg L.E.* Direct DNA sequencing of PCR amplified genomic DNA by the Maxam–Gilbert method // *Biotechniques.* 1990. Vol. 8. P. 366–368.
- Testoff M.A., Pless R.C.* Sequence analysis of end-labelled DNA fragments by solvolysis in aqueous solutions of different amines // *Anal. Biochem.* 1991. Vol. 197. P. 316–320.
- Volckaert G.* A systematic approach to chemical DNA sequencing by subcloning in pGV451 and derived vectors // *Meth. Enzymol.* 1987. Vol. 155. Recombinant DNA, pt F. P. 231–250.



Voss H., Schwager C., Wirkner U., Sproat B., Zimmermann J., Rosenthal A., Erfle H., Stegemann J., Ansorge W. Direct genomic fluorescent on-line sequencing and analysis using *in vitro* amplification of DNA // Nucl. Acids Res. 1989. Vol. 17. P. 2517–2527.

### Дополнительная литература

Кочетков Н.К., Будовский Э.И., Свердлов Е.Д. и др. Органическая химия нуклеиновых кислот. М.: Химия, 1970. 720 с.

Максам А.М., Гилберт У. Метод определения последовательности ДНК (секвенирования) после введения метки в конец молекулы и расщепления ее по основаниям // Молекуляр. биология. 1986. Т. 20. С. 581–637.

Методы молекулярной генетики и геной инженерии / А.В. Мазин, А.В. Кузнецов, А.С. Краев и др. Новосибирск: Наука, 1990. 248 с.

Alphey L. DNA sequencing from experimental methods to bioinformatics. Berlin etc.: BIOS sci. publ., 1997. 224 p.

Ambrose B.J.B., Pless R.C. DNA sequencing: Chemical methods // Meth. Enzymol. 1987. Vol. 152: Guide to molecular cloning techniques. P. 522–538.

Ansorge W., Voss H., Zimmermann J. (ed.). DNA Sequencing strategies. N.Y. etc. Wiley; Spektrum, 1996. 202 p.

Griffin H.G., Griffin A.M. (ed.). Methods in molecular biology. Vol. 23. DNA sequencing protocols. Totowa: Humana press, 1993. 392 p.

Maxam A.M., Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavage // Meth. Enzymol. 1980. Vol. 65. Nucleic acids, pt. I. P. 499–560.

Perbal B. A practical guide to molecular cloning. N.Y. Wiley, 1984. 554 p.

Sambrook J., Fritsch E.F., Maniatis T. Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor: CSHL press, 1989. 1626 p.

## Глава 2

# СЕКВЕНИРОВАНИЕ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ ПО СЭНГЕРУ

За время, прошедшее со дня разработки метода ферментативного секвенирования ДНК по Сэнгеру с помощью дидезокситерминаторов, он претерпел многочисленные модификации и всевозможные улучшения. В той или иной степени изменения подверглись практически все составляющие этого процесса. Так, другими стали подходы к получению матриц ДНК для их последующего секвенирования, практически не используется основной фермент первых лет – Кленовский фрагмент ДНК-полимеразы *E. coli*, изменились взгляды на затравочные молекулы, появились модифицированные дНТФ и ддНТФ, расширился выбор меченых молекул, включая и нерадиоактивные, заметно усовершенствовались этапы электрофоретического разделения, радиоавтографии и “чтения” нуклеотидной последовательности, несравненно выросло компьютерное обеспечение. Не говоря уже о появлении ПЦР и об автоматизации основных этапов определения нуклеотидной последовательности ДНК ферментативным методом, рассматриваемых, впрочем, в других главах.

Однако описание только нынешнего состояния секвенирования ДНК ферментативным методом без привлечения сведений, показывающих его последовательное становление, не способно, на наш взгляд, продемонстрировать тот огромный прогресс, который был достигнут усилиями очень многих ученых.

## 2.1. ПРИНЦИП СЕКВЕНИРОВАНИЯ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ ПО СЭНГЕРУ

Прежде чем перейти к описанию метода секвенирования ДНК, известному ныне как метод Сэнгера, видимо, стоит уделить некоторое внимание более ранним методам секвенирования ДНК, также разработанным Сэнгером и его коллегами. Обычному ныне ферментативному методу секвенирования ДНК с помощью дидезокситерминаторов предшествовал так называемый плюс/минус метод, показанный на рис. 2.1, [Sanger, Coulson, 1975]. В его основе также лежало построение второй радиоактивно меченой (за счет присутствия в реакционной смеси одного меченого наряду с тремя немечеными дНТФ) цепи ДНК по существующей матрице с помощью ДНК-полимеразы I (Кленовского фрагмента). В качестве затравки использовался короткий синтетический олигонуклеотидный десятизвенный праймер. На втором этапе гетерогенные (причем их гетерогенность была крайне желательна, поскольку в противном случае сколь-нибудь значимого результата в секвенирова-

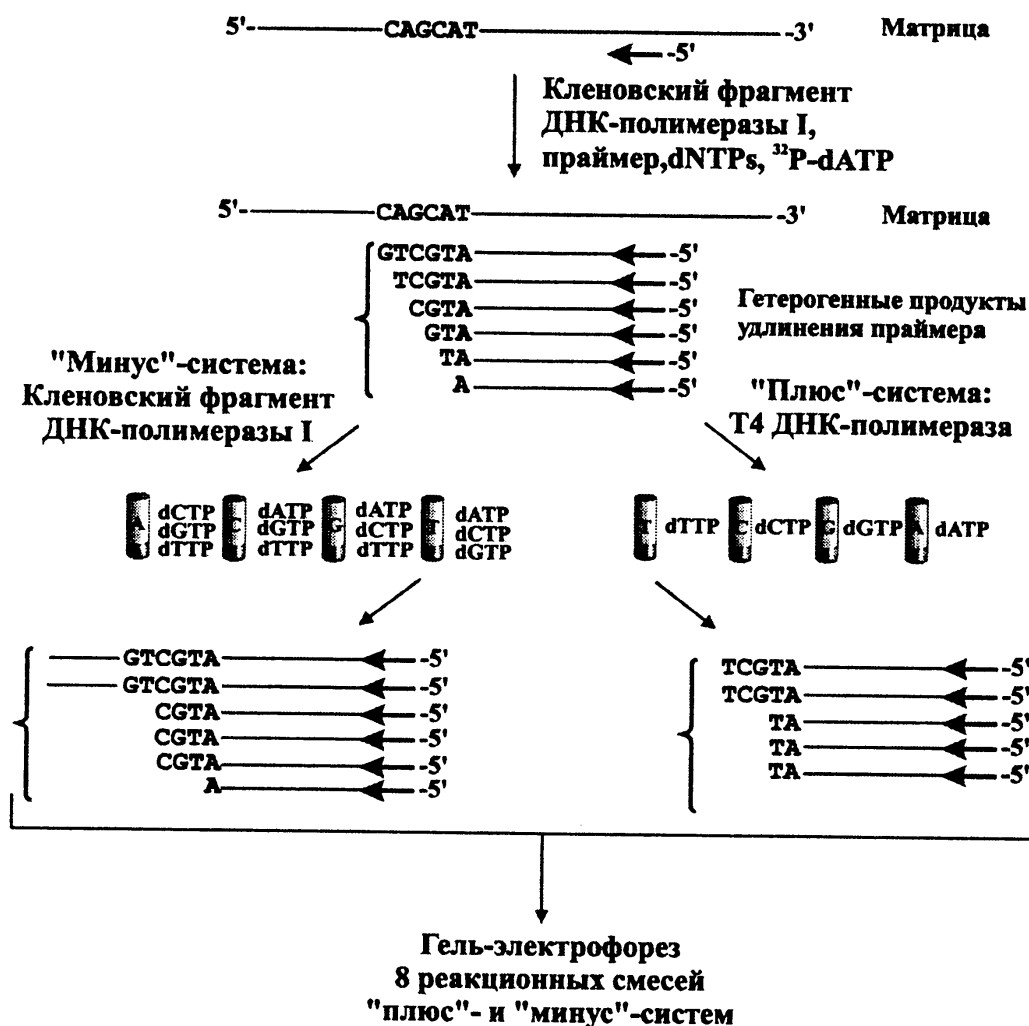


Рис. 2.1. Схема секвенирования ДНК "плюс-минус"-методом

нии достичь бы не удалось) продукты удлинения данного праймера отделяли хроматографией на колонке от непрореагировавших ингредиентов и делились на восемь частей. Четыре части были использованы в минус-системе, где в каждой реакции присутствовало по три разных дНТФ, а четвертый отсутствовал. Дальнейшее построение цепи новой порцией Кленовского фрагмента ДНК-полимеразы I приводило к удлинению цепи ДНК в каждой реакции до тех пор, пока в матрице не оказывался нуклеотид, комплементарный отсутствующему. Вторые четыре части с добавлением только одного дНТФ (в каждую – разный) выдерживались в плюс-системе, где в качестве фермента использовалась 3' → 5'-экзонуклеазная активность ДНК-полимеразы фага Т4. Так, последовательное отщепление дНМФ с 3'-конца ДНК шло до тех пор, пока не встречался дНМФ, дНТФ которого присутствовал в реакционной смеси и тогда в действие вступала уже полимеразная активность Т4 ДНК-полимеразы и фрагмент ДНК приобретал свой окончательный размер. По завершению всех этих реакций радиоактивно меченные продукты разделялись в полиакриламидном гель-электрофорезе и экспонировались на рентгеновскую пленку, по которой можно было восстановить нуклеотидную последовательность секвенируемого участка

ДНК. Следует отметить, что, несмотря на низкую производительность этого метода, с его помощью была определена почти полная (из 5386 пн) нуклеотидная последовательность фага φX174 [Sanger et al., 1977b].

В еще более ранней работе этих же авторов [Sanger et al., 1973] им удалось с помощью ДНК-полимеразы в присутствии в реакционной смеси ионов  $Mn^{2+}$  определить последовательность 50 нуклеотидов фага φ1 путем удлинения 8-мерного синтетического праймера и внедрения в растущую цепь ДНК рибонуклеотидов, по которым потом под действием щелочи происходило расщепление цепи. Этот подход лег потом в основу метода секвенирования ДНК частичным рибозамещением [Barnes, 1978]. Эти работы позволили вплотную подойти к разработке метода секвенирования с терминаторами путем синтеза комплементарной цепи ДНК, который и будет рассматриваться в дальнейшем.

В основе метода секвенирования ДНК, разработанного Сэнгером и соавт. [Sanger et al., 1977a], называемого также методом секвенирования путем терминации цепи, лежал принцип ферментативного построения комплементарной цепи ДНК по существующей одноцепочечной матрице при происходящем в разных местах цепи ДНК ингибировании ее дальнейшего роста.

Составляющими этого процесса являлись: одноцепочечная матрица ДНК; короткая затравочная молекула, комплементарная определенному участку матрицы; ДНК-полимераза (Кленовский фрагмент ДНК-полимеразы I *E. coli*); 2'-дезоксинуклеотид 5'-трифосфаты (дАТФ, дГТФ, дТТФ, дЦТФ или просто дНТФ); 2',3'-дидезоксинуклеотид 5'-трифосфаты (ддАТФ, ддГТФ, ддТТФ, ддЦТФ или просто ддНТФ); реакционный буфер с ионами  $Mg^{2+}$ .

Ключевым моментом этого процесса являлась терминация построения комплементарной цепи ДНК, происходящая при включении ДНК-полимеразой модифицированных аналогов природных субстратов дидезоксинуклеотид трифосфатов, являющихся известными ингибиторами ДНК-полимеразы, как было ранее показано в случае с ддТТФ [Atkinson et al., 1969]. Отсутствие второго гидроксильного остатка в 3'-положении рибозного кольца у дидезоксипроизводных приводила к невозможности присоединения к ним следующего нуклеотида, и рост цепи таким образом прерывался. Такое терминирование происходило одновременно как бы в двух режимах – заданном и случайном. Заданность определялась тем, что в одной отдельной пробирке наряду со всеми четырьмя дНТФ (один из которых был радиоактивно меченным) присутствовал только один какой-нибудь конкретный ддНТФ, случайность же происходила от того, что включение данного ддНМФ в растущую цепь ДНК было произвольным, но, естественно, с учетом принципа комплементарности. То есть происходила некая конкуренция между дНТФ и ддНТФ при их выборе ДНК полимеразой. Поскольку в четырех реакционных пробирках (по типу оснований – А, С, G и Т) присутствовало огромное количество молекул секвенируемой ДНК, во много раз превосходящее ту длину фрагмента, которую могла построить ДНК-полимераза, то по теории вероятности каждое положение этого типа оснований во фрагменте ДНК оказывалось представленным соответствующим ддНМФ.

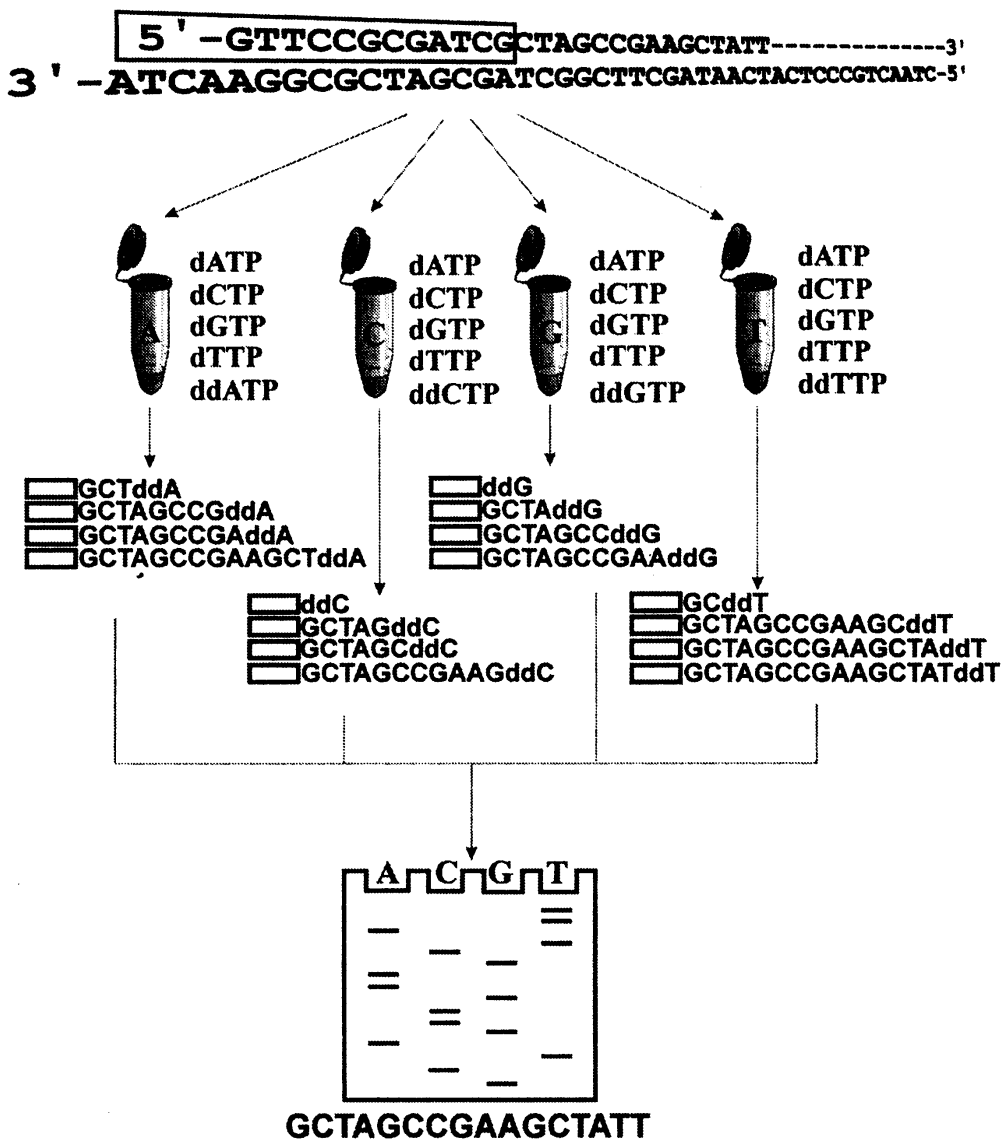


Рис. 2.2. Принцип определения нуклеотидной последовательности ДНК ферментативным методом Сэнгера с дидезокситерминаторами

Другим ключевым моментом являлся гетерогенный размер полученных фрагментов ДНК. Их гетерогенность определялась тем, что все 3'-концы фрагментов вновь синтезируемой цепи ДНК во всех четырех реакционных пробирках были терминированы соответствующими ддНМФ и, таким образом, представляли собой полный набор всех возможных длин в пределах секвенируемого участка, построенного ДНК-полимеразой. Что касается 5'-конца всех этих фрагментов ДНК, принадлежащих новой цепи, то он должен быть для всех фрагментов строго одинаковым и принадлежать 5'-концу исходной затравочной молекулы.

Продукты реакций терминирования подвергались денатурации и уже одноцепочечные меченые фрагменты, имевшие гомогенный 5'-конец и гетерогенные 3'-концы, разделялись по длине высоковольтным электрофорезом высокого разрешения в полиакриламидном геле, позволяющим разделять фрагменты ДНК, отличающиеся всего на один нуклеотид, в четырех соседних дорожках, соответствующих специфич-

ческим терминирующим реакциям, например в последовательности Т, С, G, А. По завершению электрофореза гель экспонировался на рентгеновскую пленку и по прошествии некоторого времени (обычно от одних до двух суток) с проявленной пленки можно было “читать” последовательность нуклеотидов секвенируемого участка ДНК, начиная с нижней части геля и последовательно поднимаясь вверх по этим четырем дорожкам, соответствующим одному фрагменту ДНК.

Выше на рис. 2.2 схематично показан весь описанный выше процесс, но следует отметить, что он несколько идеализирован. При его выполнении экспериментаторов подстерегает множество нюансов, которые могут свести на нет все предварительные усилия. Так, и подготовка самой матрицы к секвенированию, и выбор затравочной молекулы, и определенное соотношение дНТФ/ддНТФ, зависящее, главным образом, от используемой ДНК-полимеразы, и мечение вновь синтезируемой цепи ДНК, и параметры секвенирующего геля, и даже особенности радиоавтографии зависят от конкретных задач и требуют тщательного исполнения. В этой связи последующие разделы этой главы и некоторые другие главы будут посвящены подробному анализу составляющих процесса ферментативного секвенирования ДНК.

## 2.2. МАТРИЦЫ ДЛЯ СЕКВИНИРОВАНИЯ

Поскольку в основе метода ферментативного секвенирования лежит построение новой комплементарной цепи ДНК по уже имеющейся, то большое значение приобретает подготовка молекул ДНК для осуществления этого процесса. Так, если для метода секвенирования ДНК химической деградацией было необходимо наличие одно- или двуцепочечного фрагмента, несущего на одном (гомогенном) из своих концов единственную метку, то для секвенирования ферментативным методом необходимо наличие немеченой матрицы ДНК и при этом желательно одноцепочечной. Хотя, справедливости ради, следует отметить, что матрица ДНК, несущая какую-либо метку (включая и радиоактивную), в большинстве случаев не окажет заметного влияния на заключительный радиоавтограф из-за большой разницы в размерах самой матрицы и вновь синтезированной в условиях терминации цепи ДНК. Что касается матрицы ДНК, один из концов которой несет метку в виде молекулы биотина, то в этом случае возможна ее сорбция на твердой фазе, например, на магнитных частицах со стрептавидином и проведение так называемого твердофазного секвенирования. При этом выхода биотинилированной метки в заключительный раствор терминирующей смеси, содержащей комплементарную цепь ДНК, меченную уже как-то по-другому и предназначенную для нанесения на секвенирующий гель, даже не произойдет.

Возвращаясь к подготовке самих матриц, надо сказать, что ввиду определенных трудностей получения одноцепочечных фрагментов ДНК из двуцепочечных, первые успехи в ферментативном секвенировании были достигнуты с природными одноцепочечными ДНК, такими

как бактериофаги  $\phi$ X17 и f1 [Sanger et al., 1977a; 1973]. Значительные усилия, которые приходилось прикладывать экспериментаторам для получения одноцепочечных матриц ДНК, пригодных для секвенирования ферментативным методом, первое время сдерживали широкое распространение этого метода, поскольку время, затрачиваемое на их получение, являлось основной составляющей всего процесса. Так, в ранних экспериментах по секвенированию ДНК предлагалось клонирование представляющего интерес фрагмента в бактериофаге лямбда с последующим разделением цепей [Davies, 1982] путем весьма длительно-го ультрацентрифугирования в градиенте плотности хлористого цезия с поли(U, G), известного достаточно давно [Szybalski et al., 1971]. Другим способом получения матриц для секвенирования, также основанном на делении цепей, но уже хроматографией, был путь клонирования фрагментов ДНК в специальном векторе рКН47, сконструированном на основе плазмиды рBR322, и содержащем поли(dA):поли(dT) дуплексный фрагмент протяженностью около 100 пн [Hayashi, 1980]. Это позволяло после денатурации плазмидной ДНК осуществлять достаточно эффективное разделение цепей клонированной вставки вместе с векторными последовательностями, одна из которых несла поли(dA)-, а другая – поли(dT)-участки на колонках с олиго(dT)- и олиго(dA)-целлюлозами соответственно.

Ферментативное приготовление одноцепочечной ДНК, пригодной для секвенирования, довольно успешно осуществлялось с помощью экзонуклеазы III [Smith, 1979]. В этом случае фрагмент ДНК расщеплялся одной или двумя подходящими рестрикционными эндонуклеазами и в результате его последующей обработки экзонуклеазой III происходило одновременное расщепление обеих цепей ДНК в направлении  $3' \rightarrow 5'$  и образование или двух одноцепочечных фрагментов, не расщепляемых далее экзонуклеазой III, или структур с оставшимся двуцепочечным участком (в случае неполного расщепления). Поскольку экзонуклеаза III отщепляет различные нуклеотидмонофосфаты с неодинаковой скоростью [Linxweiler, Horz, 1982], то при полном гидролизе некоторого фрагмента ДНК образуются, как правило, несколько отличающиеся между собой по размеру одноцепочечные субфрагменты, составляющие около половины исходного фрагмента ДНК каждый.

Как видно из рис. 2.3., в условиях неполного расщепления цепей ДНК экзонуклеазой III сохраняется двуцепочечный участок, где его  $3'$ -концы также будут служить стартовыми точками для ДНК-полимеразы. Однако, ввиду того, что только специфический секвенирующий праймер будет нести концевую метку, лишь продукты построения новой цепи ДНК путем его удлинения будут видны на радиоавтографе. В дальнейшем этот подход получил продолжение и совместное использование экзонуклеазы III, имеющей  $3'$ -ОН активность, и  $5'$ - $PO_4$ -экзонуклеазной активности белка  $\phi$  фага T7 позволило в условиях *in vitro* получать для секвенирования ДНК ферментативным методом весьма протяженные одноцепочечные матрицы, не прибегая к процедуре субклонирования [Strauss, Zagursky, 1991].

### А

5' -GTCCGACGTCGGGTACAATAAGCGTTTCGCCCTTTTGGGATGGGACTGGGATGACGTCAAAAACGTTTCGCCCTTTGG-3'  
3' -CAGCTGCACGGCATGTTTACGCAAGCGGAAAACGCTACGGCTGACGCTACTGCAGTTTTGGCAAGCGGAAAACC-5'

### Б

5' -GTCCGACGTCGGGTACAATAAGCGTTTCGCCCTTTTGGGATGGC-3'  
3' -CTGACGCTACTGCAGTTTTTCGCAAGCGGAAAACC-5'

### В

5' -GTCCGACGTCGGGTACAATAAGCGTTTCGCCCTTTTGGGATGGGACT-3'  
3' -TACGCTGACGCTACTGCAGTTTTTCGCAAGCGGAAAACC-5'

### Г

5' -GTCCGACGTCGGGTACAATAAGCGTTTCGCCCTTTTGGGATGGGACT-3'  
3' -CGCATGTTTACGCA-5'      3' -TACGCTGACGCTACTGCAGTTTTTCGCAAGCGGAAAACC-5'

### Д

5' -GTCCGACGTCGGGTACAATAAGCGTTTCGCCCTTTTGGGATGGGACT-3' →  
←- 3' -CGCATGTTTACGCA-5'      ←- 5' -TACGCTGACGCTACTGCAGTTTTTCGCAAGCGGAAAACC-5'

Рис. 2.3. Принцип получения с помощью расщепления экзонуклеазой III одноцепочечной матрицы ДНК, пригодной для секвенирования путем отжига специфического праймера и последующего удлинения новосинтезируемой цепи ДНК

Подчеркнут специфический праймер, стрелками указано направление синтеза новой цепи ДНК в секвенирующей реакции с дидезоксирибонуклеотидами

А – исходный фрагмент ДНК; Б – превращение исходного фрагмента ДНК в два отдельных одноцепочечных фрагмента; В – превращение исходного фрагмента ДНК в новую молекулу, содержащую два одноцепочечных участка, удерживаемых вместе оставшимся двуцепочечным участком; Г – отжиг специфического праймера; Д – построение новых цепей ДНК в направлении 5' → 3' от 3'-концов двуспиральных участков, которые могут служить заправочными



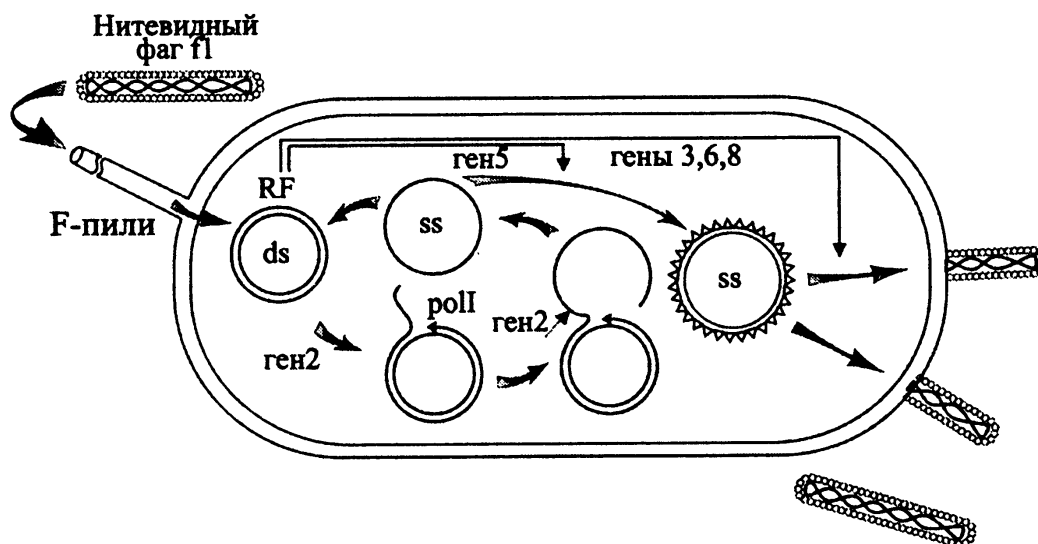


Рис. 2.4. Жизненный цикл нитевидного одноцепочечного фага f1

Весьма удобным оказалось получение одноцепочечных матриц ДНК после их клонирования в векторах на основе нитевидного фага M13 [Messing et al., 1981], чему посвящено огромное количество оригинальных статей и обзоров [Sanger et al., 1980; Davies, 1982; Bankier, Bartell, 1983, 1988; Messing, Bankier, 1988]. Данный подход сделал практически ненужными методы, кратко описанные выше, и благодаря своей простоте резко увеличил доступность, следовательно, и популярность метода ферментативного секвенирования ДНК с дидезокситерминаторами.

Биология одноцепочечного нитевидного фага f1 такова, что в своем жизненном цикле он проходит репликативную двуцепочечную стадию и выделенные на этом этапе его кольцевые молекулы используются в качестве векторов, подобно плазмидным. Наибольшее распространение получила целая серия фаговых векторов M13mp, сконструированных Мессингом и соавт. [Messing et al., 1981; Messing, Vieira, 1982; Vieira, Messing, 1982; Messing, 1983; Norrander et al., 1983; Yanisch-Perron et al., 1985].

Фаговые частицы векторов содержат одноцепочечную ДНК протяженностью около 7 тысяч нуклеотидов, покрытую фаговыми белками. Белковая оболочка фага состоит из основного белка, представленного 2700 копиями (для дикого фага) и кодируемого геном рVIII фага. Остальные белки оболочки рIII, рIV, рVII и рIX относятся к минорным белкам и представлены 5 копиями каждый. Эти белки расположены на концах вариона и отвечают за заражение клетки *E. coli* и за выход созревших фаговых частиц наружу (рис. 2.4). Заинтересованный читатель может почерпнуть более подробную информацию об особенностях организации одноцепочечных нитевидных фагов и их жизненном цикле из обзорных статей [Смит, 1991; Messing, 1983]. Краткая история создания векторов на основе фага M13 изложена в обзорной статье Мессинга [Messing, 1991], а также в разделе 7.3. данной книги.

Фаги f1, будучи не изометрическими, позволяют клонировать в них вставки чужеродной ДНК, однако их размер обычно не превышает 1 тпн ввиду нестабильности вставок большего размера. Созревание фаговых частиц сопровождается их выходом наружу в культуральную среду без лизиса бактериальной клетки. Поскольку ДНК в фаговых частицах представлена одиночной (+)-цепью, то и содержащаяся в рекомбинантных фагах вставка также будет одноцепочечной. Остается только собрать фаговые частицы и отделить от белка фаговую ДНК. После завершения этих процедур одноцепочечная матрица ДНК готова для секвенирования ферментативным методом. Детальное описание всего этого процесса изложено в многочисленных статьях [Messing et al., 1981; Messing, 1983; Sanger et al., 1980; Davies, 1982; Bankier, Barrell, 1983, 1989; Kristensen et al., 1987; Краев, 1988] и практических руководствах [Маниатис и др., 1984; Sambrook et al., 1989; Promega Guide, 1996]. Кроме этого, существует описание отдельных подходов, имеющих целью как-то усовершенствовать процедуры выращивания фага M13, выделения и очистки его одноцепочечной ДНК [Reddy, McKenney, 1996]. Довольно много статей посвящено автоматизации процесса выделения одноцепочечной ДНК фага M13 из большого числа фаговых бляшек, что имеет прямое отношение к автоматическим методам секвенирования ДНК и поэтому будет рассмотрено в соответствующем разделе.

Среди других подходов отметим предложенный оригинальный метод выделения и очистки одноцепочечной ДНК фага M13, основанный на аффинном захвате специальным образом модифицированных векторов M13-101 и M13-102, путем образования трехцепочечных гибридных конструкций [Johnson et al., 1996; Chen et al., 1996]. Данные вектора несут полипуриновый блок из двадцати нуклеотидов, который образует трехцепочечную структуру с дважды комплементарным ему (т.е. несущим два гомологических участка) 50-звенным полипиримидиновым олигонуклеотидом, и поскольку этот последний олигонуклеотид биотинилирован, то имеется возможность сорбции всего этого “сэндвича” на магнитных частицах, покрытых стрептавидином. Показанный на рис. 2.5 весь этот процесс был предназначен для сорбции фаговой ДНК из осветленной культуральной жидкости, удаления белков и элюции “чистой” ДНК, готовой для ферментативного секвенирования. Требуемый специальных векторов, специальных биотинилированных олигонуклеотидов, причем довольно большого размера, магнитных частиц со стрептавидином этот метод никак не может быть универсальным, хотя авторы цитируемой работы говорят о возможности его масштабирования и пригодности для крупномасштабного секвенирования. Следует отметить, что подобный подход с формированием трехцепочечных структур и аффинного захвата был использован этими же авторами также и для выделения вставки из космидного клона с целью ее последующего секвенирования [Ji et al., 1994].

Последующие усовершенствования векторов для молекулярного клонирования, подробно рассмотренные в седьмой главе данной книги, привели к тому, чтобы были созданы фагмиды, представляющие собой химерные конструкции, являющиеся по сути плазмидами, но несущими

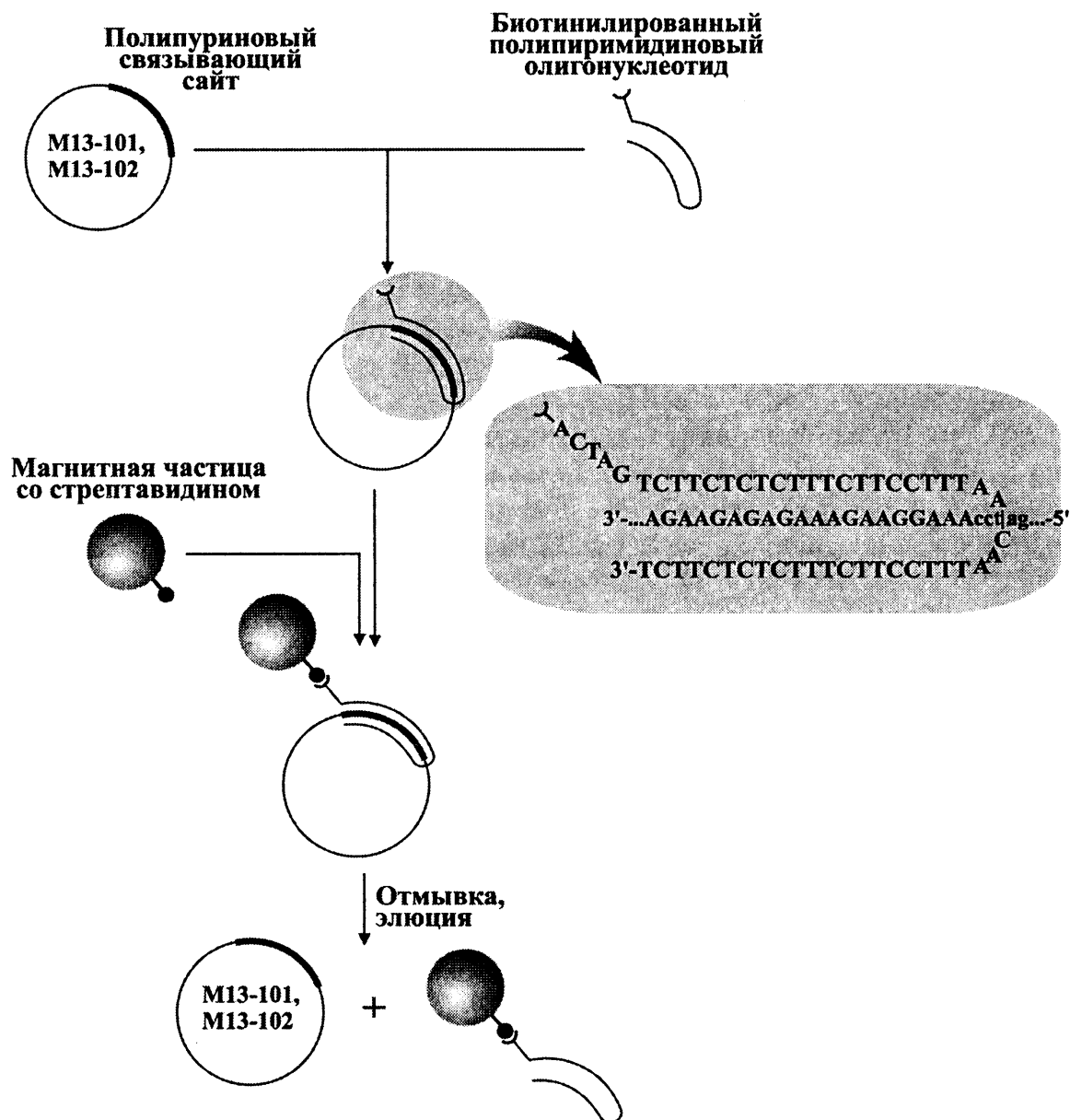


Рис. 2.5. Выделение и очистка одноцепочечной ДНК векторов M13-101 или M13-102, производных от бактериофага M13, путем их аффинного захвата и формирования трехцепочечных конструкций

участок одноцепочечного фага f1, ответственный за репликацию [Dente et al., 1984]. При обычном выращивании этих конструкций они реплицировались по типу плазмиды, но после добавления специального фага-помощника, называемого еще хелперным, они переходили в режим функционирования фага и секретировали наружу, кроме хелперного фага, еще и псевдофаговые частицы с (+)- или (-)-цепью ДНК (рис. 2.6). Упаковка той или иной цепи ДНК в фаговые частицы зависела от ориентации f1 ориджина репликации одноцепочечного фага. Дальнейшие процедуры по сбору псевдофаговых частиц и очистке одноцепочечной ДНК ничем не отличались от таковых при работе с настоящими одноцепочечными фагами. В литературе встречаются упоминания о некоторых улучшениях выделения фагмидных ДНК как одноцепочечных ма-

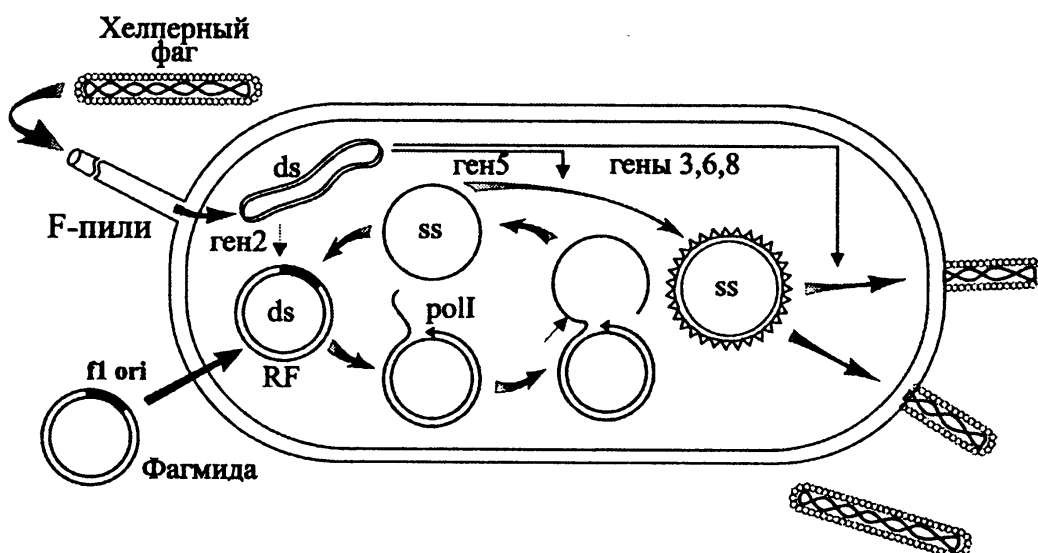


Рис. 2.6. Схема получения одноцепочечной ДНК с помощью образования под действием хелперного фага псевдофаговых частиц, несущих вставку, клонированную в фагмидном векторе

триц для секвенирования [Blondel, Thillet, 1991; Tang et al., 1996]. Интересна работа, в которой говорится о легко воспроизводимом методе получения одноцепочечной ДНК фагмид после трансформации ими компетентных клеток *E. coli*, уже содержащих хелперный фаг [Jurin, Gronenborn, 1995].

Несмотря на то, что с помощью фаговых или фагмидных векторов получение одноцепочечных молекул ДНК не представляло особых проблем, желание не тратить на очистку одноцепочечной ДНК от прочно связанных фаговых белков заставляло исследователей разрабатывать подходы к секвенированию двуцепочечных молекул ДНК, не останавливаясь даже перед фактом, что качество получаемых результатов при секвенировании двуцепочечной ДНК всегда заведомо хуже, чем одноцепочечной. (Здесь понятия “качество” секвенирования и “количество” определенных при этом нуклеотидов очень тесно связаны, поскольку при более высоком качестве полос на радиоавтографе секвенирующего геля “прочитать” можно будет гораздо большее их количество.) Так, не взирая на потенциально худшие результаты, большое число работ посвящено описанию различных подходов по подготовке двуцепочечных матриц ДНК, пригодных для ферментативного секвенирования дидезокси-методом. Такой, казалось бы, неразумный подход объясняется повышением общей производительности всего процесса секвенирования ДНК за счет экономии времени как на очистку одноцепочечных матриц ДНК, так и на проведение необходимого переклонирования вставки в специальный фаговый или фагмидный вектор.

Большое значение при таком секвенировании двуцепочечной ДНК приобретает качество выделенного препарата. И одним из главных требований является отсутствие как линейных форм плазмидной ДНК, так и релаксированных, несущих одиночные разрывы цепей ДНК – “ники”. В противном случае значительное число новосинтезированных

цепей ДНК может оказаться терминированными не специфически, а по причине разрывов цепи в исходной матрице ДНК. Высокую чистоту препаратов плазмидной ДНК, представленной в суперскрученной форме, обеспечивает метод ультрацентрифугирования в градиенте плотности хлористого цезия, широко использовавшийся в первое время [Маниатис и др., 1984; Sambrook et al., 1989]. Однако значительные усилия, затрачиваемые при очистке плазмидной ДНК с помощью ультрацентрифугирования в хлористом цезии, и все большее число плазмид, которые надо было секвенировать, заставляли исследователей разрабатывать новые, более быстрые, простые и дешевые подходы. Тем более что проведенные сравнительные исследования секвенирования двуцепочечных плазмидных ДНК, очищенных в градиенте плотности CsCl и выделенных мини-методом, показали сравнимое качество [Lim, Pene, 1988; Tneguzzo et al., 1988; Zhang et al., 1988; Riley, 1989]. Весьма оригинален метод очистки плазмидной ДНК для ее дальнейшего секвенирования с использованием хлористого цезия и бромистого этидия, не требующий ультрацентрифугирования [Saunders, Burke, 1990]. Этот подход основан на факте образования агрегатов между полисахарами, белками и бромистым этидием в присутствии высокой концентрации CsCl, удаляемых в микроцентрифуге при 10000g. Авторы приводят результаты секвенирования одного и того же образца ДНК, выделенного классическим ультрацентрифугированием в градиенте плотности хлористого цезия и этим предложенным методом, причем сравнение оказывается даже в пользу последнего, что несколько удивляет.

В большинстве случаев выделения плазмидной ДНК для ее последующего секвенирования за основу был взят быстрый щелочной метод [Birnboim, Doly, 1979; Chen, Seeburg, 1985; Hattori, Sakaki, 1985; Kraft et al., 1988; Hisao, 1991, Xiang et al., 1994]. В одной работе было показано, что депротенинизация выделенной щелочным методом плазмидной ДНК кислым фенолом вместо обычно используемой смеси фенол-хлороформ, приводит к высокому качеству картины полос секвенированной ДНК [Weikert, Chambliss, 1989]. В других работах [Murphy, Kavanagh, 1988; Wang et al., 1988; Jones, Schofield, 1990; Seto, 1990] применялся метод выделения ДНК плазмид кипячением [Holmes, Quigley, 1981]. Сообщается об использовании достаточного длительного метода выделения плазмидной ДНК мягким лизисом с помощью смеси слабых детергентов – бридж-58 и дезоксихолат натрия [Clewel, Helinski, 1969], показавшего при этом хорошее качество секвенирования [Lim, Pene, 1988]. В одной работе сообщается о быстром, также путем кипячения, “турбо-преп” выделении плазмидных ДНК, использованных для последующего секвенирования без какой-либо дополнительной очистки [Woodford, Usdin, 1991]. Недавно предложен “турбопреп II”, позволяющий с помощью кипячения осуществлять минивыделение большого количества образцов (до 1920 за 8 ч) плазмидной ДНК одним экспериментатором без использования приборов-автоматов [Konecki, Phillips, 1998]. Другая быстрая процедура выделения плазмидной ДНК, пригодной для ферментативного секвенирования, и проводимая всего в одной пробирке, основана на использовании в качестве осаждающего агента цетилтри-

метиламмоний бромид [Del Sal et al., 1988]. Сообщается об использовании ионообменной смолы Qiagen для быстрого выделения плазмидной ДНК [Zimmermann et al., 1990].

В большинстве упомянутых выше работ [Chen, Seeburg, 1985; Hattori, Sakaki, 1985; Kraft et al., 1988; Wang et al., 1988; Jones, Schofield, 1990; Seto, 1990; Hsiao, 1991, Woodford, Usdin, 1991; Xiang et al., 1994] одноцепочечные матрицы ДНК для секвенирования готовились путем щелочной денатурации уже выделенной плазмидной ДНК и некоторые различия заключались лишь в применяемых концентрациях щелочи и особенностях нейтрализации с помощью или HCl, или уксуснокислого аммония, или уксуснокислого натрия. Еще в одной работе щелочная денатурация при помощи 1M NaOH осуществлялась при повышенной (80°C) температуре в течение 5 мин и после снижения до 37°C проводили нейтрализацию 1M HCl [Zimmermann et al., 1990]. Такая же высокотемпературная щелочная денатурация плазмидной ДНК использовалась еще в одной работе [Martsen et al., 1990]. Другими авторами для удаления щелочи была применена ультрафильтрация в виде спин-диализа в микроцентрифуге [Murphy, Kavanagh, 1988]. Имеется сообщение о превращении суперскрученной формы ДНК плазмиды в одноцепочечную матрицу посредством тепловой денатурации [Agelton, Chen, 1986]. Несмотря на то что наличие в реакционной смеси молекул РНК не оказывает серьезного влияния на проведение терминирующих реакций, тем не менее их отсутствие может только улучшить качество секвенирования ДНК. Так, описан быстрый способ удаления большого количества загрязняющей РНК без использования РНКазы, основанный на селективном осаждении молекул РНК высокой концентрацией (100 мМ) хлористого магния [Yamamoto, Horikoshi, 1995].

Помимо секвенирования вставок в плазмидных векторах такое возможно и с другими векторными системами, и в литературе сообщается о быстрых и простых способах выделения ДНК фага лямбда, пригодной для ее непосредственного двуцепочечного секвенирования ферментативным методом с использованием в качестве затравок стандартных олигонуклеотидных праймеров [Manfioletti, Schneider, 1988; Malik et al., 1990].

Был описан интересный способ получения одноцепочечной матрицы ДНК, легший позднее в основу такого высокопроизводительного метода как твердофазное секвенирование (рис. 2.7) [Stahl et al., 1988]. В цитируемой работе фрагмент ДНК, меченный биотином, сорбировался на твердой фазе в виде авидин-агарозы. После этапов денатурации щелочью и последующей отмывки на оставшейся одной цепи ДНК, связанной через биотин-авидиновый комплекс с агарозой, отжигался секвенирующий праймер и шло построение комплементарной цепи ДНК Кленовским фрагментом ДНК-полимеразы в условиях терминации. Затем еще один раунд щелочной денатурации приводил к высвобождению новосинтезированных цепей ДНК, пригодных для их разделения секвенирующим гелеэлектрофорезом. Справедливости ради следует отметить, что авторы в данной статье упоминают о принципиальной возможности повторения циклов отжига праймера и построения цепи по сорбированной матрице, что

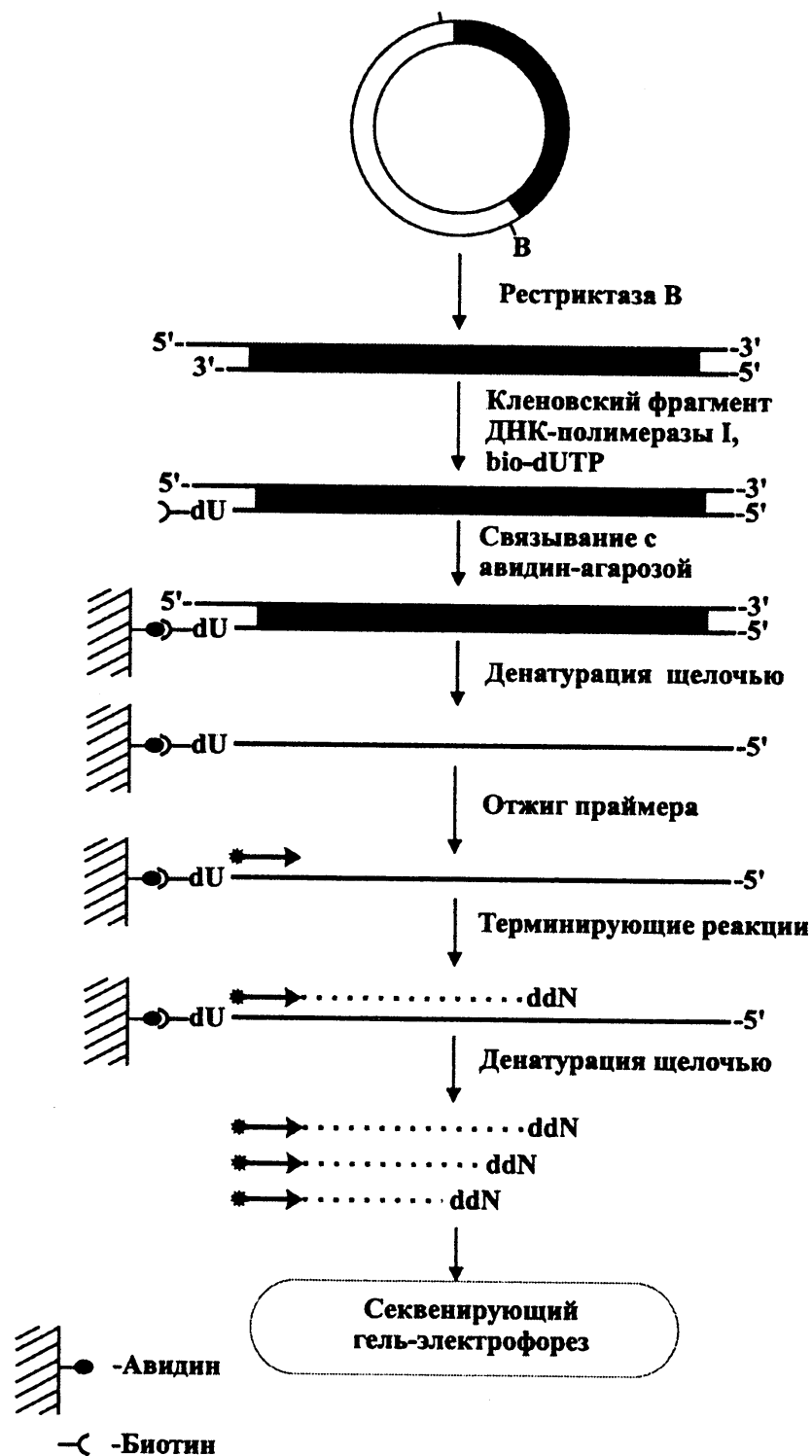


Рис. 2.7. Схема проведения твердофазного секвенирования ДНК

является, по сути, циклическим секвенированием ДНК, рассмотрению вариантов которого будет посвящен специальный раздел главы 3.

Другой способ секвенирования двуцепочечной матрицы ДНК заключается в вытеснении одной цепи за счет самого фермента Кленовского фрагмента ДНК-полимеразы I, для которого характерна подобная активность [Fu et al., 1997]. С этой целью к определенному концу рестриктазного фрагмента ДНК (предварительно дефосфорилированно-

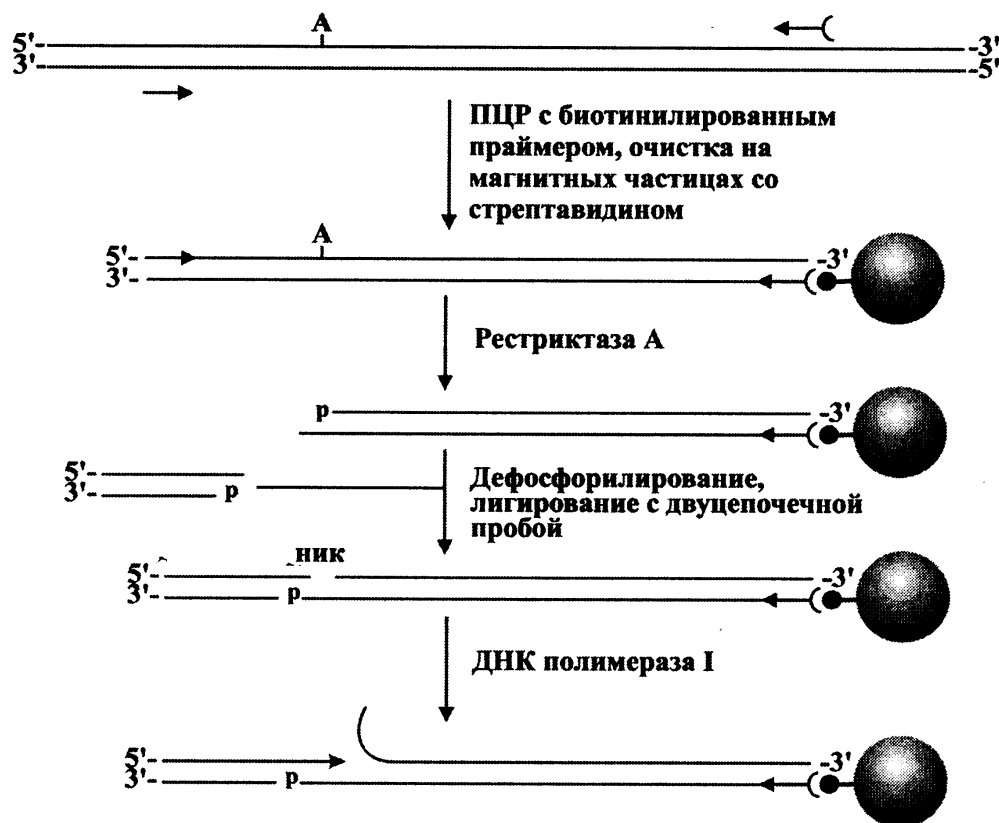


Рис. 2.8. Секвенирование двуцепочечной матрицы ДНК путем вытеснения одной цепи Кленовским фрагментом ДНК-полимеразы I

го) присоединялся путем лигирования фосфорилированный двуцепочечный олигонуклеотидный “праймер”. Образовавшийся в результате “ник” служил стартовой точкой для построения новой комплементарной цепи ДНК в условиях терминации, вытесняя ранее существующую цепь (рис. 2.8). Однако следует отметить, что количество определяемых нуклеотидов таким способом было невелико отчасти из-за низкой процессивности самой используемой ДНК-полимеразы и составляло всего около 50 нуклеотидов, впрочем добавление в реакционную смесь специфического к однонитевой ДНК SSB-белка позволило осуществить “чтение” уже 180 нуклеотидов.

С открытием процесса избирательной амплификации какого-либо фрагмента ДНК с помощью ПЦР и дальнейшего развития этого метода появилась возможность специально готовить двуцепочечные матрицы для секвенирования ДНК ферментативным методом. Причем, прибегая к некоторым ухищрениям, более подробно рассмотренным в главе 3, имеется возможность получать матрицы ДНК, или обогащенные по одной цепи, или содержащие только одну цепь.

### 2.3. ЗАТРАВОЧНЫЕ МОЛЕКУЛЫ

Для построения новой комплементарной цепи ДНК по существующей одноцепочечной матрице фермент ДНК-полимераза нуждается в двуцепочечном участке, выступающем в качестве затравки. Такие за-



травочные молекулы ДНК, называемые праймерами, в настоящее время представляют собой химически синтезированные олигонуклеотиды с определенной последовательностью оснований. Ранее, когда химический синтез был весьма дорог и ограниченно доступен, в качестве затравочных молекул использовались преимущественно короткие олигонуклеотиды, протяженностью 10–12–13 звеньев [Winter et al., 1981], а также рестриктазные фрагменты ДНК-вектора, прилегающие к сайту клонирования вставки. Одним из главных недостатков таких биологических праймеров были их чрезмерная длина, следствием чего являлось уменьшение количества “читаемых” нуклеотидов с одного геля. Рекомендовалось даже проведение дополнительного этапа рестрикции после отжига самого праймера и завершения терминирующих реакций [Godson, 1980; Smith, 1980]. В случае, если сайт узнавания рестрикционной эндонуклеазы, использованной для приготовления биологического праймера и расположенный на его 3'-конце в процессе построения новой цепи восстанавливался, праймер можно было исключить полностью из электрофоретического разделения. В противном случае было необходимо использовать какой-либо пригодный сайт рестрикционной эндонуклеазы внутри праймера (если таковой имелся). Протяженный праймер предлагалось также удалять обработкой щелочью или РНКазой после введения на его 3'-конец рибонуклеотида [Brown, 1978]. Есть примеры использования в качестве затравочной молекулы всей последовательности клонирующего вектора pBR322 [Smith et al., 1979]. В описываемом примере вставка была клонирована по сайтам рестрикционных эндонуклеаз *EcoRI* и *HindIII* и для осуществления отжига, приводящего к секвенированию самой вставки, рекомбинантная плазмида расщеплялась рестрикционной эндонуклеазой *EcoRI*, а исходный вектор – *HindIII*. По завершению терминирующих реакций был обязателен этап расщепления полученных продуктов ферментов *HindIII*, чтобы удалить векторный фрагмент.

Другим типом биологических праймеров, не получившим широкое распространение, являлись довольно короткие структуры (до 13 звеньев), синтезируемые с помощью полинуклеотидфосфорилазы *E. coli* [Gillam et al., 1978; Smith et al., 1979]. Данный фермент в бактериальной клетке ответственен за перенос рибонуклеотидных остатков от рибонуклеозид-5'-дифосфатов в олиго- или полирибонуклеотидам, однако при определенных условиях он может в качестве донора и акцептора использовать и дезоксирибонуклеотидные производные.

Обычно сами исследователи готовили для своих экспериментов определенные праймеры путем расщепления какого-либо клонированного фрагмента ДНК с установленной рестриктазной картой соответствующими рестрикционными эндонуклеазами с последующей очисткой полученных фрагментов. Однако некоторые биологические праймеры были доступны из коммерческих источников. Так, фирма “Bethesda Research Laboratories” (США) поставляла биологические двуцепочечные праймеры длиной 24 пн и 96 пн. 96 пн фрагмент ДНК представлял из себя субклонированный в плазмидном векторе pBR325 исходный 92 пн *EcoRI*–*AluI* участок вектора M13mp2 [Heidecker et al., 1980]. Также

для секвенирования вставок в фаговом векторе M13mp2 был предложен относительно короткий биологический 24 пн *EcoRI*-*Bam*HI праймер (общей протяженностью 30 пн), 6 нуклеотидов которого на 5'-конце не имели гомологии с ДНК матрицы [Anderson et al., 1980]. Для большего удобства авторами цитируемой статьи этот праймер был клонирован по сайтам рестрикционных эндонуклеаз *EcoRI* и *Bam*HI в плазмидном векторе pBR322.

Другим общим недостатком биологических праймеров являлась их изначальная двуцепочечная структура и, следовательно, имеющая место конкуренция за отжиг с комплементарной цепью праймера и матрицей ДНК. Для исключения такой конкуренции необходимо было прибегать к такой весьма трудоемкой процедуре, как деление цепей ДНК праймера денатурирующим гель-электрофорезом. Происходящее восстановление двуцепочечной структуры биологических праймеров во время этапа отжига на стадии проведения терминирующих реакций ДНК-полимеразой приводило к их мечению и в результате на радиоавтографе во всех дорожках выявлялись жирные полосы, соответствующие праймеру. В то же время следует отметить, что в некоторых случаях при использовании двуцепочечных матриц такие праймеры в виде рестриктазных фрагментов ДНК позволяли осуществить секвенирование обеих цепей [Valenzuela, Siddiqui, 1989; Akiyama et al., 1992; Chretien, Shen, 1992].

Кроме биологических праймеров, в литературе встречается упоминание о праймерах, как бы синтезируемых самой ДНК-полимеразой в процессе удлинения исходного праймера [Speek et al., 1986]. Данные праймеры представляли собой некий гибрид синтетического универсального праймера (о котором речь пойдет ниже) и биологического. Так, в результате построения второй цепи ДНК Кленовским фрагментом ДНК-полимеразы I с использованием в качестве затравки универсального праймера образовывались двуцепочечные участки ДНК, протяженность которых зависела от времени инкубации. После этого предлагалось, зная рестриктазную карту клонированного фрагмента ДНК, расщеплять вновь построенный двуцепочечный участок соответствующими рестрикционными эндонуклеазами и уже далее проводить непосредственно само секвенирование в виде терминирующих реакций. Следует отметить, что о подобном подходе говорилось и ранее [Godson, 1980], но в силу своей трудоемкости и плохой воспроизводимости он не нашел широкого распространения.

Совершенствование химического синтеза олигонуклеотидов и его удешевление привели к тому, что биологические праймеры были практически полностью вытеснены из процесса секвенирования ДНК за исключением каких-то особых случаев. Разработанная Мессингом и соавт. система для секвенирования ДНК в одноцепочечном фаге M13 включала и синтетический универсальный праймер длиной 15 нуклеотидов с последовательностью 5'-СССAGTCACGACGTT-3', участок отжига которого был локализован внутри *lacZ'* региона, располагаясь за 17 нуклеотидов до полилинкерного участка [Messing et al., 1981]. (Несовпадение указанного количества нуклеотидов до полилинкера с рис. 2.9.

CGCCAGGGTTTTCCSAGTCAACGAC 24-мерный праймер  
 GTTTTCCSAGTCAACGAC 17-мерный праймер  
 GTTTTCCSAGTCAACGACG 18-мерный праймер  
 CCSAGTCAACGACGTT 15-мерный праймер  
 GTAAAAACGACGGCCAGTG 18-мерный праймер  
 ТААТАСГАСТСАСТАТАГГГ Т7 праймер  
 ААТАСГАСТСАСТАТАГГГ Т7 праймер  
 CGCCAGGGTTTTCCSAGTCAACGACGTTGTAAAAACGACGGCCAGTGAATTGТААТАСГАСТСАСТАТАГГГCGAATTC вектор

Рис. 2.9. Часть нуклеотидной последовательности плазмидного, фагового или фагмидного вектора, содержащего *lacZ'* регион и промотор фага Т7 с указанными местами отжига олигонуклеотидных праймеров различной протяженности

Последовательность полилинкера, совпадающая в данном случае с сайтом гексануклеотидной рестрикционной эндонуклеазы *EcoRI*, подчеркнута

вызвано тем, что те вектора, разработанные Мессингом и соавт., не содержали участок, несущий промотор фага Т7.) Поскольку все вектора серий М13 и рUC несли в своем составе *lacZ'* регион, то, как следствие, они имели и участок для отжига данного праймера, названного поэтому универсальным. Предложенный способ секвенировать вторую цепь клонированного фрагмента ДНК в одноцепочечном векторе серии М13 привел к необходимости разработать новый, в то время более редко используемый праймер, названный обратным, и расположенный с противоположной стороны полилинкера [Hong, 1981]. Позднее была разработана и поставлялась разными фирмами целая линейка аналогичных праймеров (как прямых, так и обратных), несколько различающихся своим расположением относительно полилинкера, как это видно из рис. 2.9 для прямых праймеров. С созданием фагмидных векторов, унаследовавших от своих предшественников данный *lacZ'* участок, оба типа праймеров приобрели равноценное значение. Так, прямой, или форвард-праймер, используются с фагмидами, образующими фаговые частицы с (+)-цепью вектора, тогда как обратный, или реверс-праймер, – с (–)-цепью. Добавление участков фаговых промоторов Т3, Т7 или SP6, фланкирующих полилинкер, позволило разработать специальные праймеры для их отжига на этих участках, так как они были расположены несколько ближе к клонированному фрагменту ДНК, чем стандартные универсальный прямой и обратный. Наличие полилинкеров большой протяженности в векторах серии рBluescript заставило фирму “Stratagene” (США) предложить новые SK- и KS-праймеры, отжигающиеся непосредственно на полилинкере.

Менее универсальный набор из семи праймеров был разработан для плазмидного вектора рBR322 [Wallace et al., 1981]. Эти праймеры представляли собой гомологичные последовательности для участков, фланкирующих некоторые основные сайты узнавания гексануклеотидных рестрикционных эндонуклеаз, таких как *EcoRI*, *HindIII*, *BamHI*, *SalI* и *PstI* и в настоящее время поставляются рядом фирм для секвенирования разных цепей ДНК. Для определения пригодности праймера к секвенированию той или иной цепи ДНК они указываются относительно рестриктазной карты вектора как по часовой, так и против часовой стрелок.

Разработка различных стратегий секвенирования ДНК, рассматриваемых более подробно в разделе 8, и дальнейшее усовершенствование химического синтеза олигонуклеотидов легли в основу стратегии секвенирования, названной подходом с использованием внутренних или прогрессивных праймеров, или праймерной “прогулкой” [Sanchez-Pescador, Urdea, 1984; Brenner, Shaw, 1985; Strauss et al., 1986]. Поскольку данная стратегия требует, кроме универсального праймера, использования и специально синтезированных праймеров, гомологичных “прочитанным” участкам ДНК, то существует необходимость постоянного синтеза все новых и новых олигонуклеотидов. Вся забота, которую должны проявлять экспериментаторы по отношению к универсальным праймерам для секвенирования, так это только беспокоиться об их постоянном наличии в лаборатории, что можно обеспечить как специальным синтезом того или иного праймера или приобретением их из различных коммерческих источников (первое, впрочем, значительно дешевле). Что касается праймеров для секвенирования ДНК праймерной “прогулкой” (или, правильнее сказать, ее классическим вариантом), то здесь выбор каждого праймера индивидуален и в связи с этим надо придерживаться некоторых правил, кратко рассмотренных ниже.

Нуклеотидная последовательность фрагмента ДНК, в пределах которой предполагается выбрать участок для синтеза праймера, должна быть уверенно “прочитана”. Следует исключить сомнительные последовательности, допускающие их неоднозначное “прочтение”. Преследуя цель синтеза минимального количества праймеров, необходимых для завершения определения нуклеотидной последовательности протяженного фрагмента ДНК, надо располагать новый праймер возможно близко к 5'-концу уже известной последовательности, но, как правило, чтобы его 3'-конец находился не ближе 20–50 нуклеотидов от края. Конкретное расположение места для отжига праймера во многом зависит и от особенностей самой последовательности ДНК, где надо принимать в расчет и ее GC-состав, и возможные повторяющиеся элементы, и участки со значительной (потенциальной) вторичной структурой. Не последнюю роль при выборе праймера играет и температура отжига синтезируемых олигонуклеотидов, которая обычно рекомендуется в пределах 40°–50°С при предполагаемой использовании в качестве ДНК полимеразы секвеназы в 60°–70°С для термостабильных ДНК-полимераз. Кроме GC-состава, температура отжига праймеров в значительной степени зависит и от их длины, обычно выбираемой в пределах 15–22 нуклеотидов. Для более точного определения температуры плавления олигонуклеотидов в расчет принимаются не только сами основания, но и их ближайшее окружение, что может быть проведено с помощью специализированных компьютерных программ, подробнее рассмотренных в соответствующей главе. Однако грубый подсчет температуры плавления праймеров может быть проведен вручную с помощью следующей формулы:

$$T_m = 4^{\circ}\text{C} \cdot (\text{G} + \text{C}) + 2^{\circ}\text{C} \cdot (\text{A} + \text{T}) - 5^{\circ}\text{C},$$

хотя, считается, что данная формула справедлива лишь для относительно коротких праймеров длиной до 20 звеньев.

В литературе встречаются рекомендации при выборе длины секвенирующих праймеров исходить из стандартной длины в 18 нуклеотидов для праймера с 50%-ным GC-составом и прибавляя по одному звену с понижением GC-состава на каждые два процента [Gerischer, Durte, 1993]. В то же время эти авторы сообщают об успешном применении различных 17-мерных праймеров, с содержанием G- и C-нуклеотидов от 2 до 9 для секвенирования AT-богатого генома *Clostridium acetobutylicum* [Gerischer, Durte, 1990, 1993].

Исходя из теории вероятности встречаемости отдельных участков нуклеотидной последовательности разной протяженности в ДНК, можно считать, что уже 15-мерный праймер будет (при условии его 100%-ной гомологии с мишенью) уникальным для генома человека, оцениваемого обычно в  $3 \times 10^9$  пн. Вторым условием является непринадлежность данного олигонуклеотида к каким-либо повторяющимся элементам генома. Таким образом, вероятность фальш-праймирования данными олигонуклеотидами такой протяженности с какого-то другого места генома крайне низка. В то же время нельзя не учитывать возможность отжига праймера с не полностью спарившимися основаниями, что как раз может привести к фальш-праймированию. Чтобы хоть как-то уменьшить вероятность этого, необходимо провести анализ уже известной последовательности секвенируемого фрагмента ДНК и самого вектора, уделяя максимальное внимание выявлению участков 100%-ной гомологии с последними 5–7 нуклеотидами праймера на его 3'-конце. В случае все же "нечитаемого" радиоавтографа геля, ввиду наличия в матрице ДНК дополнительных мест отжига праймера, повышение температуры отжига праймера теоретически может помочь преодолеть эту проблему. В противном случае необходимо синтезировать новый праймер с другой последовательностью.

Весьма опасным может быть формирование шпилечных структур внутри самого праймера и образование димеров праймеров, 3'-концы которых образуют прочные связи. Примеры таких праймеров приведены ниже на рис. 2.10.

Как видно из рис. 2.10., для праймера А характерно значительное присутствие комплементарных нуклеотидов на 3'-конце. По всей вероятности для части праймерных молекул (этот процесс не относится к абсолютно количественным и даже при одной и той же температуре олигонуклеотиды могут существовать в разных вариантах) такая структура может сохраниться во время этапа отжига и, таким образом, определенное количество праймера будет не способно служить в качестве затравки для ДНК-полимеразы. Так, для первого праймера существуют различные варианты образования димерных молекул, два из которых (А1 и А2) характеризуются достаточно прочной структурой, чтобы существовать в растворе во время этапа отжига данного олигонуклеотида на матрице ДНК. Результатом этого могут явиться более слабые сигналы с радиоавтографа геля. Что касается варианта А3, то его существование крайне маловероятно. Для праймера Б, незначительно отличающегося по своей последовательности от праймера А, приведенные возможные димеры при температуре отжига, скорее всего, существовать не будут.

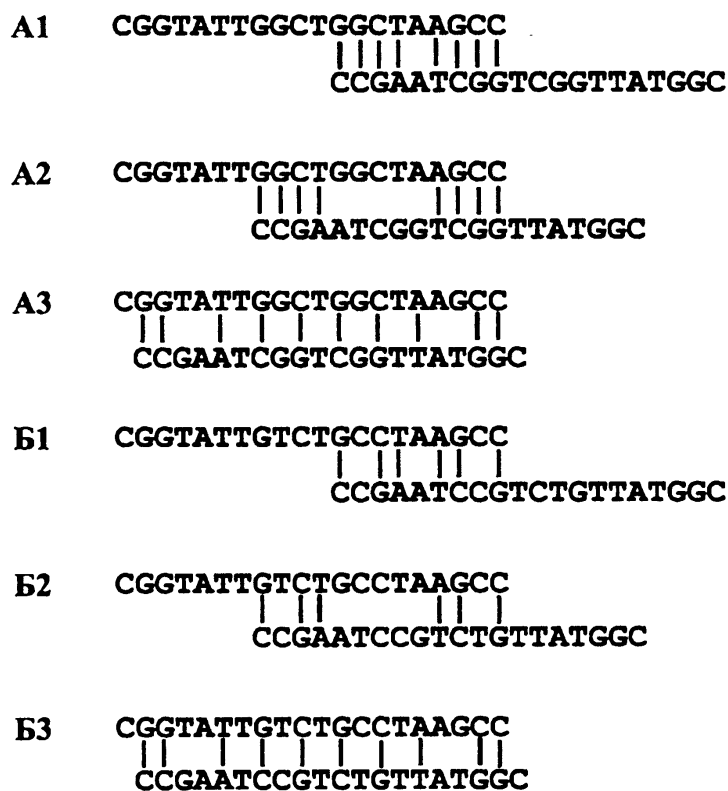


Рис. 2.10. Образование различных вариантов димеров олигонуклеотидными праймерами

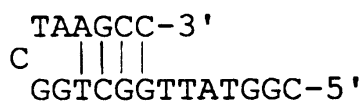


Рис. 2.11. Внутримолекулярная шпильчатая структура олигонуклеотидного праймера

Кроме образования димеров, для праймера А возможно предсказать образование внутримолекулярной шпильчатой структуры, изображенной на рис. 2.11. Такая достаточно прочная структура также способна исключить часть праймера из реакции отжига на матрице ДНК и соответственно ухудшить конечные результаты.

Особый тип праймеров для секвенирования ДНК представляет собой короткая двуцепочечная молекула ДНК, состоящая из двух комплементарных олигонуклеотидов, сконструированных так, что при их отжиге друг на друга образуется несколько выступающих нуклеотидов (обычно 4 или 5) на 3'- или 5'-конце. Данная короткая молекула ДНК за счет образующегося таким образом так называемого липкого конца при помощи ДНК-лигазы лигируется с комплементарным ему концом матрицы ДНК, образованным под действием соответствующей рестрикционной эндонуклеазы. В дальнейшем участок самого лигированного олигонуклеотида служит местом для отжига секвенирующего праймера (которым может быть один из данных комплементарных олигонуклеотидов) [Fuguyama et al., 1994]. В другом варианте при лигировании праймерной двуцепочечной молекулы ДНК специально формируется "ник" на границе этих фрагментов ДНК, с которого и начинается построение новой цепи или за счет смещения прежней цепи ДНК или ее

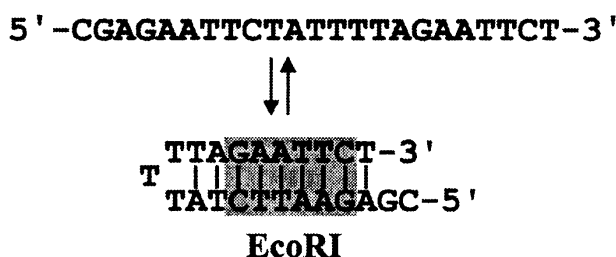


Рис. 2.12. Нуклеотидная последовательность и образование вторичной структуры сплинкера

удаления 5' → 3'-экзонуклеазной активностью ДНК-полимеразы [Fu et al., 1995; 1997]. Также было найдено интересное решение по объединению в одной молекуле праймера и линкера, получившего название сплинкер (splinker – sequencing primer linker) [Kalisch et al., 1986]. Особенностью такого сплинкера было то, что его последовательность содержала внутренний участок гомологии, что при определенной температуре приводило к формированию вторичной структуры в виде “шпильки” с выступающим или тупым концами.

Из рис. 2.12, на котором приведен пример такого сплинкера, видно, что данный олигонуклеотид может существовать также и в форме “шпильки” с выступающим на 5'-конце нуклеотидами, пригодными для их лигирования с подобными же. Необходимым элементом сплинкера является наличие в его двуцепочечном участке доступного сайта узнавания относительно редкошпящей рестрикционной эндонуклеазы, позволяющей на определенном этапе удалить концевую петлю. Более подробно варианты с использованием подобных одно/двуцепочечных праймеров будут рассмотрены в главе 8.

Даже самый малый (обычный) масштаб синтеза олигонуклеотидов в 0,05 мкмоль (существуют синтезаторы с масштабом синтеза и 0,03 мкмоль) позволяет синтезировать в тысячи раз превышающее количество олигонуклеотидов, используемых в качестве затравки в реакциях секвенирования. Остаток праймера, как правило, больше нигде не применяется, что крайне неэкономно. Данное обстоятельство заставляло искать какие-то новые варианты стратегии секвенирования ДНК праймерной “прогулкой”.

Весьма существенное влияние на дальнейший процесс совершенствования праймеров для секвенирования протяженных фрагментов ДНК праймерной “прогулкой” оказала статья Штудиера [Studier, 1989], в которой впервые говорится о необходимости существования предсинтезированной библиотеки олигонуклеотидных праймеров. В данной статье был проведен подробный анализ теоретически возможных частот встречаемости праймеров разной длины в случайной ДНК, клонированной в космидах. (Под случайной ДНК здесь понимается ДНК с приблизительно равным соотношением всех четырех нуклеотидов в одной цепи и не содержащей блоков каких-либо простых повторов.) Были проанализированы праймеры длиной от 6 до 12 нуклеотидов и показано, что 8-мерные праймеры вполне пригодны для их использования в каче-

стве затравок, тогда как ранее считалось, что секвенирующий праймер должен быть не короче 15–17 нуклеотидов. Что касается 6- или 7-мерных праймеров, то они также могут служить в качестве затравок, но для них, как правило, имеются множественные участки отжига на анализируемой в цитируемой работе рекомбинантной космиде (общей длиной около 45 тпн) и поэтому такие короткие праймеры не пригодны для целей секвенирования. Вероятность наличия уникальных единичных сайтов в рекомбинантной космиде для окта-, нона- и декамеров значительно выше и, исходя из этого, Штудierer предположил, что существование библиотеки из 5900 октануклеотидных, 8400 нонануклеотидных и 26000 декануклеотидных праймеров будет вполне достаточным для осуществления проекта секвенирования генома человека. Экономический расчет показал, что стоимость таких праймеров будет составлять очень незначительную долю стоимости всего процесса секвенирования ввиду того, что для каждой секвенирующей реакции необходима 1/1000 или даже 1/10000 часть синтезированного праймера.

Однако синтез библиотеки праймеров такого размера все равно был невозможен и в связи с этим проводился более тщательный анализ секвенирующих праймеров на предмет сокращения их общего количества. Было продемонстрировано высокое качество секвенирования ДНК с использованием октануклеотидных праймеров [Burbelo, Iadarola, 1994]. Авторы считают, что нет необходимости в синтезе и применении для секвенирования рекомбинантной ДНК, клонированной в плазмидных векторах, полной библиотеки октамеров, состоящей из 65536 вариантов, а вполне можно обойтись 1000 вариантов, причем среди них не должно быть участков, гомологичных обычным векторным последовательностям, палиндромным участкам или праймеров с очень низким GC-составом.

Хотя и существует возможность праймирования ДНК гекса- и даже пентануклеотидными праймерами, они образуют множественные участки связывания с ДНК матрицы и поэтому как единичные праймеры не пригодны для секвенирования. Несмотря на то, что была показана принципиальная возможность использования в качестве затравки октануклеотидных праймеров, все же их частота встречаемости в случайной ДНК не позволяет считать их достаточно уникальными. В этом отношении нонамерные праймеры являются более предпочтительными и как показано рядом авторов, минимально возможно короткими для эффективного секвенирования ДНК. Однако, полная библиотека 9-мерных праймеров должна уже состоять из 262 144 различных вариантов ( $4^9$ ) и содержать 2 359 296 нуклеотидных звеньев. Синтез такого количества праймеров нереален и поэтому ряд исследователей пошли по пути компьютерного анализа всех теоретически возможных вариантов 9-меров, с тем чтобы сильно ограничить число действительно пригодных и необходимых [Siemieniak, Slightom, 1990]. Так, этими авторами была предложена библиотека всего из 3342 нонамерных праймеров. Проведенный компьютерный анализ на предмет распределения всех этих праймеров по всей длине клонированных фрагментов ДНК крупного размера из генома человека с известными



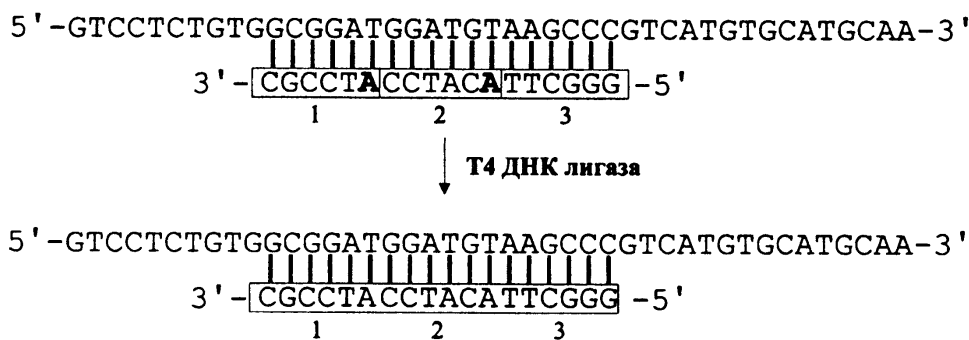


Рис. 2.13. Схема отжига встык на матрице ДНК гексануклеотидных праймеров, показанных в виде прямоугольников и их последующего лигирования с помощью Т4 ДНК-лигазы

Выделены фосфорилированные нуклеотиды на 5'-концах двух гексануклеотидов. Водородные связи, образующиеся при спаривании комплементарных оснований, показаны вертикальными черточками

нуклеотидными последовательностями, имеющимися на тот момент в 63-м выпуске Генбанка, общей протяженностью около 800 тпн показал их довольно равномерное распределение и пригодность для секвенирования ДНК с помощью подхода праймерной “прогулкой”. Столь сильное ограничение числа нонамерных праймеров было достигнуто за счет ряда следующих требований предъявленных к этим праймерам. Так, предполагалось расположение на 3'-конце праймера G или C нуклеотидов и в соседней позиции A или T нуклеотидов; общий GC-состав должен быть в пределах от 45 до 60%; не должно быть полинуклеотидных повторов (таких как AA, CC, GG, TT); не должно быть динуклеотидных повторов (например, GCGCGCGC); и, наконец, праймер не должен быть гомологичен каким-либо коротким повторяющимся элементам, широко представленным в геномах различных организмов, как, например *Alu*-повторам. Разработанный компьютерный алгоритм с этими заложенными требованиями выдал последовательности 3342 нонамерных праймеров, удовлетворяющих всем этим условиям. Однако авторы признавали, что и этот размер библиотеки праймеров все же великоват. Позднее они ограничили размер подобной библиотеки праймеров до всего 2391 варианта, но при этом, к сожалению, не указали, за счет каких дополнительных ограничений им это удалось сделать [Slightom et al., 1994]. В этой работе авторы провели не только компьютерный анализ, но и осуществили определение нуклеотидных последовательностей двух клонированных фрагментов ДНК общей протяженностью около 15 тпн с помощью 59 праймеров из разработанной ими библиотеки в 2391 нонамерный праймер.

Другим подходом для секвенирования ДНК праймерной “прогулкой” с помощью коротких предсинтезированных праймеров явилось составление 12-, 18- или даже 24-членных праймеров из библиотеки гексануклеотидных фрагментов посредством их лигирования после отжига встык на одноцепочечной матрице ДНК (рис. 2.13) [Szybalski, 1990, 1993].

Серьезным преимуществом данного процесса, получившего название SPEL-6 (Sequential Primer Elongation by Ligation with 6-mers), стал от-

носителем небольшой размер библиотеки гексануклеотидных праймеров, равный всего  $4^6 = 4096$ . После низкотемпературного лигирования 6-меров на матрице ДНК и их превращения в полноразмерный праймер с этого участка шло построение новой цепи ДНК, тогда как при повышении температуры непролигировавшие 6-членные фрагменты диссоциировали с матрицы ДНК. Еще меньшую библиотеку из всего  $10^{24}$  гексануклеотидных праймеров удалось создать за счет использования вырожденных по одному нуклеотиду праймеров и успешно применить ее для секвенирования ДНК [Kaczorowski, Szybalski, 1994]. Эти авторы показали, что 18-мерный (3 гексануклеотидных фрагмента) праймер формируется за 10 мин лигированием отожженных встык на матрице ДНК данных 6-меров. В дальнейшем этими же авторами был проведен детальный анализ условий лигирования и секвенирования ДНК с помощью SPEL-6 подхода [Kaczorowski, Szybalski, 1996a]. Было показано, что лигирование 6-членных праймеров носит кооперативный характер и лучшие результаты по секвенированию с помощью лигированных на матрице ДНК праймеров обнаружили 18–60 нуклеотидные праймеры, составленные из 3–10 гексамерных фрагментов. Более длинные праймеры позволили лучше преодолеть проблемы, связанные со вторичной структурой. Различная эффективность и некоторые особенности лигирования гексамеров выявились при использовании ДНК лигазы фага T4 и термостабильной ДНК лигазы *Rodothermus marinus*. Так, Rm-лигаза требовала не менее 6 гексануклеотидных фрагментов, но при этом, как было выяснено, не нуждалась, в отличие от T4 ДНК-лигазы, в SSB-белке. Последнее весьма важно, так как удаление этого, иногда добавляемого в реакцию и способствующего лигированию, белка [Kaczorowski, Szybalski, 1994] перед электрофоретическим разделением секвенируемых образцов представляет значительную проблему [Kaczorowski, Sektas, 1996].

Пригодность SPEL-6 подхода для автоматического секвенирования флуоресцентно меченой ДНК была продемонстрирована разными авторами [Hou, Smith, 1994; Johnson et al., 1996; Kaczorowski, Szybalski, 1996b; Lodhi, McCombie, 1996]. Интересно отметить, что последние авторы обнаружили высокую пригодность лигированных гексамерных праймеров для целей секвенирования ДНК праймерной “прогулкой” и только один такой лигированный праймер показал плохой результат. Контрольное секвенирование этого участка с помощью соответствующего обычного 18-мерного праймера, синтезированного в олигонуклеотидном синтезаторе, показало, что и он также оказался не способен однозначно определить последовательность этого участка ДНК (Lodhi, McCombie, 1996).

Был предложен интересный способ создания 12-, 18-, 24- или 30-мерных праймеров путем лигирования в растворе коротких гексамерных блоков с помощью T4 или T7 ДНК лигаз [Dunn et al., 1995]. Такие блоки для лигирования подбирались так, что каждый из них образовывал по три комплементарные пары оснований с двумя другими. Особенностью этого комплекса праймеров являлось то, что фосфатные группы на 5'-конце, необходимые для лигирования, несли лишь гексамерные блоки, располагающиеся на одной цепи. В результате лигирования и последующей денатурации образовывались полноценные 12-, 18-, 24-

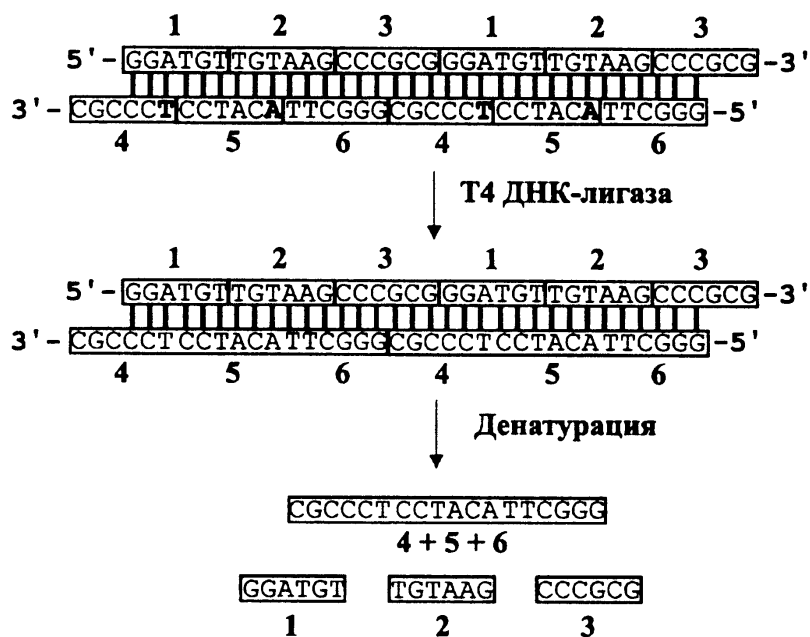


Рис. 2.14. Схема формирования 18-мерного олигонуклеотидного праймера в растворе из гексануклеотидных блоков путем лигирования с помощью T4 ДНК-лигазы

Отдельные гексануклеотиды изображены в виде прямоугольников. Выделены фосфорилированные нуклеотиды на 5'-концах двух гексануклеотидных звеньев. Вертикальные черточки указывают на спаривание комплементарных нуклеотидов

или 30-нуклеотидные праймеры и не пролигированные 6-меры от противоположной цепи (рис. 2.14). Авторы обнаружили, что более эффективное лигирование гексануклеотидов и формирование, например, 18-звенного олигонуклеотидного праймера за счет существующих стэкинг-взаимодействий происходит при условии образования шестью гексамерами выступающих комплементарных концов, способных к отжигу друг с другом, приводящему к увеличению общей протяженности отрезка двуцепочечной сегментированной ДНК. Ранее этой же группой исследователей был предложен способ создания сегментированного праймера из пула гексануклеотидных фрагментов путем их тандемного отжига на матрице ДНК [Kieleczawa et al., 1992].

Главным отличием от SPEL-6 подхода здесь явилось то, что гексануклеотидные фрагменты ДНК не лигировали между собой и они удерживались на матрице ДНК с помощью SSB-белка, а также за счет стэкинг-взаимодействий между соседними нуклеотидными основаниями из разных 6-меров (рис. 2.15). Независимо и почти одновременно с данной работой Улановским и соавт. было проведено детальное исследование особенностей праймирования ДНК с помощью так называемых модульных праймеров [Kotler et al., 1993]. Отдельными модулями служили пентамерные, гексамерные, гептамерные и октамерные праймеры. Используемые поодиночке пентамерные и гексамерные праймеры образуют множественные участки праймирования, однако они же в составе сегментированного праймера (например, 5 + 5 + 7 или 6 + 6 + 6 или 6 + 8) отжигаются в уникальных местах как если бы они представляли собой цельный праймер и вполне пригодны для секвенирования. Проявляющийся кооперативный эффект заключается в том, что для каждого модуля повышается



Рис. 2.15. Схема отжига встык на матрице ДНК гексануклеотидных праймеров, показанных в виде прямоугольников и удерживаемых на ней кроме водородных связей, образующихся при спаривании комплементарных оснований (показанных вертикальными черточками), еще и стэкинг-взаимодействиями соседних оснований разных гексануклеотидных праймеров и специфическим SSB-белком

температура ассоциации/диссоциации при наличии соседних праймеромодулей. Авторам удалось исключить использование нежелательного SSB-белка и отдельные модули при этом взаимодействуют друг с другом за счет стэкинг-взаимодействия. Как известно, пурин-пуриновые взаимодействия прочнее вариантов с другими нуклеотидами и, таким образом, соседние модули, несущие в своих граничных участках пурины, в большей степени пригодны для секвенирования ДНК с помощью таких модульных праймеров. Две проанализированные этими авторами ДНК-полимеразы обнаружили некоторые различия при использовании одних и тех же модульных праймеров в качестве затравок. Дальнейшее развитие семейство модульных праймеров получило в следующей работе этой же группы авторов [Kotler et al., 1994]. В цитируемой работе модульные праймеры составлялись из предсинтезированной библиотеки в тысячу 5- и 7-мерных олигонуклеотидов. Сложным нововведением явилось то, что для гептануклеотидных праймеров, предназначенных быть или средними, или задними блоками составного модульного праймера во время химического синтеза, специально добавлялась фосфатная группа к 3'-концу. Это исключало нежелательное построение цепи от этих неосновных модулей. Проблема выявления другого, вновь синтезированного фрагмента ДНК от среднего модуля, оказавшегося в силу каких-то причин крайним, не возникала при использовании в качестве метки радиоактивного фосфора на 5'-конце основного модуля. Однако при другом способе введения метки, в частности флуоресцентной метки для автоматического секвенирования, это приобретает серьезное значение. Следует отметить, что с целью уменьшения предсинтезированной библиотеки гептамеров (16384 или  $4^7$  теоретически возможных) два последних нуклеотида на 5'-конце синтезировались как вырожденные. Авторы данной статьи, исходя из своего опыта, отмечают, что неудачными праймерами оказываются те, которые характеризуются весьма низким GC-составом или в их состав входят модули, несущие только один G- или C-нуклеотиды или вообще не несущие оных, причем включая вырожденные участки, состоящие всего из одного или двух типов нуклеотидных оснований, не имеющие Pu/Pu-участков, образующих стэкинг-взаимодействия на границе 5-мерного переднего модуля и следующего за ним гептамерного. Таким образом, если исключить олигонуклеотидные модули с нежелательными параметрами, то библиотеки 5- и 7-меров могут быть ограничены по 500 вариантов для каждого типа праймеров.

Заинтересовавшись обнаруженным эффектом образования единст-

венного уникального места отжига модульного праймера, тогда как используемые поодиночке составляющие его пентамерные и гексамерные праймеры образуют множественные участки праймирования, эта же группа авторов провела специальное исследование, направленное на выяснение механизмов этого процесса [Beskin et al., 1995]. Различные комбинации модульных и цельных праймеров с разным количеством фермента показали наличие эффекта конкуренции за “захват” молекулы фермента комплексами матрица–затравка, причем с заметным преимуществом более протяженных модулей. Практический вывод из результатов этой работы заключается в неиспользовании значительных количеств ДНК-полимеразы при секвенировании с помощью модульных праймеров, ввиду опасности образования множественных участков отжига и построения с них новых комплементарных цепей ДНК.

Значителен вклад российских ученых в разработку сегментированных праймеров для секвенирования ДНК праймерной “прогулкой”. Свердловым и соавторами получены интересные результаты по секвенированию ДНК с помощью трех пентамерных праймеров, образующих последовательные дуплексы с матрицей ДНК [Ажикина и др., 1993б; Azhikina et al., 1993]. Преимуществом использования библиотеки пентамерных праймеров для секвенирования ДНК праймерной “прогулкой” является довольно малый размер последней. Так, полная библиотека всех возможных пентануклеотидов состоит всего из 1024 ( $4^5$ ) вариантов. Особенностью данных пентамеров  $(m^5C)(m^5C)(am^2A)GT$ ,  $G(am^2A)(m^5C)GG$  и  $(am^2A)(am^2A)(am^2A)(am^2A)(m^5C)$  (где  $m^5C$  – 5-метилдезоксцитидин,  $am^2A$  – 2-аминодезоксаденозин) было то, что они содержали модифицированные цитозиновые и адениновые основания, что способствовало большей прочности образуемых дуплексов, как это было показано в их предыдущей работе [Ажикина и др., 1993а]. Исследованные температуры плавления ( $T_{пл}$ ) дуплексов, образованных ДНК фага M13mp18, обычным 17-мерным праймером и двумя другими, характеризующимися разным набором модифицированных оснований, заметно различались. Так, для немодифицированного праймера 5'-СТААААСGACGGCCAGT-3'  $T_{пл}$  была равна 68 °С, тогда как для несущего только модифицированные цитозины она была уже 73 °С и для полностью модифицированного данного праймера по всем цитозинам и аденинам одновременно  $T_{пл}$  составила 78 °С. Подобное увеличение  $T_{пл}$  происходит за счет того, что аминогруппа 2-аминодезоксаденозина участвует в образовании третьей водородной связи в А–Т паре, тогда как при образовании двойной спирали дополнительная метильная группа 5-метилцитозина попадает в большую бороздку ДНК [Ажикина и др., 1993а]. Использование всех этих 17-мерных праймеров в секвенировании ДНК показало, что модифицированные праймеры приводят к образованию более сильных сигналов, что может иметь важное значение для экономии реагентов. Было также выявлено, что степень модификации праймеров прямо коррелирует с величиной сигнала на рентгеновской пленке [Ажикина и др., 1993а].

Анализ различных факторов, влияющих на эффективность праймирования ДНК короткими модифицированными праймерами, расположенными встык на матрице ДНК, показал, что наиболее критичным

является прочность связывания сегмента, расположенного на 3'-конце [Azhikina et al., 1996]. Этой же группой авторов проведено исследование особенностей распределения гексануклеотидов вдоль матрицы ДНК с помощью электронного микроскопа [Lagutina et al., 1996]. Ранее ими же была предложена схема взаимодействия матрицы, затравочной молекулы и ДНК-полимеразы, заключающаяся в том, что сама молекула фермента выполняет активную роль в образовании данного комплекса [Ажикина и др., 1993б]. Так, в цитируемой работе авторы предположили, что стабильность иницирующего комплекса формируется как взаимодействием затравки с матрицей, так и вкладом самой ДНК-полимеразы, причем для разных ДНК-полимераз он может различаться, оставаясь при этом весьма значительным.

Еще более короткие блоки, состоящие из модифицированных тетра-нуклеотидов, были применены Кнорре и соавт. в качестве сегментированных праймеров для секвенирования ДНК [Кнорре и др., 1996]. Проведенное ими исследование эффективности праймирования такими составными праймерами из четырех модифицированных феназином тетра-нуклеотидных блоков показало в отдельных случаях результаты, сравнимые с такими при использовании стандартного универсального праймера. Причем наилучшие результаты были получены с праймером, у которого только два блока из четырех несли производные феназиния 5'-PhnGTAA PhnAACG ACGG CСAG-3'. Следует отметить, что данный составной праймер является укороченным (на один нуклеотид) вариантом 17-мерного универсального праймера. Продемонстрированная возможность использования таких коротких модифицированных модулей в качестве секвенирующих праймеров позволяет создать относительно небольшую библиотеку предсинтезированных олигонуклеотидов, состоящую всего из 256 (4<sup>4</sup>) вариантов, причем феназиниевые производные тех или иных тетрамеров, не зависящие от их последовательности, могут готовиться по мере необходимости. Этим данный подход обеспечивает дополнительное преимущество перед модифицированными пентамерами, полученными Ажикиной и соавт. [Ажикина и др., 1993б; Azhikina et al., 1993], поскольку последние (пентамеры) могут нести только модифицированные аденины и цитозины.

Интересным подходом к секвенированию ДНК праймерной "прогулкой" является способ ферментативного дифференциального удлинения короткого праймера с помощью неполного набора дезоксинуклеотидтрифосфатов, получившего наименование DENS (Differential Extension with Nucleotide Subsets) [Raja et al., 1997a]. Сущность данного способа заключается в том, что октануклеотидный праймер отжигается на предполагаемой к секвенированию матрице ДНК сразу в нескольких местах, но в результате ограниченного построения новой цепи только в тех местах, где на матрице будут находиться нуклеотиды комплементарные добавленным в реакционную смесь дезоксинуклеотидтрифосфатам, будет происходить удлинение праймера. Следующий этап проведения терминирующих реакций при повышенной температуре (60 °С) приведет к диссоциации с матрицы неупродлинённых 8-мерных праймеров и, таким образом, при построении новой цепи множественных вновь синтезированных фрагментов ДНК образовываться не будет (рис. 2.16).

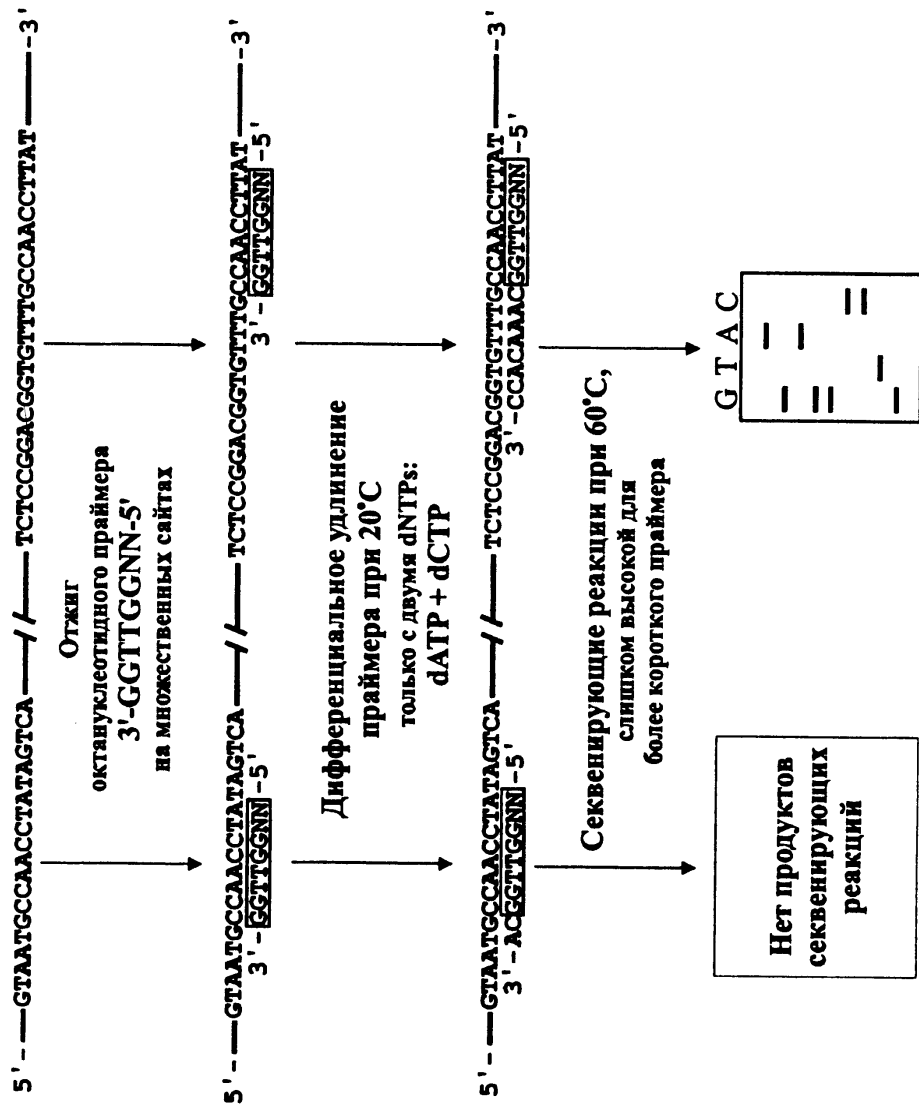


Рис. 2.16. Схема использования октамерных олигонуклеотидов в качестве уникальных праймеров для секвенирования ДНК праймерной “прогулкой” при помощи дифференциального удлинения данного праймера неполным набором дезоксирибонуклеотидтрифосфатов

Важность DENS-механизма праймирования заключается в том, что октануклеотидные праймеры служат гораздо лучшей затравкой, чем любые более короткие модульные праймеры. Хотя, как отмечают авторы, DENS-механизм праймирования в сочетании с гептамерными модульными праймерами дает еще лучший эффект (естественно, первым праймером на 3'-конце служит октануклеотидный). Все же полная библиотека октануклеотидных праймеров слишком велика ( $65536$  или  $4^8$ ) и авторами предложено использовать октануклеотидные праймеры с двумя полностью вырожденными нуклеотидами на 5'-конце, что позволяет резко ограничить число возможных вариантов данных октануклеотидов [Raja et al., 1997a]. Авторы считают, что вероятность наличия второго дополнительного сайта отжига данного праймера, где он также будет удлинен в ферментативной реакции хотя бы на 5 нуклеотидов (что может оказаться достаточным для дальнейшего праймирования при высокой температуре) весьма низка и составляет для клонированных в плаزمиде фрагментов ДНК длиной 7–10 тпн всего около 10%. Что касается векторной молекулы, то зная ее последовательность, такую возможность множественного праймирования с ней можно исключить, так как при секвенировании ДНК праймерной “прогулкой” обычно существует достаточно большой выбор участков среди уже секвенированной ДНК в качестве дальнейших праймеров с отличающейся последовательностью для удлинения самого праймера этим способом. Другой возможностью является выбор той или иной смеси дезоксинуклеотидтрифосфатов (A + C; A + G; A + T; C + G; C + T и G + T), наиболее оптимально подходящих к каждому конкретному случаю. Еще одним приложением DENS-механизма удлинения праймеров является возможность использования коротких 8-мерных олигонуклеотидов в циклическом секвенировании, где только на втором этапе используется термостабильная ДНК-полимераза [Raja et al., 1997a]. Подобное применение обычных модульных праймеров в циклическом секвенировании невозможно из-за их преждевременной диссоциации с матрицы ДНК при повышении температуры.

Отдельную группу затравочных молекул представляют собой так называемые вырожденные праймеры, составленные или на основе известного участка аминокислотной последовательности какого-либо белка, или на основе относительно консервативной последовательности нуклеотидов какого-либо гена из мультигенного семейства или ортологичного гена. В качестве примера можно привести некий условный участок аминокислотной последовательности Asp Gln Glu His Lys и соответствующий ему вырожденный праймер GAY-CAR-GAR-CAY-AAR. Показано эффективное использование подобных праймеров для скрининга бактериальных колоний на предмет конкретных вставок и их последующее секвенирование [Nichols, Dixon, 1988; Javed, Joshi, 1990]. Известно, что разные аминокислоты кодируются неодинаковым числом триплетов и поэтому для выбора подобных праймеров необходимо стремиться выбирать такие блоки аминокислот, для которых будет характерна минимальная вырожденность. Однако в случае неизбежного присутствия в выбранном участке аминокислот, кодируемых большим чис-



лом триплетов, такой праймер будет представлять собой весьма гетерогенную смесь олигонуклеотидов. Применение остатков инозина, образующего комплементарные пары с любым из четырех дезоксирибонуклеотидов, в вырожденных местах таких праймеров в значительной степени решает эту проблему [Knoth et al., 1988]. Использование таких праймеров с инозином вместо переменных нуклеотидов показало их высокую пригодность для секвенирования ДНК [Shen et al., 1993, 1996]. Сообщается также о синтезе праймеров для секвенирования ДНК и ПЦР, содержащих или 3-нитропиррол, или 5-нитроиндол в качестве универсальных оснований [Loakes et al., 1995].

Как уже отмечалось выше, для успешного секвенирования ДНК ферментативным методом необходимо осуществить отжиг затравочной молекулы на матрице ДНК в строго определенном месте. Последнее достигается как правильным подбором последовательности самого праймера, точного его синтеза и последующей очистки (если необходимо), так и определенной температурой отжига. В ранних рекомендациях предлагалось смесь одноцепочечной ДНК и двуцепочечного (биологического) праймера подвергать кипячению в течение 3 мин и затем выдерживать при температуре 65°C в течение 60–120 мин. Более короткие синтетические праймеры отжигали на матрице ДНК в течение 30–60 мин при температуре около 55–65°C. Позднее стал применяться подход с постепенным понижением температуры отжига посредством помещения пробирки с матрицей ДНК и специфическим праймером в стакан с водой (100–150 мл), нагретой до 65°C, давая ей возможность произвольно остыть до 30°C. Отжиг более коротких (8–12 мерных), а также сегментированных или модульных праймеров требует понижения температуры даже до 3–5°C.

Завершая рассмотрение различных олигонуклеотидных праймеров, следует упомянуть о предъявляемых требованиях к их качеству. Праймеры являются довольно критичным компонентом секвенирующих реакций. Как уже отмечалось выше, “читабельность” радиоавтографа геля или однозначность результатов, получаемых в автоматическом секвенаторе ДНК, будет зависеть от гомогенности 5'-конца праймера. Поскольку при химическом синтезе олигонуклеотидов, начинающегося с 3'-конца, эффективность каждой реакции присоединения очередного звена никогда не составляет 100%, то неочищенный продукт будет представлять собой смесь основного продукта и более коротких олигонуклеотидов. Легко подсчитать, что чем длиннее будет синтезируемый олигонуклеотид, тем ниже содержание основного продукта. В связи с этим очищенные праймеры будут давать более лучшие результаты, но очистка праймеров требует дополнительных усилий и значительно увеличивает их стоимость. Что касается коротких модульных праймеров (5-, 6-, 7- или 8-меров) то они, как правило, используются в неочищенном виде [Burbelo, Iadarola, 1994; Slightom et al., 1994; Kaczorowski, Szybalski, 1996a] по причине довольно высокого содержания основного продукта. Имеются специальные работы, посвященные определению возможности использования неочищенных праймеров как в ручном, так и в автоматическом режимах. Так, было показано, что при эффектив-

ности реакции присоединения при синтезе праймеров более чем 98% очистка праймеров становится необязательной [Sanchez-Pescador, Urdea, 1984]. Проведенное детальное исследование большого числа олигонуклеотидных праймеров длиной 25 и 50 звеньев, синтезированных на различных приборах разными операторами, позволило выявить интересные факты [Pon et al., 1996]. Так, было показано, что при эффективности реакции присоединения свыше 98,5% очистки даже таких протяженных праймеров (25-меров) для целей секвенирования не требуется. В то же время, несмотря на то, что производителями синтезаторов и реагентов для синтеза ДНК гарантируется эффективность присоединения не ниже 98%, отмечаются и более низкие эффективности синтеза, получающиеся как за счет низкого качества реагентов, проблем с самими инструментами, так и ошибок персонала, приводящие к получению праймеров с выходом основного продукта 70% и менее. Весьма удивительным оказалось то, что даже такие праймеры показали вполне приемлемые для чтения секвенирующего геля результаты [Pon et al., 1996]. Это можно объяснить только тем, что вклад в гетерогенность получаемых полос ДНК при использовании неудовлетворительных праймеров вносит тот олигонуклеотид, который короче, чем основной продукт, всего лишь на одно-два звена, тогда как другие, более короткие, возможно, диссоциируют с матрицы ДНК на этапе построения новой цепи.

Одним из способов выявления гетерогенности синтезированных нуклеотидов может служить их разделение в денатурирующем высокопроцентном (12–20%) полиакриламидном геле после мечения по 5'-концу радиоактивным фосфором с помощью полинуклеотидкиназы фага T4. Существует также возможность анализа праймеров без применения радиоактивной метки тонкослойной хроматографией на силикагеле и наблюдения их в ультрафиолетовом свете [Ansorge et al., 1996]. В исключительных случаях можно рекомендовать определение нуклеотидной последовательности самого праймера методом химической дегградации в модификации Данилюка и соавт. [Данилюк и др., 1986].

Автоматический синтез олигодезоксирибонуклеотидов в настоящее время осуществляется с помощью специальных приборов – ДНК-синтезаторов как фосфитамидным [Beaucage, Caruthers, 1981], так и Н-фосфонатным [Froehler, Matteucci, 1986; Froehler et al., 1986] методами, пришедшими на смену используемым ранее фосфотриэфирному методу [Gait et al., 1980] и еще более давнему фосфодизэфирному [Gait, Sheppard, 1977]. Оба современных метода легко масштабируемы и обеспечивают высококачественный скоростной синтез (суммарное усредненное время одного цикла синтеза составляет 40–45 с для Н-фосфонатного метода и от 3 до 4,5 мин для фосфитамидного). Выбор того или иного метода не диктуется синтезатором ДНК, поскольку доступные ныне из коммерческих источников приборы позволяют осуществлять синтез олигонуклеотидов любым из этих методов. Справедливости ради следует отметить, что Н-фосфонатный метод синтеза дает несколько большее число ошибок при синтезе, чем фосфитамидный. Прове-

денное секвенирование различных праймеров общим количеством 8937 нуклеотидов, синтезированных Н-фосфонатным методом, выявило 37 ошибок или в среднем 1 на 241 нуклеотид [Vasser et al., 1990]. Ошибки по типу распределились следующим образом: 20 делеций, 12 вставок и только 2 замены. Однако в данной работе осуществлялось клонирование и секвенирование преимущественно протяженных олигонуклеотидов, имеющих в длину до 85 звеньев, а, как известно, с увеличением длины олигомера эффективность и точность синтеза снижаются. В то же время при секвенировании двадцати клонов, содержащих два более коротких олигонуклеотида (31- и 36-меры), было выявлено значительное число ошибок, составившее 4 на 680. Схожая работа была проведена и с олигонуклеотидами, синтезированными фосфитамидным методом [McClain et al., 1986]. Для секвенированных олигонуклеотидов общим количеством 3152 звена было выявлено 10 ошибок, часть которых представляла собой инсерции, делеции и замены нуклеотидов, и последних было все же несколько больше, чем для праймеров, синтезированных Н-фосфонатным методом. Среднее число ошибок составило одну на 315 нуклеотидов. Также различные типы ошибок выявлены при секвенировании 249 звеньев с целью анализа точности синтеза данных олигонуклеотидов фосфитамидным методом [Hecker, Rill, 1998].

Многие приборы позволяют осуществлять различный масштаб синтеза олигонуклеотидов, варьирующий обычно от 0,05 до 5 или даже 10 мкмоль. Крупные масштабы синтеза применяются при синтезе праймеров для диагностических целей. Для секвенирования же ДНК масштаб синтеза в 0,05 мкмоль часто оказывается избыточным. В связи с этим был разработан мультиплексный ДНК-синтезатор, позволяющий в автоматическом режиме осуществлять одновременный синтез 96 независимых образцов в полипропиленовом микропланшете с масштабом синтеза всего 0,02 мкмоль или 20 нмоль, что позволило резко снизить стоимость получаемых олигонуклеотидов [Lashkari et al., 1995]. Одновременный синтез 10 независимых образцов позволяет производить EMBL ДНК-синтезатор, характеризующийся к тому же экономным расходом реактивов [Ansorge et al., 1996].

Видимо, здесь, в этом разделе, будет целесообразным упомянуть об относительно недорогом отечественном синтезаторе ДНК ASM-102U, выпускаемым ТОО "БИОССЕТ" (Новосибирск), который вполне способен обеспечить олигонуклеотидными праймерами для секвенирования ДНК и проведения ПЦР любую лабораторию. Однако в настоящее время весьма развит так называемый заказной синтез олигонуклеотидов, включая синтез модифицированных олигонуклеотидных праймеров, несущих какие-либо функциональные группы, что является дешевой альтернативой собственному синтезатору ДНК. Более подробную информацию о синтезаторе ДНК ASM-102U, а также о фирмах, занятых в заказном синтезе олигонуклеотидов, заинтересованный читатель может найти в разделе **Отечественные поставщики материалов и оборудования для секвенирования ДНК**.

В заключение данного раздела, видимо, стоит упомянуть о химерных олигонуклеотидах, так называемых PNA (Peptide Nucleic Acids), где

скелет такой молекулы образуется благодаря наличию N-(2-амино-этил)глициновых остатков и фосфодиэфирные связи между азотистыми основаниями таким образом заменены амидными. Синтезированные достаточно давно [Nielsen et al., 1991], они уже нашли весьма широкое применение в молекулярной биологии и биотехнологии [Buchardt et al., 1993; Corey, 1997; Uhlmann, 1998], поскольку оказались способными формировать двойную спираль с комплементарными им обычными олигонуклеотидами, образуя Уотсон–Криковские пары [Egholm et al., 1993; Wittung et al., 1994]. Однако до последнего времени в литературе не встречалось сообщений об использовании таких PNA в качестве затравочных молекул для ДНК-полимераз. В недавней работе американских авторов была показана принципиальная возможность использования подобных несколько модифицированных PNA в качестве праймера для разных полимераз, включая Кленовский фрагмент ДНК-полимеразы I [Misra et al., 1998]. Так, первые 19 азотистых оснований этой химеры являлись PNA, тогда как 3'-конец этой молекулы был представлен обычным динуклеотидом со свободной ОН-группой в 3'-положении дезоксирибозы. Удлинение данного праймера сопровождалось включением радиоактивно меченного дЦТФ, приводя к образованию меченых продуктов. Возможность применения подобных праймеров для практических целей в виде затравочных молекул для секвенирования ДНК выглядит несколько проблематично ввиду чрезвычайно прочного связывания PNA с ДНК-матрицей [Cherny et al., 1993]. Однако продолжающийся синтез различным образом модифицированных PNA [Efimov et al., 1998], может быть, со временем решит эту проблему. Некоторым преимуществом подобных затравочных молекул перед обычными олигонуклеотидными праймерами является их устойчивость к действию различных нуклеаз, что может позволить использовать химерные PNA-праймеры с различными ДНК-полимеразами, обладающими по крайней мере 5' → 3'-экзонуклеазной активностью. В то же время, исходя из позиций сегодняшнего дня, это не представляется очень актуальным.

## 2.4. ДНК-ПОЛИМЕРАЗЫ

Большинство ДНК-полимераз являются матрицезависимыми ферментами, т.е. синтезирующими вторую комплементарную цепь по уже существующей. Примером других может служить терминальная нуклеотидтрансфераза, осуществляющая синтез одноцепочечной ДНК произвольной последовательности путем последовательного присоединения нуклеотидов к 3'-концу<sup>3</sup>. (Собственно нематричная активность показана и для ряда других ДНК-полимераз (обратная транскриптаза, Таq ДНК-полимераза и др.), однако они присоединяют только по одному лишнему нуклеотиду на растущий 3'-конец [Clark, 1988].) Различают ДНК-зависимые и РНК-зависимые ДНК-полимеразы. Так, из группы

---

<sup>3</sup> Применение терминальной трансферазы в секвенировании ДНК ферментативным методом будет рассмотрено ниже, в разделе 2.7.

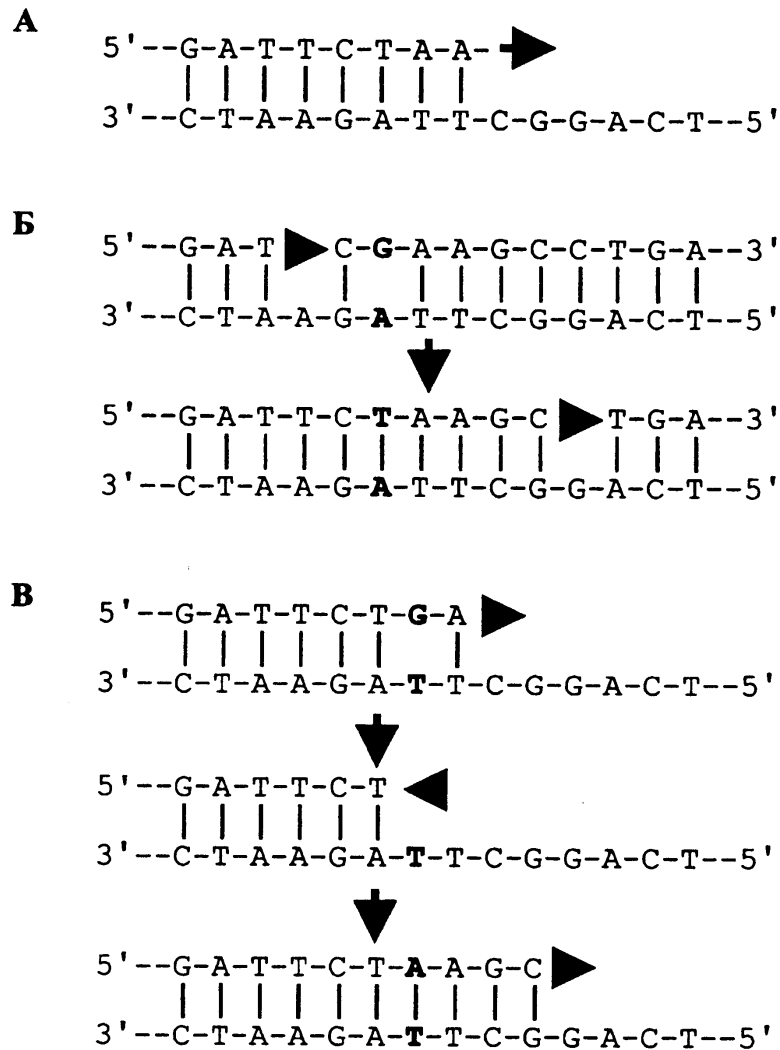


Рис. 2.17. Схематическое изображение ферментативного действия ДНК-полимеразы, обладающей 5' → 3'-полимеразной активностью (А), 3' → 5'-экзонуклеазной (редактирующей) активностью (В) и 5' → 3'-экзонуклеазной (репарирующей) активностью (Б)

ферментов, для которых матрицами являются цепи ДНК, в секвенировании применяются ДНК-полимераза *I. E. coli*, Кленовский фрагмент ДНК-полимеразы *I. E. coli*, ДНК-полимераза фага Т7, некоторые термостабильные ДНК-полимеразы и различные конструкции многих этих ферментов, полученные в том числе генно-инженерным путем. Полимеразы, синтезирующие цепь ДНК по матрице РНК, называются еще обратными транскриптазами или ревертазами и первоначально применялись для ферментативного секвенирования молекул РНК. Следует отметить, что пригодность какой-либо ДНК-полимеразы для целей секвенирования ДНК заключается в ее соответствии определенным требованиям, к ней предъявляемым, которые и будут рассмотрены ниже.

ДНК-полимеразы являются ферментами двойного действия, поскольку наряду с полимеразной активностью, выражающейся в последовательном присоединении нуклеотидов к растущей цепи ДНК в направлении 5' → 3', они катализируют реакцию пирофосфолиза (бо-

лее подробно рассмотренную при описании модифицированных T7 ДНК-полимераз), а также имеют еще и  $3' \rightarrow 5'$ -экзонуклеазную активность, заключающуюся в удалении присоединенных нуклеотидов и выполняющую редактирующие функции. Некоторые полимеразы имеют  $5' \rightarrow 3'$ -экзонуклеазную активность, призванную осуществлять в клетке репарирующие функции уже существующей цепи ДНК (рис. 2.17).

И  $5' \rightarrow 3'$ , и  $3' \rightarrow 5'$ -экзонуклеазные активности ДНК-полимераз мешают проведению секвенирующих реакций и крайне нежелательны. Так,  $3' \rightarrow 5'$ -экзонуклеазная редактирующая активность будет удалять только что включенные нуклеотиды, в том числе нуклеотиды с дидезоксирибозой (хотя и с меньшей скоростью), приводя в результате к неспецифической терминации растущих цепей ДНК и невозможности однозначного прочтения радиоавтографа геля.  $5' \rightarrow 3'$ -экзонуклеазная активность может оказывать даже более серьезное отрицательное влияние на результаты секвенирования ДНК, заключающееся в удалении нуклеотидов на  $5'$ -конце праймера, являющегося частью вновь синтезируемой цепи и приводя, таким образом, к образованию гетерогенных  $5'$ -концов разделяемых гель-электрофорезом фрагментов ДНК. Результатом будет та же нечитаемая картина полос секвенируемой ДНК или даже полное ее отсутствие при использовании в качестве метки олигонуклеотидного праймера, меченного по  $5'$ -концу.

Что касается непосредственно самой полимеразной активности, то к ее важным характеристикам, различающимся у разных ферментов, относятся скорость полимеризации и процессивность. Последняя выражается в количестве нуклеотидов, которые присоединит та же или иная ДНК-полимераза, прежде чем произойдет ее диссоциация с матрицы ДНК. Ни одна полимеразы не может строить цепь ДНК бесконечно и рано или поздно любая полимеразы диссоциирует с матрицы ДНК. Но при секвенировании ДНК, чем ниже будет процессивность используемого фермента, чем чаще будет прерываться синтез новой цепи и в результате возможного невозобновления его в том же месте будут образовываться неспецифически терминированные фрагменты. Скорость полимеризации измеряется в количестве присоединенных нуклеотидов за 1 с при оптимальной температуре для каждого фермента, и, хотя этот показатель при секвенировании ДНК и не является критическим, для синтеза новых цепей ДНК все же удобнее использовать ферменты, завершающие реакцию за непродолжительное время. Еще одной важной для секвенирования ДНК чертой ДНК-полимераз является способность того или иного фермента включать в растущую цепь модифицированные основания, как, например, дидезоксинуклеотидтрифосфаты.

Таким образом, некий идеальный фермент для секвенирования ДНК должен был бы быть, во-первых, относительно дешевым и при этом характеризоваться высокой полимеразной активностью, включая и процессивность и скорость полимеризации, отсутствием всяких экзонуклеазных активностей, не привередливостью к различным модифицированным аналогам нуклеотидов, а также устойчивостью к повышенным температурам. Забегая вперед, можно сказать, что и секвеназа (версия 2.0) и термосеквеназа во многом отвечают приведенным вы-

Таблица 2.1

*Ферментативные (эксонуклеазные) активности некоторых ДНК-полимераз*

Фермент	Эксонуклеазная активность	
	5' → 3'	3' → 5'
ДНК-полимераза I <i>E. coli</i>	+	+
Кленовский фрагмент ДНК-полимеразы I <i>E. coli</i>	-	+
Кленовский фрагмент ДНК-полимеразы I <i>E. coli</i> , мутантная форма, экзо <sup>-</sup>	-	-
AMV-ревертаза (обратная транскриптаза вируса миелобластома птиц)	-	-
MMLV-ревертаза (обратная транскриптаза Молони вируса лейкемии мышей)	-	-
Bst ДНК-полимераза <i>Bacillus stearothermophilus</i>	+	-
T7 ДНК-полимераза (нативная)	-	+
Секвеназа (версия 1.0) (модифицированная T7 ДНК-полимераза)	-	-
Секвеназа (версия 2.0) (модифицированная T7 ДНК-полимераза)	-	-
Taq ДНК-полимераза <i>Thermus aquaticus</i> (термостабильная)	+	-
Tth ДНК-полимераза <i>Thermus thermophilus</i> (термостабильная)	+	-
Tub ДНК-полимераза <i>Thermus flavus</i> (термостабильная)	+	-
Taq ДНК-полимераза <i>Thermus aquaticus</i> (модифицированная)	-	-
Tth ДНК-полимераза <i>Thermus thermophilus</i> (модифицированная)	-	-
Deep Vent <sub>R</sub> (exo <sup>-</sup> ) ДНК-полимераза <i>Ryococcus</i> sp. (модифицированная)	-	-
Термосеквеназа (мутантная форма Taq ДНК-полимеразы)	-	-
AmpliTaq FS ДНК-полимераза (мутантная форма Taq ДНК-полимеразы)	-	-

ше требованиям, различаясь по своим температурным оптимумам ферментативной активности, причем справедливости ради следует отметить, что для секвенирования ДНК далеко не всегда необходимо проведение реакций при повышенных температурах и даже некоторые стратегии секвенирования ДНК возможны только с использованием ферментов, работающих при комнатных или слегка повышенных (37 °С) температурах. Более того, как это будет видно из дальнейшего изложения, некоторые, казалось бы, нежелательные черты тех или иных ферментов в различных стратегиях секвенирования ДНК могут оборачиваться во благо.

В табл. 2.1 приведены отдельные сведения о свойствах некоторых ДНК-полимераз, наиболее широко применявшихся и применяющихся ныне для секвенирования ДНК. Кроме этих характеристик, определен-

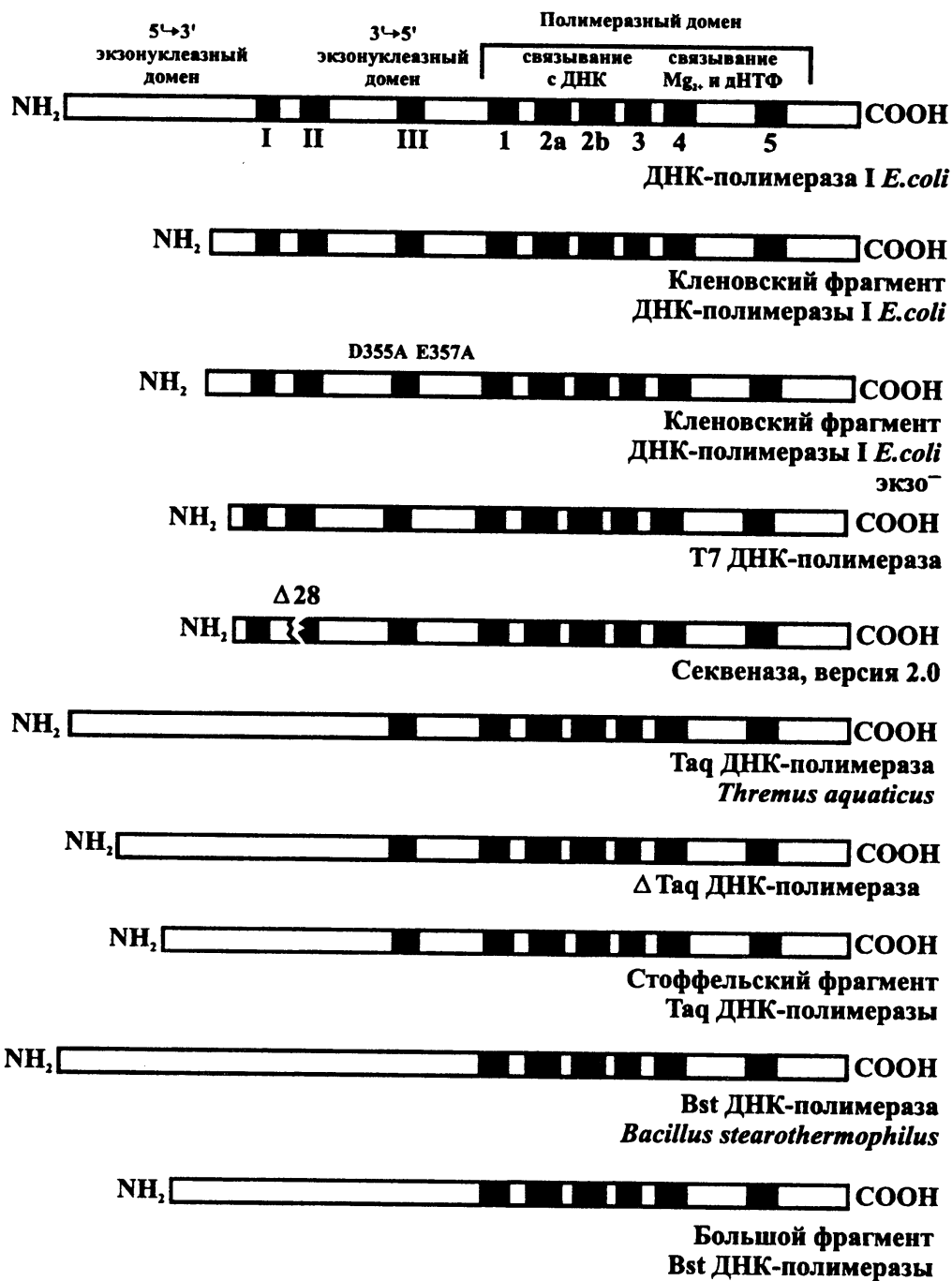


Рис. 2.18. Схемы организации различных ДНК полимераз PolI-семейства ферментов и их модифицированных форм, полученных генно-инженерным путем  
 Прямоугольниками показаны консервативные участки ДНК-полимераз, функционально важные в каталитическом отношении. Поскольку протяженность различных ферментов и их отдельных доменов неодинакова, то в целях упрощения схемы масштаб не соблюден

ное значение имеет и термостабильность фермента, более подробно рассмотренная в процессе дальнейшего изложения.

Несмотря на значительные различия упомянутых выше ДНК-зависимых ДНК-полимераз, все они относятся к так называемому PolI-семейству ферментов и имеют определенное сходство в организации белковой молекулы [Bernard et al., 1989; Blanco et al., 1990, 1992; Reha-Krantz, 1992; Astatke et al., 1995]. Аналогично ДНК-полимеразе I *E. coli*



все эти ферменты состоят из трех доменов, два из которых ответственны за экзонуклеазные активности, тогда как третий домен выполняет полимеразные функции (рис. 2.18). Сравнительный анализ аминокислотных последовательностей различных ДНК-полимераз позволил выявить несколько высококонсервативных участков в этих доменах [Blanco et al., 1991].

Как можно видеть из приведенных схем, наибольшее число консервативных участков локализовано в имеющемся у всех ферментов (что вполне естественно) домене, отвечающем за полимеразную активность. Некоторые из этих участков связываются с ДНК, а другие с дНТФ и ионами магния. Этому домену предшествует домен, выполняющий редактирующие функции в виде  $3' \rightarrow 5'$ -эксонуклеазной активности. У термостабильных ДНК-полимераз в этом домене имеется только один консервативный участок (Taq ДНК-полимераза) или они все отсутствуют (Bst ДНК-полимераза), вследствие чего оба этих фермента лишены подобной редактирующей активности.  $\text{NH}_2$ -терминальный домен у ДНК-полимеразы I несет репарирующую  $5' \rightarrow 3'$ -эксонуклеазную активность. Подобную активность проявляют обе упомянутые выше термостабильные ДНК-полимеразы, что, вероятно, объясняет обнаруженную у них некоторую гомологию этого участка с аналогичным у ДНК-полимеразы I. Что касается большинства ДНК-полимераз, модифицированных с целью их большей пригодности для ферментативного секвенирования ДНК либо ограниченным протеолизом, либо специально созданных генно-инженерным путем, то их объединяет общая черта, заключающаяся в удалении имеющейся у их предшественников  $5' \rightarrow 3'$ -эксонуклеазной активности. Путем делеции нескольких аминокислотных остатков во втором экзонуклеазном домене T7 ДНК-полимеразы была убрана имевшаяся  $3' \rightarrow 5'$ -эксонуклеазная активность. Сайтнаправленный мутагенез двух аминокислот в экзонуклеазном домене Кленовского фрагмента ДНК-полимеразы I привел к удалению  $3' \rightarrow 5'$ -эксонуклеазной активности этого фермента. Как можно представить из приведенных схем, столь разные структурные организации рассматриваемых ферментов не могли не сказаться на особенностях их ферментативного действия, чему и будет посвящено дальнейшее изложение.

ДНК-полимераза I *E. coli*, имеющая весьма низкую процессивность (10–30 нуклеотидов) и невысокую скорость полимеризации (45 нуклеотид/с) и обладающая, кроме полимеразной ( $5' \rightarrow 3'$ ), еще и  $5' \rightarrow 3'$ - и  $3' \rightarrow 5'$ -эксонуклеазными активностями, использовалась для секвенирования ДНК весьма ограниченно [Maat, Smith, 1978]. Некоторым преимуществом этого подхода с применением ДНК-полимеразы I в качестве секвенирующего фермента являлось отсутствие необходимости приготовления одноцепочечной матрицы ДНК, учитывая, что в то время (как это уже описывалось в разделе 2.2) эта процедура была весьма трудоемкой. В качестве затравочной молекулы, если так можно выразиться, служила имеющаяся вторая цепь ДНК. Принцип секвенирования ДНК с помощью нативного фермента ДНК-полимеразы I заключался в том, что после внесения ДНКазой I “ников” в секвенируемый фраг-

мент, меченный по 5'-концу с помощью полинуклеотидкиназы фага T4, 5' → 3'-экзонуклеазная активность полимеразы удаляла нуклеотиды на 5'-конце "ника", а полимеразная присоединяла их к его растущему 3'-концу и, поскольку в реакционной смеси присутствовали дидезоксинуклеотид трифосфаты, при их включении происходило ингибирование дальнейшего синтеза части "ник-транслируемых" фрагментов ДНК. Так что при секвенировании ДНК этим методом обычно мешающая 5' → 3'-экзонуклеазная активность фермента была просто необходима. Авторы утверждают, что значительной потери радиоактивной метки на 5'-конце секвенируемого фрагмента ДНК они не отмечали. Существовала модификация этого подхода, где специальное добавление ДНКазы I оказалось необязательным [Seif et al., 1980].

В ДНК-полимеразе I с помощью частичного протеолиза субтилизином была убрана 5' → 3'-экзонуклеазная активность и ее большой фрагмент, называемый Кленовским [Klenow et al., 1971], нашел широкое применение в секвенировании ДНК и на протяжении более чем десяти лет он оставался основным ферментом при проведении терминирующих реакций. К сожалению, в некоторых препаратах Кленовского фрагмента ДНК-полимеразы I иногда сохранялась незначительная 5' → 3'-экзонуклеазная активность и получаемая с помощью такого фермента картина полос ДНК на радиоавтографе секвенирующего геля часто бывала неудовлетворительной. Значительное улучшение качества секвенирования ДНК удалось добиться при использовании рекомбинантного фермента, представляющего собой клонированный фрагмент гена ДНК-полимеразы I, соответствующий Кленовскому, и заведомо не несущего 5' → 3'-экзонуклеазной активности [Joyce, Grindley, 1983]. Однако такой фермент, нарабатываемый с клонированного фрагмента гена в штамме-суперпродуценте, несмотря на свою более высокую степень чистоты и отсутствие 5' → 3'-экзонуклеазной активности, сохранил свои прочие прежние (не лучшие для секвенирования ДНК) свойства. Позднее с помощью сайтнаправленного мутагенеза была произведена замена двух аминокислотных остатков аспарагиновой кислоты в 355 положении и глутаминовой кислоты в 357 положении в обоих случаях на остатки аланина в экзонуклеазном домене Кленовского фрагмента ДНК-полимеразы I [Derbyshire et al., 1988]. Такой мутантный (D355A, E357A) фермент стал характеризоваться отсутствием оставшейся у него до этого 5' → 3'-экзонуклеазной активности и получил наименование Кленовский фрагмент, экзо-. Однако отсутствие у него редактирующей 5' → 3'-экзонуклеазной активности привело к повышенному уровню нематричного синтеза, заключающегося в добавлении на 3'-конец одного дополнительного выступающего нуклеотида [Clark et al., 1987]. Хотя и сам Кленовский фрагмент ДНК-полимеразы I обладает такой же, но менее ярко выраженной способностью [Clark, 1988]. Для целей секвенирования ДНК отсутствие обоих экзонуклеазных активностей может приводить только к улучшенным результатам, но в то же время при секвенировании коротких фрагментов ДНК необходимо учитывать возможность добавления дополнительного нуклеотида на их 3'-конец.

Как уже отмечалось выше, к отрицательным чертам ДНК-полимеразы, с помощью которой предполагается осуществлять секвенирование (кроме упомянутой экзонуклеазной активности), относятся следующие: низкая процессивность, малая скорость полимеризации, высокая избирательность по отношению к дНТФ и их аналогам, включаемым в среднем в тысячу раз хуже природных. Все эти черты в полной мере относятся к Кленовскому фрагменту ДНК-полимеразы I, получаемому как путем протеолиза субтилизином, так и нарабатываемому с помощью штамма-суперпродуцента. Дополнительным отрицательным качеством этого фермента является его невысокая термостабильность и, как следствие, невозможность достоверного секвенирования GC-богатых участков с ярко выраженной вторичной структурой.

Справедливости ради следует отметить, что относительно недавно с помощью сайтнаправленного мутагенеза проведена модификация гена ДНК-полимеразы I *E. coli*, вернее Кленовского фрагмента, путем замены аминокислоты фенилаланин в 762 положении (нумерация приведена для нативного фермента) на тирозин, что позволило резко снизить избирательность такого генно-инженерного фермента по отношению к включаемым нуклеотидным аналогам [Tabor, Richardson, 1995]. Ранее также было показано, что присутствие в реакционной смеси ионов  $Mn^{2+}$  вместо  $Mg^{2+}$  в 100 раз уменьшает избирательность Кленовского фрагмента ДНК-полимеразы I по отношению к различным дНТФ и их модифицированным аналогам [Tabor, Richardson, 1989a]. Однако в настоящее время для Кленовского фрагмента ДНК-полимеразы I как секвенирующего фермента это уже не является актуальным и имеет скорее теоретический интерес. Необходимо признать, что в последние годы Кленовский фрагмент ДНК-полимеразы I находит весьма ограниченное применение в секвенировании ДНК, поскольку и обнаружены и созданы новые ДНК-полимеразы, превосходящие его по всем параметрам. Недавняя работа Ричардсона и соавт., выявивших тиоредоксинсвязывающий домен, протяженностью 76 аминокислот, у T7 ДНК-полимеразы и создавших на его основе новый рекомбинантный фермент Klenow-TBD – Кленовский фрагмент ДНК-полимеразы I с добавленным в него данным доменом, показала заметное увеличение процессивности этой химерной ДНК полимеразы [Bedford et al., 1997]. Однако, как отмечают сами авторы, теоретически гораздо больший эффект можно ожидать от добавления этого тиоредоксинсвязывающего домена в термостабильную Taq ДНК-полимеразу или ее производные, поскольку, хорошо известно, что тиоредоксин способствует весьма заметному увеличению процессивности T7 ДНК-полимеразы [Tabor et al., 1987] и в связи с этим можно ожидать подобного эффекта и для других ДНК-полимераз, относящихся к этому семейству PolI-ферментов. Это, кстати, подтверждают и результаты по созданию химерного фермента Klenow-TBD с данным доменом. Но при этом важным условием является сохранение термостабильности этих ферментов после введения в них дополнительного домена.

Другим ферментом для секвенирования ДНК являются обратные транскриптазы, первоначально использовавшиеся для ферментативно-

го секвенирования молекул РНК методом терминации строящейся цепи ДНК с помощью дидезоксинуклеотидтрифосфатов [McGeoch, Turnbull, 1978; Zimmermann, Kaesberg, 1978; Vina-Stein et al., 1979]. Поскольку фермент способен к использованию в качестве матрицы и цепи ДНК, то позднее он стал применяться и для секвенирования некоторых матриц ДНК с гомополимерными G/C-участками, которые были трудноразрешимыми с помощью Кленовского фрагмента ДНК-полимеразы I [Mierendorf, Pfeffer, 1987]. Секвенирование протяженных нуклеотидных участков из одних цитозиновых остатков также лучше удавалось с помощью обратных транскриптаз. Обратная транскриптаза Молони вируса лейкемии мышей была менее пригодна для этих целей, нежели обратная транскриптаза вируса миелобластоза птиц [Mierendorf, Pfeffer, 1987]. Для обратных транскриптаз, в целом, характерны те же недостатки, что и для Кленовского фрагмента ДНК-полимеразы I. Как то, низкая термостабильность, даже еще более низкая скорость элонгации, чем у Кленовского фрагмента ДНК-полимеразы I. Однако процессивность обратной транскриптазы вируса миелобластоза птиц несколько выше и составляет несколько сотен нуклеотидов при скорости полимеризации всего 5 нуклеотид/с. Что касается избирательности включения в растущую цепь ДНК нуклеотидных аналогов, то ревертазы включают их лишь в десять раз хуже истинных немодифицированных нуклеотидов.

Введение в практику секвенирования модифицированных T7 ДНК-полимераз (известных также под названием секвеназа, версия 1.0 и секвеназа, версия 2.0), являющихся химически и генетически измененными T7 ДНК-полимеразами соответственно [Tabor, Richardson, 1987a; Tabor, Richardson, 1989b], заметно изменило сложившиеся представления о производительности метода ферментативного секвенирования ДНК и окончательно оттеснило на второй план метод секвенирования ДНК путем химической дегградации по Максаму–Гилберту. Этому, однако, предшествовали серьезные исследования фермента T7 ДНК-полимеразы и составляющих ее белков.

Известно, что данный фермент вырабатывается в клетках *E. coli* после инфицирования ее бактериофагом T7<sup>4</sup>. Собственно сама полимеразная активность, да и 3' → 5'-экзонуклеазная активность фермента заключена в продукте гена 5 бактериофага (80 кДа), но процессивность его крайне низка, зависит от многих факторов и составляет не более 50 нуклеотидов, однако образуемый очень прочный комплекс с маленьким (около 12 кДа) белком тиоредоксином, кодируемый бактериальным геномом, резко увеличивает этот показатель [Tabor et al., 1987]. Что касается самого белка тиоредоксина, то он не имеет ни ДНК-полимеразной, ни экзонуклеазной активностей и только стабилизирует комплекс T7 ДНК-полимеразы и матрицы ДНК с затравкой [Huber et al., 1987]. Есть сведения о том, что тиоредоксин защищает связываемые им

---

<sup>4</sup> Под T7 ДНК-полимеразой, в дальнейшем изложении, а также под уже упоминавшейся выше, подразумевается комплекс двух полипептидов, представляющих собой белковый продукт гена 5 бактериофага T7 и продукт гена *trxA* *E. coli*.

регионы T7 ДНК-полимеразы от протеолиза [Yang, Richardson, 1996]. Проведенное детальное исследование “ошибок”, совершаемых T7 ДНК-полимеразой отдельно и в комплексе с тиоредоксином, свидетельствует о серьезном влиянии этого белка на свойства самой полимеразы [Kunkel et al., 1994]. Была показана зависимость скорости полимеризации фермента от наличия SSB-белка, из которой видно, что при добавлении данного белка скорость элонгации возрастает от 110 до 750 нуклеотид/с. [Tabor et al., 1987]. Другим важным свойством фермента явилась его “неразборчивость” по отношению к включаемым нуклеотидам. Так, в отличие от ревертаз и Кленовского фрагмента ДНК-полимеразы I T7 ДНК-полимераза включала модифицированные нуклеотиды (дидезоксинуклеотидтрифосфаты и прочие аналоги) с эффективностью лишь в два раза меньшей, чем немодифицированные дНТФ [Tabor, Richardson, 1987a].

Считается, что в связи с высокой 3' → 5'-экзонуклеазной активностью (значительно превосходящей таковую Кленовского фрагмента ДНК-полимеразы I) нативная T7 ДНК-полимераза не может использоваться для проведения секвенирующих реакций. Однако в литературе встречаются сведения об успешном использовании этого фермента для целей секвенирования [Kristensen et al., 1988]. Кроме этого, шведская фирма Pharmacia на протяжении ряда лет поставляла особым образом очищенную нативную T7 ДНК-полимеразу, пригодную для секвенирования ДНК ферментативным методом. Помимо высокой очистки фермента, значительную роль в его пригодности для проведения терминирующих реакций, по-видимому, играет состав реакционной смеси с достаточно высокими концентрациями всех дНТФ.

Несмотря на различие в скорости протекания реакции полимеризации и экзонуклеазного отщепления присоединенных нуклеотидов, все же позволяющее использовать нативный фермент T7 ДНК-полимеразу для секвенирования ДНК, удаление экзонуклеазной активности сделало фермент еще более пригодным для этих целей. Первая модифицированная T7 ДНК-полимераза, называемая секвеназой, версия 1.0 была создана путем избирательного окисления ряда аминокислот (часть из которых, вероятно, гистидины) в экзонуклеазном домене фермента под действием молекулярного кислорода, солей железа и редуцирующего агента [Tabor, Richardson, 1987b]. В результате такого селективного окисления экзонуклеазная активность модифицированной T7 ДНК-полимеразы сохранилась на уровне лишь 1,3% от таковой у нативного фермента [Tabor, Richardson, 1989b]. Модифицированная таким образом T7 ДНК-полимераза характеризовалась такой же высокой процессивностью, что и нативный фермент, но при этом скорость полимеризации, измеренная без SSB-белка, выросла в три раза и составила около 300 нуклеотид/с. [Tabor, Richardson, 1987a]. Благодаря всем этим замечательным свойствам секвеназы, резко повысилась производительность метода ферментативного секвенирования ДНК. Такими составляющими повышения производительности явились эффективное построение новой цепи ДНК (свыше 2000 нуклеотидов против 200–500 для Кленовского фрагмента ДНК-полимеразы I или обратных транс-

криптаз), выходявшее далеко за рамки возможности их разделения обычным секвенирующим электрофорезом; значительно меньшее количество “стопов” ввиду высокой процессивности фермента; легкое “безошибочное” “чтение” радиоавтографов секвенирующих гелей благодаря достаточно равномерному включению всех дидезоксинуклеотидтрифосфатов (в отличие от Кленовского фрагмента ДНК-полимеразы I, для которого были разработаны даже специальные правила по чтению отдельных слабых полос на радиоавтографе). Использование секвеназы заставило пересмотреть некоторые стратегии секвенирования протяженных фрагментов ДНК, рассчитанные на получение субклонов с фрагментами ДНК определенной длины в сторону заметного увеличения последней.

Все же остаточная экзонуклеазная активность, присущая химически модифицированной T7 ДНК-полимеразе, в отдельных случаях затрудняла секвенирование и в связи с этим были приложены значительные усилия по ее практически полному удалению [Tabor, Richardson, 1989b]. С этой целью был проведен сайтнонаправленный мутагенез клонированного гена белка 5 бактериофага T7 с помощью целого ряда олигонуклеотидных праймеров. Полученные при этом результаты свидетельствуют, что множественные замены отдельных аминокислот (преимущественно, гистидинов, аргининов и лизинов) в экзонуклеазном домене фермента снижают уровень экзонуклеазной активности до 0,001% от таковой нативной T7 ДНК-полимеразы. Однако еще большим снижением экзонуклеазной активности (<10<sup>-6</sup>%) характеризовалась конструкция с делетированным участком гена белка 5 длиной 84 пн или 28 аминокислотных остатков, обозначения как Δ28 (Lys<sup>118</sup>-Arg<sup>145</sup>) [Tabor, Richardson, 1989b]. Дополнительным основанием для введения в практику секвенирования именно этого варианта генетически модифицированной T7 ДНК-полимеразы, широко известной сейчас под названием секвеназа, версия 2.0, явилась ее наиболее высокая полимеразная активность среди прочих исследуемых генно-инженерных конструкций и превосходящая таковую нативного фермента в 9 раз.

Продолжающиеся исследования особенностей ферментативного действия генетически модифицированной T7 ДНК-полимеразы позволили добиться эффекта уменьшения избирательности фермента по отношению к нуклеотидным аналогам. Было показано, что присутствие в реакционной смеси ионов Mn<sup>2+</sup> вместо Mg<sup>2+</sup> в среднем в 4 раза уменьшает избирательность включения тех или иных модифицированных оснований в растущую цепь ДНК [Tabor, Richardson, 1989a]. В то же время ионы Mn<sup>2+</sup> несколько уменьшали процессивность модифицированной T7 ДНК-полимеразы, что, впрочем, было легко преодолеть, весьма незначительно меняя состав реакционной среды. В цитируемой работе сравнительный анализ высоты пиков, соответствующих флуоресцентно меченым фрагментам ДНК, выявляемым при их секвенировании в автоматическом секвенаторе, убедительно продемонстрировал еще более равномерное включение в новую цепь ДНК всех четырех дидезоксинуклеотидтрифосфатов в присутствии ионов Mn<sup>2+</sup>. В этих условиях разброс высоты соседних пиков не превышал 20%, что оказалось наи-

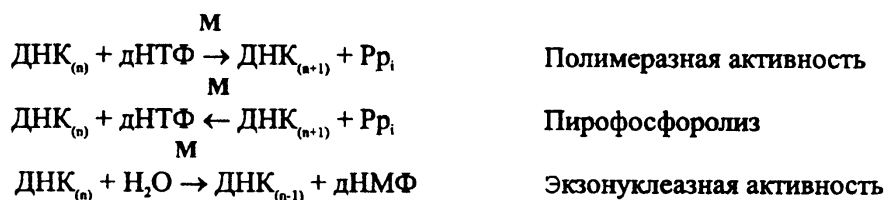


Рис. 2.19. Упрощенная схема полимеразной активности, пирофосфоролиза и экзонуклеазной деградации, происходящих в присутствии двухвалентных катионов ( $\text{Mg}^{2+}$  или  $\text{Mn}^{2+}$ ), обозначенных как М

более важным для автоматического секвенирования ДНК. Исходя из такой равномерной высоты пиков было предложено при секвенировании геномной (не клонированной) ДНК детектировать гетерозиготы, в которых бы высота какого-либо пика была бы в два раза меньше остальных [Tabor, Richardson, 1990].

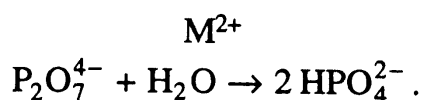
Поскольку в присутствии ионов марганца секвеназа включала в растущую цепь ДНК дНМФ и ддНМФ в приблизительно равном соотношении, это послужило основой для разделения продуктов всех 4 терминирующих реакций в одной дорожке секвенирующего геля в автоматическом секвенаторе ДНК. Ступенчатое изменение в разных смесях концентрации соответствующих ддНТФ (как например 3 мкМ для ддГТФ; 1,5 мкМ для ддАТФ; 0,75 мкМ для ддТТФ и 0,37 мкМ для ддЦТФ) приводило в результате электрофоретического разделения к получению пиков разной высоты, которые укладывались в 4 группы по числу ддНТФ. Сопоставляя высоты пиков, можно было с высокой достоверностью восстановить нуклеотидную последовательность секвенируемого участка ДНК [Tabor, Richardson, 1990].

Отмечая эффект ослабления отдельных полос ДНК и даже полное их исчезновение на радиоавтографе геля при секвенировании секвеназой (не имеющей экзонуклеазной активности), Табор и Ричардсон впервые указали на возможный вклад пирофосфоролиза в этот процесс [Tabor, Richardson, 1990]. Известно, что реакция пирофосфоролиза, протекающая в присутствии неорганического пирофосфата представляет реакцию обратную полимеризации. В свою очередь сам пирофосфат образуется при реакции полимеризации в процессе присоединения дНТФ путем превращения последнего в дНМФ и отщепления от него пирофосфата. Пирофосфоролиз находится в равновесии, когда концентрация присутствующих в реакционной смеси дНТФ достаточно высока, чтобы поддерживать скорость синтеза равной скорости образования новых трифосфатов. Хотя и реакции пирофосфоролиза и экзонуклеазной деградации приводят к одинаковому эффекту – удалению оснований на 3'-конце растущей цепи ДНК, они принципиально отличаются по субстратной зависимости и по конечному продукту. Так, как видно из упрощенной схемы (рис. 2.19) протекания пирофосфоролиза и экзонуклеазной деградации, происходящих в присутствии двухвалентных катионов ( $\text{Mg}^{2+}$  или  $\text{Mn}^{2+}$ ), пирофосфоролиз приводит к образованию вновь одного дНТФ и укорачивает растущую цепь ДНК на один нуклеотид, а под действием экзонуклеазы также укорачивается цепь ДНК, но при этом образуется дНМФ. В случае, если в данном примере пос-





мые соотношения этих ингредиентов. Так, ввиду высокой скорости элонгации новой цепи ДНК под действием секвеназы секвенирующие реакции, включая этап мечения, вполне могут быть завершены в 5–6 мин. Что касается возможного увеличения концентрации дНТФ и ддНТФ в реакционной смеси, то это приведет к их неэкономному расходованию и потому нежелательно. Другим способом исключения пирофосфоролиза является удаление из реакционной смеси образующегося пирофосфата, что достаточно просто сделать, добавив в реакцию еще один фермент – неорганическую пирофосфатазу. Как видно из приведенной ниже формулы, этот фермент в присутствии двухвалентных катионов (магния или марганца) катализирует превращение молекулы пирофосфата в две молекулы ортофосфорной кислоты



Так, было показано, что добавление незначительных количеств дрожжевой неорганической пирофосфатазы в реакционную смесь при проведении терминирующих реакций полностью исключает пирофосфоролиз [Ruan et al., 1990; Tabor, Richardson, 1990]. При этом надо иметь в виду, что неорганическая пирофосфатаза должна быть высоко очищена и не содержать каких-либо возможных примесей нуклеаз.

Таким образом, генетически модифицированная T7 ДНК-полимераза или секвеназа (версия 2.0) является практически идеальным ферментом для секвенирования ДНК с помощью дидезокситерминаторов, что подтверждается просто огромным количеством работ, определение нуклеотидных последовательностей в процессе выполнения которых осуществлено именно с помощью этого фермента.

Пожалуй, единственным недостатком генетически модифицированной T7 ДНК-полимеразы является ее низкая термостабильность. Причем, этот недостаток становится заметен лишь при секвенировании GC-богатых последовательностей за счет формирования участков с сильной вторичной структурой. Для преодоления этого недостатка встречались рекомендации увеличить концентрацию фермента и несколько повысить температуру инкубации (до 45 °С) при проведении терминирующих реакций. И то и другое в отдельных случаях способствовало нормальному “прохождению” ферментом таких сложных участков. Однако участки с сохраняющейся при этой температуре вторичной структурой были неразрешимы. Становилось очевидным, что для определения нуклеотидных последовательностей с высоким содержанием GC-пар необходимо использование термостабильных ферментов.

Одним из таких достаточно широко применяемых ферментов была термостабильная ДНК-полимераза, выделенная из почвенной бактерии *Bacillus stearothermophilus*. Полученный путем расщепления субтилизинном ее большой фрагмент, аналогичный Кленовскому, с успехом использовался при секвенировании ДНК с сильной вторичной структурой, поскольку довольно высокая температура инкубации (65 °С) способствовала расплетанию возникающих шпилечных структур [Ye,

Hong, 1987]. Этими же авторами показано успешное использование данного фермента при секвенировании противоположной цепи ДНК посредством построения на первом этапе комплементарной без участия терминаторов [Lu et al., 1992], как это ранее было предложено для Кленовского фрагмента ДНК-полимеразы I [Hong, 1981]. Довольно высокая скорость полимеризации этого фермента, составляющая около 120 нуклеотид/с позволяла проводить мечение и терминацию в два этапа [McClary et al., 1991]. Не менее важной оказалась способность Bst-полимеразы (ее большого фрагмента) использовать при построении новой цепи модифицированные аналоги, такие как дИТФ и 7-деаза-дГТФ, что уменьшало компрессию полос ДНК при их разделении секвенирующим гель-электрофорезом [McClary et al., 1991]. К другим важным свойствам этого фермента можно отнести то, что он достаточно стабилен при хранении при комнатной температуре и после 14 сут при 25 °С Bst ДНК полимеразы образывала картины полос секвенируемой ДНК высокого качества [Lu et al., 1991]. Также было показано, что Bst-полимераза может храниться в высушенном виде вместе с остальными ингредиентами для ферментативного секвенирования ДНК в микротитраторных планшетах при -20 °С в течение недели и по крайней мере в течение ночи при комнатной температуре, что делало этот фермент пригодным для массовых анализов при помощи лабораторного робота [Earley et al., 1993]. Фермент оказался пригоден и для секвенирования нанограммовых количеств ДНК матрицы при использовании флуоресцентной метки [Mead et al., 1991].

Произведенные замены двух аминокислотных остатков тирозина в 5' → 3'-экзонуклеазном домене клонированного гена Bst ДНК-полимеразы на фенилаланин и аланин привели к удалению этой экзонуклеазной активности фермента, как и в случае субтилизинового протеолиза, и измененная таким образом Bst ДНК-полимераза была пригодна для проведения секвенирующих реакций [Riggs et al., 1996]. Замены двух остатков глицина на аспарагиновую кислоту у схожего фермента Bca ДНК-полимеразы привели к аналогичному эффекту [Ishino et al., 1995]. Здесь, видимо, следует отметить, что в настоящее время эта ДНК-полимераза из термофильного микроорганизма *Bacillus caldotenax*, ген который ранее был клонирован [Uemori et al., 1993], также используется в секвенировании ДНК и поставляется фирмой "PanVera Corporation" (США).

Анализ нуклеотидной последовательности гена Bst ДНК-полимеразы показал, что все три высококонсервативных участка, типичных для ДНК-полимераз, относящихся к PolI-семейству ферментов, в ее 3' → 5'-экзонуклеазном домене отсутствуют [Aliotta et al., 1996]. Проведенные биохимические тесты на наличие экзонуклеаз у нативной Bst ДНК-полимеразы и ее большого фрагмента показали, что только полноразмерный фермент содержал 5' → 3'-экзонуклеазную активность [Aliotta et al., 1996]. Авторы делают вывод, что редактирующая 3' → 5'-экзонуклеазная активность отсутствует у ряда ДНК-полимераз (и в том числе у термостабильной ДНК-полимеразы, выделенной из бактерии *Thermus aquaticus*, к рассмотрению которой мы как раз переходим), как бы в обмен на их повышенную термостабильность.

Термостабильная Taq ДНК-полимераза, выделенная из термофильной бактерии *Thermus aquaticus*, имеет температурный оптимум ферментативного действия около 72 °С и при этом сохраняет свою активность при 95 °С довольно продолжительное время. Значительная процессивность фермента (7600 нуклеотидов) и высокая скорость полимеризации (около 60 нуклеотид/с) делали Taq ДНК-полимеразу достаточно удобным ферментом для секвенирования ДНК, особенно ДНК с высоким содержанием GC-пар и сильной вторичной структурой. Обстоятельное исследование пригодности Taq ДНК-полимеразы для секвенирования ДНК и особенностей действия фермента были проведены Иннисом и соавт. [Innis et al., 1988]. К сразу замеченным недостаткам фермента можно было отнести его довольно высокую избирательность включения нуклеотидных аналогов, в связи с чем в реакционной смеси требовались весьма высокие концентрации последних. В то же время Taq ДНК-полимераза была способна включать 7-деза-дГТФ, что способствовало исключению компрессии при разделении продуктов секвенирующих реакций гель-электрофорезом. Авторами было отмечено наличие ошибочно включаемых Taq ДНК-полимеразой нуклеотидов и даже неспецифической терминации в отсутствие ддНТФ при низкой концентрации какого-либо из дНТФ. Так, в результате проведения данной работы Иннисом и соавт. была показана принципиальная возможность секвенирования ДНК с помощью Taq ДНК-полимеразы как клонированных фрагментов ДНК, так и получаемых путем асимметричной ПЦР. Практически одновременно с цитируемой работой Петерсон [Peterson, 1988] опубликовал краткое сообщение о применении Taq ДНК-полимеразы и ее преимуществе по сравнению с секвеназой при секвенировании GC-богатых участков.

Кроме Taq ДНК-полимеразы, поставляемой многими фирмами, для секвенирования ДНК стали применять и другие термостабильные ферменты, выделенные из бактерий рода *Thermus* – *T. thermophilus* и *T. flavus* [Bechtereva et al., 1989; Rao, Saunders, 1992]. Последняя поставляется фирмой “Amersham International plc” (Англия), но уже под торговым именем Tub DNA Polymerase. Для Tth ДНК-полимеразы определена скорость полимеризации, составляющая около 25 нуклеотид/с [Carballeira et al., 1990]. Некоторые из этих ферментов до сих пор выделяются из природных штаммов, но, учитывая весьма высокую температуру выращивания, были клонированы гены некоторых термостабильных ДНК-полимераз *T. aquaticus*, *T. thermophilus*, *T. flavus*, *T. caldophilus*, [Lawyer et al., 1989; Akhmetzjanov, Vakhitov, 1992; Kwon et al., 1997]. Из других термостабильных ДНК-полимераз, не относящихся к PolI-семейству, для циклического секвенирования использовалась модифицированная Deep Vent<sub>R</sub><sup>®</sup>(exo<sup>-</sup>) ДНК-полимераза [Mariame, 1996], обнаруженная у археобактерии *Pyrococcus* sp. из глубоководного термального источника и поставляемая фирмой “New England Biolabs, Inc.”, а также некоторые другие.

Несмотря на доступность из коммерческих источников целого ряда термостабильных ДНК-полимераз бактерий рода *Thermus*, характеризующихся при этом близкими свойствами, все же из наиболее широко применяемым ферментом является Taq ДНК-полимераза как по исто-

рическим причинам, так и в силу своей большей изученности. Первое время при секвенировании ДНК использовали нативную Taq ДНК-полимеразу, характеризующуюся отсутствием 3' → 5'-экзонуклеазы, но имеющую 5' → 3'-экзонуклеазную активность. Несмотря на присутствующую экзонуклеазную активность качество секвенирования ДНК таким ферментом было вполне удовлетворительным и лишь незначительное двоение полос давало о себе знать. Поскольку 5' → 3'-экзонуклеазная активность у ферментов, относящихся к PolI-семейству, достаточно легко удаляется, то были созданы многочисленные специальные генно-инженерные конструкции Taq ДНК-полимеразы с делетированными участками экзонуклеазного домена, приходящегося на NH<sub>2</sub>-конец фермента. После удаления первых 235 аминокислот образовался фермент, названный Klen Taq ДНК-полимеразой и известный еще как ΔTaq ДНК-полимераза, для которого характерно отсутствие заметной экзонуклеазной активности [Barnes, 1992; Samols, Fuller, 1995]. Удаление 288 аминокислот с NH<sub>2</sub>-конца фермента Taq ДНК-полимеразы привело к аналогичному результату [Lawyer et al., 1993]. Такая укороченная Taq ДНК-полимераза или ее большой фрагмент длиной 544 аминокислотных остатка получила название Стоффельского фрагмента по имени одного из ее создателей. При этом, сохранив свои прежние свойства (кроме, естественно, 5' → 3'-экзонуклеазной активности), фермент стал даже более термостабилен. Так, если нативная Taq ДНК-полимераза имела время полужизни всего 9 мин при температуре 97,5 °С, то ее Стоффельский фрагмент мог выдерживать уже 21 мин при тех же условиях [Lawyer et al., 1993]. Еще более укороченная Stof delta 12 Taq ДНК-полимераза показала лишь незначительное ухудшение термостабильности, тогда как делеция в 47 аминокислот (288 + 47 = 335) привела к значительной потере как полимеразной активности фермента, так и его термостабильности [Vainshtein et al., 1996]. Другие авторы удалили еще больший фрагмент Taq ДНК-полимеразы размером 413 аминокислотных остатков и при этом получили фермент с заметно ухудшенными свойствами [Villbrandt et al., 1997]. Инактивация 5' → 3'-экзонуклеазной активности Taq ДНК-полимеразы была осуществлена и путем сайтнаправленного мутагенеза отдельных нуклеотидов, приведших к замене некоторых аминокислот в экзонуклеазном домене [Merkens et al., 1995]. Так, мутантные формы R24C и R74H Taq ДНК-полимеразы сохранили лишь незначительную экзонуклеазную активность, но термостабильность фермента и его процессивность превосходят таковые Стоффельского фрагмента Taq ДНК-полимеразы. Фирма "Pro-mega" (США) производит среди прочих термостабильных ДНК-полимераз два варианта Taq ДНК-полимеразы (с молекулярными массами 80 кДа и 85 кДа), выделяемых из природного источника, однако один фермент с меньшей молекулярной массой характеризуется очень низкой 5' → 3'-экзонуклеазной активностью. В данном случае, скорее всего за счет несколько отличающихся условий выделения того и другого фермента, имеет место ограниченный протеолиз, хотя, как подчеркивают авторы, природа этой модификации до конца не ясна [Bugke et al., 1992]. В последнее время, кроме модифицированной Taq ДНК-по-

T7 ДНК-полимераза	518	RDNAKTFI <u>Y</u> GFLYG	531
<i>E.coli</i> ДНК-полимераза I	754	RRSAKAIN <u>F</u> GLIYG	767
Taq ДНК-полимераза	659	RRAAKTIN <u>F</u> GVLYG	672

Рис. 2.21. Сравнение углеводсвязывающего участка различных ДНК-полимераз. Нумерация указана для полноразмерных белков. Аминокислоты приведены в однобуквенном коде. Выделены критичные аминокислоты-тирозин (Y) и фенилаланин (F)

лимеразы, для секвенирования ДНК довольно широко используется и модифицированная Tth ДНК-полимераза, также характеризующаяся отсутствием 5' → 3'-экзонуклеазной активности [Ikeda et al., 1995; Aikawa et al., 1996]. Показано преимущество данного модифицированного фермента над его диким предшественником в реакции циклического секвенирования с мечеными биотином дидезоксинуклеотидтрифосфатами.

Однако все эти термостабильные ДНК-полимеразы, характеризующиеся отсутствием экзонуклеазной активности и пригодные для секвенирования как обычным способом, так и в циклической реакции, сохраняли все же один существенный недостаток – высокую избирательность по отношению к включаемым нуклеотидным аналогам и, в частности к дидезоксинуклеотидтрифосфатам. Вследствие этого в реакционную смесь необходимо было вносить весьма высокие концентрации ддНТФ, что, во-первых, было относительно дорого и, во-вторых, заметно увеличивало фон при использовании флуоресцентно меченных ддНТФ. В то же время T7 ДНК-полимераза, как уже упоминалось выше, значительно менее требовательна к включению нуклеотидных аналогов, но не обладает нужной высокой термостабильностью. Поскольку задача по превращению нетермостабильного белка T7 ДНК-полимеразы в термостабильный является практически нерешаемой, Табор и Ричардсон задались целью выяснить, что лежит в основе подобной как избирательности, так и неизбирательности разных ферментов по отношению к дидезоксипроизводным [Tabor, Richardson, 1995]. Для этого ими была проведена огромная работа по выявлению сначала участков белковых молекул, а затем и отдельных аминокислот, влияющих на этот процесс. Так, были получены многочисленные гибридные конструкции T7 ДНК-полимераз с отдельными участками ДНК-полимеразы I *E. coli* и наоборот. В результате проведенных исследований по ингибированию роста цепи с помощью ддТТФ полученными гибридными конструкциями ДНК-полимераз удалось выявить, что относительно консервативный для многих ДНК-полимераз участок, взаимодействующий с углеводным компонентом нуклеотидов, содержит у T7 ДНК-полимеразы аминокислоту тирозин, тогда как для ДНК-полимеразы I *E. coli* и Taq ДНК-полимеразы в этом месте находится близкий тирозину фенилаланин и именно эти аминокислоты ответственны за распознавание углеводного компонента включаемого нуклеотида (рис. 2.21).

Предложенная Табором и Ричардсоном гипотетическая схема взаимодействия дНТФ и белка с некоторыми изменениями приведена на

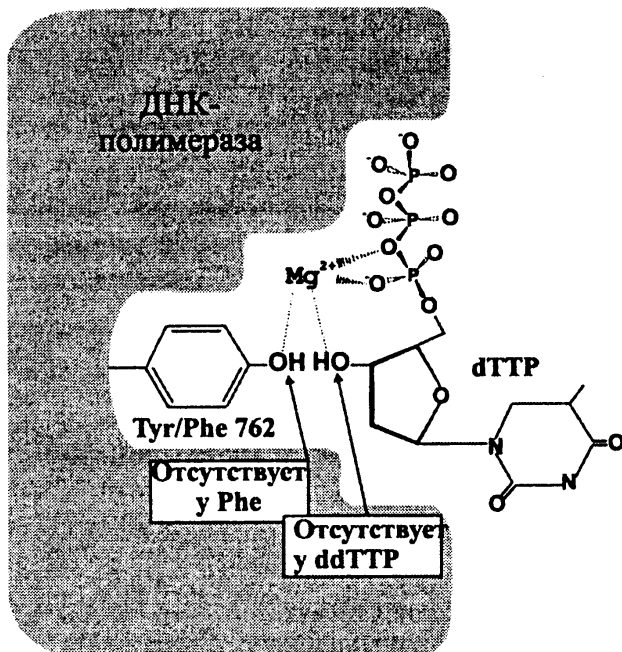


Рис. 2.22. Предположительная схема взаимодействия углеводного компонента (дезокси- и дидезокси-) нуклеотидтрифосфатов и участка ДНК-полимеразы I *E. coli* и T7 ДНК-полимеразы

рис. 2.22, из которой видно, что гидроксильные остатки тирозина и дНТФ взаимодействуют с ионом  $Mg^{2+}$  и, таким образом, обеспечивают большее сродство фермента и субстрата. В отсутствие одного или обоих гидроксильных групп процесс узнавания, видимо, проходит гораздо хуже и, как следствие, имеет

место более редкое включение в растущую цепь ДНК дидезоксипроизводных.

Созданные в процессе этой работы измененные ферменты T7 ДНК-полимеразы (Y526F), ДНК-полимеразы I *E. coli* (F762Y), Taq ДНК-полимеразы (F667Y) коренным образом изменили свои свойства касательно избирательного включения ддНТФ. (Хотя нумерация здесь приведена для полноразмерных белков, в действительности в работе были использованы ранее клонированные гены:  $\Delta 28$  T7 ДНК-полимераза с делецией 28 аминокислот или секвенза 2.0 [Tabor, Richardson, 1989b]; Кленовский фрагмент ДНК-полимеразы I *E. coli* [Joyce, Gridley, 1983]; Стоффельский фрагмент Taq ДНК-полимеразы с отсутствием 288 аминокислот [Lawyer et al., 1993].

Как можно видеть из табл. 2.2, секвенза с присущим ей тирозином демонстрирует в среднем лишь трехкратную дискриминацию при включении дидезоксипроизводных. Для Y526F-секвензы после произведенной замены тирозина на фенилаланин это соотношение резко возросло и даже превысило таковые для неизмененных Кленовского фрагмента ДНК полимеразы I *E. coli* и Стоффельского фрагмента Taq ДНК-полимеразы, изначально характеризующихся низким уровнем включения ддНМФ. В то же время Кленовский фрагмент F762Y и Стоффельский фрагмент F667Y, несущие тирозин вместо фенилаланина, стали включать дидезоксинуклеотиды более эффективно, чем сама секвенза.

Наиболее важным результатом этой работы в практическом отношении можно считать создание на основе Стоффельского фрагмента Taq ДНК-полимеразы нового фермента термостабильной ДНК-полимеразы с улучшенными свойствами благодаря высокой эффективности включения в растущую цепь ДНК дидезоксинуклеотидмонофосфатов [Tabor, Richardson, 1995]. Данный фермент получил коммерческое название ТермоСеквенза и в настоящее время поставляется фирмой "Amersham Pharmacia Biotech" [Reeve, Fuller, 1995; Samols et al., 1995].

Таблица 2.2

*Эффективность включения отдельных нуклеотидов  
и их дидезоксипроизводных различными ДНК-полимеразами*

ДНК-полимераза	Соотношение эффективности включения отдельных дезокси- и дидезоксинуклеотидов				
	dA/ddA	dC/ddC	dG/ddG	dT/ddT	dN/ddN
Секвеназа, версия 2.0	3,3	3,7	3,2	2,8	3,0
Секвеназа Y526F	7300	11000	6400	8400	8000
Кленовский фрагмент	720	250	140	1100	600
Кленовский фрагмент F762Y	0,72	0,75	0,56	0,54	0,6
Стоффельский фрагмент	4700	2600	1400	4500	3000
Стоффельский фрагмент F667Y	0,59	0,32	0,45	0,56	0,5

Равномерное включение всех ддНМФ термосеквеназой оказалось наиболее важным для циклического секвенирования ДНК, поскольку использовавшиеся ранее различные варианты Taq- или Tth-полимераз приводили к формированию весьма различающихся пиков флуоресценции при их детекции в автоматическом секвенаторе ДНК, что заметно затрудняло распознавание секвенируемой нуклеотидной последовательности. Впоследствии семейство Taq F667Y термостабильных ДНК-полимераз с улучшенными свойствами несколько увеличилось за счет появления фермента AmpliTaq, FS [Parker et al., 1996] и Taquenase [Voss et al., 1997].

Можно было бы предположить, что создан идеальный фермент для секвенирования ДНК, однако, несмотря на все свои преимущества, термосеквеназа приобрела и важный недостаток, выражающийся в высокой активности пирофосфоролиза. Увеличенная активность пирофосфоролиза термосеквеназы вполне объяснима и даже могла прогнозироваться, поскольку ДНК-полимеразы, являясь ферментами двойного действия, катализируют включение дНМФ/ддНМФ или их отщепление при прочих равных условиях со скоростью, зависящей от своего сродства к субстрату. Так, известно, что Кленовский фрагмент включает в растущую цепь ДНК, например, ддТМФ приблизительно в 1000 раз хуже, чем его природный аналог дТМФ. Но в этом случае и пирофосфоролиз ддТМФ будет таким же весьма редким событием. Поэтому при секвенировании ДНК с использованием Кленовского фрагмента ДНК-полимеразы *E. coli* и Стоффельского фрагмента Taq ДНК-полимеразы отрицательный вклад пирофосфоролиза не столь заметен и обычно не принимается во внимание. Однако уже при секвенировании ДНК с помощью секвеназы, характеризующейся лишь трехкратной дискриминацией в отношении дидезоксипроизводных, эффект пирофосфоролиза дает себя знать особенно при продолжительной инкубации реакционной смеси. Для термосеквеназы же, обладающей еще меньшей избирательностью по отношению к ддНТФ, вклад пирофосфоролиза может

быть столь заметен, что необходимо прилагать усилия по удалению из реакционной смеси образующегося пирофосфата. Поскольку оптимум температуры для термосеквенызы около 70 °С и она применяется обычно в циклическом секвенировании ДНК со сменой температур, то использование для удаления пирофосфата обычной дрожжевой пирофосфатазы становится невозможным и тогда для этой цели было предложено использовать термостабильную пирофосфатазу [Reeve, Fuller, 1995; Samols et al., 1995; Vander Horn et al., 1997]. Такой термостабильной пирофосфатазой является фермент, обнаруженный ранее у археобактерии *Thermoplasma acidophilum* и клонированный в *E. coli* [Richter, Schafer, 1992a, b].

Завершая рассмотрение различных ДНК-полимераз, находящихся применение в секвенировании ДНК, следует отметить, что, несмотря на создание термостабильных ДНК-полимераз с улучшенными свойствами, использование секвенызы, работающей при 37 °С, не потеряло своей актуальности и эти две системы удачно дополняют друг друга [Ranu, 1995].

## 2.5. МЕЧЕНИЕ НОВОСИНТЕЗИРУЕМОЙ ЦЕПИ ДНК

За некоторыми исключениями электрофоретическому разделению подвергаются меченые продукты терминирующих реакций. В качестве меченых молекул могут выступать различные соединения. Так, в первые годы секвенирования ДНК ферментативным методом для мечения использовались исключительно дНТФ, несущие  $^{32}\text{P}$  в  $\alpha$ -положении. Высокая энергия  $\beta$ -частиц данного радионуклида способствовала непродолжительной экспозиции геля на рентгеновскую пленку (что более детально рассмотрено в главе 5. **Радиоавтография**), но такая высокая энергия  $^{32}\text{P}$  не в полной мере отвечала всем требованиям и не способствовала увеличению разрешающей способности метода секвенирования ДНК из-за невозможности различения отдельных полос в верхней части радиоавтографа секвенирующего геля. Короткий период полураспада этого радионуклида также ставил экспериментаторов в жесткие рамки. Для преодоления этих недостатков было предложено использовать дНТФ, меченные радиоактивным изотопом  $^{35}\text{S}$ , обладающим меньшей энергией  $\beta$ -частиц (рис. 2.23) [Biggin et al., 1983]. Относительно большой период полураспада данного радионуклида, достигающий почти 3 мес, и низкая энергия его  $\beta$ -частиц оказались предпочтительнее.

Еще более удобным оказалось использование изотопа  $^{33}\text{P}$  с периодом полураспада немногим менее месяца, максимальная энергия  $\beta$ -частиц (0,249 МэВ) которого несколько выше, чем у  $^{35}\text{S}$ , и при этом достаточно для относительно короткого времени экспозиции и дает такие же четкие полосы ДНК в верхней части геля, как и в случае с  $^{35}\text{S}$  [Краев и др., 1990; Zagursky et al., 1991; Barnard et al., 1994]. Весьма важным преимуществом радионуклида  $^{33}\text{P}$  является то, что благодаря его невысокой энергии, обращение с ним гораздо менее опасно, чем с  $^{32}\text{P}$  и не требует применения свинцовых контейнеров или специальных толстостенных



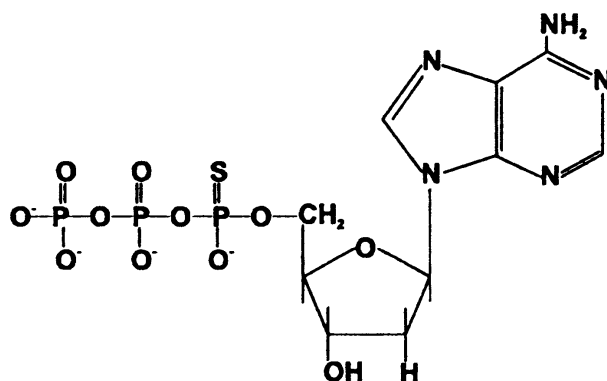


Рис. 2.23. Структурная формула тиопроизводного дНТФ

экранов из органического стекла. В то время как для безопасной работы с радионуклидом  $^{32}\text{P}$  при проведении молекулярно-биологических экспериментов даже предложены различные защитные приспособления из оргстекла, включая специальную прямоугольную прозрачную насадку на автоматическую пипетку [Merrifield, 1987]. Низкая атомная масса компонентов органического стекла не приводит к возникновению весьма опасного тормозного  $\gamma$ -излучения, при этом эффективно задерживая  $\beta$ -частицы  $^{32}\text{P}$ , и, таким образом, экраны и прочие конструкции из оргстекла заметно уменьшают вредное воздействие радионуклида  $^{32}\text{P}$ .

Нельзя не отметить заметно улучшившиеся за эти годы секвенирования ДНК транспортные формы радиоактивных предшественников для мечения нуклеиновых кислот. Так, раньше радиоизотопные препараты поставлялись для лучшей их сохранности в 50%-ном этаноле, что вынуждало исследователей тратить значительные усилия на испарение спирта и подвергать себя облучению в течение лишнего времени. Теперь данные радиопрепараты поставляются, как правило, в трициновом буфере в высококонцентрированном виде. Для большего удобства исследователей при обращении с малыми объемами радиоактивного вещества в раствор стали добавлять специальные инертные красители, не мешающие протеканию ферментативных реакций. Претерпели изменения и сами внутренние и внешние транспортные упаковки препаратов, ранее поставлявшихся только в сухом льду. На протяжении уже ряда лет фирма "Amersham" отказалась от поставок в сухом льду, тем не менее ее радиоизотопные препараты нареканий не вызывают.

В настоящее время существует три основных подхода к введению радиоактивной метки в растущую цепь ДНК, показанных на рис. 2.24. Первый классический подход [Sanger et al., 1977a] заключается в проведении терминирующих реакций с дидезоксинуклеотидтрифосфатами и одновременном включении радиоактивной метки в виде одного из четырех дНТФ, присутствующего в реакционной смеси наряду с его назначенным вариантом. К недостаткам этого способа можно отнести конкуренцию между "горячим" и "холодным" дНТФ и подбор соответствующих концентраций и того и другого предшественника для достижения оптимального результата и, как следствие, необходимость присутствия в реакционной смеси довольно значительного количества радиоактив-

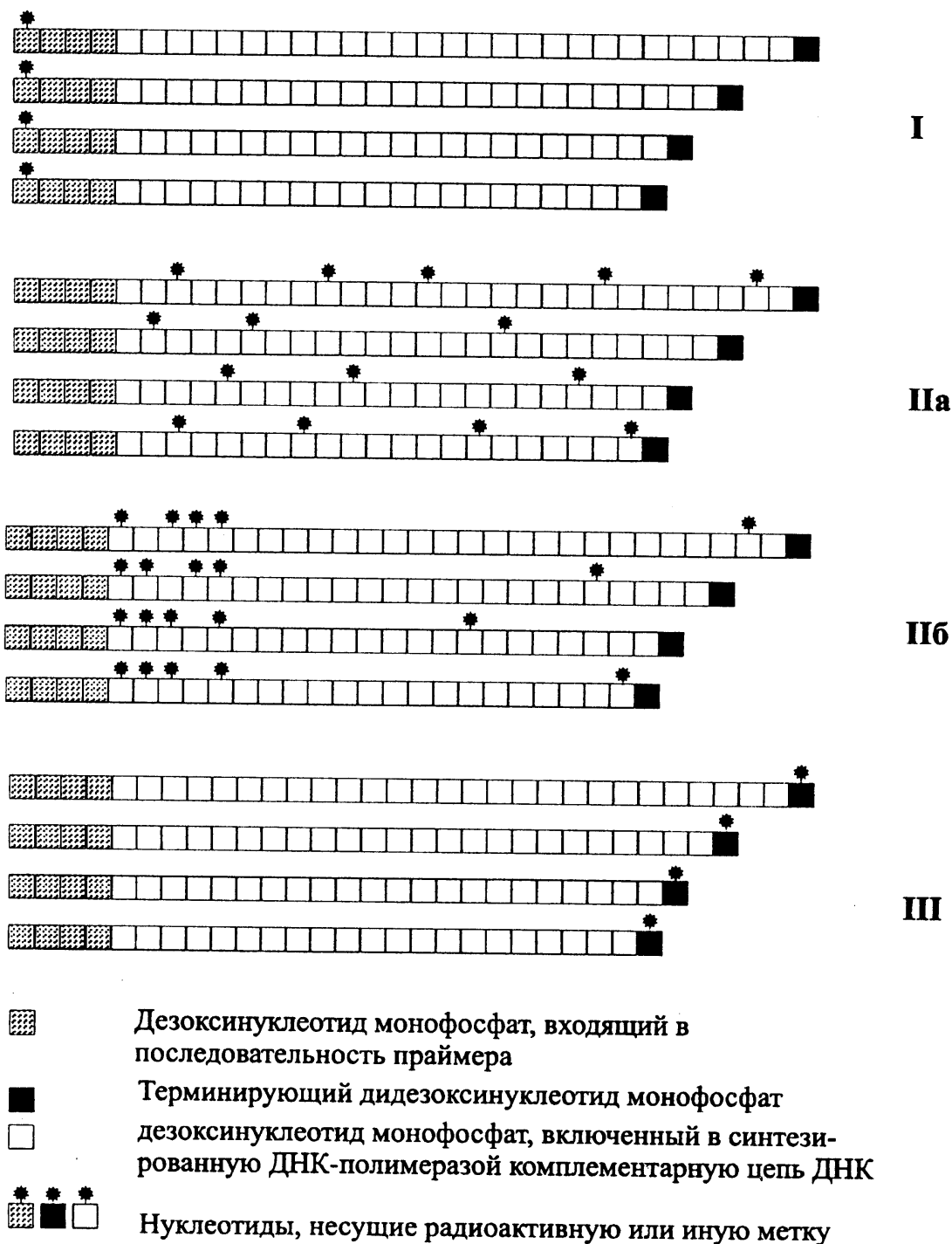


Рис. 2.24. Различные подходы к мечению новосинтезируемой цепи ДНК

I – концевое мечение (5'-конца новосинтезируемой цепи ДНК) посредством использования олигонуклеотидного праймера, предварительного меченного по 5'-концу с помощью  $[\gamma\text{-}^{32}\text{P}]$  АТФ и полинуклеотидкиназы фага T4; II(a) – внутреннее мечение с помощью присутствующего в терминирующей смеси меченного  $[\alpha\text{-}^{32}\text{P}]$  дНТФ

II(б) – квазиконцевое мечение (около 5'-конца) растущей цепи ДНК с помощью присутствующего в смеси для мечения меченного  $[\alpha\text{-}^{32}\text{P}]$  дНТФ, достигаемое путем разделения этапов мечения и терминации;

III – концевое мечение (3'-конца) новосинтезируемой цепи ДНК посредством включения  $[\alpha\text{-}^{33}\text{P}]$  ддНТФ

(Использование в качестве метки того или иного радионуклида в каждом рассмотренном случае не принципиально и он может быть заменен на любую другую приемлемую метку, например флуоресцентную.)

ного дНТФ, достигающего 20 мкКи и более. Другим отрицательным моментом являлось то, что более протяженные фрагменты несли большее количество радиоактивной метки, чем короткие, и результатом была разная интенсивность полос ДНК на радиоавтографе в нижней и верхней частях одного и того же геля и, как следствие, невозможность достоверного прочтения иногда части нуклеотидной последовательности. Это заставляло в подобных случаях прибегать к повторному этапу радиоавтографии соответственно с увеличенным или уменьшенным временем экспозиции геля на рентгеновскую пленку. (Разная интенсивность полос в нижней и верхней частях радиоавтографа зависит и от многих других причин, рассматриваемых в следующем разделе.)

Этих недостатков практически полностью был лишен предложенный Табором и Ричардсоном вариант данного подхода, заключающийся в разделении этапов мечения и терминирования [Tabor, Richardson, 1987a]. В этом случае праймер, отождествленный на матрице ДНК, удлинялся высокопроцессивной ДНК-полимеразой секвеназой в присутствии очень низких концентраций трех немеченных дНТФ и четвертого меченного в течение короткого промежутка времени и не при оптимальной температуре инкубации для фермента. Таким образом, происходило лишь весьма незначительное удлинение праймера и за счет включения во все молекулы ДНК приблизительно равного количества меченых нуклеотидов достигалась равная интенсивность полос по всей длине геля. (В то же время интенсивность полос в нижней и верхней частях геля зависит, главным образом, от соотношения дНТФ/ддНТФ в терминирующей реакции, что будет детально рассмотрено в следующем разделе.) Однако такое “квазиконцевое” мечение имеет свои особенности. Так, при подобном мечении большое значение приобретает уже не соотношение “горячего” и “холодного” дНТФ, а концентрация трех немеченных дНТФ, присутствующих в реакционной смеси или, еще точнее, соотношение их концентраций с общим количеством матрицы ДНК с “отождествленным” на ней праймером. Ввиду высокой скорости полимеризации секвеназы при достаточных концентрациях дНТФ цепь ДНК может удлиниться на весьма значительное расстояние и в связи с этим первые нуклеотиды уже невозможно секвенировать и поэтому требуется уменьшенная концентрация дНТФ. В то же время, если в эксперименте предполагается осуществлять “чтение” фрагмента ДНК в диапазоне от 300 до 800 нуклеотидов, то значительное удлинение и одновременно мечение цепи ДНК даже обоснованно. Выделение мечения новосинтезируемой цепи ДНК в отдельный этап обычно характерно при секвенировании секвеназой, Таq ДНК-полимеразой или Bst ДНК-полимеразой, то есть ферментами с высокой скоростью полимеризации, однако была показана принципиальная возможность использования такого способа и для Кленовского фрагмента ДНК-полимеразы I [Краев, Миронов, 1990].

Возможность получения квазиконцевых меченых фрагментов ДНК в циклическом секвенировании ДНК путем ограниченной элонгации праймера Таq ДНК-полимеразой также была показана [Samols, Fuller, 1995]. Поскольку секвенируемая молекула ДНК в реакционной

смеси при циклическом секвенировании представлена обычно в весьма низкой концентрации, то даже незначительные концентрации всех четырех дНТФ приведут к образованию в первых же циклах мечения довольно протяженных молекул, что совсем нежелательно. В связи с этим на этапе мечения в реакционную смесь добавляется всего 3 дНТФ, один из которых несет метку. При этом необходимо знать хотя бы небольшой участок нуклеотидной последовательности, следующей за местом отжига праймера, что позволит выбрать три наиболее подходящие дНТФ (и в том числе меченый), а также исключить определенный дНТФ. По завершению 30 циклов, в результате которых происходит незначительная (9 нуклеотидов в цитируемой работе) элонгация праймера, он оказывается меченым и после разделения на четыре соответствующие пробирки, содержащие конкретный ддНТФ, может быть готов к дальнейшему использованию в качестве затравочной молекулы в циклическом секвенировании. Подобный подход был осуществлен и ранее, однако в реакционную смесь добавлялся только меченый нуклеотид, что приводило к удлинению праймера всего на одно звено и он из 20-членного становился 21-членным [Kim, Kim, 1994]. Подобное мечение/удлинение праймера в цитируемой работе проводилось всего в течение 8 циклов и затем добавлялись соответствующие количества дНТФ/ддНТФ.

Независимо от используемой ДНК-полимеразы на этапе терминации уже меченых фрагментов ДНК в реакционной смеси также продолжает присутствовать тот же радиоактивно меченый предшественник и при этом происходит, хотя и со значительно меньшей эффективностью (из-за существенного уменьшения концентрации данного "горячего" дНТФ по сравнению с его немеченым вариантом), случайное включение в строящуюся цепь ДНК радиоактивного дНМФ. Это дает основание считать такое квазиконцевое мечение растущей цепи ДНК не самостоятельным подходом к мечению цепи ДНК, а лишь вариантом внутреннего мечения. В то же время в некоторых обзорах мечение новосинтезируемой цепи ДНК этим способом считают как самостоятельный подход к мечению, что, впрочем, лишено особых оснований, поскольку при этом способе сохраняется, как уже отмечалось выше, теоретическая возможность включения радиоактивной метки и на этапе терминации на всем протяжении строящейся цепи ДНК.

Второй самостоятельный подход к мечению новосинтезируемой цепи ДНК состоит в использовании в качестве затравки уже меченого олигонуклеотидного праймера, что также приводит к одинаковой интенсивности мечения всех фрагментов ДНК, не зависящее от их длины. Детальное исследование особенностей применения для секвенирования ДНК праймеров, меченных по 5'-концу радиоактивным фосфором с помощью  $\gamma$ -АТФ и полинуклеотидкиназы, проведено МакГроу [McGraw, 1984]. Хотя в литературе встречаются и более ранние упоминания об использовании при ферментативном секвенировании ДНК меченых праймеров [Hong, 1982]. К недостаткам этого способа можно отнести наличие лишь единичного атома радионуклида в каждом секвенируемом фрагменте ДНК, тогда как в предыдущих вариантах мечения коли-

чество включаемой метки было в несколько раз выше. Как следствие, меченные таким способом секвенируемые фрагменты ДНК требовали, как правило, большее время экспозиции. Другим недостатком можно считать необходимость проведения дополнительного предварительного этапа мечения с использованием другого фермента Т4 полинуклеотидкиназы, тогда как при мечении описанными выше способами все происходило в одной реакционной пробирке с использованием одного и того же фермента ДНК-полимеразы как для мечения, так и для дальнейшего построения цепи ДНК в условиях специфического терминирования. Некоторым положительным моментом такого мечения по 5'-концу является то, что при имеющем место радиолизе происходит разрыв цепи ДНК в участке, расположенном преимущественно поблизости от включенного радионуклида, а поскольку радиоактивная метка при таком способе мечения локализована только на 5'-конце цепи ДНК, то, следовательно, будут образовываться достаточно короткие фрагменты ДНК, продолжающие нести радиоактивную метку на своем 5'-конце. Результатом этого будет некий повышенный фон только в нижней части геля, тогда как при использовании квазиконцевого и/или внутреннего мечения могут образовываться радиоактивно меченные неспецифические обрывки новосинтезированных цепей ДНК довольно крупного размера, что неминуемо приведет к увеличенному фону на протяжении значительной части геля. (Отрицательный вклад радиолиза можно избежать, разделяя секвенирующим гель-электрофорезом свежеполученные продукты терминирующих реакций и не прибегая к их длительному хранению. В то же время следует отметить, что препараты ДНК, меченные более низкоэнергетическими радионуклидами  $^{33}\text{P}$  или  $^{35}\text{S}$ , могут храниться значительно дольше, чем таковые после их мечения  $^{32}\text{P}$ .)

Все эти описанные выше способы мечения новосинтезируемой цепи ДНК объединяет то, что они имеют один общий существенный недостаток, заключающийся в том, что на радиоавтографе геля детектируются все радиоактивные фрагменты, будь то специфически терминированные молекулы ДНК соответствующими ддНМФ или являющиеся артефактами построения комплементарной цепи ДНК. Известно, что наряду с терминацией строящейся комплементарной цепи, происходящей при включении дидезоксипроизводных нуклеотидов, среди продуктов терминирующих реакций всегда присутствуют фрагменты ДНК, не несущие на своем 3'-конце соответствующего ддНМФ. Последнее происходит по целому ряду причин, главными из которых могут считаться низкое качество матрицы ДНК, вызванное ее плохой очисткой, либо сильная вторичная структура какого-либо участка секвенируемой ДНК. ДНК-полимераза не может эффективно "пройти" эти участки ДНК и в результате на заключительном радиоавтографе секвенирующего геля в этих местах образуются "нечитаемые" полосы, называемые "стопами" или BAFs (Bands in All Four Lanes), которые присутствуют в нескольких или сразу во всех четырех дорожках. Для того чтобы избежать эти нежелательные события, разработан целый ряд всевозможных улучшений проведения терминирующих реакций, которые подробно будут рассматриваться в следующем разделе. Однако оказа-

лось, что нежелательный эффект выявления все же возникающих “стопов”, а также исключение просто высокого фона, который может иметь место за счет неспецифической терминации, затрудняющего “чтение” или делающего его невозможным, можно избежать и на стадии мечения благодаря недавно появившемуся третьему подходу к получению меченых фрагментов ДНК.

Заметного повышения качества ферментативного секвенирования ДНК ручным методом удалось добиться за счет способа мечения растущей цепи ДНК с помощью ддНТФ, несущих радиоактивную метку  $^{33}\text{P}$  в остатке фосфорной кислоты, расположенной в  $\alpha$ -положении [Fan et al., 1996; Gimmert et al., 1997]. Так, в случае произошедшего неспецифического терминирования растущей комплементарной цепи ДНК на ее 3'-конце будет находиться какой-либо нерадиоактивный дНМФ. В случае использования любого из описанных выше способов мечения данный секвенируемый фрагмент ДНК все равно будет нести метку и на заключительном радиоавтографе ему будет принадлежать соответствующая полоса. Однако, если в качестве метки используются  $[\alpha\text{-}^{33}\text{P}]$  ддНТФ, то неспецифически терминированный фрагмент ДНК, не несущий на своем 3'-конце соответствующего  $[\alpha\text{-}^{33}\text{P}]$  ддНМФ, на радиоавтографе выявлен не будет.

Как и в случае с использованием меченого праймера, каждая молекула секвенируемого фрагмента ДНК будет нести только единственную метку в виде ддНМФ на своем 3'-конце. Из этого следует, во-первых, что время экспозиции такого секвенирующего геля с мечеными дидезокситерминаторами будет также несколько увеличено, во-вторых, количество радиоактивной метки, присутствующей в реакционной смеси, может быть уменьшено до всего 1 мкКи (против 20 или 5 мкКи для  $[\alpha\text{-}^{33}\text{P}]$  дНТФ, необходимых для их включения в строящуюся цепь ДНК разными описанными выше способами), что позволяет обращаться со значительно меньшим количеством радиоактивного материала.

Обычно перед нанесением продуктов секвенирующих реакций на полиакриламидный гель не требуется удаления невключившейся радиоактивной метки, ввиду того, что благодаря своему относительно малой молекулярной массе  $\gamma$ -АТФ,  $\alpha$ -дНТФ или  $\alpha$ -ддНТФ будут двигаться значительно быстрее фрагментов ДНК и на заключительном радиоавтографе будут не видны. Однако в случае использования флуоресцентно меченных ддНТФ повышение уровня фона может быть заметным и рекомендуется удаление непрореагировавшей метки. Был предложен довольно простой и быстрый способ удаления метки с помощью смолы Wizard, выпускаемой фирмой “Promega” (США) [Kaczorowski, Sektas, 1996].

В литературе встречается упоминание об использовании в секвенировании ДНК в качестве метки стабильного изотопа  $^{57}\text{Fe}$  [Jacobson et al., 1991]. Авторы специально синтезировали органический компаунд ферроцен, несущий изотоп  $^{57}\text{Fe}$  и присоединили его к 5'-концу олигонуклеотидного праймера. В модельном эксперименте с помощью радиоактивной квазиконцевой метки, включаемой секвеназой, они продемонстрировали принципиальную возможность построения комплементарной

цепи ДНК с такого  $^{57}\text{Fe}$ -содержащего праймера и при этом подвижность таких железосодержащих фрагментов ДНК замедлялась совсем незначительно. По мнению авторов цитируемой статьи, гораздо больший интерес в качестве метки для секвенирования ДНК представляет олово, поскольку оно имеет сразу 10 стабильных изотопов. Получение большинства из них в высокоочищенном виде позволит проводить мультиплексное секвенирование, поскольку каждый изотоп может детектироваться независимо. Однако серьезным ограничением для широкого использования данного подхода секвенирования ДНК с помощью стабильных изотопов является необходимость детекции, разделенных гель-электрофорезом полос ДНК с помощью дорогостоящего оборудования, каковым является обладающий очень высокой чувствительностью резонансный ионизационный масс-спектрометр. Учитывая высокую скорость анализа, авторы цитируемой статьи подсчитали, что данный подход теоретически может обеспечить "чтение"  $1,44 \times 10^7$  полос секвенируемой ДНК в сутки. Дальнейшее развитие подхода в секвенировании ДНК с использованием стабильных изотопов олова получило в другой работе этих авторов [Arlinghaus et al., 1997], более подробно рассматриваемой в главе, посвященной методу секвенирования ДНК гибридизацией.

Желание экспериментаторов хоть как-то оградить себя от вредного воздействия радиоактивности повлекло за собой выявление полос ДНК как колориметрическими, так и хемилюминесцентными методами. Однако эффективное проведение таких реакций возможно только с фрагментами ДНК, сорбированными на поверхности какого-нибудь твердого носителя, а не находящимися внутри полиакриламидного геля. Чисто технические трудности переноса фрагментов ДНК из секвенирующего полиакриламидного геля на нейлоновые, нитроцеллюлозные или какие-то другие мембранные фильтры заметно сдерживали развитие этого подхода. Разработка специальных приборов для секвенирующего электрофореза с одновременным переносом выходящих фрагментов ДНК из нижнего среза геля на находящийся в соприкосновении с ним движущийся мембранный фильтр заметно упростила процедуру переноса (более подробно такие приборы вместе с особенностями переноса фрагментов ДНК, а также другие способы переноса рассмотрены в главе, посвященной электрофоретическому разделению секвенируемых фрагментов ДНК).

Для выявления перенесенных из полиакриламидного геля на мембранный фильтр полос ДНК колориметрическим методом с помощью красителей нитротетразолиевого голубого и 5-бромо-4-хлоро-3-индолфосфата необходимо введение особой метки в строящуюся комплементарную цепь ДНК. Наиболее широкое применение нашли олигонуклеотидные праймеры, несущие на своем 5'-конце молекулу биотина или дигоксигенина [Beck, 1987; Pohl, Maier, 1995]. Для получения внутренне меченных молекул ДНК обычно используется дУТФ, меченный биотином или дигоксигенином [Pohl, Maier, 1995].

Биотин (витамин Н) образует прочные комплексы с авидином или стрептавидином и находит благодаря этому широкое применение в мо-

лекулярной биологии [Wilchek, Bayer, 1988]. Дигоксигенин, являющийся стероидным соединением растения наперстянки, используется также достаточно широко. В случае с дигоксигениновой меткой дальнейшая схема выявления полос заключается во взаимодействии фрагментов ДНК после их разделения секвенирующим гель-электрофорезом и переноса на мембранный фильтр с конъюгатом щелочная фосфатаза/антитела к дигоксигенину. Для биотинмеченных фрагментов ДНК требуется инкубация в растворе конъюгата стрептавидин/щелочная фосфатаза. Таким образом, после той или иной процедуры каждая полоса ДНК на мембранном фильтре несет щелочную фосфатазу и после непродолжительной инкубации (менее 1 ч) в указанных выше красителях за счет отщепления под действием данного фермента остатка фосфорной кислоты начинают проявляться полосы, имеющие синюю окраску. Для развития окраски слабых полос может потребоваться более продолжительное время – до 24 ч. Следует отметить достаточно высокую трудоемкость метода при тщательном проведении всех этих этапов, поскольку требуются многочисленные инкубации в блокирующих и прочих растворах и последующие отмывки, предшествующие стадии проявления окрашивания. Другим недостатком является то, что при хранении мембранных фильтров с проявившимися на них полосами ДНК красители на свету имеют тенденцию к выцветанию. И, наконец, главный недостаток метода заключается в его не очень высокой чувствительности при выявлении полос ДНК, значительно уступающей радиоактивному.

Применение другого субстрата, такого как 1,2-диоксетан, хемилюминесценцию которого также вызывала обработка щелочной фосфатазой, показало его высокую чувствительность и пригодность для молекулярно-биологических целей [Schaap et al., 1987]. Так, выявление полос ДНК, меченных биотином или дигоксигенином, с применением той же щелочной фосфатазы, конъюгированной со стрептавидином или с антителами к дигоксигенину, соответственно с помощью хемилюминесцентного субстрата 1,2-диоксетана, называемого еще АМРРД (полное название которого 3-(2'-спироадамantan)-4-метокси-4-(3''-фосфорилокси) фенил-1,2-диоксетан) (рис. 2.25), по чувствительности вполне сравнимо с радиоактивной меткой. Однако трудоемкость выполнения подготовительных процедур весьма высока, поскольку аналогична таковой при колориметрическом окрашивании [Beck et al., 1989; Tizard et al., 1990; Cate et al., 1991; Chee, 1991; Creasey et al., 1991; Holtke et al., 1992; Debuire et al., 1993]. Указанное выше соединение АМРРД весьма стабильно и может храниться на протяжении даже нескольких лет, но при удалении под действием щелочной фосфатазы остатка фосфорной кислоты, выступающего в качестве заместителя, происходит быстрая декомпозиция этого соединения с выделением квантов света, для регистрации которых может быть применена обычная рентгеновская пленка. АМРРД имеет довольно низкий квантовый выход ( $1,3 \cdot 10^{-5}$ ), что требует декомпозиции 77000 молекул для генерации одного фотона [Beck, Koster, 1990]. В связи с этим фирма "Tropix, Inc." (США) стала производить специальные энхансеры "Сапфир", "Изумруд" и "Рубин", отлича-



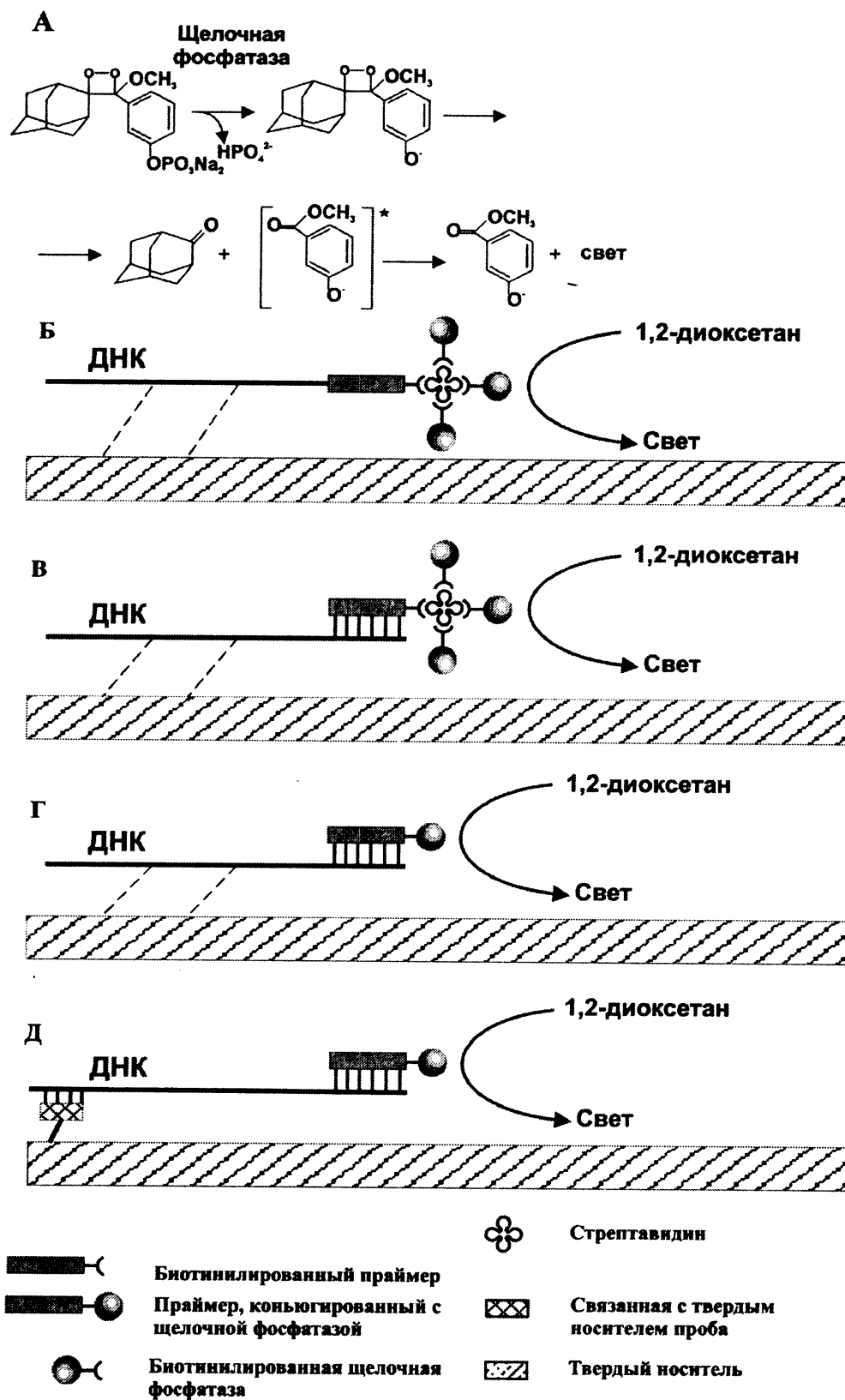


Рис. 2.25. Схема выявления полос ДНК с помощью хемилюминесцентного субстрата 1,2-диоксетана

ющие соответственно длиной волны эмиссии и производящие усиление сигнала в 10–500 раз. Однако проведенное специальное исследование по их применению в секвенировании ДНК показало неважные результаты, главным образом, из-за неравномерного распределения этих соединений по поверхности мембранного фильтра [Cate et al., 1991]. Детальное рассмотрение свойств 1,2-диоксетана и его производных, а также применение в молекулярной биологии системы диоксетан/щелочная фосфатаза проведено в обзоре Бека и Костера [Beck, Koster, 1990].

Позднее было получено другое производное диоксетана – CSPD (3-(4-метокси-спиро[1,2-диоксетан-3,2'-(5'-хлоро)-трицикло [3.3.1.1.3.7] декан]-4-ил) фенил фосфат), характеризующееся улучшенным квантовым выходом, и с успехом применено в качестве субстрата для щелочной фосфатазы при секвенировании ДНК [Olesen et al., 1993]. В этой же работе продемонстрирована возможность использования в качестве метки как биотина для системы стрептавидин/щелочная фосфатаза, так и различных гаптенных (дигоксигенина, 2,4-динитрофенила и флуоресцеина) за счет формирования комплексов антител к этим соединениям/щелочная фосфатаза с целью их дальнейшего выявления 1,2-диоксетаном. При этом оказалось, что чувствительность детекции для системы биотин–стрептавидин/щелочная фосфатаза в 8 раз выше по сравнению с таковой для флуоресцеина и в 16 раз – по сравнению с дигоксигенином и 2,4-динитрофенилом [Olesen et al., 1993]. Поскольку цитируемая работа была посвящена возможностям мультиплексного секвенирования, то авторы используемый ими подход вполне справедливо назвали мультиплексным мечением в связи с измененной схемой этого подхода, более детально рассмотренной в разделе 8.6.

Особо следует отметить очень короткое время экспозиции, обычно составляющее от 1 до 30 мин против 12–24 ч для радиоактивной метки, что послужило дополнительным основанием для использования диоксетановой химии при проведении мультиплексного секвенирования [Beck et al., 1989; Gillevet, 1990; Olesen et al., 1993]. Несколько позже, чтобы облегчить труд экспериментаторов, фирмой “Hoefler Scientific Instruments, Inc.” (США) стал выпускаться даже специальный прибор, позволяющий автоматически проводить все необходимые процедуры по образованию комплекса ДНК-биотин/стрептавидин–щелочная фосфатаза и заключительному этапу декомпозиции диоксетана [Patel, Nash, 1995].

В настоящее время при секвенировании ДНК также широко применяются различные флуоресцентные красители, однако детекция такой ДНК предполагает наличие сложного дорогостоящего оборудования, каковыми являются автоматические секвенаторы ДНК и поэтому подробное рассмотрение флуоресцентных меток будет проведено в специальной главе, посвященной автоматическому секвенированию ДНК.

Хотя и данный раздел имеет целью описать существующие способы внесения радиоактивной или какой-либо прочей метки во вновь синтезируемую цепь ДНК, секвенирование ДНК, не несущей никакой метки, также возможно, а поскольку метка служит для выявления фрагментов ДНК после их разделения в секвенирующем геле, то упомина-

ние или даже краткое описание в данном разделе подходов к секвенированию ДНК, где процедуры по ее детекции проводятся по завершению электрофореза, вполне логично.

Так, фрагменты ДНК, разделяемые с помощью гель-электрофореза при мультиплексном секвенировании, на первый взгляд, не несут какой-то ставшей уже привычной метки, однако определенная последовательность нуклеотидов вектора(ов), предшествующая вставке, может рассматриваться как своеобразная метка, в сущности таковой и являющаяся. А вот при секвенировании ДНК с последующей визуализацией полос в полиакриламидном геле с помощью окрашивания серебром [Wade-Evans, 1996], действительно, можно считать ДНК немеченной. Разработанный высокочувствительный и относительно простой метод окрашивания серебром фрагментов ДНК в полиакриламидном геле [Bassam et al., 1991] позволил выявлять весьма незначительные количества молекул ДНК, сравнимые с теми, которые обычно разделяются секвенирующим электрофорезом. Проведение соответствующих процедур после завершения электрофоретического разделения не занимает много времени и уже через 90 мин можно “читать” последовательность нуклеотидов ДНК с окрашенного серебром геля. Однако из-за высокой чувствительности метода к различным загрязнениям необходимо исключить отпечатки пальцев на геле и поэтому все операции рекомендуется проводить в перчатках. К другим обязательным требованиям следует отнести точное выполнение всех стадий, включая их временные интервалы и температуру применяемых растворов, высокую чистоту используемых реагентов и тщательно вымытые стекла для геля, причем, как указывает производитель специального набора для секвенирования ДНК с помощью серебряного окрашивания (Promega, США), некоторые детергенты приводят к увеличению общего фона геля. К недостаткам метода можно отнести то, что обычно выявляемые полосы ДНК видны на геле, приклеенном к стеклу с помощью  $\gamma$ -метакрилоксипропилтриметоксисилана, долговременное хранение которого (стекла с гелем) просто неразумно. Так как обращение с гелем такого большого размера при окрашивании серебром весьма затруднительно, то некоторой альтернативой стеклу может быть специальная Gel-Fix пленка для фиксирования на ней полиакриламидного геля, пригодная и для хранения на ней высушенного геля. В то же время фирма “Promega” поставляет специальную пленку EDF (Electrophoresis Duplicating Film), на которой с помощью кратковременного освещения окрашенного серебром геля можно получить постоянный отпечаток.

## 2.6. ТЕРМИНИРУЮЩИЕ РЕАКЦИИ

Пожалуй, ключевым моментом ферментативного метода секвенирования ДНК, разработанного Сэнгером и соавт. [Sanger et al., 1977], является образование специфически терминированных меченых фрагментов вновь синтезированной ДНК. Такая терминация построения комплементарной цепи ДНК происходит при включении ДНК-полимеразой в растущую цепь ДНК-модифицированных аналогов природных

субстратов, не допускающих присоединения дальнейших нуклеотидов. Наиболее широко применяемыми среди таких модифицированных аналогов являются 2',3'-дидезоксинуклеотид-5'-трифосфаты, характеризующиеся отсутствием второго гидроксильного остатка в 3'-положении рибозного кольца и превращению ее, таким образом, в дидезоксирибозу и невозможностью поэтому образовывать фосфодиэфирные связи между нуклеотидами в цепочке ДНК. В оригинальной статье Сэнгера и соавт., наряду с ддНТФ, для терминации цепи также был применен арабинодезоксицитозинтрифосфат [Sanger et al., 1977a], однако широкого применения арабинопроизводные не получили. Надо заметить, что механизм ингибирования дальнейшего роста цепи ДНК путем включения арабинопроизводных не совсем ясен [Davies, 1982]. Предполагается, что ДНК-полимераза (Кленовский фрагмент) имеет низкую аффинность к 3'-гидроксилу  $\beta$ -D-арабинофуранозильного сахара.

Как уже отмечалось выше, используемые в секвенировании ДНК ДНК-полимеразы обладают разным сродством к дидезоксипроизводным и соответственно различной скоростью включения последних в растущую цепь ДНК. Таким образом, для эффективного построения цепи ДНК в условиях терминации необходимо использовать определенные, для каждого фермента свои, соотношения дНТФ и ддНТФ.

Как можно видеть из приведенных в табл. 2.3 данных, Кленовский фрагмент ДНК-полимеразы I обладает заметной дискриминацией в отношении ддНТФ, причем эффективность включения каждого из них в свою очередь также различается. Последнее, впрочем, вполне объяснимо, поскольку известно, что во взаимодействие с тем или иным дНТФ вступают разные аминокислоты активного центра Кленовского фрагмента [Astatke et al., 1995] и для их ддНТФ-аналогов скорее всего мало что меняется. В связи с этим соотношение различных ддНТФ/дНТФ в реакционной смеси достаточно уникально и в оригинальных статьях встречаются иногда еще более дробные величины. То же самое отчасти справедливо и для AMV-ревертазы, Bst ДНК-полимеразы и Taq ДНК-полимеразы.

Что касается секвеназы, то в табл. 2.3 для нее приведены базовые соотношения дНТФ/ддНТФ (10:1), рассчитанные на эффективное построение в условиях терминации цепи ДНК довольно малой протяженности. Таким образом, при указанных соотношениях дНТФ/ддНТФ наибольшее число терминированных фрагментов окажется в диапазоне 50–500 нуклеотидов (при этом концентрации дНТФ на этапе мечения также должны быть соответствующими). Однако, учитывая высокую процессивность этого фермента, возможен сдвиг наибольшей представленности терминированных фрагментов в область более крупных молекул ДНК. Например, такое соотношение дНТФ/ддНТФ как 30:1 позволит получать специфически терминированные фрагменты ДНК длиной до 2000 нуклеотидов. Замена в реакционном буфере ионов  $Mg^{2+}$  на ионы  $Mn^{2+}$  повышает эффективность включения в растущую цепь ДНК ддНМФ и, таким образом, сдвигает область фрагментов ДНК превалирующих размеров [Tabor, Richardson, 1989a]. Особая актуальность такой замены ионов в реакционной смеси отмечается в случаях ограниченного количества матрицы ДНК.

Таблица 2.3

Концентрации дНТФ и соответствующих ддНТФ в терминирующих реакционных смесях для некоторых ДНК-полимераз, применяемых в секвенировании ДНК

Фермент	Тип реакции	Концентрация дНТФ и соответствующего ддНТФ в реакционной смеси (в микромолях)							
		dAT P	dCT P	dGT P	dTT P	ddAT P	ddCT P	ddGT P	ddTT P
Кленовский фрагмент	A	25	250	250	250	1000	–	–	–
	C	250	25	250	250	–	750	–	–
	G	250	250	25	250	–	–	500	–
	T	250	250	250	25	–	–	–	2000
AMV-ревертаза	A	80	800	800	800	160	–	–	–
	C	800	80	800	800	–	80	–	–
	G	800	800	80	800	–	–	80	–
	T	800	800	800	80	–	–	–	160
Bst ДНК-полимераза	A	25	250	250	250	480	–	–	–
	C	250	25	250	250	–	200	–	–
	G	250	250	25	250	–	–	240	–
	T	250	250	250	25	–	–	–	800
Секвеназа	A	80	80	80	80	8	–	–	–
	C	80	80	80	80	–	8	–	–
	G	80	80	80	80	–	–	8	–
	T	80	80	80	80	–	–	–	8
Taq ДНК-полимераза	A	25	250	250	250	100	–	–	–
	C	250	25	250	250	–	750	–	–
	G	250	250	25	250	–	–	750	–
	T	250	250	250	25	–	–	–	2000
Термосеквеназа	A	120	120	120	120	20	–	–	–
	C	120	120	120	120	–	20	–	–
	G	120	120	120	120	–	–	20	–
	T	120	120	120	120	–	–	–	20

Высокая процессивность секвеназы способствует эффективному построению цепи ДНК и за счет этого неспецифическая терминация и образование фрагментов, не несущих на своем 3'-конце соответствующего ддНТФ, в диапазоне длин, которые могут быть разделены секвенирующим гель-электрофорезом, являются довольно редкими событиями. Особенностью же проведения терминирующих реакций с помощью Кленовского фрагмента ДНК-полимеразы I является часто имеющаяся место неспецифическая терминация растущей цепи ДНК ввиду низкой процессивности этого фермента. Поскольку такая неспецифическая терминация приводит к увеличенному фону, то для ее исключения по завершению реакций терминации применяется так называемый чейз, заключающийся в дополнительной инкубации со всеми 4 немеченными дНТФ в высокой концентрации еще в течение 15 мин после 20 мин основной реакции при комнатной температуре. Высокая кон-

центрация дНТФ в “чейз”-растворе способствует дальнейшему построению цепи ДНК и неспецифически терминированные (т.е. не несущие на своем 3'-конце соответствующий ддНМФ) молекулы ДНК удлиняются настолько, что при их разделении секвенирующим электрофорезом они останутся в верхней части геля и, таким образом, уже не окажут отрицательного влияния на картину полос секвенируемой ДНК.

При использовании для проведения реакций мечения и терминирования высокопроцессивной секвеназы, обладающей к тому же высокой скоростью полимеризации, необходимо ограничивать время протекания этих реакций по двум причинам. Поскольку на стадии мечения дНТФ присутствуют в низких концентрациях при чрезмерно продолжительной инкубации (свыше 5 мин) секвеназа может удлинить праймер настолько, что при этом начнут истощаться содержащиеся в растворе дНТФ. Следствием может быть неспецифическое терминирование новосинтезируемых цепей ДНК и при условии невозобновления их дальнейшего синтеза на стадии терминации увеличенный фон при радиоавтографии. Вторая причина заключается в отрицательном вкладе пирофосфоролиза, который становится заметен при продолжительной инкубации и истощении нуклеотидтрифосфатов в реакционной смеси.

Известно, что отдельные участки молекул ДНК иногда не удается “прочитать” обычным способом из-за того, что на заключительном радиоавтографе наблюдаются полосы в двух-трех или сразу во всех четырех соседних дорожках. Причин тому бывает несколько. Считается, что одной из легко устранимых причин может быть недостаточная степень очистки матрицы ДНК, предполагаемой к секвенированию. В этом случае рекомендуется провести дополнительные этапы депротенинизации и/или переосаждение этанолом. Другая причина может заключаться в формировании в этом месте ДНК участка с сильной вторичной структурой и образовании структур, подобных “шпилькам”, что наиболее характерно для молекул ДНК с высоким GC-составом. ДНК-полимераза не может эффективно осуществлять синтез комплементарной цепи в этих местах, в связи с чем образуются неспецифически терминированные фрагменты ДНК. Преодолеть это иногда можно простым повышением температуры инкубации при проведении терминирующих реакций. Так, Кленовский фрагмент ДНК-полимеразы I при 42°C или даже 55°C или секвеназа при 45°C бывают способны преодолеть это препятствие за счет того, что при такой повышенной температуре участок вторичной структуры может уже “расплестись”. Другая возможность заключается в добавлении в реакционную смесь диметилсульфоксида [Seto, 1990; Seto et al., 1995] или формамида [Zhang et al., 1991], способствующих понижению температуры плавления участков со вторичной структурой. Правда, было показано, что присутствие в реакционной смеси даже 30% формамида (добавляемого из расчета 0,7% формамида на каждый 1% GC-пар, превышающий их 50%-ное содержание, взятое за отправную точку) не способно ликвидировать отдельные “стопы”, причем на фоне уже более слабых из-за формамида сигналов, принадлежащих остальной последовательности ДНК [Zhang et al., 1991]. Были рекомендации уменьшить в реакционной смеси концентрацию NaCl,

что дестабилизирует потенциальную вторичную структуру фрагментов ДНК [Ornstein, Kashdan, 1985]. Однако в случае “шпильки”, обогащенной GC-парами, такие ухищрения могут не привести к желаемому результату. Тогда необходимо использовать термостабильные ДНК-полимеразы (Bst, Taq или аналогичные), оптимум температуры которых составляет 65°, 72°C или даже несколько выше [Innis et al., 1988; McClary et al., 1991; Ranu, 1994]. При таких температурах участки со вторичной структурой обычно перестают существовать и ДНК-полимеразы беспрепятственно их “проходят”.

В литературе встречаются рекомендации последовательного использования двух ДНК-полимераз, одна из которых, как правило, термостабильна. В целом ряде статей сообщается об успешном последовательном применении секвеназы и Taq ДНК-полимеразы при оптимальных температурах инкубации для каждого фермента при секвенировании GC-богатых ДНК [Khambaty, Ely, 1990; Beck et al., 1993; Austin, 1995]. Одновременное использование двух не термостабильных ДНК-полимераз – Кленовского фрагмента ДНК-полимеразы I и секвеназы – позволило исключить характерные для секвенируемой матрицы “стопы” [Redston, Kem, 1994]. Независимое секвенирование одной и той же матрицы разными полимеразы (Кленовским фрагментом ДНК-полимеразы I и обратной транскриптазой) и их последующее одновременное разделение гель-электрофорезом позволило “прочитать” отдельные “трудные” участки ДНК, поскольку эти полимеразы формировали “стопы” в несколько отличающихся местах [Ornstein, Kashdan, 1985].

Осуществление “чейз”-реакции после построения комплементарной цепи ДНК T7 ДНК-полимеразой с помощью высоких концентраций дНТФ и терминальной нуклеотидтрансферазы, являющейся нематричной ДНК-полимеразой, позволило исключить образовавшиеся “стопы” [Fawcett, Bartlett, 1990; Kho, Zarbl, 1992]. Ранее подобный эффект был получен при секвенировании молекул РНК с помощью ревертазы с последующим “чейзом” с терминальной нуклеотидтрансферазой [DeBorde et al., 1986]. Терминальная нуклеотидтрансфераза присоединяет значительное количество нуклеотидов к свободному 3'-ОН-концу, не содержащему ддНМФ, и в результате образуются фрагменты весьма крупного размера, при электрофоретическом разделении остающиеся в самой верхней части геля. Однако другие авторы указывают на инверсию отдельных нуклеотидов, выделенных (подчеркнуто) в участке ACGCCGGCA, приведшую, как они считают, в результате действия терминальной нуклеотидтрансферазы к последовательности ACCGCGCGA, причем картина секвенируемых полос не вызывала каких-либо серьезных подозрений [Simard, Langlois, 1995]. Впрочем, весьма трудно представить, что же реально произошло в данном случае с фрагментом ДНК под действием терминальной трансферазы. В то же время внимательное рассмотрение указанных нуклеотидных последовательностей позволяет предположить, что произошли следующие события. Так, при секвенировании этого участка ACGCCGGCA, скорее всего, имеет место двойная компрессия полос (хотя авторы при проведении терминирующих реакций использовали 7-деза дГТФ, способст-

вующий недопущению компрессии) и он стал виден на радиоавтографе как ACCGCGCGA. Следует отметить, что ранее подобный эффект (инверсии GC в CG, правда, без участия терминальной нуклеотидтрансферазы) был отмечен и в другой работе [Odagiri, 1994]. Здесь автор решил, что инверсия этих нуклеотидов (GC, следующих за последовательностью GAAA в CG) в участке AGCGAAAGCAAGG образуется за счет того, что GAAA служит неким сигналом для фермента секвеназы для “неправильного” включения этих нуклеотидов при построении ею комплементарной цепи. В то же время похожий участок AGCAAAAGCAGG в этой же работе “читался” правильно. Можно предположить, что причиной такой инверсии нуклеотидов также является компрессия, несмотря на то, что при проведении терминирующих реакций автор использовал ДИТФ, способный в большинстве случаев ликвидировать это явление.

Другим способом преодоления ДНК-полимеразами возникающих “шпилек” считается подход с использованием специфических белков, связывающихся с одноцепочечной ДНК и ликвидирующих вторичную структуру матриц ДНК. Присутствие в реакционной смеси специфического SSB-белка *E. coli* заметно улучшает эффективность построения комплементарной цепи ДНК и исключает образование “стопов”. Также было показано, что проведение терминирующих реакций с двуцепочечной матрицей посредством секвеназы в присутствии другого связывающегося с одноцепочечной ДНК белка, кодируемого геном 32 фага T4, значительно улучшало картину полос секвенируемой ДНК на заключительном радиоавтографе [Kaspar et al., 1989]. Однако главное неудобство применения таких специфических к однонитевой ДНК белков заключается в необходимости их обязательного удаления перед нанесением продуктов терминирующих реакций на полиакриламидный гель. Предложены различные, подчас весьма сложные способы их удаления, включая обработку протеиназой K. Быстрый метод удаления таких белков основан на использовании очень сильного катионообменника 50W-X2 (Bio-Rad, США), или специальной смолы Wizard фирмы “Promega” (США) [Kaczorowski, Szybalski, 1994; Kaczorowski, Sektas, 1996].

Успешное “прохождение” ДНК-полимеразой “трудных” участков за счет тех или иных описанных выше ухищрений является только половиной дела, поскольку в новосинтезированной комплементарной цепи ДНК также будут присутствовать эти участки с потенциальной вторичной структурой. После денатурации цепей ДНК в этих местах одноцепочечной ДНК образуются “шпилечные” структуры, заметно меняющие электрофоретическую подвижность такой молекулы ДНК. Результатом будет видимая на конечном радиоавтографе компрессия отдельных полос, не позволяющая их однозначное “прочтение”. Проведение гель-электрофореза при повышенных температурах или с добавлением в буфер формамида в качестве дополнительного денатурирующего агента, кроме мочевины, в ряде случаев позволяет избежать эффект компрессии. Был также предложен способ ликвидации эффекта компрессии при электрофоретическом разделении продуктов секвенирующих реакций, заключающийся в проведении дополнительного эта-



па модификации цитозинового остатка в новосинтезированной цепи ДНК с помощью химических реагентов, приводившей к невозможности формирования участков с сильной вторичной структурой [Ambartsumyan, Mazo, 1980]. Применение для этой цепи метоксиамиона и бисульфита было основано на раннем наблюдении других авторов [Sverdlov et al., 1974] по превращению цитозинового основания в N<sup>4</sup>-метокси-5,6-дигидро-6-сульфоцитидин под действием метабисульфита натрия и O-метилгидрооксиламингидрохлорида. Однако такой подход нельзя было назвать достаточно удобным, поскольку он требовал дополнительных нежелательных усилий по осаждению, растворению продуктов секвенирующих реакций.

В то же время потенциальную проблему компрессии полос при их разделении в секвенирующем полиакриламидном геле, характерную, главным образом, для GC-богатых ДНК, можно разрешить уже на этапе проведения терминирующих реакций путем использования некоторых аналогов природных субстратов, формирующих более слабые связи со своими комплементарными основаниями. Поскольку GC-пары формируют, как известно, между собой три водородные связи против двух для AT-пар, то, следовательно, исключения или уменьшения компрессии возможно добиться, используя при построении комплементарной цепи ДНК аналогов дГТФ. Дезоксиинозинтрифосфат является одним из таких наиболее часто применяемых для секвенирования GC-богатых ДНК нуклеотидов [Mills, Kramer, 1979; Gough, Murray, 1983]. Благодаря отсутствию экзоциклической NH<sub>2</sub>-группы он не способен к формированию третьей водородной связи с остатком цитозина, что приводит к заметному уменьшению температуры плавления таких цепей ДНК. Таким образом, новосинтезированные фрагменты ДНК, содержащие вместо дГМФ дИМФ, формировать структуры типа "шпилек" во время их разделения в полиакриламидном геле при повышенной температуре и в присутствии мочевины как денатурирующего агента не будут. Результатом явится четкая картина полос секвенируемой ДНК, пригодная для их однозначного "прочтения". Некоторыми отрицательными моментами использования дИТФ является его плохое узнавание некоторыми ДНК-полимеразами и соответственно относительно слабое включение, что приводит к неэффективному построению комплементарной цепи ДНК и уменьшает количество нуклеотидов, которое можно определить в одном эксперименте, а также иногда образованию "пауз", следующих за включенным дИМФ и видимых как пустое место на заключительном радиоавтографе. В связи с не столь эффективным включением дИМФ в ДНК реакционная смесь при проведении терминации по G-типу с участием дИТФ содержит весьма высокую концентрацию последнего, превышающую таковую для дГТФ в 2–4 раза (для разных ДНК-полимераз), а также 100 мкмоль и более его дидезоксипроизводного ддИТФ или значительно уменьшенное количество ддГТФ. Кроме этого, из-за "пауз" (не являющихся результатом компрессии) в отдельных случаях рекомендуется пятидорожечное секвенирование, где G-тип реакции проводится параллельно как с гуанином, так и с инозином, причем следует иметь в виду, что подвижность фрагментов с дИМФ несколько выше таковых с дГМФ.

Другим изостерическим аналогом дГТФ при проведении терминирующих реакций служит его деазапроизводное 7-деаза-2'-дезоксигуанозин-5'-трифосфат, успешное применение которого для секвенирования ДНК с очень высоким GC-составом было почти одновременно продемонстрировано двумя группами исследователей [Mizusawa et al., 1986; Barr et al., 1986]. Несмотря на то, что 7-деаза-2'-дезоксигуанозин-5'-монофосфат образует с комплементарным ему дЦМФ также три водородные связи и характеризуется лишь чуть более низкой температурой плавления цепи ДНК, отсутствие азота в 7-м положении пуринового кольца препятствует формированию Хугстиновских пар и, таким образом, ослабляет взаимодействие между основаниями. Использование 7-деаза дГТФ, а не дИТФ в некоторых случаях предпочтительнее, так как он лучше узнается большинством ДНК-полимераз и не требует включения в реакционную смесь 7-деаза ддГТФ, тогда как ддИТФ должен обязательно присутствовать при проведении терминирующих реакций. В то же время дИМФ образует более слабые связи с комплементарным ему цитидином и в ряде случаев оказывается все же предпочтительнее [Tabor, Richardson, 1987a].

Также было показано исключение компрессии полос ДНК за счет использования при построении комплементарной цепи нового производного цитидина – N<sup>4</sup>-метил-2'-дезоксцитидин-5'-трифосфата [Li et al., 1993]. Как утверждают авторы, это соединение, эффективно включаемое Кленовским фрагментом ДНК-полимеразы I, Taq ДНК-полимеразой и секвеназой, в некоторых случаях было способно ликвидировать компрессию, с которой не справлялся 7-деаза-дГТФ. В то же время N<sup>4</sup>-метил-2'-дезоксцитидин-5'-трифосфат довольно часто вызывал образование ложной терминации строящейся цепи ДНК и одновременное присутствие в реакционной смеси обычного дЦТФ и этого его производного приводило из-за их разной подвижности к некоторому расширению полос, ухудшающему общую разрешающую способность метода.

Заметное улучшение качества секвенирования ДНК как в ручном, так и в автоматическом режимах удалось достичь путем проведения реакций терминирования в присутствии 7-деаза-дГТФ, а также 7-деаза-дАТФ, что позволило исключить аномалии в подвижности фрагментов ДНК из-за компрессии полос, вызванных вкладом обоих (дГМФ и дАМФ) нуклеотидов [Jensen et al., 1991]. При этом авторы отметили формирование более равномерных по высоте пиков фрагментов ДНК при их секвенировании в автоматическом секвенаторе с помощью флуоресцентной метки.

Недавно было отмечено, что обычно (более 70% случаев) компрессия вызывается определенными последовательностями нуклеотидов, консенсус которых был выведен как 5'-YGNAR-3' или 5'-YGNNAAR-3' [Yamakawa, Ohara, 1997]. Причем во многих случаях в указанной консенсусной последовательности под N скрываются остатки аденина. В связи с этим в цитируемой работе для устранения эффекта компрессии проведение терминирующих реакций осуществлялось с использованием как 7-деаза дГТФ, так и 7-деаза дАТФ. Отсутствие азота в положении 7 пуринового кольца у 7-деаза дАТФ не позволяет образовываться

неуотсон–криковским парам G–A, которые, как считается, играют важную роль в формировании вторичной структуры. Авторы сообщают об успешном преодолении проблемы компрессии после разделения продуктов реакции в автоматическом секвенаторе. Также при автоматическом флуоресцентном секвенировании ДНК было обнаружено, что замена обычных дНТФ на их тиопроизводные ((5'-O-(1-тиотрифосфат) дНТФ) при проведении терминирующих реакций позволяет частично разрешить проблему компрессии [Lee et al., 1992].

Несмотря на то, что классические терминаторы построения цепи ДНК ддНТФ вполне соответствуют запросам экспериментаторов, продолжается поиск и синтез новых производных нуклеотидов, узнаваемых ДНК-полимеразами и способных вызывать терминацию новосинтезируемой цепи ДНК. Было показано, что 2',2'-дидезокси-3'-аминонуклеозид-5'-трифосфаты являются сильными ингибиторами построения цепи ДНК и могут использоваться в качестве специфических терминаторов при секвенировании ДНК с помощью ДНК-полимеразы I [Chidgeavadze et al., 1984].

Можно отметить также синтез 8 производных дНТФ с модификацией 3'-положения сахара, среди которых только 3 удовлетворяли основным требованиям, предъявляемым к терминаторам [Metzker et al., 1994]. Первые два представляли собой 3'-O-метил-дАТФ и 3'-O-метил-дТТФ, узнаваемые большинством применяемых для секвенирования ДНК-полимераз и показали вполне сравнимую картину секвенируемых полос ДНК после их разделения в секвенирующем гель-электрофорезе. Тогда как уникальность третьего терминатора 3'-O-(2-нитробензил)-дАТФ заключается в его чувствительности к УФ-свету, в результате воздействия которого отщепляется блокирующая группа на 3'-конце дезоксирибозы и дальнейший синтез цепи ДНК может быть продолжен. Как считают авторы цитируемой работы, разработка для каждого азотистого основания подобных терминаторов, которые после определенных воздействий (например, УФ-светом) будут поддерживать дальнейшее удлинение цепи ДНК, может лечь в основу нового подхода к секвенированию ДНК, более подробно рассматриваемого в специальной главе 12.

Завершая данный раздел, видимо, необходимо отметить, что долгое время терминирующие реакции проводились в запаиваемых стеклянных капиллярах, как и в оригинальной работе Сэнгера и соавт. [Sanger et al., 1977a]. Так, например, еще в 1982 г. Хонг в своей методической статье упоминает о проведении секвенирующих реакций в стеклянных капиллярах [Hong, 1982]. В это, наверное, сейчас трудно поверить, но, такое, казалось бы, незначительное событие как переход к полипропиленовым реакционным 1,5 мл пробиркам, осуществленный Хонгом и соавт. в 1983 г. [Biggin et al., 1983], оказало значительное влияние на дальнейшее развитие метода ферментативного секвенирования, упростив проведение центрального этапа этого метода [Hong, 1987] и заложив тем самым основы автоматизации процесса секвенирования ДНК. Следующий шаг, заключавшийся в проведении терминирующих реакций уже в микротитраторных планшетах, был более ожидаем [Bankier et al., 1987; Koop et al., 1990].

## 2.7. ДРУГИЕ ПОДХОДЫ К ФЕРМЕНТАТИВНОМУ ОПРЕДЕЛЕНИЮ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ДНК

В середине 1980-х годов возник метод ферментативного секвенирования ДНК без использования уже привычных дидезоксинуклеотидов в качестве терминаторов строящейся цепи ДНК [Labeit et al., 1986; Labeit et al., 1987]. Более того, никакой заданной терминации растущей цепи ДНК на этапе ее построения даже не требовалось. Вместо дидезоксипроизводных использовались дНТФ $\alpha$ S, у которых в остатке фосфорной кислоты в  $\alpha$ -положении один из кислородов был замещен на атом серы. Данные соединения не останавливали синтез цепи ДНК, а служили в качестве своеобразных ограничителей для фермента экзонуклеазы III, с помощью которой и осуществлялась специфическая деградация секвенируемых фрагментов ДНК, несущих радиоактивную метку на одном из концов, благодаря использованию праймера с  $^{32}\text{P}$  на 5'-конце. Как известно, экзонуклеаза III последовательно удаляет нуклеотиды с 3'-конца цепи ДНК, но, как было выяснено, не способна к узнаванию и соответственно удалению тиопроизводных нуклеотидов [Putney et al., 1981]. Основываясь на этом факте, были составлены реакционные А-, С-, G- и Т-смеси, в которых наряду с обычными присутствовали и их соответствующие для каждой реакции тиопроизводные, что приводило к их произвольному включению в процессе построения новой цепи ДНК. После завершения полимеразной реакции проводилась обработка экзонуклеазой III, и конечные продукты разделялись обычным секвенирующим гель-электрофорезом с последующей радиоавтографией. Однако предел "чтения" секвенируемых данным подходом фрагмента ДНК был невелик и отчасти из-за того, что экзонуклеаза III, обладая разным сродством к различным нуклеотидам [Linxweiler, Horz, 1982], все же расщепляла тиопроизводные цитозина.

Некий гибрид методов химической деградации и ферментативного определения нуклеотидной последовательности ДНК был положен в основу подхода к секвенированию ДНК с теми же тиопроизводными [Gish, Eckstein, 1988; Nakamaya et al., 1988]. Отличия от описанного выше подхода с экзонуклеазой III заключались, во-первых, в том, что реакционные смеси содержали только три обычных дНТФ и по одному тио-дНТФ, и, во-вторых, химическая фрагментация цепей ДНК проводилась посредством алкилирующего агента, которым обычно служили 2-йодоэтанол или 2,3-эпокси-1-пропанол, причем последний, характеризуясь большей реакционной способностью для проведения реакции, требовался в меньшем количестве. Поскольку в каждой реакционной смеси один из четырех дНТФ был представлен только своим тиопроизводным, то, следовательно, все нуклеотиды данного типа в синтезированной цепи ДНК несли атом серы в остатке фосфорной кислоты. Посредством ограниченного гидролиза алкилирующим агентом более лабильных тиофосфатных диэфиров происходило их расщепление и радиоактивно меченные продукты реакции разделялись секвенирующим гель-электрофорезом.

Здесь, по-видимому, нельзя не отметить тот факт, что в основе уже упоминавшегося в самом начале этой главы метода секвенирования

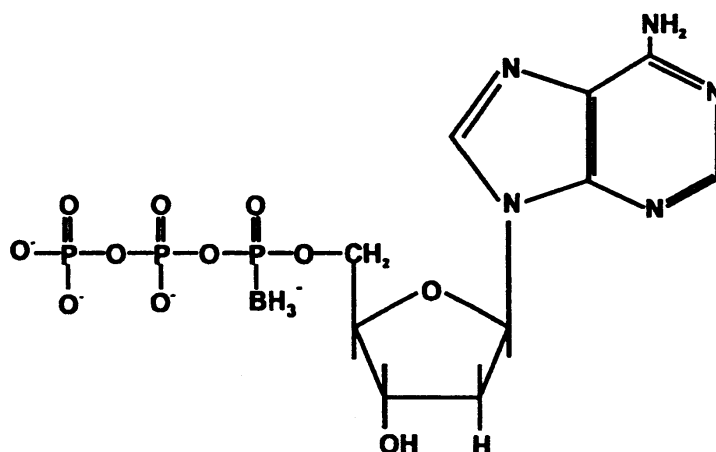


Рис. 2.26. Структурная формула борпроизводного дНТФ

ДНК путем рибозамещения [Barnes, 1978] лежит схожий принцип не терминации растущей цепи ДНК, а ее полного построения с происходящим произвольным включением соответствующих рибонуклеотидов, выступающих в роли своеобразных ограничителей. Поскольку реакционные смеси при проведении основного этапа содержали все 4 дНТФ и по одному рибоНТФ (в каждой пробирке – своего и в определенном соотношении к аналогичному дезоксипроизводному), то это приводило к образованию комплементарной полинуклеотидной цепи ДНК, в которой статистически случайно были представлены и рибопроизводные. Далее следовал этап расщепления новосинтезированной цепи ДНК по таким рибонуклеотидам с помощью или щелочи, или неспецифической РНКазы. Разделение продуктов реакции секвенирующим гель-электрофорезом позволяло определять последовательность нуклеотидов исследуемого фрагмента ДНК.

Следующим этапом развития определения нуклеотидной последовательности ДНК с тиопроизводными, как ограничителями расщепления ДНК, явился способ секвенирования путем амплификации определенного фрагмента Таq ДНК-полимеразой в присутствии двух олигонуклеотидных праймеров, один из которых нес  $^{32}\text{P}$  метку на 5'-конце, и четырех дНТФ, один из которых также присутствовал и в качестве ограничителя в виде тиопроизводного, что позволило потом провести расщепление новой цепи ДНК фосфодиэстеразой змеиного яда [Olsen, Eckstein, 1989]. В этой работе авторы подбирали условия амплификации таким образом, чтобы Таq ДНК-полимераза успевала включать тиопроизводные и обеспечивала дальнейший рост цепи и соответственно амплификацию фрагмента ДНК, протяженность которого составляла немногим менее 1 тпн. Однако следует отметить, что какого-либо значительного увеличения производительности секвенирования ДНК этим методом авторам добиться не удалось.

Новое развитие метод секвенирования ДНК с ограничителями получил после создания нового класса нуклеотидных аналогов, так называемых 2'-дезоксинуклеотид-5'- $\alpha$ -P-борано]-трифосфатов (рис. 2.26), в которых один из атомов кислорода у  $\alpha$ -остатка фосфорной кислоты замещен борпроизводным [Sood et al., 1990].

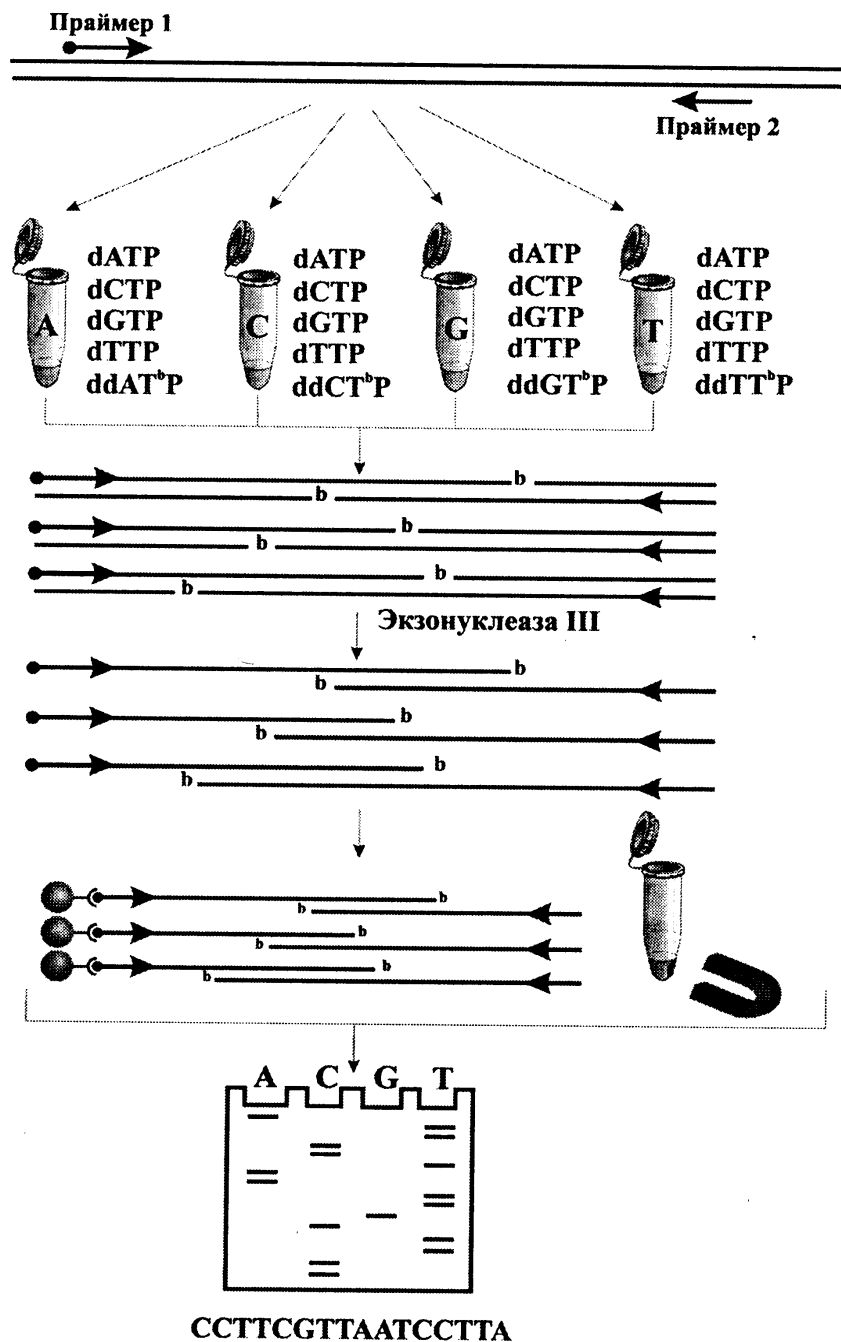


Рис. 2.27. Схема определения нуклеотидной последовательности ДНК с использованием боранофосфатных производных дНТФ

А – отжиг меченого (показанного звездочкой) и немеченого праймеров на матрице ДНК; Б – включение соответствующих борпроизводных дНМФ (показанных латинской буквой b) в процессе амплификации ДНК в четырех типах реакций (А, С, G и Т); В – расщепление построенных цепей ДНК под действием экзонуклеазы III и фосфодиэстеразы змеиного яда; разделение продуктов реакции секвенирующим гель-электрофорезом.

К важным чертам борпроизводных нуклеотидов следует отнести их термостабильность, что позволяет использовать их в ПЦР, а также их нераспознавание экзонуклеазой III [Li et al., 1995]. Последнее легло в основу разработанного подхода к секвенированию ДНК при помощи ПЦР и данных борпроизводных нуклеотидов [Porter et al., 1997].

Как можно видеть из принципиальной схемы данного подхода определения нуклеотидной последовательности, приведенной на рис. 2.27, в ходе осуществления секвенирования ДНК происходит отжиг двух праймеров, один из которых является меченым. (В цитируемой работе праймер был флуоресцентно меченным и для большего удобства при переходе от одной реакции к другой авторы ввели в этот праймер и биотиновую метку.) Затем происходит амплификация фрагмента ДНК, ограниченного данными праймерами, посредством специфических А-, С-, G- и Т-реакций в присутствии четырех обычных дНТФ и одного соответствующего борпроизводного, представленного в меньшей концентрации. В результате проведения ПЦР конечные продукты данных реакций будут нести полноразмерные фрагменты ДНК, в которые произвольным образом оказались встроеными борпроизводные нуклеотидов. На данном этапе было осуществлено удаление всех ставших уже ненужными реакционных компонентов путем связывания биотинилированного ПЦР продукта (за счет исходного биотинилированного праймера) с магнитными частицами, покрытыми стрептавидином. После необходимой промывки и суспендирования в экзонуклеазном буфере продукты реакций были подвергнуты расщеплению экзонуклеазой III и фосфодиэстеразой I и после денатурации нанесены на секвенирующий гель. Полученные результаты сравнительного секвенирования в двух автоматических секвенаторах (один цвет, четыре дорожки и четыре цвета, одна дорожка) с обычными дидезокситерминаторами и с борпроизводными показали высокое качество последнего.

### *Цитируемая литература*

*Ажикина Т.Л., Потапов В.К., Веселовская С.В., Мясников В.А., Свердлов К.Д.* Комбинации коротких олигонуклеотидов с повышенной прочностью образования дуплексов в качестве объединенных праймеров в секвенировании // Докл. РАН. 1993. Т. 331. С. 751–753.

*Ажикина Т.Л., Шевченко Ю.О., Лебедев Ю.Б., Веселовская С.В., Мясников В.А., Потапов В.К., Свердлов Е.Д.* Олигонуклеотиды, образующие высокостабильные дуплексы, их использование в качестве праймеров при секвенировании и в полимеразной цепной реакции // Там же. 1993. Т. 330. С. 642–645.

*Данилюк Н.К., Ястребов С.И., Артамонова Т.П., Попов С.Г.* Упрощенный вариант метода Максама–Гилберта для определения первичной структуры олигонуклеотидов и фрагментов ДНК // Биоорганическая химия. 1986. Т. 12. С. 1185–1188.

*Кнорре Д.Г., Пышный Д.В., Иванова Е.М., Бондарь А.А., Морозов И.В., Мертвецов Н.П., Зарытова В.Ф.* Тетрануклеотиды и их феназиниевые производные в качестве составных праймеров для секвенирования ДНК методом Сэнгера // Докл. РАН. 1996. Т. 350. С. 119–120.

*Краев А.С.* Простая система клонирования в фаге M13 и секвенирования ДНК с терминаторами // Молекулярная биология. 1988. Т. 22. С. 1164–1197.

*Краев А.С., Миронов В.М.* Экономическое секвенирование ДНК с терминаторами // Там же. 1990. Т. 24. С. 1095–1099.

*Краев А.С., Ли Н.Н., Сидоров Ю.Д., Твердохлебов Е.Н.* Новая пленка для регистрации секвенирующих гелей // Там же. 1992. Т. 26. С. 876–879.

*Смит Д.П.* Нитевидные фаги в качестве клонирующих векторов // Сельскохозяйственная биология: Векторные системы молекулярного клонирования. М.: Мир, 1991. С. 65–90.

Agelon L.B., Chen T.T. Supercoiled plasmid sequencing: One buffer for template denaturation, primer annealing, and synthesis of labeled strands by reverse transcriptase // *Gene Anal. Techn.* 1986. Vol. 3. P. 86–89.

Akhmetzjanov A.A., Vakhitov V.A. Molecular cloning and nucleotide sequence of the DNA polymerase gene from *Thermus flavus* // *Nucl. Acids Res.* 1992. Vol. 20. P. 5839.

Akiyama H., Kanai S., Ozawa H. Multi-priming sequencing: A DNA sequencing method involving restriction enzyme-digested DNA fragments as primers // *J. Biochem.* 1992. Vol. 111. P. 589–593.

Aliota J.M., Pelletier J.J., Ware J.L., Moran L.S., Benner J.S., Kong H. Thermostable Bst DNA polymerase I lacks a 3'→5' proofreading exonuclease activity // *Genet. Anal.* 1996. Vol. 12. P. 185–195.

Ambartsumyan N.S., Mazo A.M. Elimination of the secondary structure effect in gel sequencing of nucleic acids // *FEBS Lett.* 1980. Vol. 114. P. 265–268.

Anderson S., Gait M.J., Mayol L., Young I.G. A short primer for sequencing DNA cloned in the single-stranded phage vector M13mp2 // *Nucl. Acids Res.* 1980. Vol. 8. P. 1731–1743.

Ansorge W., Voss H., Zimmermann J. (ed.). DNA sequencing strategies. N.Y.; Wiley; Spektrum, 1996. P. 152–153.

Arakawa T., Jongsareejit B., Tatsumi Y., Tanaka K., Ikeda K., Komatsubara B., Inoue H., Kawakami B., Oka M., Emi S., Yomo T., Shima Y., Negoro S., Urabe I. Application of N-terminally truncated DNA polymerase from *Thermus thermophilus* (delta Tth polymerase) to DNA sequencing and polymerase chain reaction: Comparative study of delta Tth and wild-type Tth polymerases // *DNA Res.* 1996. Vol. 3. P. 87–92.

Arlinghaus H.F., Kwoka M.N., Guo X.O., Jacobson K.B. Multiplexed DNA sequencing and diagnostics by hybridization with enriched stable isotope labels // *Anal. Chem.* 1997. Vol. 69. P. 1510–1517.

Astatke M., Grindley N.D.F., Joyce C.M. Deoxynucleotide triphosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalysed by DNA polymerase I (Klenow fragment) // *J. Biol. Chem.* 1995. Vol. 270. P. 1945–1954.

Atkinson M.R., Deutcher M.P., Kornberg A., Russel A.F., Moffatt J.G. Enzymatic synthesis of deoxyribonucleic acid. 34. Termination of chain growth by a 2',3'-dideoxyribonucleotide // *Biochemistry.* 1969. Vol. 8. P. 4897–4904.

Austin C.C. A new method of bipolymerase sequencing prevents "stop-bands" // *Mol. Biotechnol.* 1995. Vol. 4. P. 100–101.

Azhikina T., Kostina M., Skaptsova N., Potapov V., Berg D., Sverdlov E. Factors affecting the priming efficiency of short contiguous oligonucleotide strings in the primer walking strategy of DNA sequencing // *DNA Sequenc.* 1996. Vol. 6. P. 211–216.

Azhikina T., Veselovskaya S., Myasnikov V., Potapov V., Ermolayeva O., Sverdlov E. Strings of contiguous modified pentanucleotides with increased DNA-binding affinity can be used for DNA sequencing by primer walking // *Proc. Nat. Acad. Sci. USA.* 1993. Vol. 90. P. 11460–11462.

Bankier A.T., Barrell B.G. Sequencing single-stranded DNA using the chain-termination method // *DNA sequencing: a practical approach* / Ed. C.T. Howe, E.S. Ward. Oxford: IRL press, 1989. P. 37–78.

Bankier A.T., Weston K.M., Barrell B.G. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 51–93.

Barnard G.F., Puder M., Begum N.A., Chen L.B. PCR product sequencing with [ $\alpha$ -<sup>33</sup>P] and [ $\alpha$ -<sup>32</sup>P] dATP // *Biotechniques.* 1994. Vol. 16. P. 572–573.

Barnes W.M. DNA sequencing by partial ribosubstitution // *J. Mol. Biol.* 1978. Vol. 119. P. 83–99.

Barnes W.M. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion // *Gene.* 1992. Vol. 112. P. 29–35.

Barr P.J., Thayer R.M., Laybourn P., Najarian R.C., Seela F., Tolan D.R. 7-Deaza-2'-deoxyguanosine-5'-triphosphate: Enhanced resolution in M13 dideoxy sequencing // *Biotechniques.* 1986. Vol. 4. P. 428–432.

Beaucage S.L., Caruthers M.H. Deoxynucleotide phosphoramidites – a new class of key intermediates for deoxypolynucleotide synthesis // *Tetrahedron Lett.* 1981. Vol. 22. P. 1859–1862.

Bechtereva T.A., Pavlov Y.I., Kramorov V.I., Migunova B., Kiselev O.I. DNA sequencing with thermostable Tet DNA polymerase from *Thermus thermophilus* // *Nucl. Acids Res.* 1989. Vol. 17. P. 10507.



- Beck K.-F., Stathopoulos I., Berninger M.G., Schweizer M.* Overcoming GC compression in nucleotide sequencing // *Biotechniques*. 1993. Vol. 14. P. 375.
- Beck S.* Colorimetric-detected DNA sequencing // *Anal. Biochem.* 1987. Vol. 164. P. 514–520.
- Beck S., Koster H.* Chemiluminescent detection of DNA: Applications of dioxetane chemiluminescent probes to molecular biology // *Anal. Chem.* 1990. Vol. 62. P. 2258–2270.
- Beck S., O'Keefe T., Coull J.M., Koster H.* Chemiluminescent detection of DNA: Application for DNA sequencing and hybridization // *Nucl. Acids Res.* 1989. Vol. 17. P. 5115–5123.
- Bedford E., Tabor S., Richardson C.C.* The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I // *Proc. Nat. Acad. Sci. USA*. 1997. Vol. 94. P. 479–484.
- Bernard A., Blanco L., Lazaro J.M., Martin J., Salas M.* A conserved 3'→5'-exonuclease active site in prokaryotic and eukaryotic DNA polymerases // *Cell*. 1989. Vol. 59. P. 219–228.
- Beskin A.D., Zevin-Sonkin D., Sobolev I.A., Ulanovsky L.E.* On the mechanism of the modular primer effect // *Nucl. Acids Res.* 1995. Vol. 23. P. 2881–2885.
- Biggin M.D., Gibson T.J., Hong G.F.* Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination // *Proc. Nat. Acad. Sci. USA*. 1983. Vol. 80. P. 3963–3965.
- Birnboim C., Doly J.* Rapid alkaline extraction procedure for screening recombinant plasmid DNA // *Nucl. Acids Res.* 1979. Vol. 7. P. 1513–1523.
- Blanco L., Bernard A., Blasco M.A., Salas M.* A general structure for DNA-dependent DNA polymerases // *Gene*. 1991. Vol. 100. P. 27–38; Errata // *Gene*. 1991. Vol. 108. P. 165.
- Blanco L., Bernard A., Salas M.* Evidence favouring the hypothesis of a conserved 3'-5'-exonuclease active site in DNA-dependent DNA polymerases // *Gene*. 1992. Vol. 112. P. 139–144.
- Blondel A., Thillet J.* A fast and convenient way to produce single stranded DNA from a phagemid // *Nucl. Acids Res.* 1991. Vol. 19. P. 181.
- Boom D. van den, Ruppert A., Jurinke C., Koster H.* Combined amplification and sequencing in a single reaction using two DNA polymerases with differential incorporation rates for dideoxynucleotides // *J. Biochem. and Biophys. Meth.* 1997. Vol. 35. P. 69–79.
- Brenner D.G., Shaw W.V.* The use of synthetic oligonucleotides with universal templates for rapid DNA sequencing: results with staphylococcal replicon pC221 // *EMBO J.* 1985. Vol. 4. P. 561–568.
- Brown N.L.* A primed-synthesis method for ribosubstitution of DNA at a single site // *FEBS Lett.* 1978. Vol. 93. P. 10–15.
- Brummet S.R., Ruan C., Hosta L.P., Nash B., Smith C., Fuller C.* [ $\alpha$ -<sup>33</sup>P]ddNTP terminators – the new standard for DNA sequencing // *Editorial comments. Amersham Life Sci.* 1997. Vol. 23. P. 1–7.
- Buchardt O., Egholm M., Berg R.H., Nielsen P.E.* Peptide nucleic acids and their potential applications in biotechnology // *Trends Biotechnol.* 1993. Vol. 11. P. 384–386.
- Burbelo P.D., Iadarola M.J.* Rapid plasmid sequencing with multiple octamer primers // *Biotechniques*. 1994. Vol. 16. P. 645–650.
- Burke T.J., Thompson D.V., Spurgeon S.L.* Method of using a Taq DNA polymerase without 5'-3'-exonuclease activity. Pat. 5108892 (USA). Publ. 1992.
- Carballeira N., Nazabal M., Brito J., Garcia O.* Purification of a thermostable DNA polymerase from *Thermus thermophilus* HB8, useful in the polymerase chain reaction // *Biotechniques*. 1990. Vol. 9. P. 276–281.
- Cate R.L., Ehrenfels C.W., Wusk M., Tizard R., Voyta J.C., Murhy O.J., Bronstein I.* Genomic southern analysis with alkaline-phosphatase-conjugated oligonucleotide probes and the chemiluminescent substrate AMPPD // *Genet. Anal. Techn. Appl.* 1991. Vol. 8. P. 102–106.
- Chee M.* Enzymatic multiplex DNA sequencing // *Nucl. Acids Res.* 1991. Vol. 19. P. 3301–3305.
- Chen D., Johnson A.F., Severin J.M., Rank D.R., Smith L.M., Guilfoyle R.A.* M13-102: a vector for facilitating construction and improving quality of M13 shotgun libraries // *Gene*. 1996. Vol. 172. P. 53–57.
- Chen E.Y., Seeburg P.H.* Supercoiled sequencing: Fast and simple method for sequencing plasmid DNA // *DNA*. 1985. Vol. 4. P. 165–170.
- Cherny D.Y., Belotserkovskii B.P., Frank-Kamenetskii M.D., Egholm M., Buchardt O., Berg R.H., Nielsen P.E.* DNA unwinding upon strand-displacement binding of a thymine-substituted polyamide to double-stranded DNA // *Proc. Nat. Acad. Sci. USA*. 1993. Vol. 90. P. 1667–1670.

*Chidgeavadze Z.G., Beabealashvili R.Sh., Atrazhev A.M., Kukhanova M.K., Azhayev A.V., Krayevsky A.A.* 2',3'-Dideoxy-3'-aminonucleoside 5'-triphosphates are the terminators of DNA synthesis catalyzed by DNA polymerases // *Nucl. Acids Res.* 1984. Vol. 12. P. 1671–1686.

*Chretien P., Shen S.H.* Multiplex random priming of internal restriction fragments for DNA sequencing // *DNA Cell Biol.* 1992. Vol. 11. P. 337–343.

*Clark J.M.* Novel non-template nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases // *Nucl. Acids Res.* 1988. Vol. 16. P. 9677–9686.

*Clark J.M., Joyce C.M., Beardsley G.P.* Novel blunt-end addition reactions catalyzed by DNA polymerase I of *Escherichia coli* // *J. Mol. Biol.* 1987. Vol. 198. P. 123–127.

*Clewell D.B., Helinski D.R.* Supercoiled circular DNA-protein complex in *E. coli*: Purification and induced conversion to an open circular DNA form // *Proc. Nat. Acad. Sci. USA.* 1969. Vol. 62. P. 1159–1166.

*Corey D.R.* Peptide nucleic acids: Expanding the scope of nucleic acid recognition // *Trends Biotechnol.* 1997. Vol. 15. P. 224–229.

*Creasey A., D'Angio L.Jr., Dunne T.S., Kissinger C., O'Keeffe T., Perry-O'Keeffe H., Moran L.S., Roskey M., Schildkraut I., Sears L.E., Slatko B.* Application of a novel chemiluminescence-based DNA detection method to single-vector and multiplex sequencing // *Biotechniques.* 1991. Vol. 11. P. 102–109.

*Davies R.W.* DNA sequencing // *Gel electrophoresis of nucleic acids: A practical approach.* Oxford: IRL press, 1982. P. 117–172.

*DeBorde D.C., Naeve C.W., Herlocher M.L., Maassab H.F.* Resolution of a common RNA sequencing ambiguity by terminal deoxynucleotidyl transferase // *Anal. Biochem.* 1986. Vol. 157. P. 275–282.

*Debuire B., Chabli A., Frenoy N.* Fast, manual, nonradioactive method for DNA sequencing // *Clin. Chem.* 1993. Vol. 39. P. 1682–1685.

*Del Sal G., Manioletti G., Schneider C.* A one-tube plasmid DNA mini-preparation suitable for sequencing // *Nucl. Acids Res.* 1988. Vol. 16. P. 9878.

*Dente L., Cesareni G., Cortese R.* pEMBL: a new family of single stranded plasmids // *Ibid.* 1984. Vol. 11. P. 1645–1655.

*Derbyshire V., Freemont P.S., Sanderson M.R., Beese L., Friedman J.M., Joyce C.M., Steitz T.A.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I // *Science.* 1988. Vol. 240. P. 199–201.

*Dunn J.J., Butler-Loffredo L.L., Studier F.W.* Ligation of hexamer on hexamer templates to produce primers for cycle sequencing or the polymerase chain reaction // *Anal. Biochem.* 1995. Vol. 228. P. 91–100.

*Earley J.J., Kuivaniemi H., Prockop D.J., Tromp G.* Efficient DNA sequencing on microtiter plates using dried reagents and Bst DNA polymerase // *DNA Sequenc.* 1993. Vol. 4. P. 79–85.

*Efimov V.A., Choob M.V., Buryakova A.A., Kalinkina A.A., Chakhmacheva O.G.* Synthesis and evaluation of some properties of chimeric oligomers containing PNA and phosphono-PNA residues // *Nucl. Acids Res.* 1998. Vol. 26. P. 566–575.

*Egholm M., Buchardt O., Christensen L., Behrens C., Freier S.M., Driver D.A., Berg R.H., Kim S.K., Norden B., Nielsen P.E.* PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules // *Nature.* 1993. Vol. 365. P. 566–568.

*Fan J., Ranu R.S., Smith C., Ruan C., Fuller C.W.* DNA sequencing with [ $\alpha$ -<sup>33</sup>P]-labelled ddNTP terminators: A new approach to DNA sequencing with ThermoSequenase DNA polymerase // *Biotechniques.* 1996. Vol. 21. P. 1132–1137.

*Fawcett T.W., Bartlett S.G.* An effective method for eliminating "artifact banding" when sequencing double-stranded DNA templates // *Ibid.* 1990. Vol. 9. P. 46–48.

*Froehler B.C., Matteucci M.D.* Nucleoside H-phosphonates: Valuable intermediates in the synthesis of deoxyoligonucleotides // *Tetrahedron Lett.* 1986. Vol. 27. P. 469–472.

*Froehler B.C., Ng P.G., Matteucci M.D.* Synthesis of DNA via deoxynucleoside H-phosphonate intermediates // *Nucl. Acids Res.* 1986. Vol. 14. P. 5399–5407.

*Fry G., Lachenmeier E., Mayrand E., Giusti B., Fisher J., Johnston-Dow L., Cathcart R., Finne E., Kilaas L.* A new approach to template purification for sequencing applications using a paramagnetic particles // *Biotechniques.* 1992. Vol. 13. P. 124–131.

*Fu D.-J., Broude N.E., Koster H., Smith C.L., Cantor C.R.* A DNA sequencing strategy that requires only five bases of known terminal sequencing for priming // *Proc. Nat. Acad. Sci. USA.* 1995. Vol. 92. P. 10162–10166.

- Fu D.-J., Koster H., Smith C.L., Cantor C.R.* A sequencing double-stranded DNA by strand displacement // *Nucl. Acids Res.* 1997. Vol. 25. P. 677–679.
- Furuyama H., Okano K., Kambara H.* DNA sequencing directly from a mixture using terminal-base-selective primers // *DNA Res.* 1994. Vol. 1. P. 231–237.
- Gait M.J., Sheppard R.C.* Rapid synthesis of oligodeoxyribonucleotides: A new solid-phase method // *Nucl. Acids Res.* 1977. Vol. 4. P. 1135–1158.
- Gait M.J., Singh M., Sheppard R.C.* Rapid synthesis of oligodeoxyribonucleotides. IV. Improved solid phase synthesis of oligodeoxyribonucleotides through phosphotriester intermediates // *Ibid.* 1980. Vol. 8. P. 1081–1096.
- Gerisher U., Durre P.* Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum* // *J. Bacteriol.* 1990. Vol. 172. P. 6907–6918.
- Gerisher U., Durre P.* Sequencing using custom designed oligonucleotides // *Meth. Mol. Biol.* 1993. Vol. 23: DNA sequencing protocols. P. 75–81.
- Gillam S., Jahnke P., Smith M.* Enzymatic synthesis of oligodeoxyribonucleotides of defined sequence // *J. Biol. Chem.* 1978. Vol. 253. P. 2532–2539.
- Gillevet P.M.* Chemiluminescent multiplex DNA sequencing // *Nature.* 1990. Vol. 348. P. 657–658.
- Gish G., Eckstein F.* DNA and RNA sequence determination based on phosphorothioate chemistry // *Science.* 1988. Vol. 240. P. 1520–1522.
- Godson G.N.* Primed synthesis methods of sequencing DNA and RNA // *Fed. Proc.* 1980. Vol. 39. P. 2822–2829.
- Golumbeski G.S.* A rapid miniprep procedure used to facilitate sequencing // *Editorial Comments. US Biochem. Corp.* 1990. Vol. 16. P. 4.
- Gough J.A., Murray N.E.* Sequence diversity among related genes for recognition of specific targets in DNA molecules // *J. Mol. Biol.* 1983. Vol. 166. P. 1–19.
- Hattori M., Sakaki Y.* Dideoxy sequencing method using denatured plasmid templates // *Anal. Biochem.* 1986. Vol. 152. P. 232–238.
- Hayashi K.* A cloning vehicle suitable for strand separation // *Gene.* 1980. Vol. 11. P. 109–115.
- Hecker K.H., Rill R.L.* Error analysis of chemically synthesized polynucleotides // *Biotechniques.* 1998. Vol. 24. P. 256–260.
- Heidecker G., Messing J., Gronenborn B.* A versatile primer for DNA sequencing in the M13mp2 cloning system // *Gene.* 1980. Vol. 10. P. 69–73.
- Holmes D.S., Quigley M.* A rapid boiling method for preparation of bacteria plasmids // *Anal. Biochem.* 1981. Vol. 114. P. 193–197.
- Holtke H.J., Sagner G., Kessler C., Schmitz G.* Sensitive chemiluminescent detection of digoxigenin-labeled nucleic acids: A fast and simple protocol and its applications // *Biotechniques.* 1992. Vol. 12. P. 104–113.
- Hong G.F.* A method for sequencing single-stranded cloned DNA in both directions // *Biosci. Rep.* 1981. Vol. 1. P. 243–253.
- Hong G.F.* Sequencing of large double-stranded DNA using the dideoxy sequencing technique // *Ibid.* 1982. Vol. 2. P. 907–912.
- Hong G.F.* The use of DNase I, buffer gradient gel, and <sup>35</sup>S label for DNA sequencing // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 93–110.
- Hou W., Smith L.M.* DNA sequencing with a hexamer string primer and dye-labelled terminators // *Anal. Biochem.* 1994. Vol. 221. P. 136–141.
- Hsiao K.-C.* A fast and simple procedure for sequencing double stranded DNA with sequenase // *Nucl. Acids Res.* 1991. Vol. 19. P. 2787.
- Hsu S.I.-H.* Rapid purification of mini-prep plasmid DNA with silica matrix for supercoil sequencing using sequenase // *Editorial Comments. US Biochem. Corp.* 1989. Vol. 16. P. 8, 17–18.
- Huber H.E., Tabor S., Richardson C.C.* *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates // *J. Biol. Chem.* 1987. Vol. 262. P. 16224–16232.
- Hultman T., Stahl S., Hornes E., Uhlen M.* Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support // *Nucl. Acids Res.* 1989. Vol. 17. P. 4937–4946.

*Ikeda K., Inoue H., Oka M., Kawakami B., Kawamura Y.* A non-radioactive DNA sequencing method using biotinylated dideoxynucleotide triphosphates and delta Tth DNA polymerase // *DNA Res.* 1995. Vol. 2. P. 225–227.

*Innis M.A., Myambo K.B., Gelfand D.H., Brow A.D.* DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA // *Proc. Nat. Acad. Sci. USA.* 1988. Vol. 85. P. 9436–9440.

*Ishino Y., Takahashi-Fujii A., Uemori T., Imamura M., Kato I., Doi H.* The amino acid sequence required for 5'→3'-exonuclease activity of *Bacillus caldotenax* DNA polymerase // *Protein Eng.* 1995. Vol. 8. P. 1171–1175.

*Jacobson K.B., Arlinghaus H.F., Schmitt H.W., Sachleben R.A., Brown G.M., Thonnard N., Sloop F.V., Foote R.S., Larimer F.W., Woychik R.P., England M.W., Burchett K.L., Jacobson D.A.* An approach to the use of stable isotopes for DNA sequencing // *Genomics.* 1991. Vol. 9. P. 51–59.

*Javed A.A., Joshi S.* Targeted DNA sequencing: Rapid identification of DNA clones by sequencing DNA using mixed oligodeoxynucleotide probes as a primers // *Biotechniques.* 1990. Vol. 9. P. 28–32; Erratum – *Biotechniques.* 1990. Vol. 9. P. 421.

*Jensen M.A., Zagursky R.J., Trainor G.L., Cocuzza A.J., Lee A., Chen E.Y.* Improvements in the chain-termination method of DNA sequencing through the use of 7-deaza-2'-deoxyadenosine // *DNA Sequenc.* 1991. Vol. 1. P. 233–239.

*Ji H., Smith L.M., Guilfoyle R.A.* Rapid isolation of cosmid insert DNA by triple-helix-mediated affinity capture // *Genet. Anal. Techn. Appl.* 1994. Vol. 11. P. 43–47.

*Johnson A.F., Lodhi M.A., McCombie W.R.* Fluorescence-based sequencing of double-stranded DNA by hexamer string priming // *Anal. Biochem.* 1996. Vol. 241. P. 228–237.

*Johnson A.F., McCombie W.R.* High-quality automated DNA sequencing primed with hexamer strings // *Genome Res.* 1996. Vol. 6. P. 10–18.

*Johnson A.F., Wang R., Ji H., Chen D.* Purification of single-stranded M13 DNA by cooperative triple-helix-mediated affinity capture // *Anal. Biochem.* 1996. Vol. 234. P. 83–95.

*Jones D.S.C., Schofield J.P.* A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing // *Nucl. Acids Res.* 1990. Vol. 18. P. 7463–7464.

*Joyce C.M., Grindley N.D.F.* Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *Escherichia coli* // *Proc. Nat. Acad. Sci. USA.* 1983. Vol. 80. P. 1830–1834.

*Jupin I., Gronenborn B.* Abundant, easy and reproducible production of single stranded DNA from phagemids using helper phage-infected competent cells // *Nucl. Acids Res.* 1995. Vol. 23. P. 535–536.

*Kaczorowski T., Sektas M.* Rapid removal of unincorporated label and protein from DNA sequencing reactions // *Mol. Biotechnol.* 1996. Vol. 5. P. 177–180.

*Kaczorowski T., Szybalski W.* Assembly of 18-nucleotide primers by ligation of three hexamers: Sequencing of large genomes by primer walking // *Anal. Biochem.* 1994a. Vol. 221. P. 127–135.

*Kaczorowski T., Szybalski W.* Rapid removal of proteins from DNA sequencing mixtures using a strong cationic exchanger // *Biotechniques.* 1994b. Vol. 17. P. 34.

*Kaczorowski T., Szybalski W.* Co-operativity of hexamer ligation // *Gene.* 1996a. Vol. 179. P. 189–193.

*Kaczorowski T., Szybalski W.* Automated four-color DNA sequencing using primers assembled by hexamer ligation // *Gene.* 1996b. Vol. 179. P. 195–198.

*Kalisch B.W., Krawetz S.A., Schoenwaelder K.-H., Sande J.H. van de.* Covalently linked sequencing primer linkers (splinkers) for sequence analysis of restriction fragments // *Gene.* 1986. Vol. 44. P. 263–270.

*Karger A.E., Weiss R., Gesteland R.F.* Digital chemiluminescence imaging of DNA sequencing blots using a charge-coupled device camera // *Nucl. Acids Res.* 1992. Vol. 20. P. 6657–6665.

*Kaspar P., Zadrazil S., Fabry M.* An improved double stranded DNA sequencing method using gene 32 protein // *Ibid.* 1989. Vol. 17. P. 3616.

*Khambaty F.M., Ely B.* A method for the determination of sequence of high G + C% DNA by the sequential application of sequenase and Taq polymerase // *Biotechniques.* 1990. Vol. 9. P. 714–715.

*Kho C.J., Zarbl H.* A rapid and efficient protocol for sequencing plasmid DNA // *Ibid.* 1992. Vol. 12. P. 228–230.

*Kieleczawa J., Dunn J.J., Studier F.W.* DNA sequencing by primer walking with strings of contiguous hexamers // *Science.* 1992. Vol. 258. P. 1787–1791.

- Kim H.-J., Kim B.-K. Simple cycle sequencing by labeling primer using *Taq* DNA polymerase // *Biotechniques*. 1994. Vol. 16. P. 576–580.
- Klenow H., Overgaard-Hansen K., Patkar S.A. Proteolytic cleavage of native DNA polymerase into two different catalytic fragments // *Europ. J. Biochem.* 1971. Vol. 22. P. 371–381.
- Knoth K., Roberds S., Poteet C., Tamkun M. Highly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction // *Nucl. Acids Res.* 1988. Vol. 16. P. 10932.
- Konecki D.S., Phillips J.J. TurboPrep II: An expensive, high-throughput plasmid template preparation protocol // *Biotechniques*. 1998. Vol. 24. P. 286–293.
- Koop B.F., Wilson R.K., Chen C., Halloran N., Sciammis R., Hood L., Lindelien J.W. Sequencing reactions in microtiter plates // *Ibid.* 1990. Vol. 9. P. 32–37.
- Kotler L., Sobolev I., Ulanovsky L. DNA sequencing: Modular primers for automated walking // *Ibid.* 1994. Vol. 17. P. 554–558.
- Kotler L., Zevin-Sonkin D., Sobolev I.A., Beskin A.D., Ulanovsky L. DNA sequencing: Modular primers assembled from a library of hexamers and pentamers // *Proc. Nat. Acad. Sci. USA.* 1993. Vol. 90. P. 4241–4245.
- Kraft R., Tardiff J., Krauter K.S., Leinwand L.A. Using mini-prep plasmid DNA for sequencing double stranded templates with sequenase // *Biotechniques*. 1988. Vol. 6. P. 544–547.
- Kristensen T., Voss H., Ansorge W. A simple and rapid preparation of M13 sequencing templates for manual and automated dideoxy sequencing // *Nucl. Acids Res.* 1987. Vol. 15. P. 5507–5516.
- Kristensen T., Voss H., Schwager C., Stegemann J., Sproat B., Ansorge W. T7 DNA polymerase in automated dideoxy sequencing // *Ibid.* 1988. Vol. 16. P. 3487–3496.
- Kunkel T.A., Patel S.S., Johnson K.A. Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit // *Proc. Nat. Acad. Sci. USA.* 1994. Vol. 91. P. 6830–6834.
- Kwon S.T., Kim J.S., Park J.H., Kim H.K., Lee D.S. Cloning and analysis of the DNA polymerase-encoding gene from *Thermus caldophilus* GK24 // *Mol. Cells.* 1997. Vol. 7. P. 264–271.
- Labeit S., Lehrach H., Goody R.S. A new method of DNA sequencing using deoxynucleoside alpha-thiotriphosphates // *DNA.* 1986. Vol. 5. P. 173–177.
- Labeit S., Lehrach H., Goody R.S. DNA sequencing using  $\alpha$ -thio deoxynucleotides // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 166–177.
- Lagutina I.V., Yurkova E.V., Azhikina T.L. et al. Determination of the patterns of hexanucleotide distribution along DNA by electron microscopy // *Genet. Anal.* 1996. Vol. 13. P. 9–14.
- Lashkari D.A., Hunnicke-Smith S.P., Norgren R., Davis R.W., Brenman T. An automated multiplex oligonucleotide synthesizer: Development of high-throughput, low-cost DNA synthesis // *Proc. Nat. Acad. Sci. USA.* 1995. Vol. 92. P. 7912–7915.
- Lawyer F.C., Stoffel S., Saiki R.K., Myambo K., Drummond R., Gelfand D.H. Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus* // *J. Biol. Chem.* 1989. Vol. 264. P. 6427–6437.
- Lawyer F.C., Stoffel S., Saiki R.K., Chang S.Y., Landre P.A., Abramson R.D., Gelfand D.H. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity // *PCR Meth. Appl.* 1993. Vol. 2. P. 275–287.
- Lee J.-S. Alternative dideoxy sequencing of double-stranded DNA by cyclic reactions using *Taq* polymerase // *DNA Cell Biol.* 1991. Vol. 10. P. 67–73.
- Lee L.G., Connell C.R., Woo S.L., Cheng R.D., McArdle B.F., Fuller C.W., Halloran N.D., Wilson R.K. DNA sequencing with dye-labeled terminators and T7 DNA polymerase: Effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments // *Nucl. Acids Res.* 1992. Vol. 20. P. 2471–2483.
- Li H., Porter K., Huang F., Shaw B.R. Boron-containing oligodeoxyribonucleotide 14mer duplexes: Enzymatic synthesis and melting studies // *Ibid.* 1995. Vol. 23. P. 4495–4501.
- Li S., Haces A., Stupar L., Gebeyehu G., Pless R.C. Elimination of band compression in sequencing gels by the use of N4-methyl-2'-deoxycytidine 5'-triphosphate // *Ibid.* 1993. Vol. 21. P. 2709–2714.
- Lim H.M., Pene J.J. Optimal conditions for supercoiled DNA sequencing with the *Escherichia coli* DNA polymerase I Large fragment // *Genet. Anal. Techn.* 1988. Vol. 5. P. 32–39.
- Linxweiler W., Horz W. Sequence specificity of exonuclease III from *E. coli* // *Nucl. Acids Res.* 1982. Vol. 10. P. 4845–4859.

- Loakes D., Brown D.M., Linde S., Hill F. 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR // *Ibid.* 1995. Vol. 23. P. 2361–2366.
- Lu Y., Ye S., Hong G. Large fragment of DNA polymerase I from *Bacillus stearothermophilus* (Bst polymerase) is stable at ambient temperature // *Biotechniques*. 1991. Vol. 11. P. 464–466.
- Lu Y., Ye S., Hong G. The reverse DNA sequencing using Bst DNA polymerase // *Chin. J. Biotechnol.* 1992. Vol. 8. P. 29–32.
- Maat J., Smith A.J.H. A method for sequencing restriction fragments with dideoxynucleotide triphosphates // *Nucl. Acids Res.* 1978. Vol. 5. P. 4537–4545.
- Malik A.N., McLean P.M., Roberts A., Barnett P.S., Demaine A.G., Banga J.P., McGregor A.M. A simple high yield method for the preparation of lambda gt10 DNA suitable for subcloning, amplification and direct sequencing // *Ibid.* 1990. Vol. 18. P. 4031.
- Manfioletti G., Schneider C. A new and fast method for preparing high quality lambda DNA suitable for sequencing // *Ibid.* 1988. Vol. 16. P. 2873–2884.
- Mariame B. Cycle sequencing protocol using Deep Vent<sub>R</sub><sup>®</sup> (exo<sup>-</sup>) DNA polymerase and reduced dNTP and [ $\alpha$ -<sup>35</sup>S] dATP concentrations // *Biotechniques*. 1996. Vol. 21. P. 18–19.
- Martsen E.O., Zeng M., Lapeyre J.N. Fast plasmid DNA sequencing using a thermal cycler and high temperature alkali denaturation // *Ibid.* 1997. Vol. 22. P. 420–422.
- McClain W.H., Foss K., Mittelstadt K.L., Schneider J. Variants in clones of gene-machine-synthesized oligodeoxynucleotides // *Nucl. Acids Res.* 1986. Vol. 14. P. 6770.
- McClary J., Ye S.Y., Ilong G.F., Witney F. Sequencing with the large fragment of DNA polymerase I from *Bacillus stearothermophilus* // *DNA Sequenc.* 1991. Vol. 1. P. 173–180.
- McGeoch D.J., Turnbull N.T. Analysis of the 3'-terminal nucleotide sequence of vesicular stomatitis virus N protein mRNA // *Nucl. Acids Res.* 1978. Vol. 5. P. 4007–4024.
- McGraw R.A., III. Dideoxy DNA sequencing with end-labelled oligonucleotide primers // *Anal. Biochem.* 1984. Vol. 143. P. 298–303.
- Mead D.A., McClary J.A., Luckey J.A., Kostichka A.J., Witney F.R., Smith L.M. Bst DNA polymerase permits rapid sequence analysis from nanogram amounts of template // *Biotechniques*. 1991. Vol. 11. P. 76–87.
- Merkens L.S., Bryan S.K., Moses R.E. Inactivation of the 5'-3' exonuclease of *Thermus aquaticus* DNA polymerase // *Biochim. et biophys. acta.* 1995. Vol. 1264. P. 243–248.
- Merrifield R.K. The construction of radiation shielding for use with phosphorus-32 labeling in molecular biology // *Anal. Biochem.* 1987. Vol. 167. P. 400–405.
- Messing J. New M13 vectors for cloning // *Meth. Enzymol.* 1983. Vol. 101. Recombinant DNA, pt C. P. 20–78.
- Messing J. Cloning in M13 phage or how to use biology at its best // *Gene.* 1991. Vol. 100. P. 3–12.
- Messing J., Bankier A.T. The use of single-stranded DNA phage in DNA sequencing // *DNA sequencing: A practical approach.* Oxford: IRL press, 1989. P. 1–36.
- Messing J., Crea R., Seeburg P.H. A system for shotgun DNA sequencing // *Nucl. Acids Res.* 1981. Vol. 9. P. 309–321.
- Messing J., Viera J. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments // *Gene.* 1982. Vol. 19. P. 269–276.
- Metzker M.L., Raghavachari R., Richards S., Jacutin S.E., Civitello A., Burgess K., Gibbs R.A. Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates // *Nucl. Acids Res.* 1994. Vol. 22. P. 4259–4267.
- Mierendorf R.C., Pfeffer D. Direct sequencing of denatured plasmid DNA // *Meth. Enzymol.* 1987. Vol. 152: Guide to molecular cloning techniques. P. 556–562.
- Mills D.R., Kramer F.R. Structure-independent nucleotide sequence analysis // *Proc. Nat. Acad. Sci. USA.* 1979. Vol. 76. P. 2232–2235.
- Misra H.S., Pandey P.K., Modak M.J., Vinayak R., Pandey V.N. Polyamide nucleic acid-DNA chimera lacking the phosphate backbone are novel primers for polymerase reaction catalyzed by DNA polymerases // *Biochemistry.* 1998. Vol. 37. P. 1917–1925.
- Mizusawa S., Nishimura S., Seela F. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of GTP // *Nucl. Acids Res.* 1986. Vol. 14. P. 1319–1324.
- Murphy G., Kavanagh T. Speeding-up the sequencing of double-stranded DNA // *Ibid.* 1988. Vol. 16. P. 5198.
- Nakamaye K.L., Gish G., Eckstein F., Vosberg H.-P. Direct sequencing of polymerase chain

reaction amplified DNA fragments through the incorporation of deoxynucleoside  $\alpha$ -thiotriphosphates // *Ibid.* 1988. Vol. 16. P. 9947–9959.

Nichols R., Dixon J.E. Rapid identification of clones using the same degenerate oligonucleotide mixture for both screening and sequencing // *Anal. Biochem.* 1988. Vol. 170. P. 110–115.

Nielsen P.E., Egholm M., Berg R.H., Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide // *Science.* 1991. Vol. 254. P. 1497–1500.

Norrander J., Kempe T., Messing J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis // *Gene.* 1983. Vol. 26. P. 101–106.

Odagiri T. Reverse misreading of a GC doublet by the modified T7 DNA polymerase, sequenase // *Virus Genes.* 1994. Vol. 8. P. 271–274.

Olesen S.E., Martin C.S., Bronstein I. Chemiluminescent DNA sequencing with multiplex labelling // *Biotechniques.* 1993. Vol. 15. P. 480–485.

Olsen D.B., Eckstein F. Incomplete primer extension during *in vitro* DNA amplification by *Taq* polymerase; exploitation for DNA sequencing // *Nucl. Acids Res.* 1989. Vol. 17. P. 9613–9620.

Ornstein D.L., Kashdan M.A. Sequencing DNA using  $^{35}\text{S}$ -labeling: A troubleshooting guide // *Biotechniques.* 1985. Vol. 3. P. 476–483.

Parker L.T., Zakeri H., Deng O., Spurgeon S., Kwork P.Y., Nickerson D.A. AmpliTaq DNA polymerase, FS dye-terminator sequencing: Analysis of peak height patterns // *Ibid.* 1996. Vol. 21. P. 694–699.

Patel A., Nash B. Innovations in non-isotopic DNA sequencing: Using an electrotransfer unit to blot sequencing gels and an automated membrane processor for detecting DNA sequences // *Ibid.* 1995. Vol. 18. P. 328–333.

Peterson M.G. DNA sequencing using *Taq* polymerase // *Nucl. Acids Res.* 1988. Vol. 16. P. 10915.

Pohl T.M., Maier E. Sequencing 500 kb of yeast DNA using a GATC 1500 direct blotting electrophoresis system // *Biotechniques.* 1995. Vol. 19. P. 482–486.

Pon R.T., Buck G.A., Hager K.M., Naeve C.W., Niece R.L., Robertson M., Smith A.J. Multifacility survey of oligonucleotide synthesis and an examination of the performance of unpurified primers in automated DNA sequencing // *Ibid.* 1996. Vol. 21. P. 680–685.

Porter K.W., Briley J.D., Shaw B.R. Direct PCR sequencing with boronated nucleotides // *Nucl. Acids Res.* 1997. Vol. 25. P. 1611–1617.

Putney S.D., Benkovic S.J., Schimmel P.R. A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo* // *Proc. Nat. Acad. Sci. USA.* 1981. Vol. 78. P. 7350–7354.

Raja M.C., Zevin-Sonkin D., Shvartzburd J., Kotler L., Ulanovsky E. DNA sequencing with modular primers using a two-step protocol with thermostable polymerase at the second step // *Biotechniques.* 1997a. Vol. 23. P. 362–368.

Raja M.C., Zevin-Sonkin D., Shvartzburd J., Rozovskaja T.A., Sobolev I.A., Chertkov O., Ramanathan V., Lvovsky L., Ulanovsky E. DNA sequencing using differential extension with nucleotide subsets (DENS) // *Nucl. Acids Res.* 1997b. Vol. 25. P. 800–805.

Ranu R.S. DNA sequencing with  $\Delta\text{Taq}^{\text{®}}$  DNA polymerase // *Biotechniques.* 1995. Vol. 18. P. 390–396.

Ranu R.S. Relief of DNA polymerase stop(s) due to severity of secondary structure of single-stranded DNA template during DNA sequencing // *Anal. Biochem.* 1994. Vol. 217. P. 158–161.

Rao V.B., Saunders N.B. A rapid polymerase-chain-reaction-directed sequencing strategy using a thermostable DNA polymerase from *Thermus flavus* // *Gene.* 1992. Vol. 113. P. 17–23.

Reddy P., McKenney K. Improved method for the production of M13 phage and single-stranded DNA for DNA sequencing // *Biotechniques.* 1996. Vol. 20. P. 854–860.

Redston M.S., Kern S.E. Klenow co-sequencing: A method for eliminating “stops” // *Ibid.* 1994. Vol. 17. P. 286, 288; Erratum // *Ibid.* 1995. Vol. 18. P. 68.

Reeve M.A., Fuller C.W. A novel thermostable polymerase for DNA sequencing // *Nature.* 1995. Vol. 376. P. 796–797.

Reha-Krantz L.J. Are there highly conserved DNA polymerase 3'  $\rightarrow$  5' exonuclease motifs? // *Gene.* 1992. Vol. 112. P. 133–137.

Richter O.M., Schafer G. Purification and enzymatic characterization of the cytoplasmic pyrophosphatase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum* // *Europ. J. Biochem.* 1992a. Vol. 209. P. 343–349.

- Richter O.M., Schafer G.* Cloning and sequencing of the gene for the cytoplasmic inorganic pyrophosphatase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum* // *Ibid.* 1992b. Vol. 209. P. 351–355.
- Richterich P., Heller C., Wurst H., Pohl T.M.* DNA sequencing with direct blotting electrophoresis and colorimetric detection // *Biotechniques.* 1989. Vol. 7. P. 52–59.
- Riggs M.G., Tudor S., Sivaram M., McDonough S.H.* Construction of single amino acid substitution mutants of cloned *Bacillus stearothermophilus* DNA polymerases I which lack 5' → 3' exonuclease activity // *Biochim. et biophys. acta.* 1996. Vol. 1307. P. 178–186.
- Riley D.E.* Very rapid nucleotide sequence analysis of improved, double-stranded minipreps // *Gene.* 1989. Vol. 75. P. 193–196.
- Robertson H.D., Barrell B.G., Weith H.L., Donelson J.E.* Isolation and sequence analysis of a ribosome-protected fragment from bacteriophage  $\phi$ X174 DNA // *Nature. New Biol.* 1973. Vol. 241. P. 38–40.
- Samols S.B., Fuller C.W.*  $\Delta$ Taq DNA polymerase from DNA sequencing // *Editorial Comments. Amersham Life Sci.* 1995. Vol. 21. P. 57–78.
- Samols S.B., McArdle B.F., Ruan C.C., Vander Horn P.B., Fuller C.W.* Thermo sequenase; a new thermostable DNA polymerase for DNA sequencing // *Ibid.* 1995. Vol. 22. P. 29–36.
- Sanchez-Pescador R., Urdea M.S.* Use of unpurified synthetic deoxynucleotide primers for rapid dideoxynucleotide chain termination sequencing // *DNA.* 1984. Vol. 4. P. 339–343.
- Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes J.C., Hutchinson C.A., III, Slocombe P.M., Smith M.* Nucleotide sequence of bacteriophage  $\phi$ X174 DNA // *Nature.* 1977a. Vol. 265. P. 687–695.
- Sanger F., Coulson A.R.* A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase // *J. Mol. Biol.* 1975. Vol. 94. P. 441–448.
- Sanger F., Coulson A.R., Barrell B.G., Smith A.J., Roe B.A.* Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing // *J. Mol. Biol.* 1980. Vol. 143. P. 161–178.
- Sanger F., Donelson J.E., Coulson A.R., Kossel H., Fischer D.* Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA // *Proc. Nat. Acad. Sci. USA.* 1973. Vol. 70. P. 1209–1213.
- Sanger F., Nicklen S., Coulson A.R.* DNA sequencing with chain-terminating inhibitors // *Ibid.* 1977b. Vol. 74. P. 5463–5467.
- Saunders S.E., Burke J.F.* Rapid isolation of miniprep DNA for double strand sequencing // *Nucl. Acids Res.* 1990. Vol. 18. P. 4948.
- Seif I., Khoury G., Dhar R.* A rapid enzymatic DNA sequencing technique: Determination of sequence alterations in early simian virus 40 temperature sensitive and deletion mutants // *Ibid.* 1980. Vol. 8. P. 2225–2240.
- Seto D.* An improved method for sequencing double stranded plasmid DNA from minipreps using DMSO and modified template preparation // *Ibid.* 1990. Vol. 18. P. 5905–5906.
- Seto D., Seto J., Deshpande P., Hood L.* DMSO resolves certain compressions and signal dropouts in fluorescent dye labelled primer-based DNA sequencing reactions // *DNA Sequenc.* 1995. Vol. 5. P. 131–140.
- Schaap A.P., Sandison M.D., Handley R.S.* Chemical and enzymatic triggering of 1,2-dioxetanes. 3. Alkaline phosphatase-catalyzed chemiluminescence from an aryl phosphate-substituted dioxetane // *Tetrahedron Lett.* 1987. Vol. 28. P. 1159–1162.
- Shen Z., Liu J., Wells R.L., Elkind M.M.* Cycle sequencing using degenerate primers containing inosines // *Biotechniques.* 1993. Vol. 15. P. 82, 84–89.
- Shen Z., Liu J., Wells R.L., Elkind M.M.* Direct sequencing with highly degenerate and inosine-containing primers // *Meth. Mol. Biol.* 1996. Vol. 65: PCR sequencing protocols. P. 111–118.
- Siemieniak D.R., Slightom J.L.* A library of 3342 useful nonamer primers for genome sequencing // *Gene.* 1990. Vol. 96. P. 121–124.
- Simard C., Langlois I.* Chasing dideoxy-mediated sequencing reactions with terminal deoxynucleotidyl transferase may result in nucleotide inversions in sequencing gels // *Nucl. Acids Res.* 1995. Vol. 23. P. 2341.
- Slightom J.L., Bock J.H., Siemieniak D.R., Hurst G.D., Beattie K.L.* Nucleotide sequencing double-stranded plasmids with primers selected from a nonamer library // *Biotechniques.* 1994. Vol. 17. P. 536, 537, 540–544.
- Smith A.J.H.* The use of exonuclease III for preparing single stranded DNA for use as a template in the chain terminator sequencing method // *Nucl. Acids Res.* 1979. Vol. 6. P. 831–848.
- Smith A.J.H.* DNA sequence analysis by primed synthesis // *Meth. Enzymol.* 1980. Vol. 65: Nucleic acids, pt I. P. 560–580.



- Smith M., Leung D.W., Gillam S., Astell C.R. Sequence of the gene for iso-1-cytochrome c in *Saccharomyces cerevisiae* // Cell. 1979. Vol. 16. P. 753–761.
- Speek M., Ilves H., Lind A. Consecutive DNA terminator sequencing by using enzymatically generated primers // Anal. Biochem. 1986. Vol. 158. P. 242–249.
- Stahl S., Hultman T., Olsson A., Moks T., Unlen M. Solid phase DNA sequencing using the biotin–avidin system // Nucl. Acids Res. 1988. Vol. 16. P. 3025–3038.
- Strauss E.C., Kobori J.A., Siu G., Hood L.E. Specific-primer-directed DNA sequencing // Anal. Biochem. 1986. Vol. 154. P. 353–360.
- Strauss N.A., Zagursky R.J. In vitro production of large single-stranded templates for DNA sequencing // Biotechniques. 1991. Vol. 10. P. 376–384.
- Studier F.W. A strategy for high-volume sequencing of cosmid DNAs: Random and directed priming with a library of oligonucleotides // Proc. Nat. Acad. Sci. USA, 1989. Vol. 86. P. 6917–6921.
- Sverdlov E.D., Monastyrskaya G.S., Guskova L.I., Levitan T.L., Sheichenko V.I., Budowsky E.I. Modification of cytidine residues with a bisulfite–O-methylhydroxylamine mixture // Biochim. et biophys. acta. 1974. Vol. 340. P. 153–165.
- Szybalski W. Proposal for sequencing DNA using ligation of hexamers to generate sequential elongation primers (SPEL-6) // Gene. 1990. Vol. 90. P. 177–178.
- Szybalski W. From the double-helix to novel approaches to the sequencing of large genomes // Gene. 1993. Vol. 135. P. 279–290.
- Szybalski W., Kubinski H., Hradecna Z., Summers W.C. Analytical and preparative separation of the complementary DNA strands // Meth. Enzymol. 1971. Vol. 21: Nucleic acids, pt D. P. 383–413.
- Tabor S., Huber H.E., Richardson C.C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7 // J. Biol. Chem. 1987. Vol. 262. P. 16212–16223.
- Tabor S., Richardson C.C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase // Proc. Nat. Acad. Sci. USA. 1987a. Vol. 84. P. 4767–4771.
- Tabor S., Richardson C.C. Selective oxidation of the exonuclease domain of bacteriophage T7 DNA polymerase // J. Biol. Chem. 1987b. Vol. 262. P. 15330–15333.
- Tabor S., Richardson C.C. Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I // Proc. Nat. Acad. Sci. USA. 1989a. Vol. 86. P. 4076–4080.
- Tabor S., Richardson C.C. Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis // J. Biol. Chem. 1989b. Vol. 264. P. 6447–6458.
- Tabor S., Richardson C.C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase // Ibid. 1990. Vol. 265. P. 8322–8328.
- Tabor S., Richardson C.C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides // Proc. Nat. Acad. Sci. USA. 1995. Vol. 92. P. 6339–6343.
- Tang W., Liu J.-N., Gurewicz V. Preparation of single-stranded phagemid DNA without chromosomal DNA contamination // Biotechniques. 1996. Vol. 21. P. 53–54.
- Tizard R., Cate R.L., Ramachandran K.L., Wisk M., Voyta J.C., Murphy O.J., Bronstein I. Imaging of DNA sequences with chemiluminescence // Proc. Nat. Acad. Sci. USA. 1990. Vol. 87. P. 4514–4518.
- Toneguzzo F., Glynn S., Levi E., Mjolsness S., Hayday A. Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA // Biotechniques. 1988. Vol. 6. P. 460–469.
- Uemori T., Ishino Y., Fujita K., Asada K., Kato I. Cloning of the DNA polymerase gene of *Bacillus caldotenax* and characterization of the gene product // J. Biochem. 1993. Vol. 113. P. 401–410.
- Uhlmann E. Peptide nucleic acids (PNA) and PNA-DNA chimeras: From high binding affinity towards biological function // Biol. Chem. 1998. Vol. 379. P. 1045–1052.
- Vainshtein I., Atrazhev A., Eom S.H., Elliot J.F., Wishart D.S., Malcolm B.A. Peptide rescue of an N-terminal truncation of the Stoffel fragment of taq DNA polymerase // Protein Sci. 1996. Vol. 5. P. 1785–1792.
- Valenzuela M.S., Siddiqui K.A.I. Bidirectional sequencing of supercoiled plasmid DNA // Anal. Biochem. 1989. Vol. 183. P. 258–262.
- Vander Horn P.B., Davis M.C., Cunniff J.J., Ruan C., McArdle B.F., Samols S.B., Szasz J., Hu G., Hujer K.M., Domke S.T., Burmmet S.R., Moffett R.B., Fuller C.W. Thermo sequenase DNA

polymerase and *T. acidophilum* pyrophosphatase: New thermostable enzymes for DNA sequencing // *Biotechniques*. 1997. Vol. 22. P. 758–765.

Vasser M., Ng P.G., Jhurani P., Bischofberger N. Error rates in oligodeoxynucleotides synthesized by the H-phosphonate method // *Nucl. Acids Res.* 1990. Vol. 18. P. 3089.

Viera J., Messing J. The pUC plasmids, and M13mp7-derived system of insertion mutagenesis and sequencing with synthetic universal primers // *Gene*. 1982. Vol. 19. P. 259–268.

Villbrandt B., Sagner G., Schomburg D. Investigation on the thermostability and function of truncated *Thermus aquaticus* DNA polymerase fragments // *Protein Eng.* 1997. Vol. 10. P. 1281–1288.

Voss H., Nentwich U., Duthie S., Wiemann S., Benes V., Zimmermann J., Ansorge W. Automated cycle sequencing with Taqenase: protocols for internal labelling, dye primer and “double” simultaneous sequencing // *Biotechniques*. 1997. Vol. 23. P. 312–318.

Wade-Evans A. Silver sequencing: Nonradioactive cycle sequencing of dsDNA // *Meth. Mol. Biol.* 1996. Vol. 65: PCR sequencing protocols. P. 91–100.

Wallace R.B., Johnson M.J., Suggs S.V., Miyoshi K., Bhatt R., Itakura K. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmid vector pBR322 // *Gene*. 1981. Vol. 16. P. 21–26.

Wang L.-M., Weber D.K., Johnson T., Sakaguchi A.Y. Supercoil sequencing using unpurified templates produced by rapid boiling // *Biotechniques*. 1988. Vol. 6. P. 839–843.

Wang Y. Double-stranded DNA sequencing with T7 polymerase // *Ibid.* 1988. Vol. 6. P. 843–845.

Weickert M.J., Chambliss G.H. Acid phenol minipreps make excellent sequencing templates // *Editorial Comments. US Biochem. Corp.* 1989. Vol. 16. P. 5–6.

Wilchek M., Bayer E.A. The avidin-biotin complex in bioanalytical applications // *Anal. Biochem.* 1988. Vol. 171. P. 1–32.

Winter G., Fields S., Gait M.J., Brownlee G.G. The use of synthetic oligodeoxynucleotide primers in cloning and sequencing segment 8 of influenza virus (A/PR/8/34) // *Nucl. Acids Res.* 1981. Vol. 9. P. 237–245.

Wittung P., Nielsen P.E., Buchardt O., Egholm M., Norden B. DNA-like double helix formed by peptide nucleic acid // *Nature*. 1994. Vol. 368. P. 561–563.

Woodford K., Usdin K. ‘Turboprep’ DNA: ultraquick preparation of plasmid DNA from single colonies for DNA sequencing // *Nucl. Acids Res.* 1991. Vol. 19. P. 6652.

Xiang C., Wang H., Shiel P., Berger P., Guerra D.J. A modified alkaline lysis miniprep protocol using a single microcentrifuge tube // *Biotechniques*. 1994. Vol. 17. P. 30–32.

Yamakawa H., Ohara O. A DNA cycle sequencing reaction that minimizes compressions on automated fluorescent sequencers // *Nucl. Acids Res.* 1997. Vol. 25. P. 1311–1312.

Yamamoto T., Horikoshi M. Rapid preparation of plasmid templates suitable for a DNA sequencer without RNase treatment // *Ibid.* 1995. Vol. 23. P. 3351–3352.

Yang X., Richardson C.C. Structural and functional organization of the DNA polymerase of bacteriophage T7 // *J. Biol. Chem.* 1966. Vol. 271. P. 24207–24212.

Yanish-Perron C., Viera J., Messing J. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors // *Gene*. 1985. Vol. 33. P. 103–119.

Ye S.Y., Hong G.F. Heat-stable DNA polymerase I large fragment resolves hairpin structure in DNA sequencing // *Sci. Sin. B.* 1987. Vol. 30. P. 503–506.

Zagursky R.J., Conway P.S., Kashdan M.A. Use of <sup>33</sup>P for Sanger DNA sequencing // *Biotechniques*. 1991. Vol. 11. P. 36–38.

Zhang H., Scholl R., Browse J., Somerville C. Double stranded DNA sequencing as a choice for DNA sequencing // *Nucl. Acids Res.* 1988. Vol. 16. P. 1220.

Zhang W., Hu G., Deisseroth A. Improvement of PCR sequencing by formamide // *Ibid.* 1991. Vol. 19. P. 6649.

Ziff E.B., Sedat J.W., Galibert F. Determination of the nucleotide sequence of a fragment of bacteriophage φX174 DNA // *Nature. New Biol.* 1973. Vol. 241. P. 34–37.

Zimmermann J., Voss H., Schwager C., Stegemann J., Erfle H., Stucky K., Kristensen T., Ansorge W. A simplified protocol for fast plasmid DNA sequencing // *Nucl. Acids Res.* 1990. Vol. 18. P. 1067.

Zimmern D., Kaesberg P. 3'-Terminal nucleotide sequence of encephalomyocarditis virus RNA determination by reverse transcriptase and chain-terminating inhibitors // *Proc. Nat. Acad. Sci. USA.* 1978. Vol. 75. P. 4257–4261.

## Дополнительная литература

- Корнберг А.* Синтез ДНК. М.: Мир, 1977. 359 с.
- Методы молекулярной генетики и геной инженерии / А.В. Мазин, А.В. Кузнецов, А.С. Краев и др. Новосибирск: Наука, 1990. 248 с.
- Шарипов Г.Л., Казаков В.П., Толстиков Г.А.* Химия и хемилюминесценция 1,2-диоксетанов. М.: Наука, 1990. 288 с.
- Alphay L.* DNA sequencing from experimental methods to bioinformatics. Berlin etc: BIOS sci. publ., 1977. 224 p.
- Ansorge W., Voss H., Zimmermann J.* (ed.). DNA sequencing strategies. N.Y.: Wiley; Spektrum, 1996. 202 p.
- Beck S., Koster H.* Chemiluminescent detection of DNA: Applications of dioxetane chemiluminescent probes to molecular biology // *Anal. Chem.* 1990. Vol. 62. P. 2258–2270.
- Griffin H.G., Griffin A.M.* (ed.). *Methods in molecular biology*. Vol. 23. DNA sequencing protocols. Totowa: Humana press, 1993. 392 p.
- Haugland R.P.* *Molecular probes: Handbook of fluorescent probes and research chemicals*. Eugene. 1992–1994. 390 p.
- Promega protocols and applications guide*. 3rd ed. The source for discovery. Madison, 1996. 404 p.
- Sambrook J., Fritsch E.F., Maniatis T.* *Molecular cloning: A laboratory manual*. 2nd ed. Cold Spring Harbor: CSHL press, 1989.
- Sanger F.* Determination of nucleotide sequences in DNA: (Nobel lecture) // *Biosci. Rep.* 1981. Vol. 1. P. 3–18.

### ПЦР-СЕКВЕНИРОВАНИЕ ДНК

Открытие возможности избирательной амплификации определенных участков ДНК с помощью полимеразной цепной реакции [Saiki et al., 1985, 1988] оказало чрезвычайно сильное влияние на всю молекулярную биологию, включая и секвенирование ДНК. До появления этого метода можно было секвенировать только рекомбинантные молекулы ДНК (кроме некоторых частных случаев). Причина этого кроется в низком содержании в геноме какой-либо определенной, даже повторяющейся, последовательности и невозможности детекции столь малого количества фрагментов ДНК при их секвенировании. Теоретически можно было бы детектировать на радиоавтографе секвенирующего геля полосы ДНК какого-нибудь условного гена, если взять в реакцию около 500 мг тотальной ДНК с ее концентрацией в реакционной смеси приблизительно 50 г/мл. Поскольку такое принципиально невозможно, то для определения последовательности нуклеотидов было необходимо использовать ДНК, “обогащенную” каким-либо геном или просто определенным участком, что и достигалось ранее посредством клонирования. В связи с достаточной сложностью методов клонирования получение таких рекомбинантных ДНК и их последующее секвенирование было доступно далеко не всем молекулярно-биологическим лабораториям. Амплификация же определенных фрагментов ДНК с помощью специфических праймеров, приводящая за короткий промежуток времени к наработке значительных количеств нужных участков ДНК, явилась мощной и доступной альтернативой методу клонирования. Таким образом, многие лаборатории, ранее и не помышлявшие о секвенировании ДНК ввиду сложности подготовительных этапов, стали активно использовать этот метод в своих исследованиях. Поэтому можно смело предположить, что удельный вес ПЦР-секвенирования в общем определении нуклеотидных последовательностей различных ДНК будет несомненно возрастать. Что касается тех, кто и так раньше владел методами секвенирования ДНК, то они также взяли на вооружение метод ПЦР-секвенирования и занялись разработкой его различных модификаций, рассмотрению которых и будет посвящена данная глава.

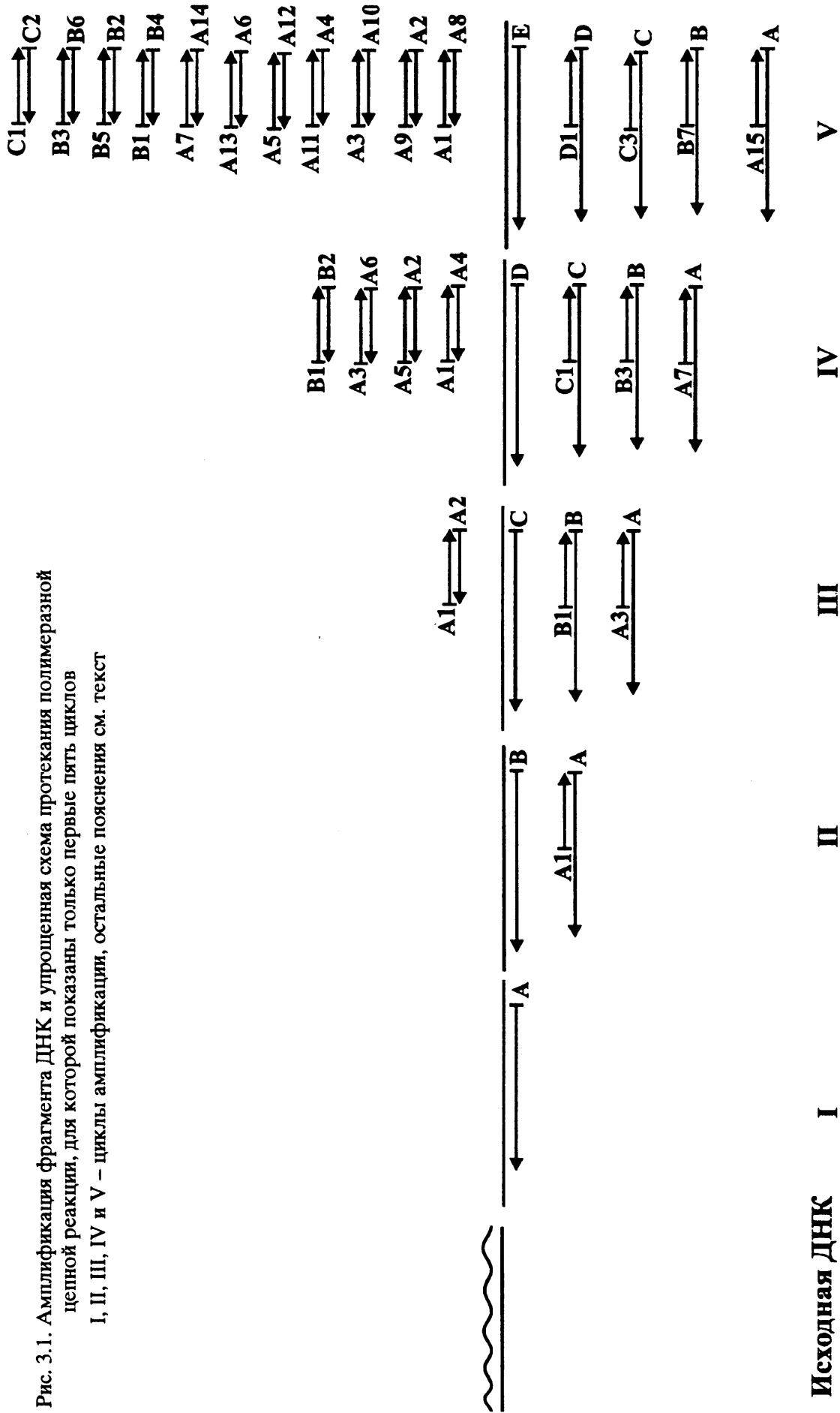
Но прежде чем перейти к подробному изложению, стоит еще немного задержаться на том “что же, собственно, подразумевается под ПЦР-секвенированием ДНК?”. Так, в большинстве случаев метод ПЦР просто служит для получения в требуемых количествах определенного участка ДНК, намечаемого к секвенированию, что в некоторой степени заменяет молекулярное клонирование или субклонирование этих же фрагментов ДНК. Определение нуклеотидных последовательностей ДНК непосредственно в ходе ПЦР в присутствии дидезокситерминаторов носит название “циклического секвенирования” и имеет свои особенности. Иногда трудно провести грань между этими

подходами, поскольку в циклическом секвенировании часто на первой стадии имеет место именно наработка материала для его дальнейшего секвенирования.

Несмотря на то, что ПЦР посвящено огромное количество статей, обзоров и целых книг (многие из которых приведены в данной главе как литература для дополнительного чтения), все же позволим себе очень кратко остановиться на самом принципе полимеразной цепной реакции, приводящей за короткий промежуток времени (от 10 мин до нескольких часов, что зависит от типа применяемого ДНК амплификатора, условий и задач эксперимента) к наработке значительных количеств целевого фрагмента ДНК.

Основными компонентами полимеразной цепной реакции, в ходе которой амплифицируется определенный участок молекулы ДНК, являются, во-первых, сама исходная ДНК, во-вторых, олигонуклеотидные праймеры, ограничивающие выбранный участок ДНК, фермент ДНК-полимеразы и “строительный материал” в виде смеси дНТФ. Дополнительные требования, предъявляемые к этим компонентам, заключаются в следующем. Исходная ДНК должна нести в своем составе тот участок, который намечено амплифицировать. Олигонуклеотидные праймеры из подобранной пары должны быть комплементарны противоположным цепям ДНК и при этом ограничивать выбранный участок так, что при их ферментативном удлинении вновь синтезируемые цепи ДНК будут расти навстречу друг другу. Из этого следует, что для выбора и синтеза самих олигонуклеотидных праймеров исследователю должна быть известна какая-то часть последовательности ДНК в выбранных регионах. (Отдельные случаи использования в качестве затравочных молекул праймеров с произвольной выбранной последовательностью нуклеотидов здесь не рассматриваются.) Кроме этого, на праймеры накладывается масса других ограничений с целью исключения их нежелательной интра- и интергомологии, формирования вторичной структуры, при этом необходимо учесть также температуры отжига каждого из них на матрице ДНК. Все это достаточно детально рассмотрено в рекомендуемых книгах и статьях [Rychlik, Rhoads, 1989; Lowe et al., 1990; Kwok et al., 1994; Rychlik, 1995], и в связи с этим нет особого смысла останавливаться на этом подробнее. Более того, существует множество специальных компьютерных программ, позволяющих выбрать наиболее оптимально подходящие праймеры для каждого эксперимента. Что касается фермента ДНК-полимеразы, то, несмотря на то, что в первых опытах по амплификации ДНК это был простой Кленовский фрагмент ДНК-полимеразы I и его использование для этой цели принципиально возможно и сейчас, в настоящее время необходимо применять какие-либо термостабильные ДНК-полимеразы, выдерживающие высокую температуру (95° С или около того) в течение значительного промежутка времени.

Периодический нагрев реакционной смеси до столь высокой температуры вызван циклическим характером протекания реакции амплификации ДНК и предназначен для денатурации цепей ДНК (как ис-



ходных, так и вновь синтезированных). Во время следующего этапа каждого цикла, проводимого обычно при температуре около  $50^{\circ}$ – $60^{\circ}$  С, происходит отжиг олигонуклеотидных праймеров на одноцепочечной ДНК. Далее следует повышение температуры до оптимальной для используемой термостабильной ДНК-полимеразы (обычно  $72^{\circ}$ – $75^{\circ}$  С), во время чего осуществляется ферментативное удлинение праймеров и построение таким образом новых комплементарных цепей ДНК. Цикл завершается. Для возобновления процесса и начала нового цикла опять необходимо наличие в реакционной смеси одноцепочечных молекул ДНК, что достигается кратковременным повышением температуры до тех же  $95^{\circ}$  С. После отжига праймеров на денатурированной одноцепочечной ДНК эти вновь образовавшиеся комплексы служат для фермента матрицами с затравкой. Цепной же данная реакция названа потому, что продукты, наработанные ДНК-полимеразой в ходе предыдущего(их) цикла(ов), используются в последующем(их) цикле(ах).

Процесс накопления продуктов в ходе ПЦР достаточно сложен, чтобы быть выраженным какой-нибудь простой (относительно, конечно) математической формулой. Так, кроме начавшейся с третьего цикла, амплификации целевых фрагментов ДНК, ограниченных с обеих сторон праймерами, с каждым новым циклом в цепную реакцию вовлекаются вновь синтезированные по гетерогенным матрицам подобные ампликоны. При этом одновременно продолжается увеличение числа самих гетерогенных матриц, комплементарных цепям исходного фрагмента ДНК. Как можно видеть из упрощенной (поскольку приведены нарабатываемые ПЦР-продукты только для одной из двух исходных цепей ДНК, обозначенной здесь прямой линией – впрочем, для комплементарной цепи, обозначенной волнистой линией, эта схема полностью справедлива) схемы протекания ПЦР, изображенной на рис. 3.1, в первом цикле образуются только гетерогенные по размеру цепи ДНК (продукт А), так как их терминация будет происходить произвольно. Только во втором цикле в реакционной смеси появляются целевые фрагменты ДНК, причем комплементарные цепи этого фрагмента (А1) на этом этапе спарены с гетерогенными матрицами и для большей наглядности расположены в нижней части схемы. В верхней части схемы, для которой на данном рисунке некой “демаркационной” линией служит одна из цепей исходного фрагмента ДНК, продолжающая также служить матрицей и в последующих циклах, изображены целевые фрагменты определенного размера, ограниченные по краям праймерами. Дальнейшее накопление таких фрагментов (А1, А2, А3, А4..., В1, В2... и т.д.) происходит экспоненциально, тогда как первый тип гетерогенных молекул (А, В, С, D, Е и т.д.) будет накапливаться только в арифметической прогрессии. Таким образом, например, после 25 циклов такие гетерогенные молекулы будут составлять ничтожную часть, которой, в подавляющем большинстве случаев, можно просто пренебречь.

### 3.1. СЕКВЕНИРОВАНИЕ КЛОНИРОВАННЫХ ФРАГМЕНТОВ ДНК, ПОЛУЧЕННЫХ С ПОМОЩЬЮ ПЦР

Одной из первых опубликованных работ, где было осуществлено секвенирование двух клонированных в векторе M13 фрагментов ДНК, полученных путем амплификации с использованием Кленовского фрагмента ДНК-полимеразы I в качестве фермента, была статья Шарфа и соавт [Scharf et al., 1986]. Чтобы упростить процедуру лигирования, последовательности праймеров, используемые в их работе, несли в своей 5'-области сайты узнавания рестрикционных эндонуклеаз, но, несмотря на это, эффективность клонирования составила только 1%. Далее полученные клоны секвенировали обычным дидезокси-методом.

Использование в ПЦР термостабильной ДНК полимеразы из бактерии *Thermus aquaticus* значительно упростило саму процедуру амплификации, поскольку отпала необходимость добавления новых порций фермента после этапов температурной денатурации, однако при этом появились другие проблемы. Одна из таких проблем заключалась в нематричном синтезе, осуществляемом Taq ДНК-полимеразой, и приводящем к образованию цепей ДНК с дополнительным выступающим нуклеотидом на 3'-конце, в большинстве случаев являющимся аденином [Clark, 1988; Hu, 1993]. Из-за этого дополнительного нуклеотида, присутствующего в подавляющем числе амплифицированных фрагментов ДНК, их лигирование с подготовленным вектором, имеющим тупые концы, протекает крайне плохо, что приводит к низкой эффективности клонирования амплификатов. Какие только ухищрения не предпринимаются исследователями, чтобы как-то обойти эту проблему! Используют праймеры с дополнительными участками, несущими сайты узнавания подходящих рестрикционных эндонуклеаз, обрабатывают концы амплификатов различными ферментами, готовят специальные вектора, осуществляют соединение вставки и вектора путем отжига их концов и пр.

Одним из активно используемых в настоящее время подходов является применение специальных, так называемых T-векторов, предварительно расщепленных какой-либо рестрикционной эндонуклеазой и несущих выступающие тимидины на своих 3'-концах, что достигается различными способами [Holton, Graham, 1991; Kovalic et al., 1991; Marchuk et al., 1991; Cease, Lohff, 1993; Cha et al., 1993; Ichihara, Kurosawa, 1993; Harrison et al., 1994; Iwahana et al., 1994]. В то же время сообщается и об успешном клонировании амплифицированных фрагментов ДНК в векторе с тупыми концами за счет присутствия среди ПЦР-продуктов и таких, у которых не имеется дополнительного нематричного нуклеотида на 3'-конце [Naqqi, 1992]. Другим решением проблемы клонирования амплификатов является "полирование" их концов с помощью Кленовского фрагмента ДНК-полимеразы I, T4 ДНК-полимеразы или использование в ПЦР термостабильных ДНК-полимераз, обладающих 3' → 5'-экзонуклеазной редактирующей активностью, приводящих к получению фрагментов ДНК с тупыми концами без дополнительных нематричных нуклеотидов [Lundberg et al., 1991; Mattila et al., 1991; Lohff, Cease, 1992].



Еще одна проблема заключается в том, что Taq ДНК-полимераза, как, впрочем, и другие ДНК-полимеразы, иногда включает “неверные” некомплементарные нуклеотиды, но ввиду отсутствия у нее 3' → 5'-экзонуклеазной редактирующей активности, они остаются в новосинтезированной цепи ДНК. Считается, что частота таких “ошибок” Taq ДНК-полимеразы достаточно высока и зависит от ряда факторов. К некоторым из них можно отнести или очень высокую концентрацию всех четырех дНТФ, или “истощение” какого-нибудь одного из трифосфатов, или очень низкую их концентрацию. Поскольку при клонировании в векторную молекулу встраивается только одна копия, то теоретически она может нести измененные нуклеотиды. Вероятность этого значительно повышается в случае, если Taq ДНК-полимераза допустила “ошибку” в ранних циклах ПЦР и соответственно тем большее число амплифицированных молекул будет нести такие измененные нуклеотиды. Таким образом, определение нуклеотидной последовательности одного такого клона не гарантирует достоверности полученных результатов. В какой-то степени проблема может быть решена секвенированием нескольких независимых клонов и лучше с клонированными амплификационными фрагментами из разных ПЦР, поскольку вероятность ошибочного включения дНМФ в одно и то же место матрицы ДНК невелика. Еще одно нежелательное явление в виде рекомбинации ДНК в ходе ПЦР, происходящей с частотой около 5% [Meyerhans et al., 1990; Odelberg, et al., 1995], также может быть преодолено секвенированием разных клонов [Raabo, Wilson, 1988]. Однако нельзя исключить определенный полиморфизм участков ДНК, особенно принадлежащих к повторяющимся элементам генома, вследствие чего выявляемые отличия между разными клонами могут быть, и не обязательно, результатом ошибок Taq ДНК-полимеразы. В то же время в литературе имеются сообщения о довольно высоком качестве построения новой комплементарной цепи ДНК Taq ДНК-полимеразой в оптимальных условиях [Eckert, Kunkel, 1990]. Так, в цитируемой работе авторы приводят доказательства, что и замена нуклеотидов, и сдвиг рамки считывания в ходе ПЦР происходят реже, чем 1 событие на 100000 нуклеотидов.

Непосредственно сама техника секвенирования клонированных амплификатов ничем не отличается от таковой для прочих клонированных фрагментов ДНК и подробно описана в предыдущей главе. В этой связи более не будем задерживаться на секвенировании клонированных фрагментов ДНК, полученных с помощью ПЦР, и перейдем к рассмотрению ПЦР-секвенирования ДНК, не предусматривающему клонирование.

### 3.2. ПРЯМОЕ СЕКВЕНИРОВАНИЕ ФРАГМЕНТОВ ДНК, ПОЛУЧЕННЫХ С ПОМОЩЬЮ ПЦР

Прямое секвенирование фрагмента ДНК, полученного в результате амплификации, в значительной мере снимало проблему ошибочно встроенных Taq ДНК-полимеразой нуклеотидов, поскольку в этом случае секвенировался сразу весь пул полученных в ПЦР молекул. Так,

при рассмотрении даже самого худшего сценария, где уже в первом цикле в одной из цепей единичной копии нужного гена произошло включение некоплементарного нуклеотида, то по завершении всего процесса ПЦР только около четвертой части всех молекул будет нести эту замену и поэтому его достоверное “прочтение” все же возможно. Допустить то, что в первом же цикле произойдут замены какого-то конкретного нуклеотида и также комплементарного ему в другой цепи ДНК, просто невероятно. Необходимо отметить, что в этом сценарии мы исходили из присутствия в реакционной смеси изначально всего одной копии амплифицируемого фрагмента, что случается крайне редко. Неправильное включение одного и того же нуклеотида в разных матрицах ДНК кажется весьма маловероятным событием и поэтому нуклеотидная последовательность, выявляемая прямым секвенированием амплификатов, вполне достоверна. Хотя, все же нельзя исключать каких-то особенностей отдельных участков одноцепочечных матриц ДНК, например, в виде прочной вторичной структуры, сохраняющейся даже при высокой температуре, которая может вызывать в таких местах “сбои” Taq ДНК-полимеразы. Если же включение “неправильного” нуклеотида произойдет в более поздних циклах, то вклад такого ошибочно встроенного нуклеотида может быть даже незаметен. Как уже отмечалось выше, при выявлении гетерогенных нуклеотидов в определенном месте фрагмента нельзя сбрасывать со счетов и возможный полиморфизм исходных молекул ДНК, особенно в случаях секвенирования каких-либо повторяющихся генов или прочих повторов ДНК.

Также следует иметь в виду, что при прямом секвенировании повторяющихся элементов генома, наработанных в ходе ПЦР, будет определяться их некая консенсусная нуклеотидная последовательность. Более того, в случае сильного полиморфизма отдельных участков этих повторов, делеций или вставок одного или нескольких нуклеотидов (что для многих повторяющихся элементов бывает обычным явлением), определение нуклеотидной последовательности таких амплификатов, скорее всего, будет невозможно. Некоторое решение проблемы в подобном случае заключается все-таки в клонировании и последующем секвенировании довольно большого числа независимых амплификатов с тем, чтобы уже на основе их последовательностей составить усредненную, если в таковой будет необходимость.

В то же время следует отметить, что в первых работах по секвенированию фрагментов ДНК, наработанных с помощью ПЦР [McMahon et al., 1987; Wrischnik et al., 1987; Engelke et al., 1988], проблемы ошибочно включенных дНМФ не возникало в связи с тем, что для амплификации использовался Кленовский фрагмент ДНК-полимеразы I, характеризующийся довольно высокой точностью синтеза комплементарной цепи ДНК. Особенностью двух последних работ было применение специального секвенирующего внутреннего или “вложенного” праймера, место отжига которого находится в амплифицированном фрагменте. Такой праймер отжигался только на “нужном” амплификате и позволял осуществлять секвенирование в присутствии гетерогенных фрагментов ДНК, полученных в ходе ПЦР и содержащих случайные после-

довательности. Данный подход в ряде случаев необходим, поскольку олигонуклеотидные праймеры могут отжигаться не только в предполагаемом месте, но и, в силу ряда причин, в каких-либо других местах матрицы ДНК. Так, например, если в первых циклах почему-то произошел отжиг праймеров с участками ДНК, не обнаруживающими с ними полной гомологии, то во всех последующих циклах гомология этого участка у новосинтезируемых цепей ДНК с праймерами будет уже полная. Однако внутренняя последовательность таких ложных амплификатов будет совсем иной, в связи с чем третий праймер на ней не отождествится. Соответственно, если секвенирующий праймер будет одним из амплификационных, то он отождествится на таком амплификате без каких-либо проблем, что приведет к невозможности определения последовательности ДНК, поскольку будет секвенироваться весь пул этих молекул, так как в каждой из них присутствует последовательность этого праймера (частные случаи экспоненциальной амплификации фрагментов ДНК с одним праймером здесь в расчет не принимаются). Надо заметить, что иногда проблема образования в ходе ПЦР гетерогенных фрагментов ДНК может быть решена весьма просто – путем повышения температуры отжига олигонуклеотидных праймеров или изменением состава реакционного буфера. Возможно применение подходов с различными вариантами “горячего старта”. Сообщается также о так называемом энхансере специфичности, заключающемся в добавлении в реакционную смесь тетраметиламмония хлорида [Hung et al., 1990]. Другое кардинальное решение может заключаться в выборе и синтезе новых праймеров.

Определенную проблему в секвенировании двуцепочечных фрагментов ДНК, полученных в ходе ПЦР, составляет то, что на этапе отжига секвенирующего праймера происходит и гибридизация комплементарных цепей данного фрагмента ДНК, приводящая к формированию исходной двуцепочечной молекулы. Этому способствует также и то, что амплификаты, предназначенные для секвенирования, представляют собой, как правило, не очень крупные молекулы ДНК, следовательно, скорость этого процесса может быть весьма значительной, и требующееся на это время может оказаться сравнимым со временем отжига короткого олигонуклеотидного праймера. Таким образом, имеет место некая конкуренция между олигонуклеотидным праймером и одной из цепей ДНК за место отжига на комплементарной им цепи. Чтобы хоть как-то уменьшить отрицательное влияние этого процесса на секвенирование ДНК, были предложены различные добавки в буфер для отжига. Так, в одной работе рекомендовалось применение 10%-ного диметилсульфоксида [Winship, 1989], в другой – различных неионных детергентов [Bachmann et al., 1990], причем наилучшие результаты показал нонидет NP-40 и твин-20.

Другой отрицательный эффект от присутствия в реакционной смеси второй комплементарной цепи состоит в образовании неспецифических “стопов”, происходящих из-за “утыкания” ДНК-полимеразы, удлиняющей нормально отождествленный праймер, в двуцепочечный участок матрицы, образовавшийся в результате отжига на ней комплементарной

цепи ДНК, тогда как остальная часть цепи оказалась неспаренной, что и дало возможность ДНК-полимеразе до поры, до времени, а точнее, до данного места, осуществлять синтез второй цепи ДНК [Kusakawa et al., 1990; Rao, Saunders, 1992; Rao, 1994]. Подобные места могут быть или произвольно распределены по всей длине матрицы ДНК, или располагаться в некоторых участках, где в силу определенной нуклеотидной последовательности восстановление двуцепочечной структуры происходит особенно быстро [Olsen, Eckstein, 1989]. На заключительном радиоавтографе для таких мест будет характерно наличие полос во всех четырех дорожках, не позволяющее осуществить достоверное “прочтение” этих нуклеотидов. В некоторой степени это можно преодолеть довольно быстрым охлаждением раствора денатурированной матрицы до 12° С в присутствии избытка олигонуклеотидного праймера, поскольку более короткий праймер все же успевает в этих условиях отжечься на комплементарном участке. В этой же работе было показано, что видимый в агарозном геле как единая полоса ПЦР-продукт в действительности, при его анализе в секвенирующем геле, представляет собой достаточно гетерогенную смесь основного продукта и более коротких полупродуктов, причем последние при секвенировании также приведут к упомянутым выше “стопам”. В работе других авторов [Casanova et al., 1990] была исследована эффективность секвенирования двуцепочечного ПЦР-продукта в зависимости от времени отжига олигонуклеотидного праймера. Наилучшие результаты были получены в случаях максимально быстрого охлаждения, занимающего всего 15 с, что достигалось путем помещения пробирки с матрицей, денатурированной нагревом в присутствии праймера, на –70° С.

Еще одна проблема секвенирования продуктов амплификации заключается в неиспользованном в ходе ПЦР большом количестве дНТФ и олигонуклеотидных праймеров, что делает невозможным прямое секвенирование ПЦР-продуктов без соответствующих процедур удаления этих ингредиентов. Обычно в ПЦР используется не более 10% праймеров и их избыток вместе со взятой аликвотой для секвенирования, таким образом, перенесется в следующую реакцию. Поскольку в качестве секвенирующего праймера может быть добавлен или третий праймер, или даже один из амплификационных, предварительно меченный по своему 5'-концу, то олигонуклеотидные праймеры из первой реакции не могут сделать секвенирование ДНК невозможным. Но они будут конкурировать с секвенирующим праймером за места отжига, занимать молекулы ДНК-полимеразы и расходовать дНТФ и ддНТФ и за счет этого снижать выход целевого продукта. Все это вместе приведет к ухудшению результатов секвенирования амплификатов.

Избыток дНТФ, перенесенный вместе с ПЦР-продуктом в реакцию секвенирования, представляет даже большую проблему. Обычная концентрация дНТФ в ПЦР составляет 0,2 мМ для каждого нуклеотида. После 25 циклов, как правило, используется около 3% от их исходного количества, что зависит от длины амплифицируемого фрагмента и эффективности всего процесса. Однако такая высокая концентрация дНТФ требуется отчасти из-за небольшой продолжительности этапа

удлинения комплементарной цепи ДНК и необходимости его полного завершения, поскольку в противном случае экспоненциальное течение реакции будет невозможно. Таким образом, присутствие очень большого количества дНТФ не позволяет на этапе секвенирования ДНК выполнить необходимые соотношения дНТФ/ддНТФ, поскольку последних придется добавить так много, что проект по секвенированию ДНК с помощью ПЦР и Taq ДНК-полимеразы станет чрезмерно дорогим.

Эффективное секвенирование ПЦР-продуктов требует удаления дНТФ и праймеров из первой реакции, остающихся в избытке после ее завершения, что может быть осуществлено с помощью различных методов, основывающихся, главным образом, на заметной разнице в размерах дНТФ (менее 600 Да), олигонуклеотидных праймеров (в среднем – около 8000 Да и обычно не превышающих 12 000 Да) и самих продуктов ПЦР в виде двуцепочечных фрагментов ДНК, характеризующихся значительно большей молекулярной массой [Gyllensten, 1989]. Достаточно простым способом является осаждение ДНК, полученной в ходе ПЦР, в условиях неосаждения дНТФ и коротких фрагментов ДНК, каковыми являются праймеры. Сообщается об использовании для этой цели полиэтиленгликоля [Kusukawa et al., 1990], количественно осаждающего фрагменты ДНК крупнее 150 пн, тогда как более мелкие он осаждаёт плохо или совсем не осаждаёт [Paithankar, Prasad, 1991]. Аналогичное действие оказывает изопропанол. Применяется также ультрафильтрация с пороговыми величинами 30 000 Да [Nichols, Raben, 1994] или 100 000 Да [Krowczynska, Henderson, 1992], ниже которых ингредиенты предыдущей реакции удаляются. Этот же принцип ультрафильтрации лежит в основе спин-диализа в микроцентрифуге с помощью специальных микроконцентраторов Centricon 30 фирмы “Amicon” (США), продемонстрировавшего эффективное удаление оставшихся в ходе ПЦР дНТФ и олигонуклеотидных праймеров, что позволило провести успешное секвенирование полученного продукта [Gal, Hohn, 1990; Wong et al., 1987].

Следует отметить описанную возможность ферментативного удаления неиспользованного количества праймеров после завершения первой стадии ПЦР с помощью экзонуклеазы VII [Li et al., 1991]. Экзонуклеаза VII проявляет свою активность по отношению к одноцепочечным фрагментам ДНК, начиная одновременный гидролиз как с 5'-, так и с 3'-концов, не оказывая при этом заметного действия на двуцепочечные продукты амплификации. Был предложен и другой подход, требующий, как и при осаждении фрагментов ДНК, только одну пробирку и заключающийся в ферментативном разрушении с помощью экзонуклеазы I и щелочной фосфатазы краба неиспользованных дНТФ и праймеров [Werle et al., 1994]. Метод относительно прост и исключает потерю ДНК во время многочисленных процедур, применяемых при очистке ПЦР-продуктов более сложными методами, описанными ниже.

Как уже отмечалось выше, в ходе ПЦР случается, что образуется смесь гетерогенных молекул ДНК, и тогда простое осаждение или ферментативная обработка, хотя и исключат из дальнейшего этапа секвенирования мешающие дНТФ и праймеры из первой реакции, однако не

решат проблему наличия посторонних ДНК. В этой связи было предложено проводить очистку амплификатов с помощью электрофореза в агарозном или полиакриламидном гелях или хроматографией, что одновременно удаляло непрореагировавшие дНТФ и праймеры, а также позволяло в дальнейшем проводить секвенирование именно того фрагмента ДНК, который и был намечен. После проведения электрофоретического разделения дальнейшие процедуры по очистке нужных фрагментов ДНК из геля могут быть практически любыми, начиная от простого замораживания/оттаивания до электроэлюции [Маниатис и др., 1984; Sambrook et al., 1989; Cao, Brosius, 1993]. Описанные в литературе способы выделения фрагментов ДНК из геля, позволяющие осуществить дальнейшее эффективное секвенирование, состоят в их элюции на ДЭАЭ-бумагу [Nickrent, 1994], в разрушении агарозы под действием агаразы [Gold, 1992] или сорбции биотинилированной ДНК (за счет одного из праймеров, несущего биотин) из расплава геля на магнитных частицах со стрептавидином [Green et al., 1990]. Использование других магнитных частиц, покрытых карбоксилем, способных в присутствии полиэтиленгликоля и NaCl обратимо сорбировать ДНК, уже не требовало наличия биотиновой метки [DeAngelis et al., 1995]. Сообщается также о разделении продуктов амплификации в легкоплавкой агарозе и использовании в дальнейшем секвенировании с помощью секвенназы данного ПЦР-продукта прямо вместе с расплавленным кусочком легкоплавкой агарозы, не прибегая к элюции, что, впрочем, не позволило на этапе мечения снижать температуру инкубации ниже 37° С, во избежание образования при этом в реакционной смеси геля [Kretz et al., 1989; Kretz et al., 1990].

Был предложен интересный подход к элюции ПЦР-фрагмента ДНК из полиакриламидного геля, заключающийся в предварительном высушивании геля, окрашенного бромистым этидием, и с наложенным на него кусочком фильтровальной бумаги, вырезанному по размеру полосы ДНК [Wu et al., 1996]. Далее, поскольку размер фильтровальной бумаги строго соответствовал размеру исследуемой полосы ДНК, то вырезался кусочек сухого полиакриламидного геля, исходя из контуров бумаги. После прогрева в течение 15 мин в микроцентрифужной пробирке этого кусочка геля вместе с бумагой, последняя удалялась центрифугированием. Есть работы, в которых элюция из полиакриламидного геля продукта ПЦР для его последующего секвенирования осуществлялась простой диффузией в солевой раствор в течение ночи [Huber, McMahon, 1988]. Недавно было сообщено об использовании специального небулайзера (Gel Nebulizer) (Millipore, США) для выделения ПЦР-продукта из агарозного геля, пригодного для последующего его секвенирования [Leonard et al., 1998]. Это устройство представляет собой своеобразную колонку с соплом, рассчитанную на центрифугирование в микроцентрифуге при 10 000g, и предназначено для элюции ДНК из образующихся в этом процессе мельчайших кусочков агарозного или полиакриламидного гелей. Причем данная процедура требует всего около 10 мин. Также весьма быстрая очистка ПЦР-продуктов, завершаемая всего за 7–20 мин, возможна с помощью хроматографии высокого давления [Katz, Dong,

1990; Kalnoski et al., Warren, Doninger, 1991], что, однако, требует наличия специального дорогостоящего оборудования.

В то же время в литературе встречаются сообщения об успешном секвенировании ПЦР-продуктов без какой-либо их очистки [Meltzer et al., 1990; Srivastava et al., 1992; Douglas et al., 1993; Liu et al., 1993]. Непременным условием этого является, как уже отмечалось выше, наличие единственного целевого продукта и отсутствие гетерогенных амплификатов. Другим условием является использование в ПЦР относительно низких концентраций в реакционной смеси как дНТФ, так и олигонуклеотидных праймеров. Так, в качестве примера можно привести работу, в которой в первую фазу (30 циклов) осуществлялась амплификация фрагмента ДНК с помощью ДНК-полимеразы *Thermus flavus* в присутствии всего 10 мкМ каждого из дНТФ, а во вторую фазу, после добавления меченого праймера (одного из амплификационных) и соответствующих количеств ддНТФ уже проводились терминирующие реакции [Rao, Saunders, 1992]. В другой работе в качестве фермента использовалась ДНК-полимераза *Deer Vent<sub>r</sub>*<sup>®</sup>(exo<sup>-</sup>) из археобактерии *Rugosoccus* sp., способная к эффективному построению цепей ДНК при достаточно низких концентрациях дНТФ, что также не потребовало удаления неключившихся ингредиентов из первой реакции [Magiame, 1996].

Хорошо известно, что гораздо лучшие результаты дает секвенирование одноцепочечной ДНК, которая не содержит в реакционной смеси комплементарной цепи. Существует много подходов к достижению этой цели как на этапе самой ПЦР, так и после ее завершения. Самым простым вариантом можно считать подход, заключающийся в проведении второй стадии ПЦР уже только с одним праймером с небольшой алиquotой, взятой из первой реакции [Kaltenboeck et al., 1992]. Хотя необходимо отметить, что вместе с ДНК-продуктом в этом случае перенесется и второй праймер, который будет присутствовать хотя и в меньшей концентрации, но все же будет принимать участие в амплификации, приводя к образованию и второй цепи ДНК. Таким образом, можно считать, что амплификация проводится с двумя праймерами, имеющими разную концентрацию, что и было реализовано в виде асимметричной ПЦР, описанной ниже. Так, достаточно простой способ наработки одной из цепей фрагмента ДНК состоит в асимметричной амплификации, которая становится возможной из-за значительной разницы (в 50–100 раз) в количестве присутствующих в реакционной смеси двух праймеров [Gyllensten, Erlich, 1988; Innis et al., 1988; Wilson et al., 1990]. В результате во время ПЦР происходит истощение первого праймера (представленного в низкой концентрации), что приводит к накоплению в последующих циклах одноцепочечных продуктов удлинения второго праймера (рис. 3.2). Серьезными недостатками этого подхода являются неоптимальные условия амплификации, заметно снижающие выход целевого продукта, и необходимость в отдельных случаях экспериментальным путем подбирать оптимальное количественное соотношение этих двух праймеров.

Подобный эффект достижим и при использовании двух праймеров, имеющих сильно отличающуюся температуру отжига. Этот про-

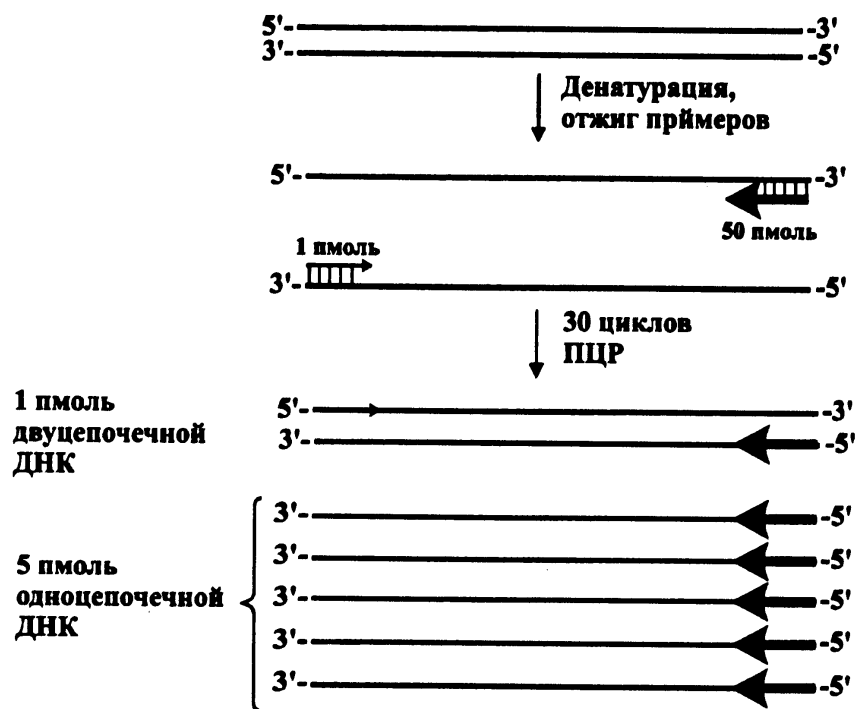


Рис. 3.2. Асимметричная ПЦР, приводящая к накоплению одной из цепей амплифицируемого фрагмента ДНК за счет “истощения” одного из праймеров

цесс, названный авторами [Mazars et al., 1991] “температурной асимметричной ПЦР”, заключается в проведении первых 25 циклов при низкой температуре отжига, характерной для одного из праймеров и приемлемой одновременно и для второго, и уже затем следует проведение дальнейших циклов при более высокой температуре, соответствующей температуре отжига второго праймера, что не позволяет, таким образом, отжечься первому (рис. 3.3). Результатом этого будет накопление в этих циклах второго этапа только одной цепи ДНК, представляющей собой продукт ферментативного удлинения второго праймера.

Своеобразный способ “удаления” из реакции одной из цепей ДНК, полученной в ходе ПЦР, осуществляется путем ее гибридизации с клонированным ранее в соответствующей ориентации подобным (гомологичным) фрагментом в одноцепочечном векторе M13 [Gal, Hohn, 1990; Gal, 1993, 1996]. Причем клонированный фрагмент не должен содержать места отжига амплификационных и секвенирующего праймеров, чтобы не вызывать конкуренцию. В цитируемой работе показано улучшение качества картины секвенируемых полос ДНК после добавления в реакционную среду возрастающих количеств одноцепочечной ДНК вектора M13, несущего вставку, что объясняется одновременным увеличением доступного для ДНК-полимеразы числа копий нужной цепи секвенируемого ПЦР-фрагмента. Однако подобный подход применим только в специальных случаях секвенирования новых членов какого-либо семейства повторяющейся ДНК или гомологичных генов других организмов, поскольку предполагает наличие заранее клонированного участка такого фрагмента ДНК.



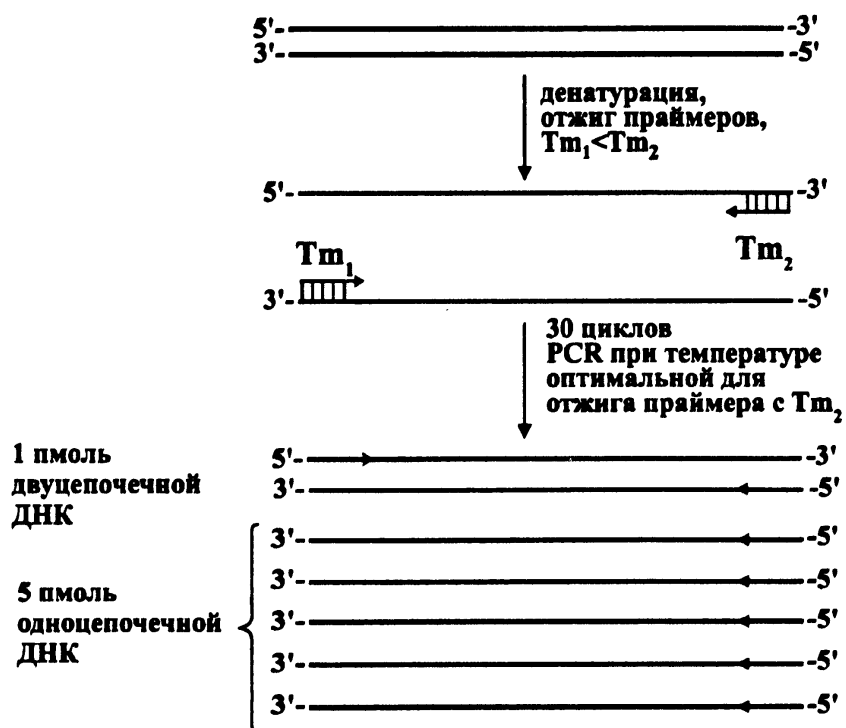


Рис. 3.3. Температурная асимметричная ПЦР, приводящая к накоплению одной из цепей амплифицируемого фрагмента ДНК

Интересен способ получения из двуцепочечного ПЦР-продукта одноцепочечной матрицы, пригодной для ее секвенирования обратной транскриптазой [Stoflet et al., 1988]. Особенностью этого метода является то, что матрица представляет собой РНК-копию ПЦР-фрагмента ДНК. Такое становится возможным благодаря специальному праймеру (одному из пары амплификационных), несущему, кроме комплементарной последовательности к секвенируемому фрагменту ДНК, участок промотора фагов T7, или T3, или SP6 [Lind et al., 1996]. Таким образом, после завершения самой ПЦР в реакционную смесь добавляется фермент T7 РНК-полимераза, способная за короткий промежуток времени наработать значительное число молекул РНК, представляющих собой, в данном случае, матрицы для секвенирования (рис. 3.4).

Другой ферментативный подход к созданию одноцепочечной матрицы заключается не в построении, а наоборот, в удалении одной из цепей ДНК (рис. 3.5). Известно, что экзонуклеаза фага лямбда разрушает цепь ДНК в направлении  $5' \rightarrow 3'$ , однако для проявления этой ее активности 5'-конец цепи ДНК должен нести остаток фосфорной кислоты. Так, использование во время амплификации фосфорилированного праймера (одного из пары) позволило затем с помощью экзонуклеазы фага лямбда удалить эту фосфорилированную цепь ПЦР-фрагмента ДНК [Higuchi, Ochman, 1989; Takagi et al., 1993; Berg, Olaisen, 1994].

Удаление одной цепи ПЦР-продукта с помощью экзонуклеазы III [Ward et al., 1989] основано на факте возможной защиты другой цепи тиопроизводным дНТФ, которое данный фермент “не узнает” [Putney et al., 1981]. Такая защита одной цепи ПЦР-продукта достигалась путем его расщепления какой-либо подходящей рестрикционной эндонуклеа-

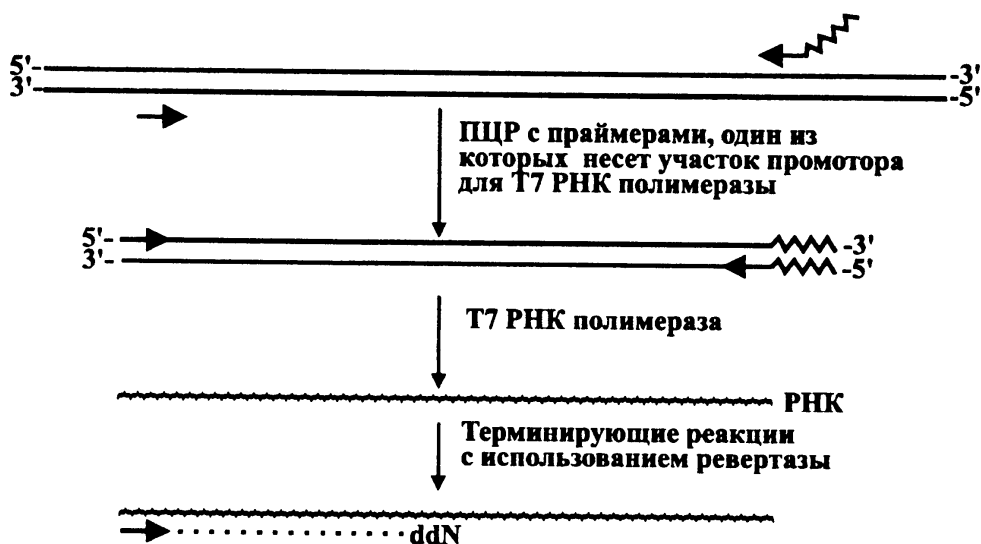


Рис. 3.4. Получение одноцепочечной матрицы в виде РНК-транскрипта продукта ПЦР с помощью T7 РНК-полимеразы

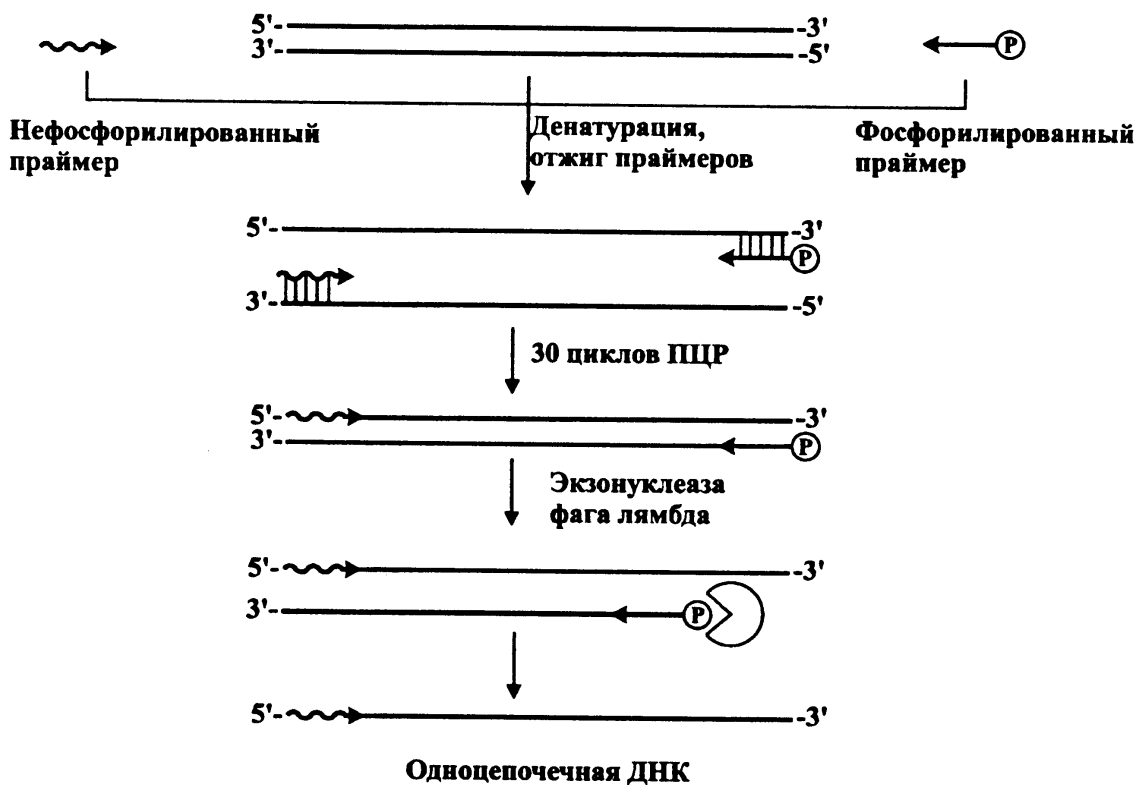


Рис. 3.5. Создание одноцепочечной матрицы за счет удаления одной из цепей ПЦР-фрагмента ДНК с помощью экзонуклеазы фага лямбда

зой и последующим ферментативным действием Кленовского фрагмента ДНК-полимеразы в присутствии трех обычных дНТФ и одного тио-дАТФ. Однако необходимо отметить, что данный способ применим далеко не всегда и, главным образом, из-за ограничений, вызванных самой нуклеотидной последовательностью амплифицируемого фрагмента ДНК. В другой работе сообщается об использовании для удаления одной из цепей фрагмента ДНК экзонуклеазы, кодируемой геном б фа-

га T7 и успешном проведении затем ПЦР-секвенировании [Somers et al., 1998].

Один из предложенных способов отделения одной цепи от другой для ПЦР-наработанных фрагментов ДНК, получивший в дальнейшем серьезное развитие, заключался в аффинной хроматографии полученных продуктов на колонке со стрептавидин-агарозой [Mitchell, Merril, 1989]. Такое становилось возможным благодаря использованию при амплификации ДНК биотинилированного праймера, поскольку в результате дальнейшей элюции с такой колонки с использованием 0,2 N NaOH как элюента происходила денатурация ДНК и выход биотиннесодержащей цепи ДНК. Что касается биотинилированной цепи ДНК, то она в этом случае оказывалась прочно связанной со стрептавидином и оставалась в колонке. Таким образом, из двух цепей ДНК только одна оказывалась доступной для секвенирования и “виновата” в этом была так называемая твердая фаза в виде не совсем удобной стрептавидин-агарозы.

### 3.3. ТВЕРДОФАЗНОЕ СЕКВЕНИРОВАНИЕ ДНК

Разработанные магнитные частицы, покрытые тем же стрептавидином, оказали революционизирующее влияние на дальнейший ход событий. Удобство обращения с ними позволило получать в чистом виде разные цепи ДНК, пригодные для их дальнейшего секвенирования (при условии, что одна из цепей несла метку в виде биотина). В литературе имеется значительное количество статей, описывающих применение магнитных частиц со стрептавидином для секвенирования ДНК [Hultman et al., 1989; Uhlen, 1989; Wahlber et al., 1990; Hultman et al., 1991; Lee, Vacquier, 1993; Verhasselt et al., 1993; Kofoid, 1995; Wang et al., 1995]. Такое секвенирование ДНК с помощью магнитных частиц получило название “твердофазного секвенирования” и используется в настоящее время довольно широко, особенно для определения нуклеотидной последовательности ПЦР-продуктов. Причем следует отметить, что почти все упомянутые в предыдущем разделе методы, направленные на получение одной цепи ДНК продуктов ПЦР, позволяют, как правило, обеспечить только обогащение, в той или иной степени, реакционной смеси какой-то одной цепью. А способ деления цепей, основанный на аффинной сорбции биотинилированной ДНК на магнитных частицах со стрептавидином, обеспечивает практически их 100%-ную очистку, благодаря, во-первых, прочному взаимодействию комплекса биотин/стрептавидин и, во-вторых, легкости разделения фаз (магнитные частицы и водный раствор), в которых и оказываются сосредоточенными разные цепи ДНК (рис. 3.6).

Так, в результате проведения ПЦР с олигонуклеотидными праймерами, один из которых несет молекулу биотина на своем 5'-конце, продукты амплификации сорбируются на магнитные частицы (например, Dynabeads M-280 Streptavidin фирмы “Dyna”, Норвегия). Далее следует этап отмывки, удаляющий все ставшие ненужными ингредиенты (дНТФ, праймеры и пр.) ПЦР. Затем в результате денатурации с помо-

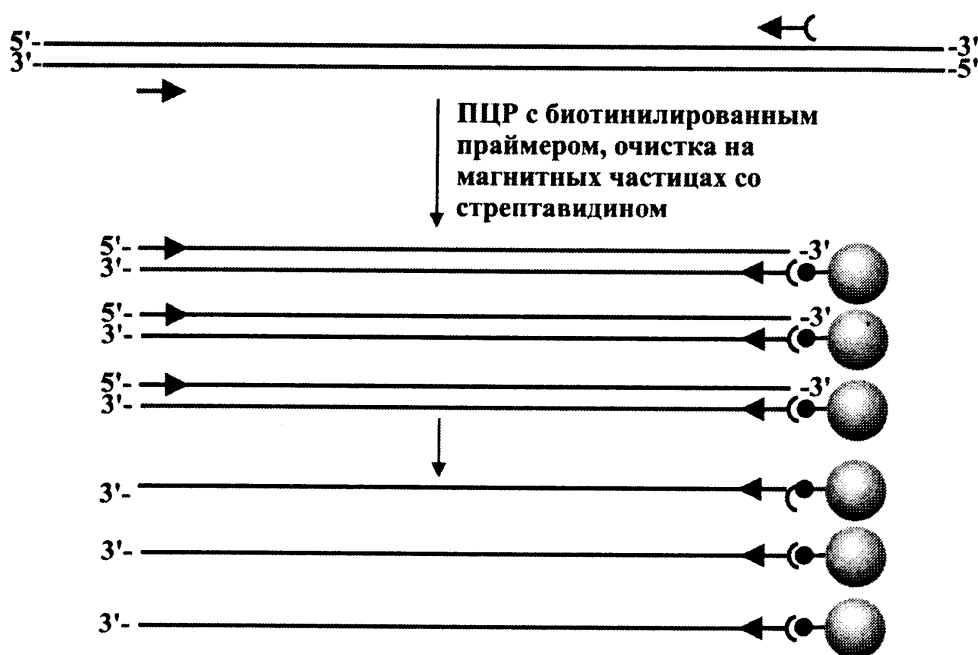


Рис. 3.6. Деление цепей двуцепочечного ПЦР-фрагмента ДНК с помощью аффинной сорбции на твердой фазе в виде магнитных частиц за счет формирования биотин/стрептавидинового комплекса

щью щелочи с магнитными частицами остается связанной только та цепь ДНК, которая была построена путем удлинения биотинилированного праймера, а вторая цепь выходит в раствор. Удерживая магнитом магнитные частицы у боковой стенки пробирки, отбирают раствор, содержащий вторую цепь ДНК. Магнитные частицы промывают от остатков щелочи и после отжига на них соответствующего праймера проводят секвенирующие реакции в присутствии ддНТФ. По их завершении проводится еще один этап денатурации и продукты терминирующих реакций наносятся на секвенирующий гель.

Что касается второй небитинилированной цепи ДНК, элюированной раньше, то после нейтрализации она может быть также секвенирована с соответствующим праймером. В качестве нейтрализующего агента предлагается использовать соляную кислоту [Hultman et al., 1991], что, впрочем, требует очень точного соблюдения нужного количества этого реагента. Менее строгие условия возможны при использовании для нейтрализации 3М ацетата натрия или 5М ацетата аммония. Другой подход заключается в проведении двух этапов спиндиализа через миниконцентратор Centricon-30 [Lee, Vacquier, 1993]. В литературе описывается использование для денатурации цепей ДНК, сорбированных на магнитных частицах со стрептавидином, гидроокиси аммония, обработка которым проводилась при комнатной температуре [van den Boom et al., 1998]. Далее, после удаления вышедшей в раствор цепи ДНК, биотинилированную цепь ДНК элюировали с магнитных частиц с помощью той же гидроокиси аммония, но при температуре 60° С.

Альтернативой этим процедурам является возможность одновременного использования в ПЦР второго также меченого амплификаци-

онного праймера, несущего молекулу дигоксигенина. Таким образом, в результате ПЦР образуются фрагменты, обе цепи которых несут меченые молекулы: одна – биотин, другая – дигоксигенин, что дает возможность после эволюции второй небiotинилированной дигоксигениновой цепи с магнитных частиц осуществить повторную сорбцию этой цепи на новые магнитные частицы с антителами к дигоксигенину и провести дальнейшие процедуры по вышеописанной для биотина схеме. В работе Лоусона и соавт. были исследованы особенности твердофазного секвенирования одноцепочечной матрицы, сорбированной на магнитных частицах, покрытых стрептавидином, разных фирм-производителей и при этом было показано, что добавление в среду 1%-ного казеина способствует уменьшению неспецифического связывания и улучшает качество получаемых результатов [Lawson et al., 1996].

В литературе описан другой оригинальный подход к проведению ПЦР с использованием двух меченых праймеров с последующим секвенированием продуктов реакции. Один из этих праймеров несет молекулу биотина, а второй, в виде своеобразной метки, содержит дополнительный участок размером 21 нуклеотид, представляющий собой *lac*-оператор [Wahlberg et al., 1990]. После иммобилизации ПЦР-продуктов на магнитных частицах со стрептавидином в реакционную смесь добавляется комплекс *lac*-репрессор (*LacI*) и  $\beta$ -галактозидаза (*LacZ*) и хромогенный субстрат О-нитрофенил  $\beta$ -D-галактозид, что позволяет выявить нужные образцы для их дальнейшего секвенирования и провести деление цепей с помощью щелочи, как описано выше.

Завершая рассмотрение твердофазного секвенирования, следует еще раз кратко остановиться на преимуществах этого подхода. Так, пожалуй, главное, что обеспечивает аффинная сорбция биотинилированной ДНК на магнитных частицах, покрытых стрептавидином, так это быстрое и высокоэффективное деление цепей секвенируемого фрагмента ДНК, обеспечивающее их практически 100%-ную очистку, и доступность обеих этих цепей в качестве одноцепочечных матриц для ферментативного секвенирования. Кроме этого, сорбция продуктов ПЦР на твердой фазе в виде магнитных частиц позволяет легко удалить ингредиенты реакции амплификации, такие как дНТФ, праймеры, соли. Некоторым недостатком можно считать более низкую эффективность сорбции биотинилированных фрагментов ДНК протяженностью свыше 3 тпн на магнитных частицах со стрептавидином [Wahlberg et al., 1995]. Для преодоления этого было предложено вводить в праймеры множественные биотинилированные метки, приводящие к упрочнению связывания матрицы ДНК с подобными магнитными частицами [Kasai et al., 1992]. В литературе встречаются указания на имеющуюся возможность повторного секвенирования (т.е. проведения терминирующих реакций) биотинилированной цепи ДНК, сорбированной на твердой фазе, после удаления второй вновь синтезированной комплементарной цепи ДНК в случае не совсем удачного результата после первой попытки. Однако, справедливости ради, следует отметить, что для проведения терминирующих реакций в растворе берут, как правило, не все количество имеющейся матрицы ДНК, что также дает возможность повторения этого эксперимента.

Другой аспект использования твердой фазы в секвенировании ДНК заключается в возможности удаления как исходной матрицы, так и прочих ингредиентов реакции, проводимой в растворе, за счет связывания новосинтезированной цепи ДНК, несущей биотинилированный праймер, с магнитными частицами со стрептавидином. Так, в одной работе было показано, что нанесение на секвенирующий гель только новосинтезированной цепи ДНК улучшает качество разделения [Tong, Smith, 1993].

Еще одной важной особенностью твердофазного секвенирования биотинилированных цепей ДНК, сорбированных на магнитных частицах, является возможность использования этого подхода не только при прямом секвенировании ПЦР-продуктов, рассмотренном выше, но и в экспериментах по циклическому секвенированию ДНК.

### 3.4. СЕКВИРОВАНИЕ ФРАГМЕНТОВ ДНК С ПОМОЩЬЮ ЛИНЕЙНОЙ ПЦР ИЛИ ЦИКЛИЧЕСКОЕ СЕКВИРОВАНИЕ ДНК

Как уже отмечалось в самом начале данной главы, в отдельных случаях очень трудно отнести тот или иной процесс к определенному типу ПЦР-секвенирования ДНК. Хотя, следует сказать, что в последнее время стали появляться работы, в которых описываются процессы, когда действительно происходит накопление продуктов секвенирующих (терминирующих) реакций и одновременное экспоненциальное увеличение числа амплифицируемых и секвенируемых молекул.

Главное отличие циклического секвенирования ДНК от большинства описанных в предыдущем разделе, пожалуй, заключается в особенностях проведения терминирующих реакций. Так, прямое секвенирование фрагментов ДНК, полученных в ходе ПЦР, может проводиться с помощью любой ДНК-полимеразы, включая и нетермостабильную секвеназу. Причиной тому служит весьма значительное количество матрицы ДНК, позволяющей осуществить ее детекцию в секвенирующем геле. А ПЦР в данном случае просто служит для наработки необходимого количества сначала двуцепочечной, а затем и одноцепочечной ДНК. Секвенирование же ДНК с помощью так называемой линейной ПЦР осуществляется посредством циклических этапов денатурации, отжига одного праймера и удлинения цепи в условиях ее терминации дидезоксинуклеотидами, причем последние присутствуют в смеси, начиная с первого цикла (рис. 3.7). Ввиду того, что в этих случаях происходит амплификация ДНК только в арифметической прогрессии (не приводящая даже к образованию полноразмерных продуктов), то исходным материалом, несущим изначально довольно большое число копий секвенируемого фрагмента, может служить, например, ранее клонированная ДНК. С помощью всего 30 циклов становится возможным получение такого количества продуктов реакции амплификации/терминирования, которое уже будет видно после электрофоретического разделения в секвенирующем геле.

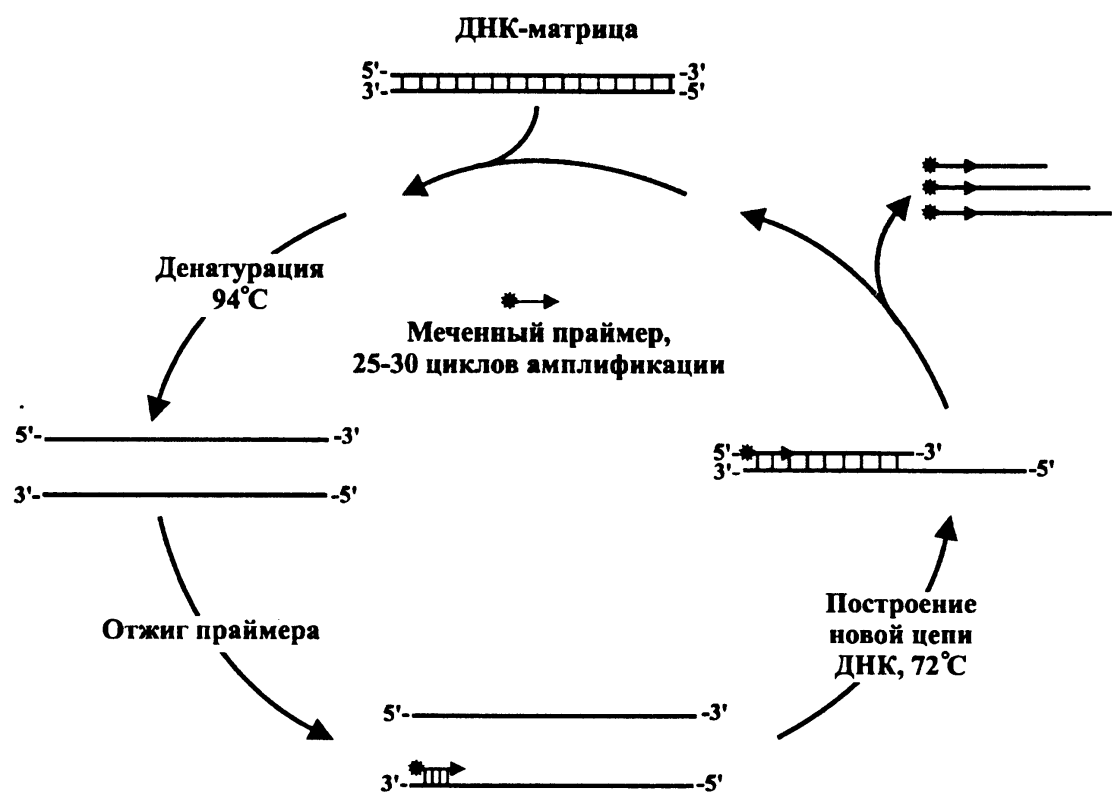


Рис. 3.7. Схема циклического секвенирования ДНК

Примером подобных работ может служить статья Мюррея, где им был секвенирован участок около 200 пн ранее клонированного крупного фрагмента ДНК, причем следует отметить, что в реакцию линейной амплификации были взяты несколько заниженные, против обычного, концентрации дНТФ, что было необходимо для снижения, в свою очередь, концентраций ддНТФ и поддержания требуемых соотношений между ними для эффективного построения цепей ДНК в условиях терминации [Murray, 1989]. Другие авторы секвенировали с помощью линейной ПЦР клонированные фрагменты ДНК, взятые из единичных фаговых бляшек (лямбда или M13) или бактериальных колоний [Krishnan et al., 1991]. Таким образом, в этих работах и им подобных, метод ПЦР был использован исключительно с одной целью – не нарабатывать значительное количество ДНК, необходимое для секвенирования, с помощью обычных процедур выращивания фагов или бактерий, что, впрочем, позволяет сильно экономить время, поскольку выращивание бактерий или фагов и последующая очистка ДНК может занять более суток, тогда как с помощью ПЦР требуемое количество ДНК можно получить через 2–3 ч. Аналогичный подход был продемонстрирован Ли, которым была проведена сначала обычная ПЦР, где в течение 30 циклов нарабатывались целевые продукты амплификации и затем после очистки в агарозном геле фрагмент ДНК, предназначенный для секвенирования, в количестве всего 0,05 пмоль подвергался новым 25 циклам линейной амплификации с одним праймером в присутствии дидезоксигтерминаторов [Lee, 1991]. Столь малого количества матрицы ДНК (приблизительно в 10 раз меньше, чем обычно) оказалось достаточно,

чтобы после 25 циклов стало возможным “чтение” нуклеотидной последовательности ДНК с радиоавтографа геля.

Интересное решение для построения только одной из присутствующих в реакционной смеси в равном соотношении цепей заключается в эксплуатации ранее показанного принципа температурной асимметричной ПЦР [Mazars et al., 1991]. Так, на первом этапе наработки ПЦР-продукта для секвенирования амплификация проводилась с использованием более коротких праймеров, тогда как этап терминации в присутствии дидезокситрифосфатов проводился путем удлинения секвенирующего праймера, комплементарного одной из цепей и имеющего большую протяженность и более высокую температуру отжига (на 10°C), что исключило на этой стадии из реакции короткие амплификационные праймеры [Liu et al., 1993]. Это позволило авторам осуществить секвенирование фрагмента ДНК, не прибегая к предварительной очистке ПЦР-продукта. Более того, поскольку в ходе секвенирования в этом случае не происходит ферментативного удлинения более коротких амплификационных праймеров из первой реакции, появляется возможность применить метод внутреннего мечения продуктов терминирующей реакции, а не готовить заранее секвенирующий праймер, меченый по своему 5'-концу.

Что касается секвенирования с помощью линейной ПЦР неклонированных фрагментов геномной ДНК без этапа их предварительной экспоненциальной амплификации, было показано, что требуется проведение более двухсот циклов для наработки необходимого количества ДНК, которое можно будет затем детектировать гель-электрофорезом [Cairns, Muggay, 1994]. В цитируемой работе потребовалось провести 213 циклов, чтобы амплифицировать фрагмент гена β-глобина человека из исходного количества 5–10 мкг тотальной ДНК, выделенной из белых кровяных клеток. При этом лестница секвенируемых полос ДНК детектировалась не стандартной радиоавтографией, а высокочувствительным прибором PhosphorImager (Molecular Dynamics, США). Известно, что большое число циклов ПЦР часто приводит к возникновению гетерогенных продуктов амплификации [Bell, DeMarini, 1991], однако Кайрнс и Мюррей связывают отсутствие значительного числа таких фрагментов в их работе с включаемыми ддНМФ в процессе синтеза новых цепей ДНК, тогда как после 50–75 циклов линейной ПЦР без присутствия ддНТФ ими отмечено увеличение числа неспецифических фрагментов ДНК.

Интересен метод “спаренной” амплификации и секвенирования, решающий сразу проблему наработки целевого продукта путем обычной ПЦР и образования специфически терминированных меченых фрагментов ДНК, пригодных для их разделения секвенирующим гель-электрофорезом [Ruano, Kidd, 1991]. В цитируемой работе на первой стадии в течение 15 циклов проводилась экспоненциальная амплификация целевого фрагмента ДНК в присутствии относительно низкой концентрации дНТФ (по 10 мкмоль каждого). Затем на второй стадии в реакционную смесь, поделенную пополам, добавлялись меченые амплификационные/секвенирующие праймеры (в каждую пробирку – по одному). В свою очередь, содержимое этих пробирок делилось еще на четыре час-



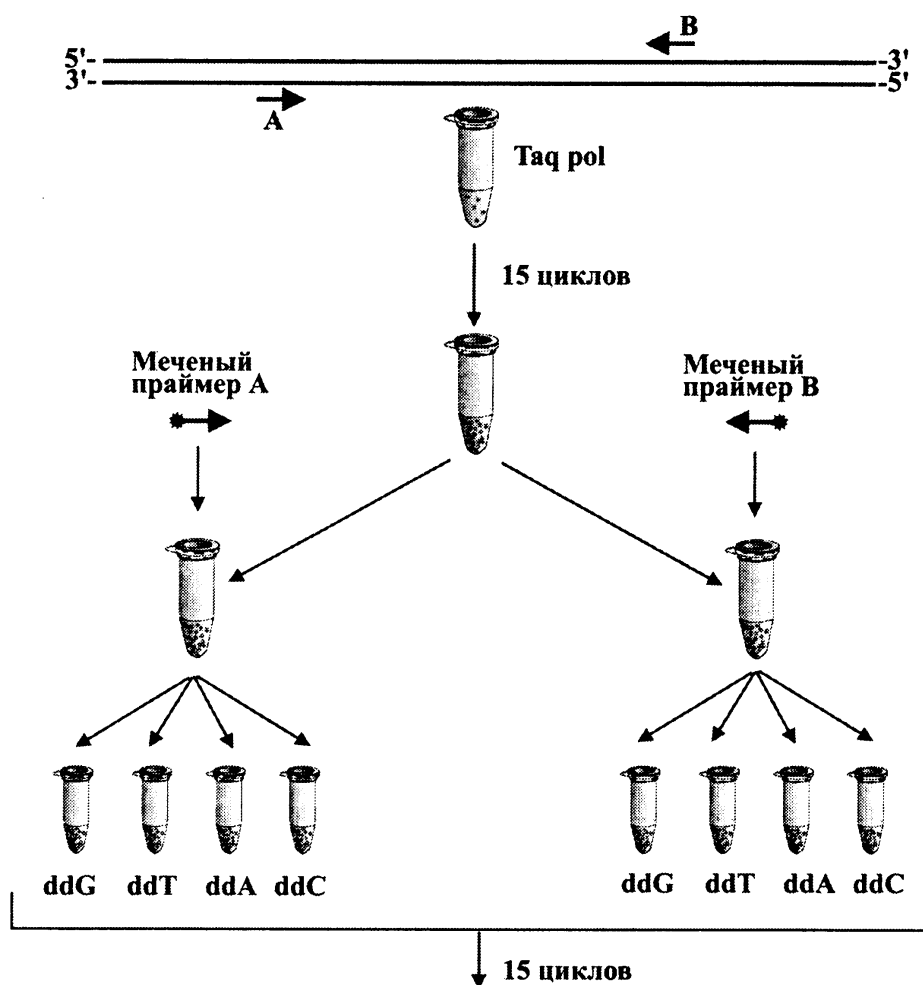


Рис. 3.8. “Спаренная” амплификация и секвенирование фрагмента геномной ДНК

ти с соответствующими дидезокситерминаторами и следующие 15 циклов были посвящены образованию специфически терминированных продуктов. В дальнейшем этот подход после некоторой модификации позволил секвенировать нанограммовые количества исходного фрагмента ДНК [Reynolds et al., 1993].

В то же время остающийся в реакционной смеси после завершения первой стадии ПЦР другой праймер из данной пары теоретически мог обеспечивать и образование нетерминированных полноразмерных фрагментов ДНК, осуществляя, таким образом, их дальнейшее накопление в реакции. Эффективность этого процесса на данной стадии может варьировать от не образования таких фрагментов вообще до заметного увеличения их доли в ходе ПЦР, что в сильной степени зависит от общей длины амплифицируемого фрагмента ДНК и соотношения дНТФ/ддНТФ. Но в любом случае экспоненциального накопления таких полноразмерных продуктов происходить не будет.

В цитируемой выше статье на стадии циклического секвенирования второй праймер специально не добавлялся, а его присутствие было вызвано тем, что он переносился вместе с ПЦР-продуктом и реакционной средой из первой стадии амплификации ДНК. Тогда как в другой работе, посвященной циклическому секвенированию ДНК, в реакционную

смесь наряду с основным секвенирующим праймером специально добавлялся и второй амплификационный праймер, что приводило, как называли его сами авторы, к “полуэкспоненциальному” течению ПЦР [Sarkar, Bolander, 1995]. На приведенной в данной работе иллюстрации наглядно продемонстрировано появление, а затем и заметное улучшение картины секвенируемых полос ДНК в зависимости от количества второго (амплификационного) праймера. Однако для достижения положительного результата в таком полуэкспоненциальном циклическом секвенировании ДНК очень важное значение приобретает правильное соотношение сразу большого числа компонентов реакции, таких как исходное количество амплифицируемого и одновременно секвенируемого фрагмента ДНК, праймеров (секвенирующего и вспомогательно-амплификационного), дНТФ и ддНТФ.

В описанном выше полуэкспоненциальном циклическом секвенировании ДНК процессы получения все новых копий полноразмерного продукта и образования терминированных цепей ДНК, предназначенных для их последующего секвенирования, происходят одновременно и оказывают очень сильное влияние друг на друга. Так, если будет превалировать второй процесс, то как таковой амплификации, приводящей к резкому увеличению числа молекул полноразмерного фрагмента, не произойдет, следовательно, недостаточное количество матриц ДНК для их копирования в условиях терминации (секвенирования) не позволит выявить эти продукты в секвенирующем гель-электрофорезе. Превалирующий первый процесс, хотя и приведет к заметной амплификации, все же не позволит “увидеть” лестницу секвенируемых полос ДНК по той же причине нехватки конечного продукта в виде специфически терминированных фрагментов ДНК. Поэтому идеальным вариантом можно было бы считать некое разобщение этих двух, одновременно протекающих в одной пробирке процессов – построение новых цепей ДНК (амплификация) и специфическое терминирование продуктов удлинения праймера (секвенирование).

Создание новой термостабильной ДНК-полимеразы–термосеквеназы – и ей подобных – позволило разработать новый подход к циклическому секвенированию ДНК и в какой-то степени разобщить эти два процесса. Это стало возможным благодаря появившемуся у термосеквеназы свойству с высокой эффективностью включать в растущую цепь ДНК ддНМФ. Таким образом, присутствие в реакционной (амплификационной и одновременно секвенирующей) среде сразу двух термостабильных ДНК-полимераз, резко отличающихся по своей способности использовать в качестве субстрата ддНТФ, сделало возможным одновременную амплификацию ДНК и ее секвенирование [van den Boom et al., 1997; Kilger, Raabo, 1997a; Kilger, Raabo, 1997b]. Так, обычная Taq ДНК-полимераза “отвечает” за наработку полноразмерного продукта (амплификацию), поскольку низкая эффективность включения ею в растущую цепь ДНК ддНМФ приводит к образованию в реакционной смеси все новых копий таких молекул (рис. 3.9). Термостабильная F667Y ДНК-полимераза эффективно включает во вновь строящуюся ДНК ддНМФ, приводя к специфической терминации цепей (секвениро-

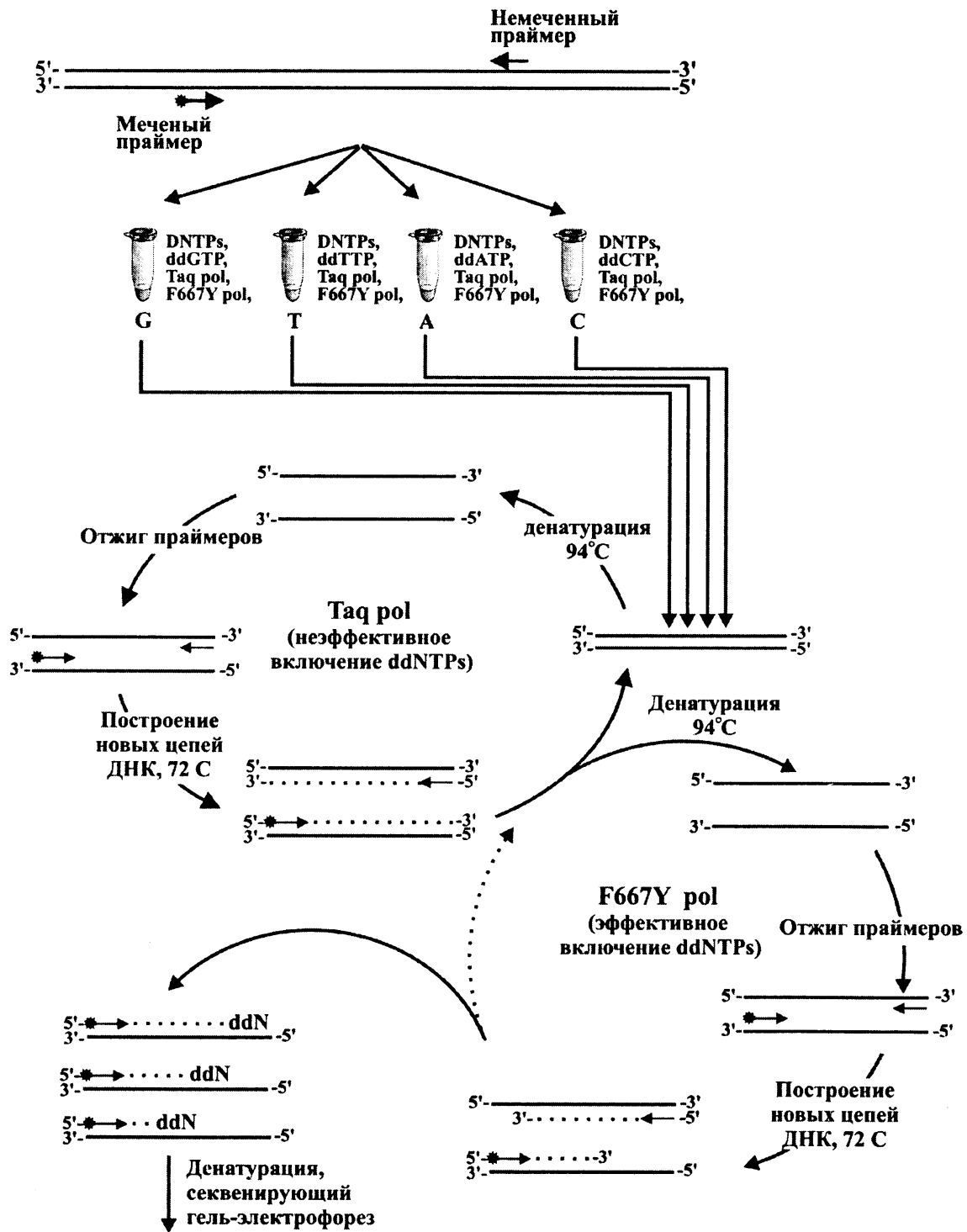


Рис. 3.9. Прямая экспоненциальная амплификация и секвенирование ДНК

ванию). Данный подход был назван авторами как DEXAS (Direct Exponential Amplification and Sequencing) – прямая экспоненциальная амплификация и секвенирование [Kilger, Raabo, 1997b]. Использование меченого секвенирующего праймера позволяет затем детектировать эти терминированные фрагменты ДНК с помощью гель-электрофореза. Возможно секвенирование и обеих цепей ДНК при использовании биотиновой метки и твердой фазы в виде магнитных частиц со стрептавидином [van den Boom et al., 1997].

Если успешное протекание всего процесса в ранее описанных подходах циклического секвенирования ДНК зависит, прежде всего, от правильного соотношения таких ингредиентов реакции, как количество исходной матрицы ДНК, праймеров, дНТФ и ддНТФ, а количество самого фермента ДНК-полимеразы не играет особой роли, то в процессе прямой экспоненциальной амплификации и секвенирования с помощью двух разных ДНК-полимераз очень большое значение приобретает именно их соотношение, что было наглядно продемонстрировано в работе Килгера и Паабо [Kilger, Paabo, 1997a].

Важным преимуществом циклического секвенирования является возможность использования в качестве матрицы гораздо меньших количеств ДНК по сравнению с таковыми при обычном секвенировании и часто даже не требующих какой-либо очистки. Так, в литературе сообщается о проведении циклического секвенирования с препаратами ДНК, выделенными из единичных колоний бактерий [Wilson et al., 1990; Hofmann, Brien, 1991; Krishnan et al., 1991; Kilger et al., 1997], из фаговых бляшек [Smith et al., 1990; Wilson et al., 1990; Krishnan et al., 1991; Wang et al., 1995], а также из грубого лизата фага лямбда [Lasham, Darlison, 1993], космидного клона [Dugaiczuk et al., 1992] или клонированных в бактериофаге P1 [Benes et al., 1997], или в специальном ВАС-векторе на основе бактериальной искусственной хромосомы (Bacterial Artificial Chromosome) [Kim et al., 1997]. Как мало надо ДНК, чтобы осуществить циклическое секвенирование, показывает работа, в которой в качестве стартового количества служили “крохи” ДНК, сорбированной на зубочистке, просто воткнутой в интересующую полосу ДНК, разделенную агарозным гель-электрофорезом [Kadokami et al., 1995]. Однако авторы рекомендуют втыкать зубочистку в полосу ДНК по крайней мере 3 раза и, учитывая все же малое количество сорбированной ДНК, увеличить число циклов при последующем секвенировании до 40.

Время, затрачиваемое на проведение терминирующих реакций при циклическом секвенировании, обычно значительно больше, чем при стандартном секвенировании. Однако есть сообщения о циклическом секвенировании ДНК, где терминирующие реакции были завершены всего за 25 мин вместо обычных 2 ч [Swerdlow et al., 1993], что, впрочем, потребовало использования специального прибора Rapidcycler, выпускаемого фирмой “Idaho Technology Inc.” (США). Данный прибор позволяет за счет высокой скорости смены температур проводить 30 циклов за 10 мин в специальных капиллярах, обдуваемых потоками воздуха [Wittwer et al., 1990, 1994]. Для определения последовательности 300 нуклеотидов путем циклического секвенирования другим авторам потребовалось всего 18 мин на 30 циклов в ДНК-амплификаторе очень оригинальной конструкции [Nakano et al., 1994]. Главной особенностью данного прибора является то, что полимеразная реакция протекает не как обычно в пробирке, а в постоянном потоке жидкости, за счет специального пневматического насоса движущейся по тефлоновому капилляру, отдельные отрезки которого помещены в масляные бани с соответствующими температурами денатурации ДНК, отжига праймеров и построения новых цепей.

В то же время быстрота смены температур при циклическом секвенировании требуется не всегда. Хорошим примером этому служит использование в циклическом секвенировании коротких праймерных молекул. Использование в циклическом секвенировании какой-либо термостабильной ДНК-полимеразы накладывает определенные ограничения на олигонуклеотидные праймеры, которые должны иметь достаточно высокую температуру отжига, чтобы формировать затравочный комплекс с матрицей ДНК в условиях проявления термостабильной ДНК-полимеразой своей ферментативной активности. Так, в циклическом секвенировании наиболее часто используемая температура отжига праймеров колеблется между 55° и 65°С, что достигается или их повышенным GC-составом, или некоторым увеличением длины. Таким образом, использование более коротких олигонуклеотидных праймеров из предсинтезированных библиотек для осуществления секвенирования ДНК путем праймерной “прогулки” становится практически невозможным. Однако в одной работе было показано, что длительный отжиг короткого нонамерного праймера при 20°С в течение 5 мин и затем медленное (в течение еще 5 мин) повышение температуры до 60°С все же позволяют термостабильной ДНК-полимеразе за все это время несколько удлинить праймер, повысив тем самым температуру его диссоциации [Bock, Slightom, 1995]. В этой работе был, таким образом, исследован 121 нонамерный праймер из предсинтезированной библиотеки и около половины из них показали хорошие результаты в циклическом секвенировании. Разработанный способ ферментативного дифференциального удлинения коротких праймеров с помощью неполного набора дезоксинуклеотидтрифосфатов, описанный в разделе 2.3, также сделал возможным использование в циклическом секвенировании коротких 8-мерных олигонуклеотидных праймеров [Raja et al., 1997].

При любом методе секвенирования ДНК серьезную проблему представляют ошибки в определении того или иного нуклеотида(ов). При нециклических методах секвенирования большинство ошибок возникает или из-за компрессии фрагментов ДНК при их разделении в секвенирующем геле, или из-за неспецифической терминации новосинтезируемой цепи ДНК, которая, в свою очередь, может иметь различные причины. В циклическом секвенировании ДНК, кроме этих причин, большое влияние на точность секвенирования оказывает и исходное количество самой матрицы ДНК. Так, было проведено специальное исследование, посвященное выявлению зависимости исходной концентрации матрицы в реакционной смеси и точности секвенирования [Hyder et al., 1994]. В этой работе было показано, что для фрагмента длиной 357 пн наименьший процент ошибок (1–3%) при циклическом секвенировании обнаруживался при изначальном количестве матрицы, варьирующим от 75 до 1125 фмоль. В другой работе, поставившей целью определение точности обычного секвенирования с использованием секвеназы и циклического, проводимого с помощью Taq ДНК-полимеразы, подсчитывали число ошибок, приходящихся на участки матрицы, поразному удаленные от праймера [Koop et al., 1993]. Так, для первых 350 нуклеотидов количество ошибок не превышало 1%, тогда как даль-

нейший участок (351–500 нуклеотидов) характеризовался уже 17% ошибок, причем эта тенденция была характерна для обоих ферментов и, следовательно, для обычного и циклического подходов, однако секвенсаза характеризовалась все же несколько меньшим числом производимых ошибок.

Циклическое секвенирование ДНК в силу особенности многократного повторяющегося протекания терминирующих реакций может приводить к значительному истощению содержания в смеси какого-либо дНТФ и ддНТФ по причине высокого содержания комплементарного ему нуклеотида в матрице ДНК (при условии изначально одинакового соотношения всех дНТФ, как обычно бывает в коммерческих наборах), приводя к ошибкам. Подобный эффект продемонстрирован при секвенировании А + Т-богатого (более 80%) региона митохондриальной ДНК *Schistosoma gregaria*, результатом чего стало включение “неправильных” нуклеотидов, когда концентрация дАТФ и дТТФ в терминирующей смеси упала ниже критической, что было выявлено параллельным секвенированием ПЦР-продукта с помощью секвенсазы [Zhang, Hewitt, 1994], поскольку преимущественное содержание какого-либо нуклеотида в цепи ДНК в обычном секвенировании практически не заметно. Другими трудными участками для секвенирования считаются гомополимерные треки и повторяющиеся элементы. Так, в одной работе было показано, что для успешного секвенирования динуклеотидных повторов (GT) и СТ-трека потребовалась повышенная температура отжига и более продолжительная денатурация [Robbins et al., 1996]. Было показано, что использование в циклическом секвенировании ДНК реакционного буфера, не содержащего даже остаточных количеств ионов  $K^+$ , приводит к исчезновению обычных со стандартным буфером “стопов”, являющихся следствием неспецифического терминирования цепи [Woodford et al., 1995]. KCl является весьма распространенным ингредиентом многих буферов хранения различных ферментов и за счет этого его следовые количества попадают в ПЦР-секвенирование. В цитируемой работе отмечается, что, впрочем, любые одновалентные катионы оказывают подобное отрицательное влияние на само течение ПЦР и циклическое секвенирование.

Следует подчеркнуть, что указанные количества ошибок при секвенировании ДНК могут быть присущи только первичному материалу в виде некой последовательности нуклеотидов, выявленной в процессе электрофореза и радиоавтографии. Однако точность определения нуклеотидной последовательности генов и прочих фрагментов ДНК должна составлять на заключительном этапе по крайней мере 99,95%, что достигается повторным (многократным) секвенированием данного фрагмента ДНК, да еще и по обеим цепям. Более строгим стандартом в последнее время считается 99,99%. Таким образом, завершенное секвенирование какого-либо фрагмента ДНК должно привести к последовательности нуклеотидов, допускающей только одну ошибку на 10 000 нуклеотидов.

Кроме наиболее часто используемой термостабильной ДНК-полимеразы из *Thermus aquaticus*, в циклическом секвенировании применяют

ся и некоторые другие ДНК-полимеразы с подобными свойствами. Так, сообщается об успешном использовании дельта Таq ДНК-полимеразы [Fan, Ranu, 1997], являющейся модифицированной версией этого же фермента с удаленной экзонуклеазной активностью; рекомбинантной AmpliТаq и SequiTherm ДНК-полимераз, поставляемых фирмами "Perkin-Elmer" и "Epicentre Technologies" (обе, США) соответственно [Fulton, Wilson, 1994]. SequiTherm ДНК-полимераза, выделенная из неохарактеризованной археобактерии, использовалась и в других работах [Woddford et al., 1994; Zimmermann et al., 1994]. В этой же статье наряду со стандартной Таq ДНК-полимеразой было проведено сравнительное исследование и других термостабильных ДНК-полимераз – Tub ДНК-полимеразы из *Thermus flavus*, Pfu ДНК-полимеразы из *Pyrococcus furiosus*, Vent<sub>R</sub><sup>®</sup> (exo<sup>-</sup>) ДНК-полимеразы из *Thermococcus litoralis*. Проведенное сравнение качества циклического секвенирования ДНК, осуществленного с помощью нативного фермента Tth ДНК-полимеразы и ее модифицированного варианта – дельта Tth ДНК-полимеразы, характеризующейся отсутствием 5' → 3'-экзонуклеазной активности, показало заметное преимущество последней, выражающееся в низком уровне фона [Arakawa et al., 1996]. Ранее этой же группой авторов показано успешное применение в циклическом секвенировании дельта Tth ДНК-полимеразы с использованием в качестве метки биотиновых производных ддНТФ для последующей хемилюминесцентной детекции [Ikeda et al., 1995]. Из других ферментов можно отметить археобактериальную ДНК-полимеразу Deep Vent<sub>R</sub><sup>®</sup> (exo<sup>-</sup>) из *Pyrococcus* sp. [Mariame, 1996]. В последней же работе приводится информация о применении после завершения реакций циклического секвенирования терминальной трансферазы, достраивающей все неспецифически (т.е. не дидезоксинуклеотидами) терминированные цепи ДНК, что заметно улучшает качество секвенирования ДНК, как это было показано ранее [Fawcett, Barlett, 1990].

### Цитируемая литература

- Маниатис Т., Фрич Э., Сэмбрук Дж.* Методы генетической инженерии: Молекулярное клонирование. М.: Мир, 1984. 480 с.
- Arakawa T., Jongsareejit B., Tatsumi Y., Tanaka K., Ikeda K., Komatsubara H., Inoue H., Kawakami B., Oka M., Emi S., Yomo T., Shima Y., Negoro S., Urabe I.* Application of N-terminally truncated DNA polymerase from *Thermus thermophilus* (delta Tth polymerase) to DNA sequencing and polymerase chain reactions: Comparative study of delta Tth and wild-type Tth polymerases // *DNA Res.* 1996. Vol. 3. P. 87–92.
- Bachmann B., Luke W., Hunsmann G.* Improvement of PCR amplified DNA sequencing with the aid detergents // *Nucl. Acids Res.* 1990. Vol. 18. P. 1308.
- Bell D.A., De Marini D.M.* Excessive cycling converts PCR products to random-length higher molecular weight fragments // *Ibid.* 1991. Vol. 19. P. 5079.
- Benes V., Kilger C., Voss H., Paabo S., Ansoerge W.* Direct primer walking on P1 plasmid DNA // *Biotechniques.* 1997. Vol. 23. P. 98–100.
- Berg E.S., Olaisen B.* Hybrid PCR sequencing: Sequencing of PCR products using a universal primer // *Ibid.* 1994. Vol. 17. P. 896–901.
- Bock J.H., Slightom J.L.* Fluorescence-based cycle sequencing with primers selected from a nonamer library // *Ibid.* 1995. Vol. 19. P. 60–64.
- Boom D. van den, Jurinke C., Ruppert A., Koster H.* Forward and reverse DNA sequencing in a single reaction // *Anal. Biochem.* 1998. Vol. 256. P. 127–129.

*Boom D. van den, Ruppert A., Jurinke C., Koster H.* Combined amplification and sequencing in a single reaction using two DNA polymerases with differential incorporation rates for dideoxynucleotides // *J. Biochem. and Biophys. Meth.* 1997. Vol. 35. P. 69–79.

*Cairns M.J., Murray V.* Dideoxy genomic sequencing of a single-copy mammalian gene using more than two hundred cycles of linear amplification // *Biotechniques.* 1994. Vol. 17. P. 910–915.

*Cao X., Brosius G.C.* Direct sequencing of double-stranded PCR products isolated from conventional agarose gels // *Ibid.* 1993. Vol. 15. P. 384–386.

*Casanova J.-L., Pannetier C., Kourilsky P.* Optimal conditions for directly sequencing double-stranded PCR products with sequenase // *Nucl. Acids Res.* Vol. 18. P. 4028.

*Cease K.B., Lohff C.J.* A vector for facile PCR product cloning and modification generating any desired 4-base 5' overhang: pRPM // *Biotechniques.* 1993. Vol. 14. P. 250–255.

*Cha J., Bishai W., Chandrasegaran S.* New vectors for direct cloning of PCR products // *Gene.* 1993. Vol. 136. P. 369–370.

*Clark J.M.* Novel non-template nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases // *Nucl. Acids Res.* 1988. Vol. 16. P. 9677–9686.

*De Angelis M.M., Wang D.G., Hawkins T.L.* Solid-phase reversible immobilization for the isolation of PCR products // *Ibid.* 1995. Vol. 23. P. 4742–4743.

*Douglas A.M., Georgalis A.M., Atchinson B.A.* Direct sequencing of double-stranded PCR products incorporating a chemiluminescent detection procedure // *Biotechniques.* 1993. Vol. 14. P. 824–828.

*Dugaiczky A., Goold R., Di Sibio G., Myers R.M.* Improved sequencing of cosmids using new primers and linearized DNA // *Nucl. Acids Res.* 1992. Vol. 20. P. 6421–6422.

*Eckert K.A., Kunkel T.A.* High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase // *Ibid.* 1990. Vol. 18. P. 3739–3744.

*Engelke D.R., Hoener P.A., Collins F.S.* Direct sequencing of enzymatically amplified human genomic DNA // *Proc. Nat. Acad. Sci. USA.* 1988. Vol. 85. P. 544–548.

*Fan J., Ranu R.S.* Direct cycle sequencing with delta Taq DNA polymerase // *DNA Sequenc.* 1997. Vol. 7. P. 285–288.

*Fawcett T.W., Bartlett S.G.* An effective method for eliminating “artifact banding” when sequencing double-stranded DNA templates // *Biotechniques.* 1990. Vol. 9. P. 46–48.

*Fulton L.L., Wilson R.K.* Variations on cycle sequencing // *Ibid.* 1994. Vol. 17. P. 298–301.

*Gal S.* Sequencing of double-stranded PCR products // *Mol. Biotechnol.* 1996. Vol. 5. P. 159–164.

*Gal S.* Sequencing of double-stranded PCR products // *Meth. Mol. Biol.* 1993. Vol. 23: DNA sequencing protocols. P. 183–190.

*Gal S., Hohn B.* Direct sequencing of double-stranded DNA PCR products via removing the complementary strand with single-stranded DNA of an M13 clone // *Nucl. Acids Res.* 1990. Vol. 18. P. 1076.

*Gold P.* Use a novel agarose-digesting enzymes for easy and rapid purification of PCR-amplified DNA for sequencing // *Biotechniques.* 1992. Vol. 13. P. 123–134.

*Green A., Roopra A., Vaudin M.* Direct single stranded sequencing from agarose of polymerase chain reaction products // *Nucl. Acids Res.* 1990. Vol. 18. P. 6163–6164.

*Gyllensten U.B.* PCR and DNA sequencing // *Biotechniques.* 1989. Vol. 7. P. 700–708.

*Gyllensten U.B., Erlich H.A.* Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus // *Proc. Nat. Acad. Sci. USA.* 1988. Vol. 85. P. 7652–7656.

*Haqqi T.M.* Direct ligation of PCR products for cloning and sequencing // *Nucl. Acids Res.* 1992. Vol. 20. P. 6427.

*Harrison J., Molloy P.L., Clark S.J.* Direct cloning of polymerase chain reaction products in an XcmI T-vector // *Anal. Biochem.* 1994. Vol. 216. P. 235–236.

*Higuchi R.G., Ochman H.* Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction // *Nucl. Acids Res.* 1989. Vol. 17. P. 5865.

*Hofmann M.A., Brian D.A.* Sequencing PCR DNA amplified directly from a bacterial colony // *Biotechniques.* 1991. Vol. 11. P. 30–31.

*Holton T.A., Graham M.W.* A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors // *Nucl. Acids Res.* 1991. Vol. 19. P. 1156.

*Hu G.* DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment // *DNA Cell Biol.* Vol. 12. P. 763–770.



- Huber L.J., McMahon G. Direct DNA sequencing of PCR amplified DNA with sequenase // Editorial Comments. US Biochem. Corp. 1988. Vol. 15. P. 3–4.
- Hultamn T., Bergh S., Moks T., Uhlem M. Bidirectional solid-phase sequencing of *in vitro*-amplified plasmid DNA // Biotechniques. 1991. Vol. 10. P. 84–93.
- Hultman T., Stahl S., Hornes E., Uhlen M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support // Nucl. Acids Res. 1989. Vol. 17. P. 4937–4946.
- Hung T., Mak K., Fong K. A specificity enhancer for polymerase chain reaction // Ibid. 1990. Vol. 18. P. 4953.
- Hyder S.M., Hu C., Needleman D.S., Sonoda Y., Wang X-Y., Baker V.V. Improved accuracy in direct automated DNA sequencing of small PCR products by optimizing the template concentration // Biotechniques. 1994. Vol. 17. P. 478–482.
- Ichihara Y., Kurosawa Y. Construction of new T vectors for direct cloning of PCR products // Gene. 1993. Vol. 130. P. 153–154.
- Ikeda K., Inoue H., Oka M., Kawakami B., Kawamura Y. A non-radioactive DNA sequencing method biotinylated dideoxynucleoside triphosphates and delta Tth DNA polymerase // DNA Res. 1995. Vol. 2. P. 225–227.
- Innis M.A., Myambo K.B., Gelfand D.H., Brow A.D. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA // Proc. Nat. Acad. Sci. USA. 1988. Vol. 85. P. 9436–9440.
- Iwahana H., Yoshimoto K., Tsujisawa T., Itakura M. T-cassette ligation: A method for direct sequencing and cloning of PCR-amplified DNA fragments // PCR Meth. Appl. 1994. Vol. 3. P. 219–224.
- Kadokami Y., Ruess S.L., Mills B.W., Lewis R.V. Cycle DNA sequencing from a gel band // Biotechniques. 1995. Vol. 18. P. 226.
- Kalnoski M.H., McCoy-Haman M.F., Hollis G.F. A rapid method for purifying PCR products for direct sequence analysis // Ibid. 1991. Vol. 11. P. 246–249.
- Kaltenboeck B., Spatafora J.W., Zhang X., Kousoulas K.G., Blackwell M., Storz J. Efficient production of single-stranded DNA as long as 2 kb for sequencing of PCR-amplified DNA // Ibid. 1992. Vol. 12. P. 164, 166, 168–171.
- Kasai H., Isono S., Kitakawa M., Mineno J., Akiyama H., Kurnit D.M., Berg D.E., Isono K. Efficient large-scale sequencing of the *Escherichia coli* genome: Implementation of a transposon- and PCR-based strategy for the analysis of ordered  $\lambda$  phage clones // Nucl. Acids Res. 1992. Vol. 20. P. 6509–6515.
- Katz E.D., Dong M.W. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography // Biotechniques. 1990. Vol. 8. P. 546–555.
- Kim J., Carver E.A., Stubbs L. Amplification and sequencing of end fragments from bacterial artificial chromosome clones by single-primer polymerase chain reaction // Anal. Biochem. 1997. Vol. 252. P. 272–275.
- Kilger C., Krings M., Poinar H., Paabo S. “Colony sequencing”: direct sequencing of plasmid DNA from bacterial colonies // Biotechniques. 1997. Vol. 22. P. 412–418.
- Kilger C., Paabo S. Direct DNA sequence determination from total genomic DNA // Nucl. Acids Res. 1997a. Vol. 25. P. 2032–2034.
- Kilger C., Paabo S. Direct exponential amplification and sequencing (DEXAS) of genomic DNA // Biol. Chem. 1997b. Vol. 378. P. 99–105.
- Kofoed E. Direct sequencing of long PCR products // Rapid Cyclist. 1995. Vol. 3. P. 7–10.
- Koop B.F., Rowan L., Chen W.-Q., Deshpande P., Lee H., Hood L. Sequenase length and error analysis of sequenase and automated *Taq* cycle sequencing methods // Biotechniques. 1993. Vol. 14. P. 442–447.
- Kovalic D., Kwak J.-H., Weisblum B. General method for direct cloning of DNA fragments generated by the polymerase chain reaction // Nucl. Acids Res. 1991. Vol. 19. P. 4560.
- Kretz K.A., Carson G.S., O'Brien J.S. Direct sequencing from low-melt agarose with sequenase® // Ibid. 1989. Vol. 17. P. 5864.
- Kretz K.A., Carson G.S., O'Brien J.S. Direct sequencing from low-melt agarose with sequenase® T7 DNA polymerase // Editorial Comments. US Biochem. Corp. 1990. Vol. 16. P. 6–7.
- Krishnan B.R., Blakesley R.W., Berg D.E. Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies // Nucl. Acids Res. 1991. Vol. 19. P. 1153.

- Krowczynska A.M., Henderson M.B.* Efficient purification of PCR products using ultrafiltration // *Biotechniques*. 1992. Vol. 13. P. 286–289.
- Kusukawa N., Uemori T., Asada K., Kato I.* Rapid and reliable protocol for directed sequencing of material amplified by the polymerase chain reaction // *Ibid.* 1990. Vol. 9. P. 66–72.
- Kwok S., Chang S.-Y., Sninsky J.J., Wang A.* A guide to the design and use of mismatched and degenerate primers // *PCR Meth. Appl.* 1994. Vol. 3. P. S39–S47.
- Lasham A., Darlison M.G.* Direct sequencing of lambda DNA from crude lysates using an improved linear amplification techniques // *Mol. Cell Probes*. 1993. Vol. 7. P. 67–73.
- Lawson V.A., McPhee D.A., Deacon N.J.* Elimination of sequence ambiguities by a single-step modification of a solid-phase, single-stranded sequencing protocol // *Biotechniques*. 1996. Vol. 21. P. 356–358.
- Lee J.-S.* Alternative dideoxy sequencing of double-stranded DNA by cyclic reaction using *Taq* polymerase // *DNA Cell Biol.* 1991. Vol. 10. P. 67–73.
- Lee Y.-H., Vacquier V.D.* A method for obtaining high-quality sequences from the non-biotinylated, free ssDNA remaining after solid-phase sequencing // *Biotechniques*. 1993. Vol. 14. P. 191–192.
- Leonard J.T., Grace M.B., Buzard G.S., Mullen M.J., Barbagallo C.B.* Preparation of PCR products for DNA sequencing // *Ibid.* 1998. Vol. 24. P. 314–317.
- Li H., Cui X., Arnheim N.* Eliminating primers from completed polymerase chain reactions with exonuclease VII // *Nucl. Acids Res.* 1991. Vol. 19. P. 3139–3141.
- Lind T., Thorland E.C., Sommer S.S.* Genomic amplification with transport sequencing (GAWTS) // *Meth. Mol. Biol.* 1996. Vol. 65: PCR sequencing protocols. P. 193–200.
- Liu Y.-G., Mitsukawa N., Whittier R.F.* Rapid sequencing of unpurified PCR products by thermal asymmetric PCR cycle sequencing using unlabeled sequencing primers // *Nucl. Acids Res.* 1993. Vol. 21. P. 3333–3334.
- Lohff C.J., Cease K.B.* PCR using a thermostable polymerase with 3' to 5' exonuclease activity generates blunt products suitable for direct cloning // *Ibid.* 1992. Vol. 20. P. 144.
- Lowe T., Sharefkin J., Yang S.Q., Diefenbach C.W.* A computer program for selection of oligonucleotide primers for polymerase chain reaction // *Ibid.* 1990. Vol. 18. P. 1757–1761.
- Lundberg K.S., Shoemaker D.D., Adams M.W.W., Short J.M., Sorge J.A., Mathur E.J.* High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus* // *Gene*. 1991. Vol. 108. P. 1–6.
- Marchuk D., Drumm M., Saulino A., Collins F.S.* Construction of T-vectors, a rapid and general systems for direct cloning of unmodified PCR products // *Nucl. Acids Res.* 1991. Vol. 19. P. 1154.
- Mariame B.* Cycle sequencing protocol using Deep Vent<sup>®</sup> (exo<sup>-</sup>) DNA polymerase and reduced dNTP and [ $\alpha$ -<sup>35</sup>S] dATP concentrations // *Biotechniques*. 1996. Vol. 21. P. 18–19.
- Martsen E.O., Zeng M., Lapeyre J.N.* Fast plasmid DNA sequencing using a thermal cycler and high temperature alkali denaturation // *Ibid.* 1997. Vol. 22. P. 420–422.
- Mattila P., Korpela J., Tenkanen T., Pitkanen K.* Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase – an extremely heat stable enzyme with proofreading activity // *Nucl. Acids Res.* 1991. Vol. 19. P. 4967–4973.
- Mazars G.-R., Moyret C., Jeanteur P., Theillet C.-G.* Direct sequencing by thermal asymmetric PCR // *Ibid.* 1991. Vol. 19. P. 4783.
- McConlogue L., Brow M.A.D., Innis M.A.* Structure-independent DNA amplification by PCR using a 7-deaza-2'-deoxyguanosine // *Ibid.* 1988. Vol. 16. P. 9869.
- McMahon G., Davis E., Wogan G.N.* Characterization of c-K-ras oncogene alleles by direct sequencing of enzymatically amplified DNA from carcinogen-induced tumors // *Proc. Nat. Acad. Sci. USA*. 1987. Vol. 84. P. 4974–4978.
- Medori R., Tritschler H.J., Gambetti P.* Production of single-stranded DNA for sequencing: An alternative approach // *Biotechniques*. 1992. Vol. 12. P. 346–350.
- Meltzer S.J., Mane S.M., Wood P.K., Johnson L., Needleman S.W.* Sequencing products of the polymerase chain reaction directly, without purification // *Ibid.* 1990. Vol. 8. P. 142–148.
- Meyerhans A., Vartanian J.-P., Wain-Hobson S.* DNA recombination during PCR // *Nucl. Acids Res.* 1990. Vol. 18. P. 1687–1691.
- Mitchell L.G., Merrill C.R.* Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction // *Anal. Biochem.* 1989. Vol. 178. P. 239–242.

- Murray V. Improved double-stranded DNA sequencing using the linear polymerase chain reaction // *Nucl. Acids Res.* 1989. Vol. 17. P. 8889.
- Nakano H., Matsuda K., Yohda M., Nagamune T., Endo I., Yamane T. High speed polymerase chain reaction in constant flow // *Biosci. Biotech. Biochem.* 1994. Vol. 58. P. 349–352.
- Nichols R.C., Raben N. Hints for direct sequencing of PCR-generated single-stranded DNA // *Biotechniques.* 1994. Vol. 17. P. 412–414.
- Nickrent D.L. From field to film: Rapid sequencing methods for field-collected plant species // *Ibid.* 1994. Vol. 16. P. 470–475.
- Odelberg S.J., Weiss R.B., Hata A., White R. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I // *Nucl. Acids Res.* 1995. Vol. 23. P. 2049–2057.
- Olsen D.B., Eckstein F. Incomplete primer extension during in vitro DNA amplification catalyzed by Taq polymerase; exploitation for DNA sequencing // *Ibid.* 1989. Vol. 17. P. 9613–9620.
- Paabo S., Wilson A.C. Polymerase chain reaction reveals cloning artefacts // *Nature.* 1988. Vol. 334. P. 387–388.
- Paithankar K.R., Prasad K.S.N. Precipitation of DNA by polyethylene glycol and ethanol // *Nucl. Acids Res.* 1991. Vol. 19. P. 1346.
- Putney S.D., Benkovic S.J., Schimmel P.R. A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo // *Proc. Nat. Acad. Sci. USA.* 1981. Vol. 78. P. 7350–7354.
- Raja M.C., Zevin-Sonkin D., Shvartzburd J., Kotler L., Ulanovsky L. DNA sequencing with modular primers using a two-step protocols with thermostable polymerase at the second step // *Biotechniques.* 1997. Vol. 23. P. 362–368.
- Rao V.B. Direct sequencing of polymerase chain reaction-amplified DNA // *Anal. Biochem.* 1994. Vol. 216. P. 1–14.
- Rao V.B., Saunders N.B. A rapid polymerase-chain-reaction-directed sequencing strategy using a thermostable DNA polymerase from *Thermus flavus* // *Gene.* 1992. Vol. 113. P. 17–23.
- Reynolds T.R., Uliana S.R., Floeter-Winter L.M., Buck G.A. Optimization of coupled PCR amplification and cycle sequencing of cloned and genomic DNA // *Biotechniques.* 1993. Vol. 15. P. 466–467.
- Robbins C.M., Hsu E., Gillevet P.M. Sequencing homopolymer tracts and repetitive elements // *Ibid.* 1996. Vol. 20. P. 862–868.
- Rolfs A., Weber I. Fully-automated, non-radioactive solid-phase sequencing of genomic DNA obtained from PCR // *Ibid.* 1994. Vol. 17. P. 782–787.
- Ruano G., Kidd K.K. Coupled amplification and sequencing of genomic DNA // *Proc. Nat. Acad. Sci. USA.* 1991. Vol. 88. P. 2815–2819.
- Rychlik W. Selection of primers for polymerase chain reaction // *Mol. Biotechnol.* 1995. Vol. 3. P. 129–134.
- Rychlik W., Rhoads R.E. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA // *Nucl. Acids Res.* 1989. Vol. 17. P. 8543–8550.
- Saiki R.K., Gelfand D.H., Stoffel S., Scharf S., Higuchi R., Horn G.T., Mullis K.B., Erlich H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase // *Science.* 1988. Vol. 239. P. 487–491.
- Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A., Arnheim N. Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia // *Ibid.* 1985. Vol. 230. P. 1350–1354.
- Sambrook J., Fritsch E.F., Maniatis T. *Molecular cloning: A laboratory manual.* 2nd edition. Cold Spring Harbor: CSHL press, 1989. Vol. 1.
- Sarkar G., Bolander M.E. Semi exponential cycle sequencing // *Nucl. Acids Res.* 1995. Vol. 23. P. 1269–1270.
- Scharf S.J., Horn G.T., Erlich H.A. Direct cloning and sequence analysis of enzymatically amplified genomic sequences // *Science.* 1986. Vol. 233. P. 1076–1078.
- Shuldiner A.R., Scott L.A., Roth J. PCR-induced (ligase-free) subcloning: A rapid reliable method to subclone polymerase chain reaction (PCR) products // *Nucl. Acids Res.* 1990. Vol. 18. P. 1920.
- Smith D.P., Johnstone E.M., Little S.P., Hsiung H.M. Direct DNA sequencing of cDNA insert from plaques using the linear polymerase chain reaction // *Biotechniques.* 1990. Vol. 9. P. 48–52.

- Somers V.A., Leimbach D.A., Murtagh J.J., Thunnissen F.B.* Exonuclease enhances hybridization efficiency: Improved direct cycle sequencing and point mutation detection // *Biochim. et biophys. acta.* 1998. Vol. 1379. P. 42–52.
- Srivastava A.K., Montanaro V., Kere J.* Simplified template preparation and improved direct sequencing using Taq polymerase // *PCR Meth. Appl.* 1992. Vol. 1. P. 255–256.
- Stoflet E.S., Koeberl D.D., Sarkar G., Sommer S.S.* Genomic amplification with transcript sequencing // *Science.* 1988. Vol. 239. P. 491–494.
- Swerdlow H., Dew-Jager K., Gesteland R.F.* Rapid cycle sequencing in air thermal cycler // *Biotechniques.* 1993. Vol. 15. P. 512–519.
- Takagi S., Kimura M., Katsuki M.* Direct sequencing of PCR products using unlabeled primers // *Ibid.* 1993. Vol. 14. P. 218–221.
- Tracy T.E., Mulcahy L.S.* A simple method for direct automated sequencing of PCR fragments // *Ibid.* 1991. Vol. 11. P. 68–75.
- Tong X., Smith L.M.* Solid phase purification in automated DNA sequencing // *DNA Sequenc.* 1993. Vol. 4. P. 151–162.
- Uhlen M.* Magnetic separation of DNA // *Nature.* 1989. Vol. 340. P. 733–734.
- Verhasselt P., Van Campenhout S., Voet M., Wellens B., Volckaert G.* A general method for routine sequencing of cloned DNA fragments using commercial dye primers and its application in sequence analysis of *Toxoplasma gondii* RH genome fragments // *DNA Sequenc.* 1993. Vol. 4. P. 71–77.
- Wahlberg J., Hultman T., Uhlen M.* Solid phase sequencing of PCR products // *PCR 2: A practical approach.* Oxford: IRL press, 1995. P. 71–87.
- Wahlberg J., Lundeberg J., Hultman T., Uhlen M.* General colorimetric method for DNA diagnostics allowing direct solid-phase genomic sequencing of the positive samples // *Proc. Nat. Acad. Sci. USA.* 1990. Vol. 87. P. 6569–6573.
- Wang S., Krinks M., Moos M., Jr.* DNA sequencing from single phage plaques using solid-phase magnetic capture // *Biotechniques.* 1995. Vol. 18. P. 130–135.
- Ward M.A., Skandalis A., Glickman B.VW, Grosovsky A.J.* Rapid generation of specific single-stranded template DNA from PCR-amplified material // *Nucl. Acids Res.* 1989. Vol. 17. P. 8394.
- Warren W., Doniger J.* HPLC purification of polymerase chain reaction products for direct sequencing // *Biotechniques.* 1991. Vol. 10. P. 216–220.
- Werle E., Schneider C., Renner M., Volker M., Fiehn W.* Convenient single-step, one tube purification of PCR products for directed sequencing // *Nucl. Acids Res.* 1994. Vol. 22. P. 4354–4355.
- Wilson R.K., Chen C., Hood L.* Optimization of asymmetric polymerase chain reaction for rapid fluorescent DNA sequencing // *Biotechniques.* 1990. Vol. 8. P. 184–189.
- Winship P.R.* An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide // *Nucl. Acids Res.* 1989. Vol. 17. P. 1266.
- Wittwer C.T., Fillmore G.C., Garling D.J.* Minimizing the time required for DNA amplification by efficient heat transfer to small samples // *Anal. Biochem.* 1990. Vol. 186. P. 328–331.
- Wittwer C.T., Reed G.B., Ririe K.M.* Rapid cycle DNA amplification // *The polymerase chain reaction.* Boston: Birkhauser, 1994. P. 174–181.
- Wong C., Dowling C.E., Saiki R.K., Higuchi R.G., Erlich H.A., Kazazian H.H.* Characterization of  $\beta$ -thalassaemia mutations using direct genomic sequencing of amplified single copy DNA // *Nature.* 1987. Vol. 330. P. 384–386.
- Woodford K.J., Howell R.M., Usdin K.* A novel K<sup>+</sup>-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes // *J. Biol. Chem.* 1994. Vol. 269. P. 27029–27035.
- Woodford K.J., Howell R.M., Usdin K.* The use of K<sup>+</sup>-free buffer eliminates a common cause of premature chain termination in PCR and PCR sequencing // *Nucl. Acids Res.* 1995. Vol. 23. P. 539.
- Wrischnik L.A., Higuchi R.G., Stoneking M., Erlich H.A., Arnheim N., Wilson A.C.* Length mutation in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA // *Ibid.* 1987. Vol. 15. P. 529–542.
- Wu S.-M., Blomberg L.A., Chan W.-Y.* Recovery of unlabeled PCR product from polyacrylamide gel for sequencing // *Biotechniques.* 1996. Vol. 21. P. 358–362.

Zhang D.-X., Hewitt G.M. Taq cycle sequencing is more sensitive to DNA base bias // *Ibid.* 1994. Vol. 16. P. 573–576.

Zimmermann J., Wieman S., Vos H., Schwager C., Ansoerge W. Improved fluorescent cycle sequencing protocol allows reading nearly 1000 bases // *Ibid.* 1994. Vol. 17. P. 302–307.

### *Дополнительная литература*

Сапки Р., Гиленстен У., Эрлих Г. Полимеразная цепная реакция // *Анализ генома: Методы* / Ред. К. Дейвис. М.: Мир, 1990. С. 176–190.

Alphey L. DNA sequencing: from experimental methods to bioinformatics. Berlin etc.: BIOS sci. publ., 1997. 224 p.

DNA sequencing: Methods and protocols / Epicentre technol. corp. Madison, 1996. 64 p.

Gibbs R.A. DNA amplification by the polymerase chain reaction // *Anal. Chem.* 1990. Vol. 62. P. 1202–1214.

Guide to cycle sequencing: Issue 1.0, July 1995 / Hybaid limited. Teddington (England), 1995. 64 p.

McPherson M.J., Hames B.D., Taylor G.R. (ed.). PCR 2: A practical approach. Oxford: IRL press, 1995. 330 p.

Mullis K.B., Ferre F., Gibbs R.A. (ed.). The polymerase chain reaction. Boston: Birkhauser, 1994. 458 p.

Rapley R. (ed.). Methods in molecular biology. Vol. 65. PCR sequencing protocols. Totowa: Humana press, 1996. 221 p.

Sambrook J., Fritsch E.F., Maniatis T. Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor: CSHL press, 1989. Vol. 1–3. 1626 p.

## Глава 4

# СЕКВЕНИРУЮЩИЙ ГЕЛЬ-ЭЛЕКТРОФОРЕЗ

Определение последовательности нуклеотидов рассмотренными выше методами предполагает разделение продуктов секвенирующих реакций (как полученных в ходе ферментативного секвенирования, так и в процессе химической деградации) с помощью высоковольтного секвенирующего электрофореза в полиакриламидном геле. Сразу следует отметить, что этот метод претерпел целый ряд всевозможных улучшений и модификаций, направленных, главным образом, на повышение его разрешающей способности, рассмотрение которых и будет основной темой данной главы, поскольку сам принцип электрофореза детально изложен во многих обзорах и целых книгах и рассматривать его здесь, видимо, будет излишним.

Описываемые ниже улучшения метода электрофоретического разделения молекул ДНК в денатурирующих условиях ставили своей целью увеличить разрешающую способность самого электрофореза и соответственно всего процесса секвенирования ДНК как за счет “чтения” большего числа полос, так и повышения его достоверности. Дальнейшее увеличение разрешающей способности секвенирующего гель-электрофореза приобрело особую актуальность в связи с использованием при построении второй цепи ДНК фермента T7 ДНК-полимеразы – секвеназы, обладающей высокой процессивностью. Так, ранее, обычный предел “чтения” секвенируемых фрагментов ДНК определялся кроме возможностей самого процесса электрофоретического разделения не в последнюю очередь и размером самих секвенируемых молекул ДНК. При секвенировании ДНК методом Максама–Гилберта получаемый размер фрагментов в результате химической деградации обычно не превышал 250–400 и в исключительных случаях 500–600 нуклеотидов, хотя в реакционной смеси всегда присутствовали продукты реакции гораздо большего размера, но количество таких молекул было недостаточно для их выявления при радиоавтографии. Использование для построения второй цепи Кленовского фрагмента ДНК-полимеразы I в ферментативном методе секвенирования ДНК также не приводило к массовому появлению в реакционной смеси молекул ДНК, имеющих размеры свыше тех же 500–600 или даже 250–400 нуклеотидов. Однако фрагменты ДНК, подготовленные к электрофоретическому разделению с помощью секвеназы, часто имеют размер свыше 2000 нуклеотидов и поэтому проблема повышения разрешающей способности метода секвенирующего гель-электрофореза с началом использования секвеназы встала особенно остро.

### 4.1. СЕКВЕНИРУЮЩИЙ ГЕЛЬ

Основным требованием, предъявляемым к методу электрофоретического разделения секвенируемых молекул ДНК, является его разрешающая способность, позволяющая в денатурирующих условиях разде-

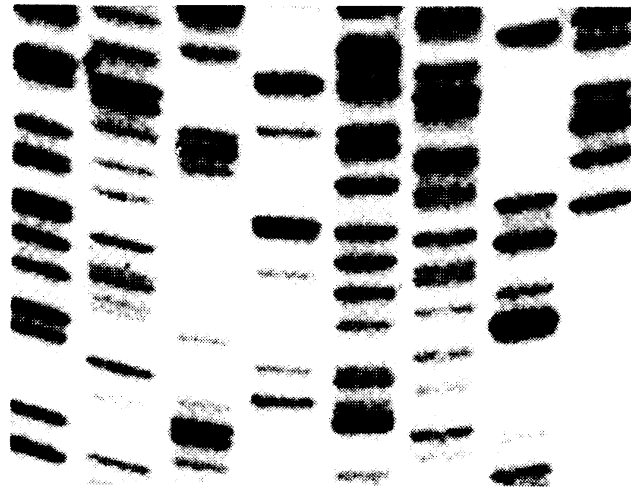


Рис. 4.1. Эффект “улыбки”, приводящей к искривлению движения полос ДНК по ширине геля

лать одноцепочечные фрагменты ДНК, отличающиеся друг от друга на один нуклеотид, в достаточно большом диапазоне. Для выполнения этого условия первоначально применялся 8–20%-ный полиакриламидный гель с использованием в качестве стандартного денатурирующего агента 7М мочевины в *трис*-боратном буфере pH 8,3 с размером гелевой пластины не менее 40 см в длину. Для эффективного разделения требовалось напряжение от 30 до 50 В на см длины геля. В первоначальных экспериментах применяли гели толщиной 0,8 и даже 1,6 мм. Приготовление таких гелей, нанесение препаратов ДНК в ячейки, образованные гребенками, и последующее обращение с гелем не представляло больших трудностей. Однако во время электрофореза происходило выделение большого количества Джоулева тепла и это, ввиду довольно высокой толщины геля, приводило к неравномерному нагреву и самого геля, и стекла. В результате возникал нежелательный эффект “улыбки”, выражающийся в виде дугового искривления полос ДНК на радиоавтографе геля из-за большей скорости продвижения молекул ДНК в средней части геля и их отставания по краям и в связи с этим использование всей ширины геля было подчас нежелательно (рис. 4.1). Таким образом, весьма сильно затруднялось “чтение” последовательности нуклеотидов. В какой-то степени преодолеть это неудобство позволял большой по ширине размер зубцов гребенки, превышающих иногда 1 см, что, впрочем, сильно уменьшало количество разделяемых образцов. Проблема вставала особенно остро при использовании достаточно узких гелей, где эффективное пространство оказывалось совсем небольшим.

Еще более неприятное последствие неравномерного нагрева заключалось в иногда имевшем место повреждении стекла в виде их расстрескивания. Частично эта проблема была решена применением гелей с толщиной 0,4 мм, не приводящим к столь значительному градиенту температур [Sanger, Coulson, 1978]. Другим важным преимуществом таких тонких гелей явилось более высокое разрешение как за счет меньшего количества наносимого препарата ДНК, так и из-за уменьшения

расстояния, которое было необходимо преодолеть радиоактивным частицам, для достижения рентгеновской пленки. Использование ультратонких 0,2 мм гелей позволило еще более повысить чувствительность метода при одновременном уменьшении необходимого количества препаратов ДНК [Ansorge, De Mayer, 1980]. Однако применение ультратонких гелей поставило перед экспериментаторами ряд проблем по их приготовлению и последующему обращению с ними. Так, был разработан способ заливки таких гелей довольно удобной техникой скольжения [Ansorge, De Mayer, 1980]. Еще более тонкие гели с толщиной всего 0,1 мм имели лишь незначительные преимущества перед 0,2 мм гелями и из-за технических трудностей не нашли тогда широкого применения [Ansorge, Barker, 1984].

Проблема неравномерного нагрева геля во время электрофореза была решена путем принудительного нагрева с помощью циркулирующей горячей воды в так называемой термоплате, являющейся одним из “стекел”, между которыми и заливался гель [Ansorge, De Maeyer, 1980; Garoff, Ansorge, 1981]. Данный принцип был положен в основу поставляемого в течение целого ряда лет фирмами LKB и затем “Pharmacia Biotech” (Швеция) и только недавно снятого с производства довольно удобного прибора для секвенирования ДНК под названием Mascophor, хотя и требующего дополнительного оборудования в виде ультратермостата, который обеспечивал циркуляцию воды. Другой подход, не требующий специальной термоплаты и дополнительного оборудования (и поэтому более экономный), основанный на равномерном распределении выделяющегося во время электрофореза Джоулева тепла благодаря конструкции самого прибора, был реализован в приборе Poker Face фирмы “Hoefler Scientific Instruments” (США). Этот эффект достигался тем, что одно из стекел между которыми заливался гель, служило одновременно передней стенкой верхней (или точнее задней) буферной камеры и буферный раствор обеспечивал поддержание равномерной температуры по всей площади геля. Выделение при электролизе пузырьков газа с расположенного на самом дне “верхней” камеры платинового электрода способствовало некоторому перемешиванию слоев буфера и более быстрому достижению одинаковой температуры всей жидкости. Надо отметить, что многие выпускаемые сегодня камеры для секвенирующего электрофореза основаны на этом принципе. Еще один подход равномерного распределения Джоулева тепла был осуществлен с использованием 2 мм пластины из алюминия (имеющего высокую теплопроводность), находящейся в тесном контакте с передним стеклом [Lang, Burger, 1990]. Однако в отличие от способа распределения тепла путем принудительного циркулирования воды в термоплате в последних случаях требовалось проведение предварительного электрофореза, имеющего, в том числе, целью достижение рабочей температуры геля. Более того, во время самого электрофореза необходим периодический контроль за температурой геля из-за опасения его перегрева.

Желание экспериментаторов “читать” с одного геля как можно больше полос ДНК привело к использованию сверхдлинных гелей, что увеличило разрешающую способность метода электрофоретического



разделения. Так, в числе первых приборов, доступных из коммерческих источников, был прибор модели S1 фирмы “Bethesda Research Laboratories” (США) с максимальной длиной геля до 85 см. В настоящее время ряд фирм производят электрофоретические камеры, приспособленные для работы со стеклами длиной 80 см. В литературе имеются сведения о применении метровых гелей [Ansorge, Barker, 1984]. “Платой” за “чтение” большего числа нуклеотидов при этом является определенная трудность обращения с такими длинными стеклами и гелями.

Другой крайностью является использование миниатюрного ультратонкого (50 мкм) геля, заливаемого между предметными стеклами для микроскопа, размерами 25 × 75 мм [Stein et al., 1998]. Электрофоретической камерой для такого геля служил обычный прибор для горизонтального агарозного электрофореза, время разделения в котором составляло около 6–7 мин. Несмотря на столь малые размеры геля авторам цитируемой работы при использовании в качестве метки радионуклида  $^{35}\text{S}$  удалось “прочитать” с рентгеновской пленки после 16-часовой экспозиции более 150 нуклеотидов.

Стекла довольно маленького размера (обычно 14 см, но с возможным варьированием от 6 см до 20 см) для секвенирующего гель-электрофореза использовались со специальным прибором, рассчитанным на прямое электрофоретическое блоттирование [Beck, Pohl, 1984]. Принцип метода заключается в постоянном движении специальной нейлоновой мембраны и ее тесном контакте с нижним срезом геля. Таким образом, фрагмент ДНК, выходя из геля, переносится на данную мембрану и сорбируется в определенном месте. Поскольку используемая длина геля оказывается достаточной для разделения одного фрагмента ДНК, отличающегося на один нуклеотид, от другого, то отпадает надобность в стеклах большого размера, требующих значительно большего времени до выхода образца из геля. После экспозиции данной мембраны с сорбированными на ней фрагментами ДНК на рентгеновскую пленку удается прочитать 500–600 и даже около 1000 нуклеотидов.

Движение молекул ДНК в электрическом поле в полиакриламидном геле имеет логарифмическую зависимость и в связи с этим в нижней части расстояние между полосами ДНК, отличающимися на один нуклеотид, значительно больше, чем в верхней части. Таким образом, другим подходом к увеличению числа “читаемых” нуклеотидов с одного геля могло стать не увеличение его длины, а уложение по всей длине геля большего количества полос ДНК за счет постепенного замедления движения молекул ДНК в его нижней части. Это одновременно приводило бы к некоторому увеличению времени электрофореза и фрагменты ДНК крупного размера могли преодолевать чуть большее расстояние, которое уже позволяло осуществить их достоверное “прочтение”. Так, применение гелей с градиентом буфера, образуемым при заливке и приводящим к локальному изменению электрического сопротивления геля, позволило с геля стандартной длины “прочитать” нуклеотидов на треть больше, чем обычно [Biggin et al., 1983]. Создаваемый градиент буфера приводил к изменению (увеличению) ионной силы раствора в нижней части геля и поэтому происходило заметное сни-

жение локального градиента напряжения и, как следствие, имело место торможение более мелких фрагментов ДНК. Однако приготовление такого геля с градиентом буфера весьма непросто. Для контроля должной линейности формирования градиента необходимо использовать маркерный краситель бромфеноловый синий, добавляемый в раствор с высокой концентрацией буфера.

Аналогичный эффект замедления движения фрагментов ДНК меньшего размера достигался также при применении гелей с электролитическим градиентом [Sheep, Seed, 1988]. В этом случае готовился обычный секвенирующий гель с однократным буфером, но нижний резервуар содержал, кроме буфера, 1М ацетат натрия, тогда как в верхней камере концентрация буфера была половинной. Известны работы, где подобный электролитический градиент формировался за счет 0,1 или 0,2М фосфата натрия [Lang, Burger, 1990].

Был реализован еще один интересный подход по замедлению движения мелких фрагментов ДНК в секвенирующем геле [Ansorge, Labeit, 1984; Olsson et al., 1984]. В однородном геле градиент напряжения является постоянным, поскольку электрическое сопротивление геля прямо пропорционально длине геля и обратно пропорционально его сечению. Применение клиновидной формы геля с толщиной 0,2 мм в верхней части и до 1 мм в нижней позволило изменить градиент напряжения и затормозить движение мелких фрагментов ДНК, что привело к уменьшению расстояния между полосами ДНК и, таким образом, увеличило число “читаемых” полос с одного геля. Однако для приготовления таких гелей в приборах некоторых типов требовались специальные спейсеры клиновидной формы. В то же время заливка гелей клиновидной формы техникой скольжения позволяла обойтись лишь дополнительным коротким фрагментом спейсера необходимой толщины, накладываемым на основной спейсер в его нижней части. Применение спейсеров с экспоненциально меняющейся толщиной позволило эффективно разделять более 500 нуклеотидов [States et al., 1991]. Некоторым недостатком таких гелей с переменным сечением является их более высокая толщина в нижней части, затрудняющая высушивание.

Увеличение числа полос, уместяющихся на геле, теоретически могло бы быть осуществлено также путем создания плавно увеличивающегося градиента концентрации полиакриламида. Однако из-за большого размера гелевой пластины и малой толщины заливка такого геля практически невозможна. Частичное решение проблемы состояло в приготовлении геля со ступенчатым изменением концентрации полиакриламида для названного авторами “4–6–8 метода” секвенирования, позволяющего успешно разделять свыше 500 нуклеотидов [Isfort, Ihle, 1988]. Используемые 6%-ные полиакриламидные гели и для протяженных фрагментов ДНК даже 3%- или 4%-ные против 8%- или 20%-ных, предлагавшихся ранее, позволили резко увеличить число разделяемых и “читаемых” полос ДНК [Ansorge, Barker, 1984].

Здесь надо отметить, что качество электрофоретического разделения в значительной степени зависит от правильности приготовления и заливки самого геля. Не последнюю роль играет и чистота используе-

мых реактивов. Мономеры – акриламид и метиленбисакриламид должны быть несколько раз перекристаллизованы и при появлении желтоватой окраски приготовленного концентрированного раствора, свидетельствующей об образовании, в том числе, акриловой кислоты, заменены новыми. Оба этих мономера являются нейротоксичными и обращение с ними, в особенности взвешивание, требует соблюдения строгих правил. В связи с этим многими фирмами поставляются концентрированные растворы этих соединений. Некоторые фирмы пошли еще дальше и для своих электрофоретических камер для секвенирующего электрофореза поставляют уже готовые гели.

Полимеризация полиакриламидного геля есть каталитический процесс, протекающий с помощью тетраметиленамина и персульфата аммония. Хорошо известно, что последний при длительном хранении своих водных растворов частично теряет реакционные свойства и в связи с этим рекомендуется использование свежеприготовленных растворов. Для исключения многократного взвешивания персульфат аммония поставляется некоторыми фирмами в виде желатиновых капсул с заранее отвешенными количествами. Удобство использования таких фасовок может оборачиваться при секвенировании ДНК нежелательным эффектом “размазывания” полос ДНК в верхней части геля [Kamps-Holtzapfel, Stanker, 1995]. Этими авторами был проведен модельный эксперимент, в котором выяснилось, что именно желатин оказывает такой отрицательный эффект, и они советуют в связи с этим просто вскрывать капсулы и высыпать их содержимое, исключив, таким образом, растворение желатины. Довольно сильным ингибитором полимеризации является растворенный в воде кислород, и для его максимального удаления рекомендуется дегазация уже готового раствора мономера перед добавлением инициатора и катализатора этого процесса.

Сам гелевый матрикс тоже не остался без изменения. Был предложен раствор геля HydroLink, имеющий целый ряд преимуществ перед стандартным полиакриламидным гелем, и мономеры которого к тому же менее токсичны по сравнению с акриламидом [Gelfi et al., 1990]. Этот гель характеризуется более высокой механической прочностью и химической стабильностью, позволяющей применять буферы с высоким рН. По существу, щелочные условия проведения секвенирующего гель-электрофореза, исключая формирование у одноцепочечных фрагментов ДНК вторичной структуры без использования мочевины, позволили сократить время обращения с таким гелем перед этапом радиоавтографии за счет исключения необходимого ранее этапа ее удаления. Отсутствие мочевины в геле имеет также еще одно небольшое преимущество, заключающееся в более удобном нанесении препаратов ДНК в ячейки геля, чему обычно препятствует диффундирующая в буфер мочевина. Результатом применения такого геля стало большее количество (свыше 600) и более высокое качество “читаемых” полос ДНК с радиоавтографа геля.

Учитывая большой размер секвенирующего геля и его крайне малую толщину, качественная заливка представляет достаточно серьезную проблему. Обеспечение герметичности пространства для геля,

формируемого двумя стеклами и спейсерами между ними, является важным условием, которое достаточно подробно рассмотрено в ряде статей [Darnay, Engel, 1992; Bollet et al., 1993; Law, Ferrell, 1994; Hengen, 1996]. Главную проблему здесь представляет нижний срез стекол, который или помещается в специальное устройство, поставляемое с некоторыми моделями секвенеров, или обклеивается липкой лентой. К другим вариантам можно отнести применение нижнего спейсера аналогичной толщины с боковыми, который потом необходимо удалить, или использование вместо такого спейсера подходящей фильтровальной бумаги, которую можно оставить. Однако, например, при заливке геля для автоматического секвенера или для прямого блоттинга последний способ не годится. Предложенный метод заливки секвенирующих гелей техникой скольжения [Ansorge, De Mayer, 1980; Garoff, Ansorge, 1981] не требовал какой-либо герметизации, поскольку раствор полимеризующегося полиакриламида удерживался между стеклами, находящимися в горизонтальном положении за счет сил поверхностного натяжения. Похожий способ описан в работе других авторов [Lang, Burger, 1990]. В этом случае верхнее стекло удерживалось над нижним, расположенным горизонтально под небольшим углом, и в образующееся пространство выливался сразу весь рассчитанный объем полиакриламидного геля, после чего верхнее стекло опускалось на нижнее с некоторым накатом вперед. В то же время следует отметить, что оба последних способа не позволяют, например, залить гель с градиентом буфера или сформировать предгель, не содержащий буфера и мочевины. В литературе встречаются упоминания, что наличие такого предгеля позволяет повысить разрешающую способность этого метода за счет некоего концентрирования фрагментов ДНК перед их вхождением в настоящий гель [Ямщиков, 1990]. Вторая проблема, сопровождающая заливку и присущая всем этим способам, состоит в недопущении образования пузырьков воздуха, которые могут исказить электрическое поле и ухудшать картину электрофоретического разделения, вплоть до ее полной непригодности. Избежать этого можно, главным образом, за счет качественно вымытых стекол, да и сам раствор мономеров не должен содержать никаких механических загрязнений.

Определенную проблему представляло “чтение” нуклеотидных последовательностей, имеющих сильную вторичную структуру. Возникающая компрессия иногда затрудняла выявление истинного положения отдельных нуклеотидов. Уже упоминавшееся выше в связи с эффектом “улыбки” термостатирование геля при повышенной температуре практически решало и эту проблему [Ansorge, De Maeyer, 1980; Garoff, Ansorge, 1981]. В отдельных, особо трудных, случаях при секвенировании ДНК с очень высоким GC-составом использовался гель с содержанием 40%-ного формамида как дополнительного денатурирующего агента, кроме стандартной мочевины, что не позволяло образовываться “шпилечным” структурам. Однако при этом часто имело место некоторое разбухание геля и было предложено применять только один 50%-ный формамид без мочевины в качестве денатурирующего агента [Khambaty, Ely, 1990]. Был разработан горизонтальный диск-электро-

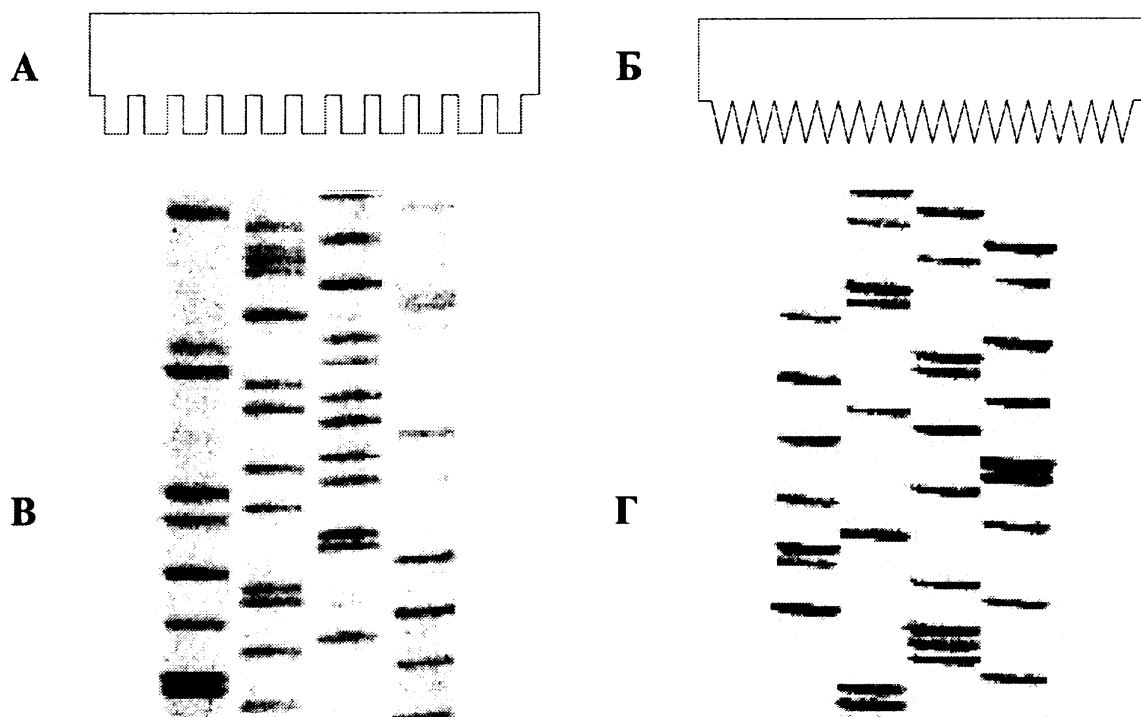


Рис. 4.2. Гребенки для формирования колодцев в секвенирующем полиакриламидном геле

А – обычная гребенка; Б – гребенка типа “зубы акулы”; В – радиоавтограф секвенирующего геля с использованием обычной гребенки; Г – радиоавтограф секвенирующего геля и с использованием гребенки “зубы акулы”

форе́з с концентрирующим предгелем 0,5 мм толщины, что позволяло осуществить легкое нанесение образцов, и ультратонким 50-микронным основным разделяющим гелем [Carninci et al., 1995]. Такой ультратонкий гель позволял прикладывать к нему высокое напряжение, что приводило к исключению компрессии полос ДНК, принадлежащих фрагментам с высоким GC-составом. Еще более тонкий (25 мкм) горизонтальный гель был предложен для эффективного разделения продуктов секвенирующих реакций [Brumley, Smith, 1991]. Авторы продемонстрировали высокое разрешение, достигаемое в их приборе и очень малое время разделения, составляющее всего 15 мин. Причиной этого было очень высокое напряжение, достигающее 250 В на см длины геля, причем размеры гелевой пластины составляли 8,2 × 25 см. Главным недостатком этой системы была ее довольно сложная конструкция.

Разработанный в начале 1980-х годов новый тип гребенки – “зубы акулы”, названный так за то, что его зубья буквально вонзались в ровную верхнюю границу геля и образовывали, таким образом, перегородки, позволил наносить образец ДНК в соседние дорожки практически без промежутков, что повысило достоверность “чтения” особенно в верхней части геля, и позволило увеличить число “читаемых” полос ДНК за счет “касания” их краев с полосами в соседних треках (рис. 4.2). Следует отметить, что у гребенки “зубы акулы” обе стороны являются рабочими, поскольку на стадии полимеризации геля его ровная верхняя граница образуется как раз с помощью гладкой стороны гребенки. Рекомендуемое иногда 8-дорожечное нанесение препарата ДНК, напри-

мер в последовательности GATCGTAC (хотя оно и менее экономно), также способствует продлению “чтения” в верхней части геля, поскольку в этом случае каждый нуклеотид вплотную соседствует со всеми остальными и, таким образом, значительно легче определить его истинное местоположение относительно других, не прибегая к экстраполяции. Уменьшение расстояния между зубьями и за счет этого уменьшение ширины ячейки для нанесения препарата (причем без заметного ухудшения “чтения”) позволили разместить на одном и том же геле большее число анализируемых образцов и, таким образом, повысить производительность каждого конкретного прибора. Стандартные гребенки также стали значительно мельче и обычно формируют ячейки с шириной около 5 мм против использовавшихся ранее 8 или даже 10 мм.

Другой способ экономии гелевого пространства был предложен Нельсоном и соавт. [Nelson et al., 1992]. К нанесению образцов ДНК на гель они подошли с математической точки зрения и записали их порядок в бинарных кодах. Так, для секвенирования ДНК ферментативным методом обычный четырехдорожечный порядок нанесения образцов ДНК может быть представлен следующим образом:

$$G = (1,0,0,0); \quad A = (0,1,0,0); \quad T = (0,0,1,0); \quad C = (0,0,0,1),$$

где 1 означает присутствие, а ноль – отсутствие. (При секвенировании ДНК химической деградацией подобная запись выглядит несколько сложнее.) Однако однозначное “прочтение” всех четырех нуклеотидов возможно осуществить после разделения их в виде определенных комбинаций всего в трех соседних дорожках. В цитируемой работе были осуществлены электрофоретические разделения продуктов терминирующих реакций в трех дорожках по двум схемам. Во втором, более приемлемом, варианте порядок нанесения был следующий – GA, GCT и CA, что соответствовало  $G = (1,1,0)$ ;  $A = (1,0,1)$ ;  $T = (0,1,0)$  и  $C = (0,1,1)$  после их перевода в бинарный код. Наличие полос одновременно во всех дорожках имело шифр (1,1,1) и свидетельствовало о неспецифической терминации. Следует отметить значительную трудность “чтения” таких радиоавтографов, и экономия четвертой части всего геля может привести к серьезным ошибкам на этапе ее “прочтения”. Однако подобное нанесение может иметь некоторый смысл при автоматическом сканировании радиоавтографа геля, где специальная компьютерная программа могла бы определять правильную последовательность нуклеотидов. Что касается автоматического флуоресцентного секвенирования ДНК, то там подобная проблема экономии дорожек, как это будет видно из дальнейшего изложения (глава 9), решается по-другому.

Еще один своеобразный способ нанесения препаратов ДНК на секвенирующей гель описан в работе других авторов [Burt et al., 1994]. Так, в предложенном ими варианте при одновременном секвенировании множественных гомологичных фрагментов ДНК схема нанесения образцов выглядела следующим образом:  $A_1, A_2, A_3, A_4, A_5, A_6, C_1, C_2, C_3, C_4, C_5, C_6, G_1, G_2 \dots$  и т.д. Однако надо отметить, что подобное нанесение не подходит для обычного секвенирования ДНК, а предназначено для выявления полиморфизма нуклеотидных последовательно-

стей секвенируемых фрагментов ДНК. При этом различия между последовательностями получаются более наглядными, что и было продемонстрировано в цитируемой выше статье. Особый эффект достигается при использовании гребешка “зубы акулы”, поскольку в случае одинаковых азотистых оснований во всех образцах соответствующие им полосы сливаются в одну широкую. Отличающиеся нуклеотиды разрывают такую линию и сразу становятся заметны.

Отсутствие промежутков между соседними матрицами ДНК, наносимыми в виде четырех реакций, как, например, G A T C, при массовых анализах затрудняло визуальное выделение конкретного образца секвенируемой ДНК. Хотя проблема разграничения образцов ДНК в секвенирующем геле решается простым пропусканием одной дорожки между матрицами, но при этом теряется значительное число дорожек, что довольно неэкономно. Этот способ вполне возможен с обеими типами гребенки, однако гребенка типа “зубы акулы” в отличие от обычной позволяет сформировать промежуток между образцами ДНК, не оставляя между ними пустые дорожки. На протяжении многих лет нами использовался подход для формирования промежутков между образцами ДНК путем их последовательного нанесения через небольшой промежуток времени (обычно около 1 мин), за который фрагменты ДНК предыдущей матрицы под действием электрического тока уже успевали войти в сам гель. Этого оказывалось вполне достаточно, чтобы на заключительном радиоавтографе между соседними образцами ДНК была видна вертикальная полоса, хороша заметная на радиоавтографе, фрагмент которого приведен на рис. 4.3.

Такое последовательное нанесение образцов ДНК с попеременным подключением/отключением источника питания занимает несколько больше времени, чем обычное, но в то же время позволяет перед самым нанесением очередного образца промывать колодцы с целью удаления диффундирующей из геля мочевины, которая несколько мешает нанесению препаратов ДНК. И, во-вторых, как уже отмечалось выше, на заключительном радиоавтографе отдельные матрицы ДНК хорошо заметны в виде четверок дорожек, принадлежащих одному образцу. Более скоростной вариант подобного нанесения, приводящего к формированию промежутков между образцами ДНК, был предложен Нельсоном и Краветцом [Nelson, Krawetz, 1994]. Главное отличие предложенного ими подхода заключалось в нанесении в первый заход нечетных матриц (1, 3, 5 и так далее) в соответствующие дорожки (1,2,3,4; 9,10,11,12; 17,18,19,20 и так далее), после чего на 5 мин подавалось рабочее напряжение. А во втором заходе наносились четные матрицы (2, 4, 6 и так далее) в оставшиеся пустыми соответствующие дорожки геля. В результате на заключительном радиоавтографе были видны те же вертикальные светлые полосы, разделяющие соседние образцы ДНК.

Стандартным буфером для нанесения образцов ДНК на секвенирующий гель служит добавляемый в терминирующие реакции в виде стоп-раствора формамид с маркерными красителями—ксиленоловым голубым и бромфеноловым синим, имеющими различную подвижность. Данные красители в полиакриламидных гелях разной concentra-

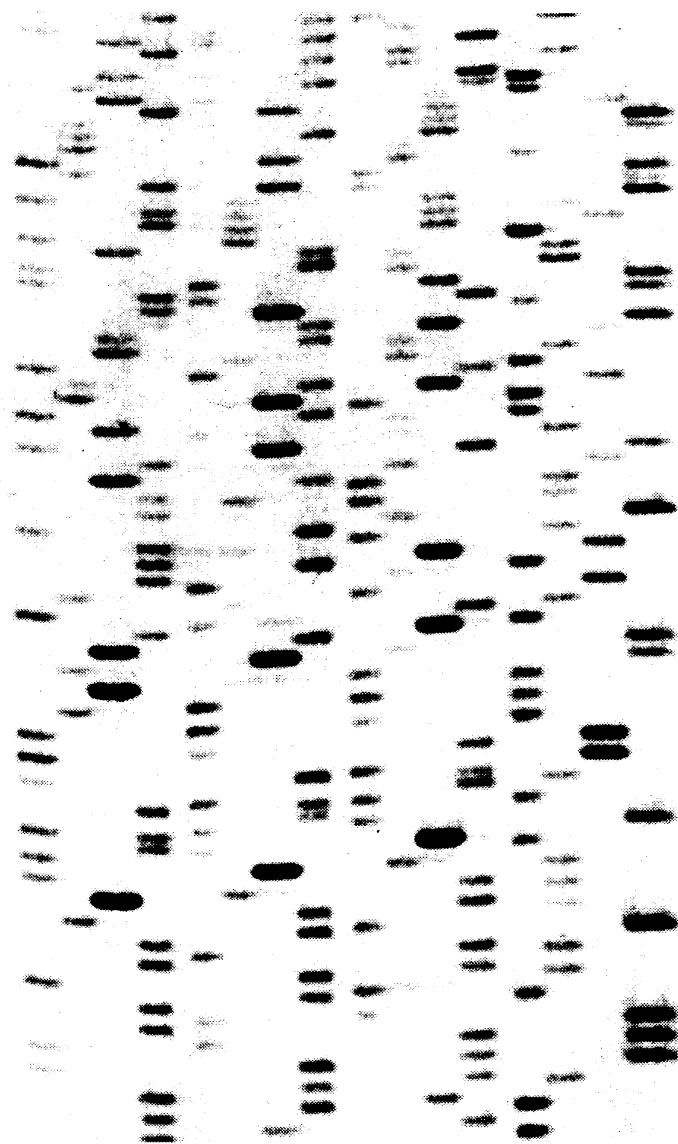


Рис. 4.3. Радиоавтограф секвенирующего геля с использованием гребенки “зубы акулы” путем последовательного нанесения образцов ДНК после того, как предыдущий образец уже вошел в гель под действием электрического тока

ции движутся с фрагментами ДНК различающихся размеров, что позволяет контролировать процесс разделения продуктов реакции. (Соответствие электрофоретической подвижности данных красителей с определенными фрагментами ДНК в полиакриламидных гелях с мочевиной приведено в Приложении.) В ранних работах рекомендовалось даже после завершения электрофоретического разделения отмечать места локализации этих красителей в геле с помощью радиоактивных чернил с тем, чтобы после радиоавтографии на рентгеновской пленке можно было оценить размер фрагмента ДНК, движущегося рядом с ними. Подобная цель преследовалась и при нанесении на гель радиоактивно меченной с помощью полинуклеотидкиназы и  $\gamma$ -АТФ Oligo (dT)<sub>4-22</sub>-лестницы, поставляемой в то время фирмой “Bethesda Research Laboratories”. В настоящее время необходимость точного определения положения какого-то конкретного нуклеотида от 5'-конца фрагмента



ДНК требуется только в определенных случаях, связанных больше не с самим секвенированием, а с соответствующим анализом фрагмента ДНК, например с помощью футпринтинга. Для этой цели экспериментаторы в соседних дорожках обычно “секвенируют” фрагмент ДНК с уже известной последовательностью. Другой альтернативой является использование специальной маркерной лестницы SequaMark, поставляемой фирмой “Research Genetics, Inc.” (США), в виде матрицы ДНК с клонированным фрагментом ДНК, в котором остатки аденина встречаются через каждые 10 нуклеотидов. Таким образом, при ферментативном построении новой цепи ДНК в условиях терминации с ддТТФ с использованием соответствующей метки образуется характерная лестница полос, имеющая размер от 10 до 500 нуклеотидов (считая от праймера) со ступеньками, равными 10 нуклеотидам. В качестве реперных зон служат множественные полосы, предшествующие сотням нуклеотидов. Так, в районе 100 нуклеотидов содержится пять полос, соответствующих 96, 97, 98, 99 и самому 100 нуклеотидам. Аналогичная ситуация только с различающимся числом полос характерна для 200, 300, 400 и 500 нуклеотидов.

Иногда при повторном нанесении препаратов ДНК через довольно большой промежуток времени бывает нужно как-то заранее отметить те дорожки секвенирующего геля, в которые уже были нанесены образцы, поскольку маркерные красители и ксиленоловый голубой и тем более бромфеноловый синий успеют отойти от старта на значительное расстояние и определение нужных колодцев “на глаз” чревато ошибкой. В этом случае в буфер для нанесения добавляется декстран голубой, представляющий собой краситель цибахроновый голубой, ковалентно сшитый с декстраном, имеющим молекулярную массу свыше 2 МДа. Таким образом, уже занятые колодцы оказываются отмеченными голубой полоской остающегося на старте декстрана голубого.

Нельзя не упомянуть о недавно предложенной уникальной гребенке, позволяющей осуществить одновременное быстрое, занимающее всего несколько минут, нанесение 200 образцов на полиакриламидный гель шириной 30 см [Erfler et al., 1997]. Эффективность данной гребенки была продемонстрирована при автоматическом секвенировании ДНК в двух различных приборах, хотя авторы утверждают о возможности ее применения как для вертикальных, так и горизонтальных гелей ультратонкой и стандартной толщины. Особенностью этой гребенки является то, что она изготовлена из пористого материала и поэтому образцы ДНК предварительно наносятся на каждый зубец гребенки, впитываются в него и уже затем гребенка с препаратами ДНК, находящимися в зубьях, втыкается в гель. Под действием электрического тока препараты ДНК выходят из гребенки и входят в полиакриламидный гель.

В последнем случае можно отметить совмещение процессов формирования колодцев и нанесения в них образцов ДНК, осуществляемое одномоментно. В то же время провести обычное нанесение в колодцы препаратов ДНК, учитывая весьма небольшую толщину секвенирующих гелей (0,4; 0,2 или даже 0,05 мм) бывает весьма непросто. Так, фирма “Drummond” (США) одной из первых начала выпуск специальных

автоматических пипеток для секвенирования с точным дозированием малых объемов и в комплекте с тонкими поликарбонатными наконечниками. Однако простое решение проблемы заключается в самостоятельном изготовлении тонких полипропиленовых наконечников путем оттягивания над пламенем спиртовки обычных желтых наконечников, что при некотором навыке получается без особых сложностей.

Возвращаясь к порядку нанесения образца ДНК на секвенирующий гель, стоит обратить внимание на последовательность расположения различных типов реакций. Так, при секвенировании ДНК методом химической дегградации, его классическим вариантом, в первой дорожке обычно разделяются продукты G-реакции, во второй – G + A, в третьей – C + T и в четвертой – C. Данная последовательность является палиндромом (GRYC), поэтому в случае необходимости “чтения” противоположной цепи ДНК можно просто повернуть радиоавтограф вокруг своей вертикальной оси на 180° и таким образом “читать” комплементарную цепь. Конечно, последовательность комплементарной цепи можно потом легко получить с помощью компьютера, однако на стадии “чтения” различных субклонов, в случае если их последовательности принадлежат разным цепям, это имеет некоторый смысл. Однако существуют некие правила “чтения” тех или иных нуклеотидов в связи с их окружением и поэтому такое “комплементарное” “чтение” “не тех” нуклеотидов может привести к увеличению числа ошибок.

При разделении продуктов секвенирующих реакций, проведенных по методу Сэнгера, существует несколько бóльшая свобода выбора порядка их нанесения на секвенирующий гель. В разных статьях приводятся различные варианты, как, например, – ACGT, AGCT, TCGA, TGCA, GATC или GTAC. Другие варианты также возможны, но непалиндромные последовательности скорее всего нежелательны. Первые четыре варианта предпочтительны, когда секвенируется ДНК с довольно высоким составом GC-пар. В этих случаях высока вероятность компрессии и соседнее расположение G и C дорожек в какой-то степени сможет нивелировать его эффект. Последовательность нанесения в виде GATC также имеет свое объяснение. Поскольку данная последовательность нуклеотидов GATC соответствует сайту узнавания рестрикционной эндонуклеазы *Sau3AI*, весьма широко используемой для получения субклонов, то лестница полос GATC на соседних дорожках позволяет находить потенциальные границы субклонов. Аналогичное применение находит последовательность нанесения GTAC (сайт узнавания рестрикционной эндонуклеазы *RsaI*) при секвенировании однонаправленных делеционных субклонов, полученных в специализированном векторе pSequoiaT12.

Вероятно, здесь следует отметить (хотя и нелегко объяснить), что палиндромная последовательность нуклеотидов GATC из-за часто применяемых некоторых тетра- и гексануклеотидных рестрикционных эндонуклеаз (*BamHI* – семейство ферментов), узнающих данную последовательность, приобрела в молекулярной биологии даже некоторое символическое значение. Видимо, поэтому в Германии имеется фирма, производящая оборудование для электрофореза и секвенирования

ДНК, название которой и состоит из этих букв/нуклеотидов – “GATC GmbH”. Более того, в одном из своих рекламных буклетов данная фирма взяла на себя смелость написать, что “Life’s alphabet – GATC”, или, в переводе на русский, “GATC – алфавит жизни”. Возможно, они правы, поскольку именно это сочетание букв GATC как никакое другое напоминает большинству молекулярных биологов о ДНК, генах, их клонировании и секвенировании.

## 4.2. ГЕЛЬ-ЭЛЕКТРОФОРЕЗ

Для проведения секвенирующего гель-электрофореза фрагментов ДНК, кроме специально изготовленной камеры, которая может быть весьма простой и способной только поддерживать стекла с залитым между ними гелем, даже не обеспечивая при этом отвод выделяющегося в процессе электрофореза тепла, необходим высоковольтный источник питания. Для эффективного разделения в полиакриламидном секвенирующем геле фрагментов ДНК подаваемое напряжение должно быть в пределах от 30 до 50 В на см длины геля и, учитывая его большой размер, источник питания должен обеспечивать стабилизацию напряжения не менее 2000 В. Для более эффективного разделения фрагментов ДНК и повышения общей производительности метода можно порекомендовать приобретение фабрично изготовленного прибора, в котором предусмотрено и определенное удобство заливки геля и дальнейшего обращения с ним, и равномерное распределение Джоулева тепла во время электрофореза, и комплект запасных спейсеров и гребешков. Выбор подобных аппаратов, включая автоматические секвенаторы ДНК, достаточно велик, но и их стоимость весьма существенна. Однако следует отметить, что для осуществления небольших проектов секвенирования ДНК не обязательно приобретать дорогостоящие автоматические секвенаторы (хотя в последнее время появились относительно недорогие модели, о которых речь пойдет в главе, посвященной автоматическому секвенированию ДНК).

Успехи в разделении крупных молекул двуцепочечной ДНК с помощью различных вариантов пульс-электрофореза в переменном поле заставили исследователей попытаться применить данный подход и к разделению одноцепочечных фрагментов ДНК в денатурирующем полиакриламидном геле. Так, при электрофоретическом разделении секвенируемых молекул ДНК были применены его наиболее простые варианты, основанные на временном прекращении подачи напряжения, а также реверсии электрического поля [Lai et al., 1989; Birten et al., 1990]. Однако существенного увеличения разрешающей способности секвенирующего гель-электрофореза добиться не удалось при одновременном значительном увеличении его длительности. В то же время математический анализ свидетельствовал о теоретическом уменьшении толщины полос ДНК в таком геле, что должно было привести к их большей четкости и большему числу достоверно “читаемых” полос [Slater, Noolandi, 1988]. Другими авторами был предпринят детальный анализ

потенциальных возможностей применения переменных электрических полей для целей секвенирования ДНК [Heller, Beck, 1992]. В результате такого исследования они пришли к выводу, что при определенных условиях можно добиться приемлемого соотношения между временем электрофоретического разделения и его разрешающей способностью. Применение более тонких гелей (0,1 мм против довольно обычных 0,2 мм) показало, что подход с использованием пульс-электрофореза все же позволяет осуществлять “чтение” в области 800–900 нуклеотидов [Noolandi et al., 1993]. Компьютерная реконструкция сканированного изображения радиоавтографа позволила этим авторам продлить границу “чтения” за 1000 нуклеотидов и отдельные индивидуальные основания могли быть “прочитаны” таким способом за пределами 2000 нуклеотидов.

Классической буферной системой для секвенирующего гель-электрофореза является трис-боратный буфер, имеющий рН 8,3 [Maxam, Gilbert, 1977; Sanger et al., 1977; Sanger, Coulson, 1978]. Стандартной однократной концентрацией этого буфера считается 89 мМ трис и 8,9 мМ борная кислота, а также 2 мМ ЭДТА. В отдельных случаях используют или половинные, или, наоборот, слегка увеличенные концентрации буферного раствора. Считается, что для разделения фрагментов ДНК большего размера необходимы увеличенная концентрация ЭДТА (до 4 мМ) и более высокий рН (до 8,8). В литературе имеются также сведения об использовании, например, *трис*-цитратного буфера [Landick et al., 1984]. Было отмечено, что ввиду присутствия в секвенируемых образцах ДНК, приготовленных ферментативным методом, крайне незначительного (но достаточного для нежелательного эффекта) количества глицерина, содержащегося изначально в самом ферментном препарате происходит деформация полос ДНК в верхней части геля, особенно при длительных электрофорезах [Fuller, 1989]. Автор подчеркивает, что причина этого явления не ясна, однако, вероятно, здесь происходит образование сложных эфиров борной кислоты и глицерина, оказывающих заметное воздействие на картину электрофоретического разделения больших фрагментов ДНК в верхней части геля. В дальнейшем было показано, что замена борной кислоты на таурин ( $\beta$ -аминоэтансульфо кислота) позволяет преодолеть это негативное явление (Pisa-Williamson, Fuller, 1992). Секвенирующий диск-электрофорез с использованием для нижней и верхней электрофоретических камер обычного *трис*-бората, а в качестве гелевого буфера *трис*-сульфата позволил несколько сократить время электрофоретического разделения при некотором увеличении его разрешающей способности [Carninci et al., 1990].

При электрофоретическом разделении радиоактивно меченной ДНК небольшие фрагменты, а также остатки непрореагировавшей метки могут выходить из геля в нижний буферный резервуар, загрязняя его. С целью освобождения нижнего буфера от радиоактивного загрязнения и превращения его в твердые радиоактивные отходы был предложен способ сорбции радиоактивных нуклеотидов и мелких фрагментов ДНК на анионно-обменной смоле путем колоночной хроматографии [Kaczorowski et al., 1994]. Эффективность данного процесса весьма

**А.**

1	...СААТАСГССАССТГддG	Нет компрессии
2	...СААТАСГССАССТГГддC	Есть компрессия
3	..СААТАСГССАССТГГСддG	Есть компрессия
4	...СААТАСГССАССТГГСГддA	Нет компрессии

**Б.**

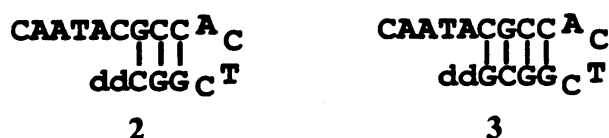


Рис. 4.4. Возникновение и исчезновение участков с потенциальной вторичной структурой на 3'-конце новосинтезированных фрагментов ДНК.

Нуклеотиды, принимающие участие в формировании потенциальной “шпильки”, подчеркнуты. А – линейная последовательность ДНК, Б – “потенциальные шпилечные” структуры для нуклеотидных последовательностей 2 и 3

высока и приближается к 100%, поэтому уже не требуется обращения с отработанным нижним буферным раствором как с жидким радиоактивным отходом и он после такой дезактивации может быть просто слит в раковину.

Уже упоминавшаяся компрессия фрагментов ДНК при их разделении секвенирующим гель-электрофорезом представляет одну из самых серьезных проблем, которая приводит, подчас, к неверному прочтению правильной последовательности отдельных нуклеотидов. Известно, что такая аномальная подвижность определенных фрагментов ДНК вызвана формированием участков со вторичной структурой на 3'-конце новосинтезированного фрагмента ДНК. Если внимательно рассмотреть процесс образования компрессии, то можно видеть, что обычно в ее формировании принимают участие 3 – 4 нуклеотида (по одной цепи), представляющие собой, как правило, G- и C-основания (рис. 4.4). Причем до какого-то участка ДНК никаких “шпилечных” структур не возникает и фрагмент ДНК имеет нормальную подвижность. Но следующий по размеру фрагмент ДНК, показывающий компрессию, отличается от предыдущего только на один терминальный нуклеотид (дидезоксинуклеотид). Проксимальным дезоксинуклеотидом (N – 1) для данного фрагмента является тот же терминальный дидезоксинуклеотид предыдущего фрагмента. Таким образом, добавление только одного очередного нуклеотида приводит к формированию участка вторичной структуры. Дальнейшее увеличение рассматриваемого здесь фрагмента ДНК еще на 1–2 нуклеотида сохраняет за ним сформировавшийся участок вторичной структуры и, как следствие, имеет место компрессия полос ДНК. Однако при присоединении последующих нуклеотидов подобные “шпилечные” структуры теряют свою прочность, исчезает эффект компрессии и фрагмент ДНК приобретает свою обычную подвижность.

Интересное исследование влияния соседних нуклеотидов на подвижность фрагментов ДНК было проведено в связи с проблемой компрессии ДНК [Bowling et al., 1991]. Авторы провели детальное исследование подвижности различных фрагментов ДНК, различающихся своими терминальными, проксимальными (N – 1) и антепроксимальным (N – 2) нуклеотидами. Ими было обнаружено, что если фрагмент ДНК несет на своем 3'-конце ddC, то он будет обладать наибольшей подвижностью и будет максимально приближен к предшествующему ему более короткому фрагменту ДНК, отличающемуся на один нуклеотид. И если у предшествующего фрагмента ДНК этот нуклеотид будет ddT, характеризующийся минимальной подвижностью, то высока вероятность плохого разделения таких фрагментов ДНК. По крайней мере, по данным этих авторов, в диапазоне до 400 нуклеотидов фрагменты ДНК с различными терминальными ддНМФ характеризуются следующей подвижностью:

быстрее —ddC < —ddA < —ddG < —ddT медленнее,

причем пурины обладают весьма близкими величинами подвижности.

В случае, когда dC и dT являются проксимальными нуклеотидами, их влияние на подвижность также весьма заметно и кардинально отличается от приведенного выше для терминальных нуклеотидов. Так, при нахождении в проксимальном положении dT такие фрагменты ДНК будут обладать наибольшей подвижностью, а dC – наоборот:

быстрее —dT-ddN < —dR-ddN < —dC-ddN медленнее.

Для “медленных” dC-ddT олигомеров присутствие антепроксимального пурина сделает подвижность таких фрагментов ДНК быстрее по сравнению с пиримидином в этом положении:

быстрее —dR-dC-ddT < —dR-dC-ddT медленнее.

В результате проведенного исследования авторы цитируемой работы пришли к заключению, что выявленные ими некоторые отличия в подвижности фрагментов ДНК зависят не только от их общей молекулярной массы и заряда, но и от весьма заметного взаимодействия различных нуклеотидов, находящихся на 3'-конце разделяемых гель-электрофорезом фрагментов ДНК.

В первоначальных экспериментах после завершения электрофореза гель во избежание диффузии полос ДНК в замороженном виде экспонировали на рентгеновскую пленку. Однако в этом случае присутствующие в геле вода и мочевины способствовали “гашению” радиоактивных сигналов и полосы ДНК на радиоавтографе получались нечеткими. Введенный впоследствии этап фиксации олигонуклеотидов в геле с помощью 10%-ной уксусной кислоты, кроме перевода нуклеиновых кислот в нерастворимую форму, удалял мочевины. Уже упоминавшийся выше маркерный краситель – бромфеноловый синий – является рН-индикатором и в тех случаях, когда он не успевает выйти и остается в геле, по изменению его окраски с фиолетово-синей на желто-коричневую можно судить об изменении кислотности среды в самом геле. Последу-

ющее кратковременное выдерживание фиксированного геля в 10%-ном метаноле или этаноле удаляло следы уксусной кислоты. После удаления избыточной влаги такой гель, завернув в кухонную пленку Saran Wrap, уже можно было экспонировать на рентгеновскую пленку. Однако присутствующая в геле вода затрудняла проведение этапа радиоавтографии. Дальнейшим шагом на пути улучшения качества получаемых сигналов на рентгеновской пленке стало высушивание фиксированных гелей в специальных аппаратах для сушки. Гель после завершения электрофореза и его фиксации переносился вместе с бумагой 3М на лист пористого полиэтилена в сушилке гелей. Несмотря на то, что многие приборы обеспечивали высушивание при повышенной температуре, продолжительность этого процесса сильно зависит от вакуумного разрежения, создаваемого различными типами насосов, и в случае использования простого водоструйного насоса весьма значительна. Отрицательным моментом является также то, что немного недосушенный полиакриламидный гель имеет свойство при разгерметизации мгновенно трескаться и съеживаться и спасти такой гель уже не представляется возможным. Разработанный подход все же применялся весьма активно и повлек за собой производство специальных сушилок гелей с рабочей площадью, пригодной для сушки больших секвенирующих гелей. Такие высушенные гели могли экспонироваться на рентгеновскую пленку без опасения их прилипания к желатиновому слою при комнатной температуре без дополнительного слоя пленки Saran Wrap, что приводило к увеличению четкости полос. Немаловажной чертой высушенных гелей, также повышающей разрешение, является, кроме отсутствия самой воды, их меньшая толщина по сравнению с влажными.

Обработка одного из стекол  $\gamma$ -метакрилоксипропилтриметоксисиланом привела к образованию на его поверхности реакционноспособных групп. В результате полимеризации полиакриламидного геля он оказывается ковалентно “сшитым” с таким стеклом и при высушивании продолжает оставаться прочносвязанным, не меняя свои размеры за исключением толщины [Ansorge, De Maeyer, 1980; Garoff, Ansorge, 1981]. После окончания электрофореза стекла разъединились и гель, приполимеризованный к одному из стекол, фиксировали в уксусной кислоте и затем после непродолжительного выдерживания в 10%-ном этаноле или метаноле высушивали. Следует отметить, что второе стекло необходимо обрабатывать диметилдихлорсиланом во избежание прилипания геля к нему. При пассивном высушивании при комнатной температуре гель может сохнуть в течение ночи. Для ускорения этого процесса можно применить или фен с теплым воздухом, или поместить стекло с гелем в термостат с температурой не выше 80°C. При использовании клиновидных гелей на этапе фиксации в уксусную кислоту иногда необходимо добавить 2–3% глицерина во избежание растрескивания геля. Метод высушивания полимеризованных на стекле гелей довольно прост и удобен, но весьма серьезным недостатком является исключение из обращения стекол на время радиоавтографии. Вторым недостатком является относительно сложное удаление геля со стекла после завершения этапа радиоавтографии. С этой целью стекло с гелем

выдерживают в воде и затем влажный гель удаляют каким-либо пластиковым скребком. Другим способом является выдерживание стекла с гелем в растворе 1М гидроокиси натрия до тех пор, пока гель сам не отстанет от стекла.

Предложенный впоследствии метод высушивания геля на стекле, не обработанном  $\gamma$ -метакрилоксипропилтриметоксисиланом, позволяет удалять гель без применения щелочи [Lang, Burger, 1990]. Однако стекло перед полимеризацией геля все же требует травления щелочью и полного исключения контакта с диметилдихлорсиланом. Еще один подход к фиксации и последующему высушиванию геля заключается в использовании специальных Gel-Fix пленок.

#### 4.3. ПЕРЕНОС ФРАГМЕНТОВ ДНК ИЗ СЕКВЕНИРУЮЩЕГО ГЕЛЯ НА МЕМБРАННЫЕ ФИЛЬТРЫ

В настоящее время перенос фрагментов ДНК из секвенирующего геля на мембранные фильтры преследует обычно несколько иные цели, чем тогда, когда этот метод только разрабатывался. Так, в большинстве случаев перенос ДНК из геля на фильтры осуществляется для последующей визуализации полос ДНК с помощью нерадиоактивной колориметрической или хемилюминесцентной детекции, описанной выше в разделе 2.6. Причем вкупе с мультиплексным секвенированием возможно последовательное многократное проведение этапов молекулярной гибридизации с определенным зондом, несущим ту или иную метку; визуализация ДНК зонда после ее гибридизации с ДНК, находящейся на фильтре; удаление зонда и проведение аналогичных этапов со следующим зондом. С фрагментами ДНК, находящимися в полиакриламидном геле, проведение подобных процедур принципиально невозможно и поэтому перенос фрагментов ДНК из секвенирующего геля на мембранные фильтры занимает важное место в арсенале современной молекулярной биологии.

Предложенный в 1984 г. метод геномного секвенирования был основан на выявлении определенных последовательностей ДНК путем их гибридизации с меченым зондом [Church, Gilbert, 1984]. Для осуществления такой гибридизации была необходима процедура переноса фрагментов ДНК из полиакриламидного секвенирующего геля на нейлоновый фильтр. Перенос даже небольших фрагментов ДНК из полиакриламидного геля на мембранные фильтры за счет капиллярных сил малоэффективен и поэтому в данной работе, учитывая большие размеры геля, была сконструирована специальная камера для электропереноса.

Оригинальным подходом к переносу фрагментов ДНК после их разделения в секвенирующем полиакриламидном геле можно считать предложенный в том же 1984 г. метод прямого электрофоретического блоттирования, требующего, впрочем, специально изготовленного аппарата для секвенирования ДНК довольно необычной конструкции [Beck, Pohl, 1984]. Принцип метода заключается в постоянном движе-



нии нейлоновой мембраны через нижнюю электрофоретическую камеру, сопровождающимся ее постоянным тесным контактом с нижним срезом геля. Таким образом, фрагменты ДНК, отличающиеся один от другого на один нуклеотид, выходя из геля, сорбируются на данной мембране каждый в своих определенных местах. Имеющее место при таком способе переноса постоянное трение мембраны о нижний срез полиакриламидного геля приводит к его истиранию и сорбции частиц геля на данном фильтре, что отрицательно сказывается на качестве последующего проявления полос ДНК. Поэтому в одной работе было предложено не допускать касания геля с мембраной, а оставлять у нижнего среза стекла зазор в 2 мм, заполняемый буфером [Dolan et al., 1995]. Поскольку скорость движения мелких и крупных фрагментов ДНК заметно отличается (что на обычном радиоавтографе заметно как уменьшение расстояния между полосами по мере увеличения размеров фрагментов ДНК, которым они соответствуют), то программируемое изменение скорости движения мембранного фильтра вдоль нижнего среза геля позволяет в некоторой степени решить эту проблему. Оказавшийся очень удобным такой способ одновременного разделения фрагментов ДНК и их переноса на мембранный фильтр привел к созданию изготовленных фабричным способом подобных приборов. Так, фирма "Hoefler Scientific Instruments, Inc." начала выпуск прибора, названного как "непосредственно блоттирующий секвенер", а фирма "GATC GmbH" выпускает две подобные модели, различающиеся размером стекол и, как следствие, количеством потенциально разрешаемых нуклеотидов.

Попутно следует отметить, что несмотря на то, что перенос ДНК из секвенирующего геля путем прямого блоттирования имеет главной целью возможность нерадиоактивной детекции полос ДНК, тем не менее нет каких-либо препятствий для разделения радиоактивно меченных фрагментов ДНК и последующей радиоавтографии уже не геля, а самой нейлоновой мембраны на рентгеновскую пленку. Более того, четкость полос на радиоавтографе обычно бывает значительно выше ввиду того, что радиоактивная ДНК находится на поверхности самой мембраны, а не в толще геля как при обычной радиоавтографии и, таким образом, не происходит дополнительного рассеяния  $\beta$ -частиц. Единственное, за чем необходимо следить экспериментатору, так это, чтобы к рентгеновской пленке была обращена именно та сторона мембраны, на которой и сорбирована ДНК, в противном случае четкость полос будет заметно хуже.

Другой способ переноса фрагментов ДНК из секвенирующего полиакриламидного геля на мембранный фильтр можно назвать вакуумным блоттированием [Gross et al., 1988], широко применяемым для обычного переноса ДНК из агарозных гелей (Medveczky et al., 1987; Olszewska, Jones, 1988). Для эффективного переноса фрагментов ДНК из секвенирующего полиакриламидного геля была применена комбинация диффузии полос ДНК, осуществляемой под действием вакуума, а также осмотического градиента, поскольку концентрация буфера в нижнем слое бумаги превышала таковую в верхнем слое в 10 раз. В ка-

честве такой конструкции с источником вакуума служил обыкновенный сушитель геля подходящего размера, на поверхности фильтра из пористого полиэтилена которого размещалось сначала несколько слоев бумаги 3ММ, смоченной в высокосолевым буфере, затем нейлоновый фильтр, сам гель, один слой бумаги 540 и несколько слоев бумаги 3ММ с низкой концентрацией буфера. Сверху весь этот “сэндвич” закрывался листом силиконовой резины и подавался вакуум. Авторы сообщают о переносе фрагментов ДНК размером до 310 нуклеотидов с 40–50%-ной эффективностью в течение 90 мин, однако эффективность переноса фрагмента размером 872 нуклеотида даже после 16 ч не превышала 6% (4% после 90 мин).

Был предложен и более быстрый способ переноса, завершаемый всего за 10 мин, но при этом требующий специального оборудования, поставляемого фирмой “Hoefel” [Patel, Nash, 1995]. Центральными элементами данного прибора, позволяющего осуществлять “полусухой” перенос фрагментов ДНК из секвенирующего геля на нейлоновую мембрану, являются катодный и анодный электроды. В качестве катодного, отрицательно заряженного, электрода служит пластина из нержавеющей стали, на которую помещается слой фильтровальной бумаги и затем пластина геля. На гель накладывается мембранный фильтр и опять слой бумаги. Верхний – анодный, положительно заряженный, электрод выполнен в виде подвижной ниобиевой проволоки, покрытой платиной. После подачи напряжения с помощью встроенного в данный прибор низковольтного источника питания, но развивающего большую силу тока (до 2А), анодный электрод передвигается по верхнему слою бумаги и, таким образом, происходит локальный “полусухой” электроперенос фрагментов ДНК. Проведенное сравнение авторами цитируемой работы полноты переноса данным методом и классическим переносом под действием капиллярных сил показало, что после 10 мин “полусухого” переноса в геле не остается никаких фрагментов ДНК, тогда как количество ДНК в геле после двухчасового обычного переноса визуально изменилось очень незначительно.

Не рассматриваемые здесь методы фиксации фрагментов ДНК на фильтрах в значительной степени зависят от последующих предполагаемых экспериментов с этим мембранным фильтром, от его типа и могут находиться в диапазоне от их не фиксирования вообще, простого выдерживания при температуре 80 °С до облучения УФ-светом или обработки в микроволновой печи. Хотя, надо заметить, что при многократных последовательных этапах гибридизации, проявления и отмывки, необходимых в методе мультиплексного секвенирования, очень важное значение приобретает прочность фиксации секвенированной ДНК на мембранном фильтре. В связи с этим в некоторых работах определялась пригодность для данных целей различных мембранных фильтров и способов фиксации на них фрагментов ДНК [Creasey et al., 1991; Olesen et al., 1993]. Отмечается, что лучшее соотношение выявляемого сигнала по отношению к фону демонстрируют заряженные нейлоновые фильтры, тогда как для нитроцеллюлозных и поливинилиден, дифлуоридных характерно гашение хемилюминесценции [Beck, 1993а, б].

## Цитируемая литература

- Ямщиков В.Ф. Метод определения последовательности ДНК на твердой фазе // Методы молекулярной генетики и геной инженерии / А.В. Мазин, А.В. Кузнецов, А.С. Краев и др. Новосибирск: Наука, 1990. С. 112–115.
- Ansorge W., Barker R. System for DNA sequencing with resolution of up to 600 base pairs // J. Biochem. and Biophys. Meth. 1984. Vol. 9. P. 33–47.
- Ansorge W., De Maeyer L. Thermally stabilized very thin (0.02–0.3 mm) polyacrylamide gels for electrophoresis // J. Chromatogr. 1980. Vol. 202. P. 45–53.
- Ansorge W., Labeit S. Field gradients improve resolution on DNA sequencing gels // J. Biochem. and Biophys. Meth. 1984. Vol. 10. P. 237–243.
- Beck S. Multiplex DNA sequencing // Methods in molecular biology / Ed. H.G. Griffin, A.M. Griffin. Totowa: Humana pres, 1993a. P. 225–234.
- Beck S. DNA sequencing by chemiluminescent detection // Ibid. 1993b. P. 235–242.
- Beck S., Pohl F.M. DNA sequencing with direct blotting electrophoresis // EMBO J. 1984. Vol. 3. P. 2905–2909.
- Biggin M.D., Gibson T.J., Hong G.F. Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination // Proc. Nat. Acad. Sci. USA. 1983. Vol. 80. P. 3963–3965.
- Birren B.W., Simon M.I., Lai E. The basis of high resolution separation of small DNAs by asymmetric-voltage field inversion electrophoresis and its application to DNA sequencing gels // Nucl. Acids. Res. 1990. Vol. 18. P. 1481–1487.
- Bollet C., Treysac F., De Micco P. A simple technique for pouring sequencing gels // Biotechniques. 1993. Vol. 15. P. 387.
- Bowling J.M., Bruner K.L., Cmarik J.L., Tibbetts C. Neighboring nucleotide interactions during DNA sequencing gel electrophoresis // Nucl. Acids. Res. 1991. Vol. 19. P. 3089–3097.
- Brumley R.L., Jr., Smith L.M. Rapid DNA sequencing by horizontal ultrathin gel electrophoresis // Ibid. 1991. Vol. 19. P. 4121–4126.
- Burt A., Carter D.A., White T.J., Taylor J.W. DNA sequencing with arbitrary primer pairs // Mol. Ecol. 1994. Vol. 3. P. 523–525.
- Carninci P., Gustincich S., Bottega S., Patrosso C., Del Sal G., Manfioletti G., Schneider C. A simple discontinuous buffer system for increased resolution and speed in gel electrophoretic analysis of DNA sequence // Nucl. Acids. Res. 1990. Vol. 18. P. 204.
- Carninci P., Volpatti F., Schneider C. A discontinuous buffer system increasing resolution and reproducibility in DNA sequencing on high voltage horizontal ultrathinlayer electrophoresis // Electrophoresis. 1995. Vol. 16. P. 1836–1845.
- Church G.M., Gilbert W. Genomic sequencing // Proc. Nat. Acad. Sci. USA. 1984. Vol. 81. P. 1991–1995.
- Creasey A., D'Angio L., Jr., Dunne T.S., Kissinger C., O'Keeffe T., Perry-O'Keefe H., Moran L.S., Roskey M., Schildkraut I., Sears L.E., Slatko B. Application of a novel chemiluminescence-based DNA detection method to single-vector and multiplex sequencing // Biotechniques. 1991. Vol. 11. P. 102–109.
- Darnay B.G., Engel M. An efficient method for pouring sequencing gels // Ibid. 1992. Vol. 13. P. 859.
- Dolan M., Ally A., Purzycki M.S., Gilbert W., Gillevet P.M. Large-scale genomic sequencing: Optimization of genomic chemical sequencing reactions // Ibid. 1995. Vol. 19. P. 264–273.
- Erfle H., Ventzki R., Voss H., Rechmann S., Stegemann J., Ansorge W. Simultaneous loading of 200 sample lanes for DNA sequencing on vertical and horizontal, standard and ultrathin gels // Nucl. Acids Res. 1997. Vol. 25. P. 2229–2230.
- Fuller C.W. The effect of excess glycerol on DNA sequencing gels // Editorial Comments. US. Biochem. Corp. 1989. Vol. 16. P. 19.
- Garoff H., Ansorge W. Improvements of DNA sequencing gels // Anal. Biochem. 1981. Vol. 115. P. 450–457.
- Gelfi C., Canali A., Righetti P.C., Vezzoni P., Smith C., Mellon M., Jain T., Shorr R. DNA sequencing in hydrolink matrices: Extension of reading ability to greater than 600 nucleotides // Electrophoresis. 1990. Vol. 11. P. 595–600.
- Gross D.S., Collins K.W., Hernandez E.M., Garrard W.T. Vacuum blotting: A simple method for transferring DNA from sequencing gels to nylon membranes // Gene. 1988. Vol. 74. P. 347–356.

- Heller C., Beck S.* Field inversion gel electrophoresis in denaturing polyacrylamide gels // *Nucl. Acids. Res.* 1992. Vol. 20. P. 2447–2452.
- Hengen P.N.* Methods and reagents: Pouring sequencing gels the old-fashioned way // *Trends Biochem. Sci.* 1996. Vol. 21. P. 273–274.
- Kaczorowski T., Furmanek B., Sektas M.* A methods for removal of radioactive nucleotides from electrophoretic buffers // *Biotechniques.* 1994. Vol. 16. P. 6–7.
- Kamps-Holtzapfle C., Stanker L.H.* Effect of encapsulated ammonium persulfate on DNA banding in sequencing gels // *Ibid.* 1995. Vol. 18. P. 1006–1007.
- Khambaty F.M., Ely B.* A method for the determination of sequence of high G+C% DNA by the sequential application of sequenase and *Taq* polymerase // *Ibid.* 1990. Vol. 9. P. 714, 716.
- Lai E., Davi N.A., Hood L.* Effect of electric field switching on the electrophoretic mobility of single-stranded DNA molecules in polyacrylamide gels // *Electrophoresis.* 1989. Vol. 10. P. 65–67.
- Landick R., Maguire D., Lutter L.C.* Optimization of polyacrylamide gel electrophoresis conditions used for sequencing mixed oligodeoxyribonucleotides // *DNA.* 1984. Vol. 3. P. 413–419.
- Lang B.F., Burger G.* A rapid, high resolution DNA sequencing gel system // *Anal. Biochem.* 1990. Vol. 188. P. 176–180.
- Law J.C., Ferrell R.E.* An improved method for pouring DNA sequencing gels // *Biotechniques.* 1994. Vol. 17. P. 850.
- Maxam A.M., Gilbert W.* A new method for sequencing DNA // *Proc. Nat. Acad. Sci. USA.* 1977. Vol. 74. P. 560–564.
- Medveczky P., Chang C.W., Oste C., Mulder C.* Rapid vacuum driven transfer of DNA and RNA from gels to solid supports // *Biotechniques.* 1987. Vol. 5. P. 242–246.
- Nelson J.E., Krawetz S.A.* Easy-reading DNA sequencing gels // *Ibid.* 1994. Vol. 17. P. 416, 418.
- Nelson M., Van Etten J.L., Grabherr R.* DNA sequencing of four bases using three lanes // *Nucl. Acids Res.* 1992. Vol. 20. P. 1345–1348.
- Noolandi J., Forsyth R., Shi A.-C.* Sequencing using pulsed field and image reconstruction // *Proc. Nat. Acad. Sci. USA.* 1993. Vol. 90. P. 10101–10104.
- Olesen S.E., Martin C.S., Bronstein I.* Chemiluminescent DNA sequencing with multiplex labelling // *Biotechniques.* 1993. Vol. 15. P. 480–485.
- Olsson A., Moks T., Uhlen M., Gaal A.B.* Uniformly spaced banding pattern in DNA sequencing gels by use of field-strength gradient // *J. Biochem. and Biophys. Meth.* 1984. Vol. 10. P. 83–90.
- Olszewska E., Jones K.* Vacuum blotting enhances nucleic acid transfer // *Trends Genet.* 1988. Vol. 4. P. 92–94.
- Patel A., Nash B.* Innovations in non-isotopic DNA sequencing: Using an electrotransfer unit to blot sequencing gels and an automated membrane processor for detecting DNA sequences // *Biotechniques.* 1995. Vol. 18. P. 328–333.
- Pisa-Williamson D., Fuller C.W.* The sequenase quick-denature plasmid sequencing kit // *Editorial Comments. Amersham Life Sci.* 1992. Vol. 21. P. 29–36.
- Sanger F., Coulson A.R.* The use of thin acrylamide gels for DNA sequencing // *FEBS Lett.* 1978. Vol. 87. P. 107–110.
- Sanger F., Nicklen S., Coulson A.R.* DNA sequencing with chain-terminating inhibitors // *Proc. Nat. Acad. Sci. USA.* 1977. Vol. 74. P. 5463–5467.
- Sheen J.-Y., Seed B.* Electrolyte gradient gels for DNA sequencing // *Biotechniques.* 1988. Vol. 6. P. 942–944.
- Slater G.W., Noolandi J.* Electric field gradients and band sharpening in DNA gel electrophoresis // *Electrophoresis.* 1988. Vol. 9. P. 643–646.
- States J.C., Patel L.R., Li Q.A.* A gel electrophoresis system for resolving over 500 nucleotides with a single sample loading // *Biotechniques.* 1991. Vol. 11. P. 46–48.
- Stein A., Hill S.A., Cheng Z., Bina M.* Simple miniaturized system for DNA sequence analysis // *Nucl. Acids Res.* 1998. Vol. 26. P. 452–455.

## *Дополнительная литература*

Методы молекулярной генетики и геной инженерии / А.В. Мазин, А.В. Кузнецов, А.С. Краев и др. Новосибирск: Наука, 1990. 248 с.

*Остерман Л.А.* Методы исследования белков и нуклеиновых кислот: Электрофорез и ультрацентрифугирование. М.: Наука, 1981. 288 с.

*Теннов А.В.* Гель-электрофорез нуклеиновых кислот // Итоги науки и техники. Молекуляр. биология. М., 1976. Т. 4: Физико-химические методы в молекулярной биологии. С. 130–175.

*Brown T.A.* DNA sequencing. Oxford: IRL press, 1994. 97 p.

*Griffin H.G., Griffin A.M.* (ed.). Methods in molecular biology. Vol. 23. DNA sequencing protocols. Totowa: Humana press, 1993. 392 p.

*Rickwood D., Hames B.D.* (ed.). Gel electrophoresis of nucleic acids: a practical approach. Oxford: IRL press, 1982. 242 p.

*Sambrook J., Fritsch E.F., Maniatis T.* Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor: CSHL press, 1989. Vol. 1.

## Глава 5

# РАДИОАВТОГРАФИЯ

Этап радиоавтографии является одной из заключительных стадий всего процесса секвенирования ДНК и последним экспериментальным этапом, повторить который не всегда возможно простым повторным экспонированием этого же геля в отличие от следующих этапов в виде “чтения” нуклеотидной последовательности с радиоавтографа и компьютерного анализа, проводить которые можно неограниченное число раз. В связи с этим, несмотря на кажущуюся простоту данного этапа, его проведение требует должного внимания и определенной подготовки гелей, уже описанной выше.

### 5.1. РАДИОАВТОГРАФИЯ СЕКВИРИУЮЩЕГО ГЕЛЯ

Возникновение фотографического изображения в результате экспонирования секвенирующего геля, содержащего радиоактивно меченные фрагменты ДНК, на рентгеновскую пленку заключается в воздействии  $\beta$ -частиц на кристаллы галогенидов серебра и их последующего превращения в металлическое серебро. Различают прямую и непрямую радиоавтографию. При прямой – фотографическое изображение возникает только благодаря  $\beta$ -частицам, испускаемым радионуклидом. Прямая радиоавтография секвенирующих гелей может проводиться при комнатной температуре с любым из применяемых для мечения ДНК радионуклидов  $^{32}\text{P}$ ,  $^{33}\text{P}$  или  $^{35}\text{S}$  (характеристики данных радионуклидов приведены в Приложении). Однако низкой энергии  $\beta$ -частиц двух последних хватает на превращение галогенидов серебра, содержащихся только в первом эмульсионном слое двусторонних пленок, непосредственно контактирующем с гелем, а второй слой для них остается недостижимым, что и показано на сильно упрощенном схематичном рис. 5.1.

При этом чувствительность процесса по сравнению с использованием  $^{32}\text{P}$  уменьшается, но одновременно увеличивается разрешающая способность метода, так как ввиду более узких зон почернения на радиоавтографе “чтение” полос ДНК может продолжаться в верхней части дальше обычного.  $\beta$ -частицы  $^{32}\text{P}$  обладают достаточно большой энергией и способны вызвать превращение галогенидов серебра в металлическое серебро в обоих эмульсионных слоях двусторонней рентгеновской пленки, но, как видно из рис. 5.1, при этом будет иметь место увеличение ширины полос ДНК и, как следствие, некоторое ухудшение разрешающей способности метода.

При непрямой радиоавтографии, кроме тех же  $\beta$ -частиц, дополнительное воздействие оказывают кванты света, излучаемые возбужденным ( $\beta$ -частицами) сцинтиллятором. Так, использование для усиления изображения специальных усиливающих экранов, содержащих сцинтилляторы и превращающих энергию  $\beta$ -частиц в кванты видимого света, позволило повысить чувствительность процедуры радиоавтографии при-

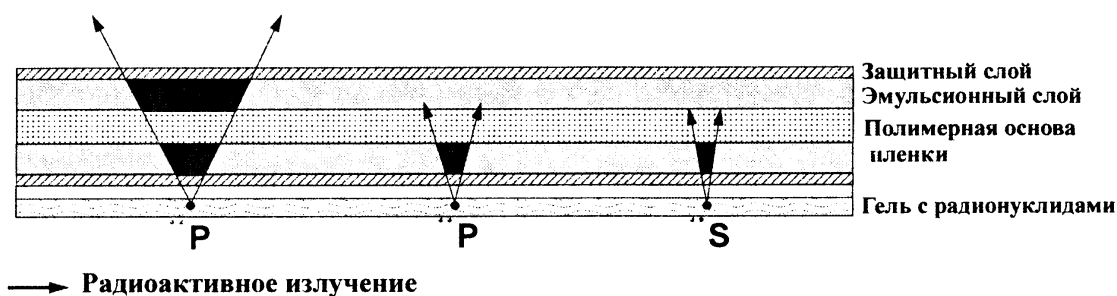


Рис. 5.1. Прямая радиоавтография секвенирующего полиакриламидного геля

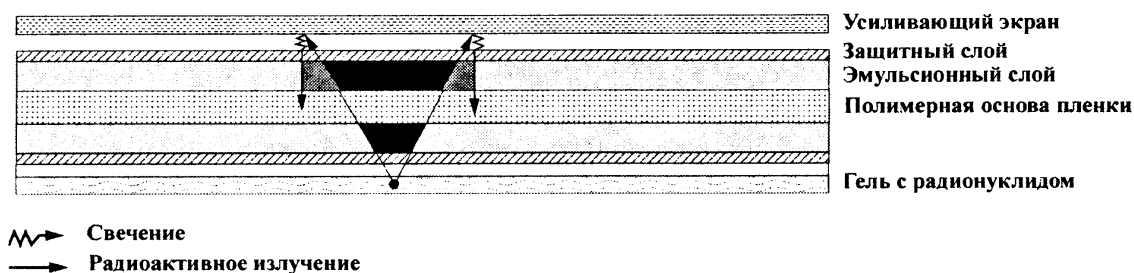


Рис. 5.2. Радиоавтография секвенирующего полиакриламидного геля с использованием усиливающего экрана

мерно на порядок за счет дополнительного засвечивания второго эмульсионного слоя, обращенного к экрану [Laskey, Mills, 1977]. Наиболее типичным экраном является экран, содержащий вольфрамат кальция, обладающий сцинтилляционными свойствами, причем испускающим свет в коротковолновой части спектра, чьи фотоны обладают большей энергией и оказывают более эффективное фотохимическое действие на микрокристаллы галогенидов серебра. Следует отметить, что использование усиливающих экранов возможно только с высокоэнергетическим  $^{32}\text{P}$ . Применение усиливающих экранов приводит к ухудшению четкости полос ДНК за счет еще большего расширения зоны образования металлического серебра под действием квантов света, что и показано на рис. 5.2. Таким образом, к сожалению, усиливающие экраны затрудняют “чтение” полос ДНК в верхней части геля и вследствие этого сокращают количество достоверно “читаемых” нуклеотидов.

Было предложено интересное применение усиливающего экрана для одновременной детекции в одном геле разных образцов ДНК, меченных сильным ( $^{32}\text{P}$ ) и слабым ( $^{33}\text{P}$ ) радионуклидами [Краев и др., 1992]. Сущность такой одновременной детекции заключается в экспозиции секвенирующего геля сразу на две рентгеновские пленки, несущих одностороннюю эмульсию. Одна из этих пленок своим эмульсионным слоем направлена к усиливающему экрану, и образование металлического серебра в ней вызывается только высокоэнергетическими  $\beta$ -частицами  $^{32}\text{P}$ , тогда как первая пленка, направленная эмульсионным слоем к гелю, регистрирует оба образца ДНК и для “чтения” нуклеотидной последовательности фрагмента ДНК, меченного  $^{33}\text{P}$ , остается только вычистить лишние полосы, совпадающие с таковыми на второй пленке.

Более редкие (и сейчас практически неприменяемые) случаи импрегнирования различных сцинтилляторов в секвенирующий гель после завершения электрофореза или в капроновой фильтр, на который перенесены меченые фрагменты ДНК из секвенирующего геля, позволяют использовать для этой цели, кроме  $^{32}\text{P}$ , и более слабые  $^{33}\text{P}$  или же  $^{35}\text{S}$ , но при этом также имеет место уменьшение четкости фотографического изображения.

Так, непрямая радиоавтография обеспечивает более высокую чувствительность при более низком разрешении и в случае секвенирования ДНК менее пригодна. Другим отрицательным моментом является необходимость проведения не прямой радиоавтографии при температуре  $-70^\circ\text{C}$ . Причиной этого является то, что пролет одной  $\beta$ -частицы приводит к превращению десятков или даже сотен ионов серебра в атомы серебра, тогда как требуется несколько квантов света для подобного превращения только единственного иона серебра. Однако, как известно, единичные атомы серебра нестабильны и могут обратно перейти в ионную форму. Понижение температуры экспозиции до  $-70^\circ\text{C}$  замедляет этот процесс и повышает вероятность образования стабильной пары атомов серебра. Радиоавтография при  $-70^\circ\text{C}$  преэкспонированной с помощью короткой вспышки рентгеновской пленки дополнительно увеличивает чувствительность процесса [Laskey 1980; Laskey, Mills, 1977], но требует весьма трудоемкого этапа подготовки подобной пленки.

Во избежание недоразумения считаем необходимым заметить, что  $\beta$ -частицы, испускаемые радиоизотопом, не летят только к рентгеновской пленке, как показано на рис. 5.1 и 5.2, а, в действительности, вылетают во всех направлениях. Ввиду разной энергии отдельных  $\beta$ -частиц и эффекта торможения их самим гелем и содержащейся в нем водой (в случае радиоавтографии влажных гелей) далеко не все частицы, вылетевшие под более острым углом к рентгеновской пленке, требующим преодоления большего расстояния, достигают эмульсионного слоя. Более того, полет  $\beta$ -частиц даже в воздухе непрямолинеен и траектория их полета представляет собой изломанную линию из-за столкновений с орбитальными электронами встречных атомов. Так, отклоняясь многократно от первоначального направления,  $\beta$ -частица может полететь даже в обратном направлении. А среди долетевших до рентгеновской пленки не все остаются способными произвести в нем необходимое число превращений микрокристаллов галогенида серебра для формирования фотоизображения при последующей проявке. Таким образом, разная энергия отдельных  $\beta$ -частиц позволяет им достичь рентгеновскую пленку на различном расстоянии от места вылета, что и определяет ширину зоны образования металлического серебра, как схематично показано на рис. 5.1 для  $^{32}\text{P}$ ,  $^{33}\text{P}$  и  $^{35}\text{S}$ .

В первоначальных экспериментах гели после завершения электрофореза без каких-либо дополнительных процедур оборачивали тонкой пленкой Sagan Wrap для предотвращения подсыхания и для исключения контакта влажного геля с желатиновым слоем рентгеновской пленки придавливали тяжелым (свинцовым) грузом и помещали в морозильную камеру с температурой около  $-20^\circ\text{C}$  [Maxam, Gilbert, 1977]. Низкая



температура требовалась для недопущения диффузии полос ДНК, которая могла ухудшить и без того не очень качественный радиоавтограф геля. Требования по повышению общей производительности метода секвенирования ДНК не могли не затронуть этап радиоавтографии и его различные составляющие.

Так, один из путей увеличения числа “читаемых” полос ДНК состоял в соответствующей подготовке полиакриламидного геля после завершения электрофореза к радиоавтографии. Исключение диффузии фрагментов ДНК, достигаемое их фиксацией с помощью уксусной кислоты, и одновременное удаление мочевины позволили несколько повысить чувствительность и разрешающую способность метода. Более значительное улучшение было достигнуто путем удаления воды и экспонированием предварительно высушенного геля.

Другой подход заключался в применении радиоактивных изотопов с меньшей энергией их  $\beta$ -частиц, таких как  $^{35}\text{S}$  и затем  $^{33}\text{P}$  [Biggin et al., 1983; Краев, Миронов, 1990; Zagursky et al., 1991; Краев и др., 1992]. Применение  $^{32}\text{P}$ , имеющего  $\beta$ -частицы с максимальной энергией 1,709 МэВ, приводило к образованию на радиоавтографе широких и диффузных полос ДНК.  $\beta$ -частицы  $^{35}\text{S}$ , максимальная энергия которых приблизительно на порядок меньше, чем у  $^{32}\text{P}$ , вызывали появление острых и четко очерченных полос, однако необходимость более длительной экспозиции являлась некоторым недостатком. Еще более удобным оказалось использование изотопа  $^{33}\text{P}$  [Краев, Миронов, 1990; Zagursky et al., 1991; Краев и др., 1992; Barnard et al., 1994]. Максимальная энергия  $\beta$ -частиц  $^{32}\text{P}$  несколько выше, чем у серы, и вполне достаточна для относительно короткого времени экспозиции, сравнимого с таковым для  $^{32}\text{P}$ , и при этом дает такие же резко очерченные полосы ДНК, как и в случае с  $^{35}\text{S}$ , что позволяет уверенно проводить “чтение” нуклеотидов в верхней части радиоавтографа.

Еще одна возможность повысить эффективность процедуры радиоавтографии была реализована путем разработки новых типов рентгеновских пленок.

## 5.2. РЕНТГЕНОВСКИЕ ПЛЕНКИ

В течение целого ряда лет для радиоавтографии секвенирующих гелей использовались обычные рентгеновские пленки медицинского назначения. Несмотря на большой выбор различных типов этих пленок они все же были разработаны не для детекции  $\beta$ -частиц и поэтому существовала необходимость в создании специализированной пленки для секвенирования ДНК. Более того, с применением низкоэнергетических радионуклидов такая потребность стала еще явственней. Так, многими ведущими фирмами были разработаны специальные односторонние пленки с повышенным содержанием серебра. На протяжении многих лет фирмой “Amersham” поставлялась имеющая высокую чувствительность рентгеновская пленка Hyperfilm  $\beta$ -max. Кроме обычной рентгеновской пленки, существует специальная рентгеновская бумага

Нуреграрег, предназначенная для получения отпечатков, которые можно легко копировать с помощью обычной копировальной техники и даже, как утверждает поставщик (Amersham), передавать по факсу. Кроме этого, подобные оригиналы радиоавтографов удобны при монтаже постерных сообщений на конференциях. К числу последних разработок можно отнести пленку BioMax MR, изготавливаемую по новой "T-grain" технологии фирмой "Kodak" (США). Утверждается, что использование этой пленки повышает разрешающую способность секвенирования на 5 – 9% для  $^{32}\text{P}$  и на 11 – 20% для  $^{33}\text{P}$  и  $^{35}\text{S}$  [Steinfeld et al., 1997]. В нашей стране рентгеновская пленка с односторонней эмульсией, предназначенная для регистрации  $\beta$ -частиц при секвенировании ДНК, также была разработана [Краев и др., 1992] и под названием РА-1 поставляется ГосНИИ "Химфотопроект" (Казань).

### 5.3. ПРОЧИЕ СПОСОБЫ РЕГИСТРАЦИИ РАДИОАКТИВНОСТИ В СЕКВЕНИРУЮЩИХ ГЕЛЯХ

Дорогой альтернативой дешевому методу радиоавтографии является подход с детекцией радиоактивных изотопов с помощью специальных приборов, выпускаемых рядом фирм. Так, приборы Cyclone ("Packard", США) и PhosphorImager ("Molecular Dynamics", США) способны детектировать энергию обычно используемых для секвенирования ДНК радионуклидов ( $^{32}\text{P}$ ,  $^{33}\text{P}$  и  $^{35}\text{S}$ , а также и более "слабых") и работают на принципе формирования на специальном экране под действием ионизирующей радиации латентного фосфоресцирующего изображения, после сканирования которого лазерным лучом регистрируется высвобождающаяся энергия в виде голубого свечения, которое улавливается, усиливается фотоэлектронным умножителем и с помощью специальных компьютерных программ оцифровывается. Таким образом, требуется только около 10 мин для получения высококачественного изображения секвенирующего геля. Причем вся работа по закладке геля в специальную кассету происходит на свету и, более того, яркий свет позволяет "стирать" оставшееся изображение предыдущего геля. Экспозиция обычно составляет одну десятую от таковой при обычной радиоавтографии на рентгеновскую пленку. Ранее было показано, что подобные экраны способны с квадратного миллиметра геля улавливать 1 распад радионуклида  $^{32}\text{P}$  и 25 –  $^{35}\text{S}$ , приводя, таким образом, к 100 – 250 кратному увеличению чувствительности по сравнению с обычной рентгеновской пленкой [Johnston et al., 1990].

Фирма "Packard" производит прибор InstantImager, способный осуществлять мгновенную радиоавтографию. Его микроканальный детектор массива данных позволяет детектировать различные радионуклиды, включая  $^{32}\text{P}$ ,  $^{33}\text{P}$  и  $^{35}\text{S}$ , обычно используемые для секвенирующего электрофореза, причем время получения окончательных результатов составляет одну десятую от описанного выше способа с использованием фосфоресцирующих экранов и одну сотую от обычной радиоавтографии.

## Цитируемая литература

*Краев А.С., Миронов В.М.* Экономичное секвенирование ДНК с терминаторами // Молекуляр. биология. 1990. Т. 24. С. 1095–1099.

*Краев А.С., Ли Н.Н., Сидоров Ю.Д., Твердохлебов Е.Н.* Новая пленка для регистрации секвенирующих гелей // Там же. 1992. Т. 26. С. 876–879.

*Barnard G.F., Puder M., Begum N.A., Chen L.B.* PCR products sequencing with [ $\alpha$ - $^{33}\text{P}$ ] and [ $\alpha$ - $^{32}\text{P}$ ] dATP // Biotechniques. 1994. Vol. 16. P. 572–573.

*Johnston R.F., Pickett S.C., Barker D.L.* Autoradiography using storage phosphor technology // Electrophoresis. 1990. Vol. 11. P. 355–360.

*Laskey R.A.* The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis // Meth. Enzymol. 1980. Vol. 65. P. 363–371.

*Laskey R.A., Mills A.D.* Enhanced autoradiographic detection of  $^{32}\text{P}$  and  $^{125}\text{I}$  using intensifying screens and hypersensitized film // FEBS Lett. 1977. Vol. 82. P. 314–316.

*Maxam A.M., Gilbert W.* A new method for sequencing DNA // Proc. Nat. Acad. Sci. USA. 1977. Vol. 74. P. 560–564.

*Steinfeld R., McLaughlin W., Vizard D., Bundy D., Attwood J.* New film technology for improved sequencing results // Editorial Comments. Amersham Life Sci. 1997. Vol. 23. P. 18–19.

*Zagursky R.J., Conway P.S., Kashdan M.A.* Use of  $^{33}\text{P}$  for Sanger DNA sequencing // Biotechniques. 1991. Vol. 11. P. 36–38.

## Глава 6

# ЧТЕНИЕ НУКЛЕОТИДНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ С РАДИОАВТОГРАФА ГЕЛЯ

Чтобы стал возможен анализ какой-либо нуклеотидной последовательности ДНК, определенной в конкретном эксперименте, необходимо “прочитать” эту последовательность, видимую на радиоавтографе секвенирующего геля как некую “лестницу” полос. Существуют различные способы (ручной, полуавтоматический, автоматический) “чтения” последовательности полос ДНК и определенные правила, рассмотрению которых и будет посвящена данная глава. Независимо от того, как проводится “чтение” последовательности нуклеотидов с радиоавтографа геля или с самого геля как ручным, так и автоматическим способом, начинается оно всегда с фрагментов ДНК меньшего размера, находящихся в нижней части радиоавтографа, последовательно поднимаясь вверх и при этом сравнивая дорожки, принадлежащие одной матрице ДНК. Одной из причин этого служит более легкая идентификация конкретных нуклеотидов за счет их лучшего вертикального отделения друг от друга, вызванного логарифмической зависимостью движения фрагментов ДНК при гель-электрофорезе. По мере увеличения размера фрагментов ДНК расстояние между полосами, их представляющими, пропорционально уменьшается и в конце концов последовательность расположения полос в соседних треках, а потом в одном и том же треке делается неразличимой, приводя к завершению “чтения”. Еще одной причиной является то, что в нижней части геля располагается как бы условное начало секвенируемого фрагмента ДНК, поскольку полосы “уменьшающегося” размера “стремятся” к своему гомогенному концу в виде конкретного первого нуклеотида. Другая причина начала “чтения” с нижней части радиоавтографа геля кроется в том, что нуклеотидная последовательность одной цепи ДНК всегда приводится в направлении  $5' \rightarrow 3'$ , совпадающем с таковым у ферментативного действия ДНК-полимераз (точнее, их полимеразной активности) и “чтение” “снизу  $\rightarrow$  вверх” как бы осуществляется в обычной принятой ориентации (впрочем, реальная ориентация какого-либо секвенируемого фрагмента ДНК в геноме может быть и противоположной, что может быть вызвано или особенностями клонирования данного фрагмента, или выбором соответствующего секвенирующего праймера). Однако следует отметить, что при секвенировании ДНК по методу Сэнгера порядок “чтения” будет всегда соответствовать направлению  $5' \rightarrow 3'$ , обусловленному ферментативным действием ДНК-полимераз и “первый” нуклеотид секвенируемого фрагмента будет всегда соответствовать  $5'$ -концу цепи ДНК. В случае же секвенирования ДНК по Максаму–Гилберту в качестве меченого “первого” нуклеотида секвенируемого фрагмента может выступить нуклеотид, находящийся как на его  $5'$ -конце,

так и на 3'-конце. Соответственно этому будет и порядок “чтения” последовательности ДНК, что необходимо учитывать при дальнейших анализах. Надо отметить, что “чтение” полос ДНК, полученных методами Максама–Гилберта и Сэнгера имеет и некоторые другие особенности, которые будут рассмотрены в процессе дальнейшего изложения.

## 6.1. РУЧНОЕ ЧТЕНИЕ

Характер расщепления фрагментов ДНК в методе секвенирования путем химической дегградации таков, что для “чтения” последовательности с радиоавтографа геля наиболее оптимальным вариантом является такой, когда при нанесении на гель одна дорожка (обычно первая) соответствует гуанину, вторая – пуринам, третья – пиримидинам и четвертая – цитозину (рис. 6.1). Таким образом, две средние дорожки содержат большее (кроме потенциально возможных частных случаев) число полос, чем крайние дорожки, поскольку в них видны фрагменты ДНК, ограниченные нуклеотидами G, A и C, T соответственно. Таким образом, эти две средние дорожки (G + A и C + T) вместе содержат все полосы частичного расщепления меченой по одному из концов молекулы ДНК. Правильное “прочтение” полос, видимых на радиоавтографе геля, заключается, во-первых, в определении вертикальной позиции каждой конкретной полосы, соответствующей тому или иному нуклеотиду, относительно полос в соседних дорожках, относящихся к одному и тому же образцу ДНК, и, во-вторых, в выявлении типа нуклеотидов, к которому принадлежит данная полоса. Так, например, если полоса имеется только во второй дорожке – то это аденин, поскольку для другого пуринового основания – гуанина будет характерно наличие двух полос: в первой G-дорожке и совпадающей с ней полосой во второй G + A-дорожке. Аналогично проводится и “чтение” пиримидиновых оснований, где полоса только в третьей дорожке указывает, что это тимин, а если эта же полоса имеется и в четвертой дорожке, то тогда это цитозин. Что касается более редко используемых реакций, приводящих к выявлению полос ДНК по типу A > C, A > G, G > A, C > T, то в таких случаях требуется более сложное сопоставление наличия или отсутствия полос в соседних треках. Анализируя таким образом полосы одну за другой и поднимаясь вверх по радиоавтографу, необходимо достигнуть места, где полосы в соседних дорожках расположены так близко друг к другу, что уже невозможно различить их вертикальную последовательность. После этого “чтение” нуклеотидной последовательности фрагмента ДНК с данного радиоавтографа можно считать законченным. Если же этот образец был нанесен на гель дважды, то для ранее нанесенного препарата на радиоавтографе будут видны фрагменты ДНК большего размера. В таком случае надо найти перекрывающиеся участки (что гораздо легче сделать, если в секвенируемой последовательности ДНК содержатся какие-нибудь характерные участки как, например, AAAAGGAAAA или им подобные, хорошо заметные глазом) и продолжить “чтение” данного образца ДНК, пока позволяет достигну-

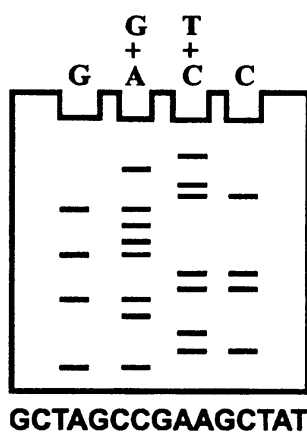


Рис. 6.1. Схема “чтения” нуклеотидной последовательности ДНК, определенной методом Максама–Гилберта

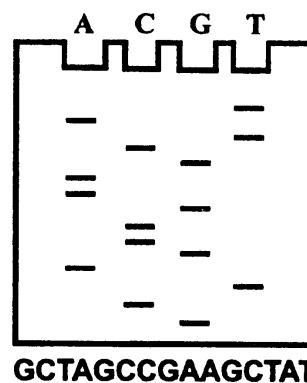
тое в этом эксперименте разрешение. При этом при поиске перекрытий необходимо уделить особое внимание совпадению нуклеотидов в случаях секвенирования фрагментов ДНК, содержащих повторяющиеся элементы.

Особенности “чтения” радиоавтографа геля при секвенировании ДНК методом химической деградации по Максому–Гилберту, включая разбор аномалий и вызвавших их причин, весьма подробно описаны в детальной статье Максама и Гилберта [Maxam, Gilbert, 1980] и в ее русском переводе [Максам, Гилберт, 1986]. Поэтому заинтересованный читатель сможет ознакомиться с ними самостоятельно. Здесь же лишь скажем, что наиболее часто некоторые неоднозначности прочтения тех или иных нуклеотидов связаны с не совсем верным проведением реакций ограниченной модификации и пиперидинового гидролиза и, в частности, с присутствием остаточных количеств солей. Так, даже незначительное количество соли может служить причиной супрессии модификации тиминнов гидразином. Возникают проблемы и при модификации других азотистых оснований. Пиперидин же в присутствии соли вызывает гидролиз не только модифицированных, но и неизмененных оснований, приводя к высокому фону, видимому на радиоавтографе как сплошная лестница полос в каждой дорожке. Однако тщательное удаление следов соли, достигаемое, в том числе, простой промывкой осадка ДНК 70%-ным и затем 96%-ным этанолом, позволяет полностью исключить эту проблему. К серьезным затруднениям или даже невозможности “прочтения” последовательности ДНК может привести загрязнение секвенируемого фрагмента ДНК другим, также оказавшимся меченым, фрагментом.

В ферментативном методе секвенирования ДНК по Сэнгеру каждая дорожка соответствует только одному конкретному типу оснований (рис. 6.2) и поэтому для правильного “прочтения” радиоавтографа секвенирующего геля необходимо следить, чтобы какая-либо полоса содержалась только в одной дорожке, поскольку в противном случае точное установление данного нуклеотида будет затруднено или скорее просто невозможно. В связи с тем, что в отличие от картины полос секвенируемой ДНК по Максому Гилберту в методе Сэнгера нет необходимости сопоставления наличия полос пуриновых или пиримидиновых оснований в одной или двух дорожках, то можно считать, что “чтение” радиоавтографа секвенирующего геля при этом будет несколько проще.

Однако в ранних экспериментах по секвенированию ДНК использование для построения комплементарной цепи Кленовского фрагмента ДНК-полимеразы I приводило к образованию на радиоавтографе секвенирующего геля полос разной интенсивности, что было следствием некоторой избирательности самого фермента при построении цепи

Рис. 6.2. Схема “чтения” нуклеотидной последовательности ДНК, определенной методом Сэнгера



ДНК и включения в нее специфических терминаторов – дидезоксинуклеотидтрифосфатов. Это сильно затрудняло “чтение” полос ДНК и иногда не позволяло провести однозначное “прочтение” некоторых нуклеотидов. В связи с этим были разработаны даже специальные правила “чтения” таких сложных участков [Bankier, Barrell, 1983]. Наибольшие трудности подстерегали экспериментатора на С-дорожке. Одиночные С-полосы часто бывают весьма слабыми, так что при слабой экспозиции геля есть вероятность пропустить такое основание. Такие же слабые С-полосы отмечаются для первого С из целого блока цитозиновых остатков, следующих друг за другом. Особую остроту вызывает тот факт, что второй С в таком участке обычно имеет наибольшую интенсивность, а поскольку вертикальное расстояние между полосами для блока цитозинов может быть уменьшено, это приводит к некоторой неясности их очертаний и, таким образом, весьма высока вероятность ошибочного “прочтения” таких участков. В то же время, чтобы не задерживать дальнейшее “чтение” радиоавтографа, для временного занесения в компьютер таких спорных нуклеотидов были придуманы специальные символы. Для примера ограничимся лишь некоторыми такими символами. Так, цифры 1, 2 3 и 4 обозначали, что это ВЕРОЯТНО С, Т, А и G соответственно. Буквой D обозначали нуклеотид(ы) С или СС в случаях, когда не могли определить это точно. Буквы V, B и H служили для тех же целей для Т, А и G нуклеотидов.

Приведенные здесь далеко не все сложности и особые случаи “прочтения” нуклеотидной последовательности ДНК с радиоавтографа секвенирующего геля на сегодняшний день полностью потеряли свою актуальность и упомянуты здесь лишь с целью рассказать о трудностях, с которыми сталкивались исследователи при секвенировании ДНК в первые годы. Применяемые сегодня ДНК-полимеразы позволяют получать достаточно однородную картину полос ДНК на радиоавтографе и описанные выше проблемы практически не возникают.

Такой нежелательный эффект, как компрессия полос при электрофоретическом разделении одноцепочечных фрагментов ДНК, зависит от последовательности нуклеотидов на 3'-конце фрагмента ДНК и не зависит от используемой ДНК-полимеразы и поэтому существует и сейчас. Несмотря на различные способы, направленные на уменьшение компрессии (рассмотренные выше в разделах 2.6. и 4.2), она все же доставляет экспериментаторам массу хлопот. Возникающая за счет формирования на 3'-конце фрагмента ДНК участка со вторичной структурой шпилька приводит к тому, что фрагменты ДНК, ее несущие, мигрируют в геле быстрее, как если бы они были короче. В результате заметно уменьшается вертикальное расстояние между последующими полосами, приводящее даже к опережению одними нуклеотидами других

и, следовательно, к неверному “прочтению” этого участка. Такие нуклеотиды, впрочем, могут быть достоверно определены при секвенировании противоположной цепи ДНК. Несмотря на то, что и для этой комплементарной цепи будет характерно формирование подобной шпильки, в этом процессе будут задействованы нуклеотиды, расположенные с другой стороны от центра шпильки и поэтому компрессия коснется именно их, а не тех, что были компрессированы при секвенировании исходной цепи ДНК.

Завершая этот раздел, хочется привести ряд неких правил, с которых начинается “чтение” нуклеотидной последовательности с радиоавтографа геля. Причем данные правила (кроме особо оговоренных случаев) действуют при секвенировании ДНК как ферментативным методом по Сэнгеру, так и методом химической дегградации по Максаму–Гилберту. Перед началом “чтения” необходимо провести беглый осмотр радиоавтографа секвенирующего геля, направленный на выявление недоэкспозиции или переэкспозиции; возможного высокого фона; размытости полос; присутствия дополнительных полос в соседних дорожках (для метода Сэнгера). Все эти факторы затрудняют чтение или делают его совсем невозможным и поэтому необходимо принять решение, что делать дальше. В результате недостаточной экспозиции некоторые слабые полосы могут быть не видны и в таком случае, если возможно, необходимо повторить этап радиоавтографии, увеличив соответственно время экспозиции (если, конечно, это допустимо и гель с радионуклидом  $^{32}\text{P}$  уже не экспонировался две недели). Переэкспозиция также может быть нежелательна, поскольку затрудняет “чтение” полос в верхней части геля (особенно меченных  $^{32}\text{P}$ ), но она относительно легко поправима. Прочие недостатки радиоавтографа (кроме размытости полос, которая может быть также следствием и плохого контакта геля с рентгеновской пленкой) этапом радиоавтографии уже не устранимы. И в случае невозможности достоверного “прочтения” нуклеотидной последовательности необходимо осуществить или повторное электрофоретическое разделение тех же продуктов секвенирующих реакций или провести эти реакции заново, что зависит от характера видимых на радиоавтографе артефактов и при этом изменить какие-то условия их проведения.

Приняв решение “читать” радиоавтограф секвенирующего геля, можно предпринять следующие действия, хотя они далеко не обязательны для опытного и внимательного экспериментатора. Как показывает опыт, значительное число ошибок при “чтении” радиоавтографа образуется за счет перескакивания взгляда на соседние дорожки, относящиеся к другой матрице, особенно в случаях не оставленной пустой дорожки между разными образцами ДНК. Избежать этого можно, проведя фломастером разграничительные линии или даже изготовив специальную маску, оставляющую открытой только четверку “читаемых” дорожек. Желательно, начав “чтение” радиоавтографа, завершить его до конца, но при каких-то вынужденных остановках необходимо сделать отметку на радиоавтографе для возобновления дальнейшего “чтения”. Это особенно важно в случаях секвенирования ДНК, содержащей повторяющиеся последовательности, о чем уже упоминалось немного



выше при описании выявления участков перекрытия. Другой тип ошибок, который может произойти при “чтении”, заключается в используемых различных порядках нанесения образцов ДНК на секвенирующий гель. Так, при постоянном нанесении, например, в порядке GATC другой характер нанесения при недостаточной опытности и внимательности экспериментатора неминуемо приведет к некоторым ошибкам, поэтому можно порекомендовать всегда придерживаться избранного порядка, кроме случаев, когда он должен быть каким-то определенным. Иногда “чтение” полос ДНК необходимо начинать с самого низа радиоавтографа, с того места, где возможно их однозначное “прочтение”. Так бывает, например, при секвенировании ДНК по методу Максама–Гилберта после элюции какого-либо фрагмента из геля или по методу Сэнгера в случае эффективного разделения крупных фрагментов ДНК, когда последовательности, принадлежащие векторной молекуле, давно вышли из геля. Однако весьма часто на радиоавтографе геля бывает видна и последовательность ДНК-вектора. В этих случаях необходимо обнаружить границы клонированной вставки с тем, чтобы исключить уже на этапе “чтения” векторную последовательность. В одной работе даже была предложена сделанная компьютером иллюстрация радиоавтографа секвенирующего геля с изображенным на нем полилинкером из вектора рUC19, как бы “прочитанная” в обеих ориентациях с помощью прямого и обратного праймеров [Cantack, 1990]. Автор предлагал использовать ее в качестве некоего шаблона для облегчения поиска участков полилинкера, но при этом необходимо заметить, что изображенные полосы ДНК соответствуют матрице как бы “нанесенной” на гель в последовательности GACT, практически никогда не применяемой (см. раздел 4.2).

Кроме самих полос ДНК, особое внимание должно уделяться и вертикальным промежуткам между ними. Так, группа сильно приближенных к друг другу полос может свидетельствовать об имеющейся в этом месте компрессии ДНК. Дополнительным аргументом того, что здесь имеется компрессия, может быть увеличенное вертикальное расстояние между полосами, наблюдаемое несколько выше. Подобное происходит вследствие того, что формирующаяся на 3'-конце новосинтезированной цепи ДНК шпилька увеличивает подвижность таких фрагментов ДНК, а уже для более крупных фрагментов, несущих дополнительные нуклеотиды после шпильки, эта структура становится менее стабильной, вторичная структура исчезает и, как следствие, межполосное расстояние возвращается к прежнему, приводя таким образом к созданию небольшого участка с увеличенным вертикальным расстоянием между соседними полосами. Выявление таких участков с компрессией полос ДНК крайне важно для правильного “прочтения” этого отрезка нуклеотидной последовательности и в случае невозможности этого – обращение повышенного внимания к этому участку при секвенировании комплементарной цепи ДНК. Увеличенное (двойное) расстояние между полосами может быть также результатом пирофосфорилиза и в этом случае установление типа отсутствующего нуклеотида возможно после секвенирования комплементарной цепи ДНК.

Наиболее типичный путь переноса нуклеотидной последовательности с радиоавтографа геля в компьютер рассматриваемым здесь ручным методом лежит через предварительную запись этой последовательности на бумагу. Последующее введение данной последовательности нуклеотидов осуществляется с помощью клавиатуры. Таким образом, подобное введение “бессмысленных” текстов в виде беспорядочного набора букв G, A, T, C и редко N в компьютер таит в себе еще источник ошибок. Чтобы исключить такие ошибки, в разных компьютерных программах предусмотрено или двойное введение нуклеотидной последовательности, с высокой вероятностью позволяющее выявить несоответствие отдельных нуклеотидов при их повторном занесении, или использовании музыкальных звуков, сигнализирующих о вводе того или иного нуклеотида [Hutchinson, 1986]. “Прочитывание” компьютером с помощью синтезатора речи занесенной последовательности нуклеотидов позволяет осуществлять ее контроль следя по сделанной записи или радиоавтографу.

Однако при осуществлении крупных проектов по секвенированию этап “чтения” нуклеотидной последовательности с радиоавтографа геля становится достаточно “узким местом”, что вызвало разработку специальных приборов, осуществляющих полуавтоматическое и автоматическое считывание последовательности нуклеотидов и ее занесение в компьютер. Рассмотрению этих систем и будет посвящен следующий раздел данной главы.

## 6.2. ПОЛУАВТОМАТИЧЕСКОЕ И АВТОМАТИЧЕСКОЕ ЧТЕНИЕ

Полуавтоматическое чтение радиоавтографа геля основано на применении планшетного дигитайзера со специальной ручкой, сопровождаемое использованием специализированных компьютерных программ [Gingeras et al., 1982; Lautenberger, 1982; Komaromy, Govan, 1984; Staden, 1984]. Сущность данного подхода заключается в размещении радиоавтографической пленки на планшетном дигитайзере с подсветкой и касании его поверхности сигнальной ручкой, передающей в компьютер оцифрованные X и Y координаты каждой полосы. Предварительно касанием четырех углов каждой дорожки, соответствующей определенному нуклеотиду, задаются ее координаты и присваивается тип нуклеотида. В дальнейшем поочередным касанием сигнальной ручки полос ДНК на радиоавтографе, исходя из тех же правил “чтения” нуклеотидной последовательности, описанной в предыдущем разделе, осуществляется занесение в компьютер соответствующего нуклеотида и всей последовательности в целом. Для исключения возможных ошибок в эту систему был добавлен синтезатор речи, проговаривающий по желанию экспериментатора занесенные нуклеотиды [Gingeras et al., 1982].

Выпуск различных вариантов подобных дигитайзеров, несколько отличающихся по своим возможностям и конструкциям, был налажен многими фирмами. Однако, учитывая высокую стоимость планшетного дигитайзера, было предложено также полуавтоматическое “чтение” последовательности нуклеотидов с радиоавтографа геля, основанное на

использовании более дешевой специализированной компьютерной мыши. Так, на протяжении некоторого времени швейцарская фирма “Integra Bioscience AG” поставляла компактную систему SequenceBoo, представляющую собой отдельный блок с небольшим дисплеем и памятью и минимальным количеством клавиш, служившую дешевой альтернативой компьютеру. Она комплектовалась специализированной компьютерной мышью Readermouse, предназначенной для ручного чтения радиоавтографа геля, лежащего на столе с подсветкой, и занесения нуклеотидной последовательности в компьютер нажатием соответствующих клавиш (ACGT и N) без применения клавиатуры самого компьютера.

Таким образом, существующие полуавтоматические системы чтения избавили экспериментаторов лишь от этапа занесения “прочитанной” последовательности нуклеотидов в компьютер, оставив ему возможность осуществлять само “чтение” все равно вручную.

Дальнейшее усовершенствование “чтения” радиоавтографа секвенирующего геля действительно сделало этот процесс полностью автоматическим [Elder et al., 1986]. Примененный крупноформатный цифровой сканер с соответствующим программным обеспечением позволил проводить довольно скоростное (5 мин на 4 трека) “чтение” нуклеотидной последовательности с точностью около 99%. Изображение секвенирующего геля на радиоавтографе не может быть идеальным, если понимать под этим равномерность полос, пропорциональное вертикальное расстояние между ними, неискривление треков и т.д. При “чтении” радиоавтографа глазами, человеческий мозг в состоянии вносить коррективы, исходя из имеющегося опыта, но простое сканирование радиоавтографического изображения без учета соответствующих поправок приводило бы к значительным ошибкам. Поэтому была разработана специальная компьютерная программа, способная корректировать возникающие искажения электрофоретического разделения фрагментов ДНК, узнавать полосы ДНК по их характерной форме, оценивать интенсивность полос и их взаимное расположение. Проведенное авторами цитируемой работы интересное сравнение точности “чтения” одной и той же последовательности с радиоавтографа геля экспериментаторами разного уровня подготовки и данной автоматической системой показало, что предложенный метод автоматического “чтения” ничуть не уступает качеству “чтения” опытным экспертом. Сравнительный анализ ошибок, происходящих на этапе “чтения” радиоавтографов геля как с помощью дигитайзера, так и автоматическим прибором, был проведен другими авторами [Khurshid, Beck, 1993]. В данной работе учитывались различные типы ошибок в виде неверно “прочитанных” нуклеотидов, вставок, делеций, отсутствующих нуклеотидов, а также неоднозначно определяемых. Однако к серьезному недостатку данного анализа можно отнести то, что во всех этих случаях секвенирование ДНК выполнялось с помощью Кленовского фрагмента ДНК-полимеразы I, а этот фермент известен тем, что включение ддНМФ в растущую цепь ДНК производит неравномерно и в сильной зависимости от окружающей последовательности ДНК. Можно предположить, что подобный

анализ мог бы дать заметно отличающиеся результаты, проводись он с радиоавтографами секвенирующих гелей, полученных с помощью секвенназы.

Дальнейшее развитие подобной системы привело к созданию еще более совершенного прибора для быстрого (5 мин на 4 дорожки) автоматического “чтения” и занесения “прочитанной” последовательности в компьютер в сочетании с удобным компьютерным обеспечением [West, 1988]. Разработка подобной техники продолжалась и в дальнейшем. Так, был создан высокопроизводительный прибор для автоматического чтения радиоавтографов секвенирующих гелей, использующий принцип обратной свертки и “читающий” последовательность ДНК со скоростью 5 нуклеотидов/с с уровнем ошибок около 1% [Ives et al., 1994].

Упомянутые выше подходы к “чтению” радиоавтографа секвенирующего геля требуют применения специализированного оборудования. В то же время существует возможность использования обычных сканеров как планшетных, так даже и ручных для сканирования радиоавтографов геля. Фирма “Jandel Scientific Software” (Германия) предоставляет компьютерную программу SigmaGel, специально предназначенную для анализа сканированного с помощью обычных сканеров изображения полос ДНК на радиоавтографе секвенирующего геля с возможностью дальнейшей обработки полученных результатов.

Нельзя не отметить одну работу по реконструированию сканированного радиоавтографа секвенирующего геля [Noolandi et al., 1993]. Особенностью данной работы было то, что разделение образцов ДНК проходило в условиях пульс-электрофореза и авторы, задавшись целью “прочитать” последовательность ДНК как можно дальше, сканировали радиоавтограф с разрешением 600 пикселей на дюйм и с помощью компьютера увеличили изображение. Они сообщают об уверенном “чтении” полос ДНК в интервале 970–1320 нуклеотидов, где их плотность составляла около 10 на 1 мм. В зоне, отошедшей от колодцев только на 13,5 см, плотность полос достигала уже около 25 на 1 см и сплошное “чтение” было уже невозможно, но отдельные нуклеотиды в районе 2332 все же могли быть распознаны.

Использование специальных CCD-камер, требующих сильного охлаждения или жидким азотом, или с помощью элементов Пельтье, как, например, в приборе SpeedReader фирмы “IntelliGenetics, Inc.” (США), регистрирующих и передающих изображение на экран компьютерного монитора, освободило экспериментатора от рутинной работы по “чтению” и последующему занесению в компьютер последовательности нуклеотидов. Соответствующее компьютерное обеспечение было способно компенсировать искажения полос ДНК с целью получения достоверных результатов [Drury et al., 1993; Elder, 1990; Sanders et al., 1991; Wu, Mislán, 1993]. Высокая чувствительность CCD-камер в сочетании с хорошей линейной зависимостью и быстротой получения результатов делает их достаточно удобными для сканирования результатов секвенирования ДНК. Подобная CCD-камера была использована для регистрации результатов секвенирования ДНК с помощью хемилюминесценции не-

посредственно с мембранного фильтра без обычного этапа радиоавтографии [Karger et al., 1992]. 30 мин экспозиции в темноте оказалось достаточно для детекции сигналов свечения, испускаемого производным 1,2-диоксетана. Их последующая оцифровка с помощью соответствующих компьютерных программ позволила “прочитать” нуклеотидную последовательность секвенируемого участка ДНК. В своей следующей работе авторы провели детальный анализ особенностей сканирования с помощью усовершенствованной CCD-камеры мембранных фильтров с секвенированной ДНК, выявляемой хемилюминесценцией [Karger et al., 1993]. В том числе было рассмотрено применение специальных фильтров для уменьшения отрицательного влияния космических лучей на заключительное изображение полос секвенируемой ДНК.

Упомянутые в предыдущем разделе специальные приборы Cyclone, InstantImager (“Packard”, США) и PhosphorImager (“Molecular Dynamics”, США), детектирующие наличие радионуклидов ( $^{32}\text{P}$ ,  $^{33}\text{P}$  и  $^{35}\text{S}$ ) прямо в геле, с помощью компьютерных программ (например, DNAscan для прибора PhosphorImager) сразу заносят считываемые полосы ДНК в виде текстового файла с последовательностью нуклеотидов в компьютер. Причем данные программы способны корректировать возможные искажения в геле, такие как “улыбка”, искривление треков и др., что позволяет получать достоверные данные.

### Цитируемая литература

Максам А.М., Гилберт У. Метод определения последовательности ДНК (секвенирования) после введения метки в конец молекулы и расщепления ее по основаниям // Молекуляр. биология. 1986. Т. 20. С. 581–637.

Bankier A.T., Barrell B.G. Shotgun DNA sequencing // Techniques in nucleic acids biochemistry / Ed. R.A. Flavell. Amsterdam etc. 1983. P. 1–34.

Carmack C.E. Multiple cloning site template for use in sequencing of plasmid clones // Biotechniques. 1990. Vol. 9. P. 45–46.

Drury H.A., Clark K.W., Hermes R.E., Feser J.M., Thomas L.J., Jr., Donis-Keller H. A graphical user interface for quantitative imaging and analysis of electrophoretic gels and autoradiograms // Ibid. 1992. Vol. 12. P. 892–901.

Elder J.K. Maximum entropy image reconstruction of DNA sequencing autoradiographs // Electrophoresis. 1990. Vol. 11. P. 440–444.

Elder J.K., Green D.K., Southern E.M. Automatic reading of DNA sequencing gel autoradiographs using a large format digital scanner // Nucl. Acids Res. 1986. Vol. 14. P. 417–424.

Gingeras T.R., Rice P., Roberts R.J. A semi-automated method for the reading of nucleic acid sequencing gels // Ibid. 1982. Vol. 10. P. 103–114.

Hutchinson C.A., III. Sequence gel reading with a portable computer // Ibid. 1986. Vol. 14. P. 1917.

Ives J.T., Gesteland R.F., Stockham T.T., Jr. An automated film reader for DNA sequencing based on homomorphic deconvolution // IEEE. Trans. Biomed. Eng. 1994. Vol. 41. P. 509–519.

Karger A.E., Weiss R., Gesteland R.F. Digital chemiluminescence imaging of DNA sequencing blots using a charge-coupled device camera // Nucl. Acids Res. 1992. Vol. 20. P. 6657–6665.

Karger A.E., Weiss R., Gesteland R.F. Line scanning system for direct digital chemiluminescence imaging of DNA sequencing blots // Anal. Chem. 1993. Vol. 65. P. 1785–1793.

Khurshid F., Beck S. Error analysis in manual and automated DNA sequencing // Anal. Biochem. 1993. Vol. 208. P. 138–143.

Komaromy M., Govan H. An expensive semi-automated sequence reader for Apple II computer // Nucl. Acids Res. Vol. 12. P. 675–678.

*Koop B.F., Rowan L., Chen W.-Q., Deshpande P., Lee H., Hood L.* Sequenase length and error analysis of sequenase and automated *Taq* cycle sequencing methods // *Biotechniques*. 1993. Vol. 14. P. 442–447.

*Lautenberger J.A.* A program for reading DNA sequence gels using a small computer equipped with a graphics tablet // *Nucl. Acids Res.* 1982. Vol. 10. P. 27–30.

*Maxam A.M., Gilbert W.* Sequencing end-labeled DNA base-specific chemical cleavage // *Meth. Enzymol.* 1980. Vol. 65: Nucleic acids, pt I. P. 499–560.

*Noolandi J., Forsyth R., Shi A.-C.* Sequencing using pulsed field and image reconstruction // *Proc. Nat. Acad. Sci. USA.* 1993. Vol. 90. P. 10101–10104.

*Sanders J.Z., Petterson A.A., Hughes P.J., Connel C.R., Raff M., Menchen S., Hood L.E., Teplow D.B.* Imaging as a tool for improving length and accuracy of sequence analysis in automated fluorescence-based DNA sequencing // *Electrophoresis*. 1991. Vol. 12. P. 3–11.

*Staden R.* A computer program to enter DNA gel reading data into a computer // *Nucl. Acids Res.* 1984. Vol. 12. P. 499–503.

*West J.* Automated sequence reading and analysis // *Ibid.* 1988. Vol. 16. P. 1847–1856.

*Wu Y., Mislan D.* Automated DNA sequencing: An image processing approach // *Appl. Theor. Electrophor.* 1993. Vol. 3. P. 223–228.

### *Дополнительная литература*

*Bankier A.T.* Electrophoresis of sequence reaction samples // *Methods in molecular biology* / Ed. H.G. Griffin, A.M. Griffin. Totowa. 1993. Vol. 23. P. 121–129.

## ВЕКТОРЫ ДЛЯ МОЛЕКУЛЯРНОГО КЛОНИРОВАНИЯ И СЕКВЕНИРОВАНИЯ

Приступая к описанию плазмидных, фаговых, фагмидных и космидных векторов, не лишним будет напомнить, что секвенирование ДНК в середине 1970-х годов во многом стало возможным именно после клонирования фрагментов ДНК в плазмидных векторах. К настоящему времени создано огромное количество разнообразных векторов различного назначения как для прокариотических, так и эукариотических организмов. Однако здесь будут рассмотрены только специализированные векторы, предназначенные для определения нуклеотидных последовательностей ДНК и просто векторы общего назначения, благодаря которым и осуществлялось секвенирование ДНК, когда специально предназначенных векторов еще не было разработано. Хотя, подчас, очень трудно провести грань между векторами, которые могли бы считаться векторами общего назначения и векторами, специально сконструированными с первоначальной целью секвенирования ДНК. Для секвенирования протяженных фрагментов ДНК, чьи последовательности не могли быть “прочитаны” в одном эксперименте, разработаны рассматриваемые в следующей главе различные стратегии, основанные, главным образом, на получении серии соответствующих субклонов в плазмидных, фаговых и фагмидных векторах. Среди плазмидных векторов для молекулярного клонирования общего назначения, включая специально сконструированные векторы для секвенирования фрагментов ДНК, можно выделить несколько поколений, характеризующихся определенными чертами.

### 7.1. ПЕРВОЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Первое поколение для молекулярного клонирования, по существу, составили естественные бактериальные плазмиды так называемого дикого типа, обнаруженные ранее в клетках кишечной палочки *E. coli* как экстрахромосомальные элементы. Никаких генно-инженерных манипуляций по их изменению не проводилось, и экспериментаторы лишь охарактеризовывали их как могли. Одним из таких векторов, который широко использовался в те первые годы молекулярного клонирования, была плазида pSC101 [Cohen, Chang, 1973; Cohen et al., 1973]. Она удовлетворяла лишь самым основным требованиям, предъявляемым к вектору для клонирования чужеродных фрагментов ДНК, обладая автономной репликацией со строгим контролем, селективным маркером в виде устойчивости к тетрациклину и уникальным сайтом гексануклеотидной рестрикционной эндонуклеазы. Низкая копияность (до пяти копий на клетку), большой размер (около 10 тпн) и единственный уникальный сайт рестрикционной эндонуклеазы *EcoRI*, по которому толь-

ко и можно было осуществить клонирование (остальные уникальные сайты приходились на функционально важные участки вектора и не могли быть использованы), являлись основными ее недостатками. (Хотя, справедливости ради, надо заметить, что в отдельных случаях низкая копияность плазмидного вектора, несущего фрагмент чужеродной ДНК, имеет определенный биологический смысл.) Следует отметить, что полная нуклеотидная последовательность плазмиды pSC101 до сих пор неизвестна, поскольку так и не определена за ненадобностью. К другим векторам первого поколения следует отнести колициногенные плазмиды, такие как ColE1 и pMB1, выделенную из клинического изолята [Bazarel, Helinski, 1968]. Данные вектора несли близкородственные репликоны и характеризовались отсутствием строгого контроля репликации и за счет этого имели более высокую копияность, но отсутствие селективного маркера в виде устойчивости к какому-либо антибиотику явилось причиной их довольно ограниченного использования. Стало очевидно, что необходимы определенные усовершенствования этих природных плазмид. Однако, потребовалось некоторое время на разработку векторных молекул следующего поколения, обладающих целым рядом преимуществ.

## 7.2. ВТОРОЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Второе поколение векторов представляло собой уже искусственно созданные вектора, объединившие в себе черты различных природных плазмид дикого типа. Так, на основе pMB1 было создано небольшое семейство векторов pMB, среди которых наиболее широкое применение нашла плазида pMB9 [Bolivar et al., 1977b]. Она характеризовалась относительно небольшим размером (около 5,3 тпн), маркерным геном устойчивости к тетрациклину и уникальными сайтами для 4 гексануклеотидных рестрикционных эндонуклеаз – *EcoRI*, *BamHI*, *HindIII* и *Sall*. Причем клонирование по 3 последним сайтам приводило к инактивации гена устойчивости к тетрациклину, что способствовало поиску рекомбинантных клонов, но одновременно требовало проводить селекцию по иммунности к колицину, что, впрочем, было не очень эффективно. Отличительной чертой продвинутых векторов второго поколения стало наличие генов, определяющих устойчивость к двум или даже трем антибиотикам, служивших удобными маркерными признаками.

Наиболее яркими представителями таких продвинутых векторов второго поколения явились векторы семейства pBR, основу которых составили плазмиды pSC101, RSF2124 и pMB1. Непосредственным предшественником плазмиды pBR312, явившейся начальной в серии векторов pBR, стала плазида pMB [Bolivar et al., 1977a]. В результате генно-инженерных манипуляций в нее были добавлены ген устойчивости к тетрациклину (из плазмиды pSC101) и ген устойчивости к ампициллину (транспозон  $Tn^3$  из вектора RSF2124). Таким образом, размер плазмиды pBR312 значительно увеличился – до 10,2 тпн. Дальнейшие процедуры по ее усовершенствованию были направлены на уменьшение размера за счет



удаления участков молекулы, не несущих важной функциональной нагрузки. За счет уменьшения размера вектора увеличилось число уникальных сайтов для гексануклеотидных рестрикционных эндонуклеаз, поскольку в то время размер плазмид и число уникальных сайтов в них (хотя и не строго) были обратно пропорциональны. (Открывшаяся в последующем возможность удаления ненужных сайтов рестрикционных эндонуклеаз нарушила это соотношение.) Другой важной причиной желаний экспериментаторов уменьшить размер векторной молекулы являлось то, что частота трансформации и размер вектора находятся также в области обратной зависимости. Так, через промежуточные плазмиды pBR313, pBR318 и pBR320 был получен вектор pBR322, с которым связана целая эпоха в молекулярном клонировании [Bolivar et al., 1977b; Balbas et al., 1986]. Дальнейшие модификации плазмиды pBR322 привели к созданию серии векторов pBR324, pBR325, pBR327, pBR328, и pBR329 [Bolivar et al., 1978]. Некоторые из них были доступны из коммерческих источников и использовались в экспериментах по молекулярному клонированию, но в гораздо меньшей степени, чем плаزمида pBR322.

Плазмидный вектор pBR322 объединил в себе такие черты своих предшественников, как репликон pMB1, ген устойчивости к тетрациклину Tc<sup>R</sup> и ген устойчивости к ампициллину Ap<sup>R</sup>. Репликон pMB1 отличается от репликона ColE1 всего одной инверсией размером 2 пн [Heusterspreute, Davison, 1983] и поэтому часто между ними не делают различий и указывают в качестве репликона вместо pMB1 ColE1. За счет ослабленного контроля репликации, характерного для этих репликонов, копияность плазмиды pBR322 составила уже около 20–50 копий на клетку. Принципиальным отличием плазмид, имеющих такой репликон с ослабленным контролем, является возможность репликации в отсутствие белкового синтеза. Это свойство использовалось экспериментаторами для увеличения копияности плазмид на клетку добавлением в культуральную среду антибиотиков хлорамфеникола или спектиномицина (для векторов, несущих ген устойчивости к хлорамфениколу), подавляющих белковый синтез, и, по некоторым оценкам, при этом число плазмидных молекул увеличивалось на два порядка [Clewel, 1972]. Следует отметить, что ген Tc<sup>R</sup> в плазмиде pSC101 является индуцибельным, тогда как при добавлении этого фрагмента в плазмиду pBR322 он был взят без участка, кодирующего один из белков-регуляторов работы данного гена, и его экспрессия стала конститутивной. Такая конститутивная экспрессия имеет некоторое отрицательное значение, однако при этом имеет место уменьшение общего размера вектора.

Первоначально размер плазмиды pBR322 был определен как равный приблизительно 4,3 тпн. За счет такого небольшого размера количество уникальных сайтов в данном векторе было уже значительным. В литературе встречаются различные сведения о числе уникальных сайтов для гексануклеотидных рестрикционных эндонуклеаз, имеющих в плазмиде pBR322, и это связано с тем, что ко времени написания той или иной статьи было известно разное количество таких ферментов. В связи с тем, что продолжается нахождение рестрикционных

эндонуклеаз с новыми сайтами узнавания, то правильнее будет ограничиться упоминанием, что таких сайтов все же просто много. Однако более важным является количество уникальных сайтов узнавания гексануклеотидных рестрикционных эндонуклеаз, расположенных в генах устойчивости к антибиотикам. Так, в гене, кодирующим устойчивость к тетрациклину, таких сайтов, по крайней мере, 7, а в гене  $\beta$ -лактамазы, обеспечивающей устойчивость к ампициллину, их 4. Достаточно большой выбор сайтов рестрикционных эндонуклеаз, расположенных в генах устойчивости к антибиотикам, по которым можно было бы осуществить клонирование и провести дальнейшую селекцию по устойчивости к другому антибиотику, сделало вектор pBR322 очень удобным для экспериментаторов. Например, при клонировании по сайту рестрикционной эндонуклеазы *Vam*HI происходило нарушение гена, кодирующего устойчивость к тетрациклину, и такие рекомбинантные плазмиды приобретали фенотип  $Ap^R$  и  $Tc^S$ . Колонии, выросшие после трансформации на чашке Петри с ампициллином, пересеивали параллельно на две чашки с ампициллином и тетрациклином соответственно, и в результате, те клоны, колонии которых не росли на тетрациклине, были рекомбинантными. Такой тест на наличие вставки, несмотря на дополнительный этап посева колоний на разные чашки Петри и большой объем рутинной работы, позволил повысить эффективность процедуры клонирования за счет упрощения поиска рекомбинантных клонов. Однако главным и, пожалуй, чуть ли не единственным недостатком плазмиды pBR322 (конечно, как вектора своего времени) было наличие сайта рестрикционной эндонуклеазы *Eco*RI, пригодного для клонирования, но не нарушающего при этом никакой ген и не позволяющего проводить подобный скрининг. Уникальность данной рестрикционной эндонуклеазы заключалась в том, что она была основным ферментом клонирования для плазмидных векторов первого поколения и большое количество клонов было получено с ее использованием. Другими важными причинами ее широкого использования была доступность этой рестрикционной эндонуклеазы из многих коммерческих источников, а также сила привычки экспериментаторов. Уже упоминавшиеся выше векторы pBR325, pBR328, и pBR329, несколько отличаясь между собой по размеру, несли одинаковый важный признак в виде устойчивости к хлорамфениколу, в котором и был расположен уникальный сайт узнавания рестрикционной эндонуклеазы *Eco*RI. Однако ни наличие этого, казалось бы, желанного, сайта узнавания рестрикционной эндонуклеазы *Eco*RI, ни три гена устойчивости к трем антибиотикам (что, впрочем, не давало особых преимуществ, так как для отбора рекомбинантов достаточно двух маркерных генов устойчивости) не позволили этим векторам превзойти по популярности классическую плазмиду pBR322.

Главной причиной того, что pBR322 в течение целого ряда лет оставалась основным плазмидным вектором для молекулярного клонирования являлось то, что она была первым вектором, полная нуклеотидная последовательность которого была определена [Sutcliffe, 1978]. Первоначально считали, что размер плазмиды pBR322 равен 4362 пн [Sutcliffe, 1978]. Позднее было обнаружено, что в тетрацикли-

новом гене пропущен один нуклеотид и длина плазмиды pBR322 была принята равной 4363 пн [Peden, 1983]. Однако уточненная последовательность плазмиды pBR322 составляет 4361 пн [Watson, 1988]. Известная нуклеотидная последовательность всего вектора в значительной мере облегчала процедуру рестриктазного картирования вставки и способствовала более точному построению рестриктазных карт клонированных фрагментов ДНК. Знание всей последовательности вектора позволило также локализовать расположение сайтов узнавания для частощепящих рестрикционных эндонуклеаз и рассчитать размеры фрагментов ДНК, получающихся в результате расщепления плазмиды pBR322 данными ферментами, и составить специальные таблицы. Фрагменты плазмиды pBR322, получающиеся в результате ее расщепления некоторыми частощепящими рестрикционными эндонуклеазами, служили в качестве маркерных молекул небольшого размера при проведении электрофореза как в агарозном, так и в полиакриламидном гелях. Наиболее оптимальное распределение полос фрагментов ДНК при разделении гель-электрофорезом характерно для рестрикционных эндонуклеаз *AluI*, *HinfI*, *MspI*, а также некоторых других, что и предопределило использование их в качестве маркеров. И если, на сегодняшний день, плазида pBR322 сдала свои позиции как вектор для молекулярного клонирования, то использование ее рестриктазных фрагментов в качестве маркерных молекул для электрофореза продолжается. Хотя следует отметить активную конкуренцию со стороны специально созданных маркеров, так называемых лестниц, образующих строгую периодичность полос в результате гель-электрофореза, за счет увеличения размера последующих “ступенек” на фиксированные величины, как, например, 50, 100 или 123 пн для разных маркерных “лестниц”.

Завершая краткое рассмотрение плазмидных векторов второго поколения, следует подчеркнуть, что это были уже искусственно созданные вектора, но без каких-либо синтетических последовательностей нуклеотидов и с естественным расположением сайтов узнавания для рестрикционных эндонуклеаз, включая гексануклеотидные.

### 7.3. ВЕКТОРЫ НА ОСНОВЕ НИТЕВИДНОГО ФАГА M13

Желание изложить развитие методов молекулярного клонирования небольших фрагментов ДНК и связанного с ним секвенирования ДНК, насколько это возможно, в хронологическом порядке, заставляет на время отвлечься от плазмидных векторов и рассмотреть небольшие одноцепочечные фаговые вектора, тем более, что они оказали, как будет видно из дальнейшего изложения, серьезное влияние на процесс создания последующих поколений векторов.

Небольшие одноцепочечные нитевидные фаги типа Ff, к которым относятся бактериофаги f1, fd, M13 и некоторые другие, были известны достаточно давно и их биология хорошо изучена. Весьма важным свойством этих фагов явилось то, что они не вызывали лизиса бакте-

рий, а только задерживали их рост, в результате чего на бактериальном газоне формировались светлые зоны фаговых бляшек. Также было известно, что при упаковке мутантных фагов в отличие от одноцепочечного изометрического фага  $\phi$ X174 образовывались фаговые частицы большего размера, которые могли инфицировать клетки *E.coli*.

Мессингом и соавт. при первой попытке использовать дикий фаг M13 в качестве вектора для молекулярного клонирования был проведен элегантный эксперимент, направленный на выявление области фага M13, в которую можно было внедрять фрагменты чужеродной ДНК без ущерба для жизнедеятельности фага [Messing et al., 1977]. В качестве маркерной молекулы был выбран *Hind*III-фрагмент  $\beta$ -галактозидазы, который, как было известно, в условиях  $\alpha$ -комплементации с аналогичным геном *LacZ* $\Delta$ M15, несущим мутацию и расположенным на хромосоме *E.coli*, мог формировать функциональный фермент и при добавлении соответствующего хромогенного субстрата и индуктора давать голубое окрашивание. Было осуществлено недорасщепление фага M13 рестрикционной эндонуклеазой *Bsu*I, имеющей в этом векторе 10 сайтов узнавания, и с помощью гель-электрофореза отобраны линейные формы с единичными разрывами цепи. В результате клонирования *Hind*III-фрагмента  $\beta$ -галактозидазы был отобран фаг, стабильно наследующий данный признак, обозначенный как M13mp1. Местом клонирования вставки оказалась межгенная область между II и IV генами фага. Следующим этапом совершенствования данного вектора для молекулярного клонирования было добавление во фрагмент гена  $\beta$ -галактозидазы, обозначаемого как *LacZ'*, уникального сайта рестрикционной эндонуклеазы *Eco*RI, осуществленное с помощью химического мутагена, в результате чего образовались вектора M13mp2 и M13mp3 [Gronenbom, Messing, 1978]. Дальнейшие улучшения векторов на основе одноцепочечного фага M13 привели к последовательному добавлению уникальных сайтов узнавания для рестрикционных эндонуклеаз *Hind*III и *Bam*HI. Однако это был еще не настоящий полилинкер, а, скорее, его прообраз. Настоящая революция в создании векторов произошла после помещения в маркерный  $\beta$ -галактозидазный ген химически синтезированного участка ДНК, содержащего расположенные друг за другом уникальные сайты узнавания для целого ряда гексануклеотидных рестрикционных эндонуклеаз, пригодные для их использования в клонировании.

Первые сконструированные полилинкеры представляли собой симметричные последовательности, несущие по два сайта узнавания некоторых рестрикционных эндонуклеаз, как, например, *Eco*RI–*Bam*HI–*Sal*I–*Pst*I–*Sal*I–*Bam*HI–*Eco*RI. Создание впоследствии асимметричных полилинкеров, таких как, например, *Eco*RI–*Sma*I–*Bam*HI–*Sal*I–*Pst*I–*Hind*III в векторе M13mp9 или ему подобных, привело к возможности использования для выделения вставки одновременно двух разных ферментов, узнающих отличающиеся последовательности и образующих как 5', так и 3'-выступающие концы, что было необходимо или для мечения только одного из них, или переклонирования в другом векторе в

заранее заданной ориентации. Последовательности полилинкеров были сконструированы таким образом, что не нарушали рамку считывания NH<sub>2</sub>-терминального фрагмента β-галактозидазы, но при клонировании какого-либо фрагмента ДНК функциональная целостность фермента нарушалась и при использовании специального хромогенного субстрата 5-бромо-4-хлоро-галактозида (x-gal) и индуктора изопропил-β-тио-галактозида (ИПТГ) легко выявлялись рекомбинантные бактериофаги, несущие вставки. Отдельные случаи сохранения голубого окрашивания рекомбинантными фаговыми бляшками, равно как плазмидами и фагмидами, происходят вследствие клонирования небольших вставок с количеством нуклеотидов кратным 3, что не приводит к нарушению рамки считывания гена β-галактозидазы (конечно, при условии отсутствия терминирующих кодонов во вставке).

В течение ряда лет была сконструирована целая серия одноцепочечных векторов M13mp, отличающихся только своими полилинкерами (M13mp6, M13mp7, M13mp8, M13mp9, ..., M13mp18, M13mp19) [Messing et al., 1981; Messing, Vieira, 1982; Vieira, Messing, 1982; Messing, 1983; Norrander et al., 1983; Yanisch-Petron et al., 1985]. Позднее на основе некоторых из них были созданы вектора m1CE с фаговыми промоторами T7 и SP6, обладающие рядом преимуществ [Eperon, 1986]. Следует отметить, что до недавнего времени продолжались работы по улучшению векторов данного семейства и им подобных [Venes et al., 1993]. Все это было вызвано тем, что одноцепочечные вектора на основе фага M13 оказались крайне удобными при выполнении проектов по секвенированию фрагментов ДНК ферментативным методом Сэнгера. Небольшое семейство векторов mWB (mWB2341, mWB2342, mWB2344) [Barnes, Bevan, 1983; Barnes et al., 1987], сконструированное на основе того же одноцепочечного фага M13, несло ряд черт, делающих их несколько отличными от векторов семейства M13mp. Так, полилинкер у этих векторов располагался в ином месте фага и за счет этого (возможно?) клонированные вставки в нем были более стабильны. Это позволило клонировать в нем крупные фрагменты ДНК длиной до 12 и даже 14 тпн. Ориджин репликации в этих векторах был расположен так, что фаг нарабатывал (-)-цепь ДНК. Еще одной особенностью было то, что все фаговые бляшки, как принадлежащие исходному вектору, так и содержащие вставки, несли голубое окрашивание и только специально осуществляемые делеции вставок, приводящие в ряде случаев к их полному удалению и затрагивающие при этом область *Lac*-оператора, расположенную уже за вставкой, приводили к исчезновению голубого окрашивания. Следует отметить, что данные векторы были специально созданы для стратегии килосеквенирования ДНК, осуществляемой путем получения однонаправленных делеций вставки.

Таким образом, одноцепочечные векторы на основе нитевидного фага M13 являются самостоятельным типом векторов, но за счет существования в жизненном цикле фага двуцепочечной репликативной формы, позволяющей работать с ней как с плазмидой, они дали "строительный материал" для создания следующих поколений плазмидных векторов.

#### 7.4. ТРЕТЬЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Третье поколение плазмидных векторов представляет собой сильно модифицированные вектора семейства pBR. Его наиболее известными представителями являются вектора семейства pUC, разработанные Мессингом и соавт. [Viera, Messing, 1982]. От pBR322 им в наследство достался репликон ColE1, маркерный ген устойчивости к ампициллину, еще меньший размер (главным образом, за счет удаления гена устойчивости к тетрациклину). Данные векторы характеризовались весьма высокой копийностью, достигающей 200 копий на клетку. Однако главными новыми чертами этих плазмидных векторов можно считать маркерный ген, кодирующий NH<sub>2</sub>-концевой фрагмент β-галактозидазы *LacZ'*, и встроенный в него полилинкер от одноцепочечных векторов серии M13mp. Дальнейшее расширение семейства векторов pUC шло за счет увеличения количества сайтов узнавания в полилинкере, называемом также MCS [Multiple Cloning Sites]. Так, данные векторы, сконструированные попарно (pUC8/pUC9, ..., pUC18/pUC19), несли в парах одинаковые MCS, но ориентированные в противоположных направлениях аналогично своим "двоюродным братьям" – векторам серии M13mp. Такая двунаправленная последовательность расположения сайтов узнавания рестрикционных эндонуклеаз в полилинкере давала определенные преимущества в отдельных случаях молекулярного клонирования. Общей чертой всех этих MCS векторов серии pUC являлось то, что они по краям были ограничены сайтами узнавания рестрикционных эндонуклеаз *EcoRI* и *HindIII*, тогда как внутренние участки у них несколько отличались. Однако, несмотря на довольно большой выбор сайтов для клонирования рестриктазных фрагментов ДНК, полилинкеры этих векторов pUC семейства все же не в полной мере удовлетворяли запросам экспериментаторов и в связи с этим шло активное конструирование подобных плазмидных векторов, характеризующихся главными чертами векторов серии pUC (ген устойчивости к ампициллину, ColE1 репликон и маркерный ген, кодирующий NH<sub>2</sub>-концевой фрагмент β-галактозидазы *LacZ'* со встроенным в него участком MCS), но с полилинкером, несущим или другую последовательность сайтов узнавания тех же рестрикционных эндонуклеаз в нем или с добавлением ранее не использовавшихся. Весьма интересно семейство векторов pMTL, характеризующееся некоторыми отличающимися чертами от pUC векторов и имеющее полилинкер, во-первых, с другим более удобным расположением сайтов узнавания для наиболее часто используемых ферментов и, во-вторых, с добавлением значительного количества новых, что сделало данные векторы более приспособленными для клонирования и последующего секвенирования фрагментов ДНК, полученных в результате ультразвуковой обработки [Chambers et al., 1988]. Следует отметить, что такой элемент практически любого современного плазмидного вектора, как полилинкер или участок MCS, подвергся за эти годы наибольшей модификации, выражающейся в постоянно продолжающемся конструировании данных участков со всевозможными вариантами сайтов узнавания для рестрикционных эндонуклеаз. Так, созданы участки

MCS, называемые суперполилинкерами и, действительно, таковыми являющиеся, поскольку самый протяженный из них SL2 векторов pSL300 и pSL301 имеет длину 353 пн и содержит все 64 возможных варианта для гексануклеотидных рестрикционных эндонуклеаз, узнающих непрерывные палиндромные последовательности, причем как для уже обнаруженных в природе, так и для еще не выявленных [Brosius, 1989].

Однако, для того чтобы все существующие в MCS сайты узнавания гексануклеотидных рестрикционных эндонуклеаз были пригодны для использования их в качестве сайтов для молекулярного клонирования, необходимо, чтобы они были уникальными, т.е. не встречались больше в остальной последовательности вектора. Кроме действительно изначально уникальных, для некоторых ферментов имелись дополнительные участки узнавания в других местах векторной молекулы. Первые векторы, как семейства одноцепочечных фаговых векторов M13mp, так и семейства pUC, подвергали химическому мутагенезу с целью изменения отдельных нуклеотидов в нежелательных сайтах узнавания тех или иных рестрикционных эндонуклеаз и в результате кропотливого анализа отбирали требуемые варианты. В последующем, с появлением и разработкой метода сайтнаправленного мутагенеза, осуществляемого с помощью специально синтезированного олигонуклеотидного праймера, несущего измененный неспариваемый нуклеотид, процедура удаления ненужных сайтов узнавания рестрикционных эндонуклеаз существенно упростилась.

Таким образом, третье поколение плазмидных векторов несет в себе кардинально новые черты, не присущие векторам предыдущих поколений, которые не могут случайно возникнуть в природе, и объединенные в одном векторе экспериментаторами. Считаем важным подчеркнуть, что для создания векторов третьего поколения потребовалось удаление сайтов узнавания некоторых гексануклеотидных рестрикционных эндонуклеаз в последовательности вектора ввиду их присутствия в полилинкере с целью использования данных ферментов для молекулярного клонирования.

## 7.5. ЧЕТВЕРТОЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Следующее усовершенствование плазмидных векторов не столь существенно, однако практически все вектора общего назначения следующих поколений содержат эти небольшие участки и по этой причине эти векторы все же можно отнести к четвертому поколению. Первыми такими векторами явились плазмиды pSP64 и pSP65, несущие участок промоторной области бактериофага SP6 из бактерии *Salmonella typhimurium*, расположенный в 5'-направлении от использованного при их конструировании полилинкера плазмид pUC12/pUC13 [Melton et al., 1984]. К сожалению, в этих плаزمиде  $\beta$ -галактозидазный ген *LacZ'* оказался неактивным. Позднее были сконструированы многочисленные векторы с промотором фага SP6, промоторами фагов T3 и T7, фланкирующими функциональный NH<sub>2</sub>-концевой фрагмент  $\beta$ -галак-

тозидазы *LacZ'* содержащимся в нем полилинкером. За счет данных промоторных участков различных фагов возможно получение в системе *in vitro* большого количества молекул РНК, транскрибируемых с линеаризованных матриц (с обеих цепей) и находящихся различного применения в молекулярной биологии. На данный тип векторов с возможностью продукции *in vitro* РНК-копий клонированного фрагмента ДНК выдан американский патент, принадлежащий фирме "Promega" [Jendrisak et al., 1988].

К определению нуклеотидной последовательности ДНК данные участки фаговых промоторов все же имеют некоторое отношение, хотя и не являются обязательными. Так, кроме широко применяемых универсальных "прямого" или "обратного" праймеров, иногда используют олигонуклеотидные праймеры, комплементарные данным фаговым промоторам. Главной причиной их использования является то, что они расположены несколько ближе к полилинкеру и соответственно уменьшается количество не нужных для чтения нуклеотидов самого вектора до начала вставки и, таким образом, увеличивается количество определенных нуклеотидов в клонированном фрагменте ДНК.

## 7.6. ПЯТОЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Как уже отмечалось выше, одноцепочечные векторы серии M13mp оказали значительное влияние на развитие и совершенствование плазмидных векторов третьего и последующих поколений. Векторы же пятого поколения представляют собой гибридные конструкции, содержащие, кроме плазмидных последовательностей, межгенный участок одноцепочечного нитевидного фага f1 (родственный фагу M13), ответственный за репликацию фага. Такие химерные векторы, содержащие два участка инициации репликации f1/ColE1, были названы фагмидами. Одной из первых фагмид была pD4, представляющая собой плазмиду pBR322 со вставкой участка фага f1 размером 1250 пн [Dotto et al., 1981]. Этот участок был просто клонирован в плазмиде pBR322 с целью изучения его функций и в качестве специализированного вектора эта конструкция не использовалась. Первыми специально сконструированными фагмидами были вектора семейства pEMBL, созданные на основе векторов pUC8/pUC9 и межгенного участка одноцепочечного фага f1 [Dente et al., 1984]. Данный участок фаговой ДНК, так называемый f1 ориджин репликации, помещенный в эти вектора в разных ориентациях, после суперинфекции специальным хелперным фагом M13K07 или R408 приводил к секреции (+)- или (-)-цепей ДНК, что потребовало соответствующего обозначения фагмид – pEMBL8<sup>+</sup>, pEMBL8<sup>-</sup>, pEMBL9<sup>+</sup> и pEMBL9<sup>-</sup>. Если плазмидные векторы семейства pUC существовали в парных комбинациях, отличаясь ориентацией полилинкера, то многие существующие ныне фагмиды представляют собой набор из четырех возможных вариантов (две ориентации полилинкера, две ориентации f1 ориджина репликации, приводящие к секреции (+)- или (-)-цепей ДНК), что значительно расширяет возможности исследователей.



Другое решение проблемы секреции разных цепей ДНК было предложено в фагмидных векторах рKUN9 и рKUN19, являющихся производными от плазмид рUC9 и рUC19 [Peeters et al., 1986]. Уникальной чертой этих векторов явилось то, что кроме f1 ориджина репликации они несли расположенный в противоположной ориентации еще один подобный участок от одноцепочечного нитевидного фага IKe, родственного фагам группы Ff. Несмотря на аналогичную организацию генома этих одноцепочечных фагов, гомология их нуклеотидных последовательностей составляет всего лишь 55% и за счет этого не происходит перекрестного узнавания их сигнальных последовательностей. Таким образом, имеется возможность секреции одной цепи ДНК в ответ на суперинфекцию хелперными фагами M13K07, R408 и другой цепи после суперинфекции хелперными фагами IKe-9, Mike Δ [Konings et al., 1986]. Следует отметить, что несмотря на привлекательность данного подхода, он все же не получил широкого развития.

Химерные фагмидные векторы обладают целым рядом преимуществ, поскольку сочетают в себе все важные черты как одноцепочечных фаговых векторов, так и плазмидных векторов. Такое взаимодействие векторов разных типов привело к ситуации, которую можно назвать “два в одном”, и эти, как бы “самостоятельные”, векторы в составе фагмиды приобрели для себя новые важные свойства и улучшения. Так, если рассмотреть, какие преимущества приобрел одноцепочечный фаговый вектор (собственно вектором и не являющийся, поскольку в фагмиде, во-первых, все же больше плазмидного и, во-вторых, требующим для формирования фаговых частиц специальный хелперный фаг в связи с отсутствием генов, кодирующих фаговые белки) в составе фагмиды, то можно видеть, что получение ДНК в одноцепочечной форме не приводит к нестабильности клонированных фрагментов ДНК большого размера по сравнению с аналогичными фрагментами, клонированными ранее в векторах серии M13mp. Главной причиной такой повышенной стабильности вставок (особенно содержащих повторяющиеся элементы) является репликация по типу плазмиды и то, что существование их в одноцепочечной форме в процессе жизнедеятельности фагмиды весьма кратковременно и даже необязательно. Второй особенностью, связанной с одноцепочечной формой, является возможность формировать фаговые частицы с (+)- или (-)-цепью ДНК в зависимости от ориентации f1 ориджина репликации. (Собственно такая возможность теоретически существовала и для самих фаговых векторов, но практически она была не реализована ввиду того, что проблема разных цепей решалась для них наличием разнонаправленных полилинкеров.) Другим важным отличием фагмид от фаговых векторов является значительно больший размер последних. Обычный размер фагмидных векторов варьирует от 3 до 4 тпн, тогда как вектора серии M13mp имеют размер около 6,5 тпн. За счет такой разницы в размерах в фагмидах возможно клонирование фрагментов ДНК гораздо большего размера (до нескольких тпн) при высокой стабильности вставки, тогда как в одноцепочечных фаговых векторах вставки крупнее 1 тпн за счет делеций часто оказывались нестабильными. Хотя, как уже отмеча-

лось выше одноцепочечные вектора серии mWB стабильно поддерживали вставки до 14 тпн [Barnes, Bevan, 1983]. Выделение двуцепочечной репликативной формы фагового вектора как для целей клонирования, так и для анализа размера вставок гель-электрофорезом и их рестриктазного картирования связано с некоторыми трудностями, тогда как ни крупномасштабное, ни мелкомасштабное выделения двуцепочечной ДНК в составе фагмиды не представляет каких-либо проблем.

Приобретенные от такого союза преимущества для “фагового вектора” в составе фагмиды значительно перевешивали некоторые недостатки количественного характера и введение дополнительного этапа, необходимого для запуска репликации по фаговому типу. Так, несмотря на несколько меньший выход одноцепочечной фагмидной ДНК по сравнению с настоящим фаговым вектором, количества ДНК, получаемого из 1 мл жидкой культуры, вполне хватало на проведение секвенирующих реакций. Дополнительным, впрочем весьма несложным, этапом была необходимость суперинфекции культуры *E.coli*, несущей фагмиду, специальным хелперным фагом, запускающим механизм репликации по фаговому типу и обеспечивающим упаковку одноцепочечной ДНК фагмиды фаговыми белками. Были разработаны упомянутые выше специальные хелперные фаги R408 и M13K07, IKe-9 и Mike Δ, секретирующиеся вместе с фаговыми частицами, содержащими одноцепочечную фагмидную ДНК, но не интерферирующие с ней на этапе отжига праймера и при проведении секвенирующих реакций. Более того, в литературе сообщается о легко воспроизводимом методе получения одноцепочечной ДНК фагмид после трансформации ими компетентных клеток *E.coli*, уже содержащих хелперный фаг, что исключает необходимость дополнительного этапа суперинфекции [Jupin, Gronenborn, 1995].

Что касается преимуществ, полученных самим плазмидным вектором от объединения с f1 ориджином репликации одноцепочечного фагового вектора, то оно всего одно, но очень значительное. Если описанные выше преимущества, полученные фаговым вектором от объединения, носят, главным образом, количественный характер, то плазмида получила качественно новое и очень важное свойство, а именно возможность существовать в одноцепочечной форме, которое ранее ей было недоступно.

За счет таких универсальных свойств фагмидные векторы общего назначения практически вытеснили для целей молекулярного клонирования обыкновенные плазмиды. Так, в каталогах некоторых ведущих молекулярно-биологических фирм почти не осталось обыкновенных плазмид общего назначения и вместо них предлагаются аналогичные фагмиды. Наибольшее распространение получили серия фагмидных векторов pGEM-Zf, поставляемых фирмой “Promega” и серия векторов pBluescript, поставляемых фирмой “Stratagene”. Они характеризуются стандартными чертами фагмидных векторов и практически все отличия между ними заключены в полилинкерах. Так, например, группа векторов pBluescript представлена всеми четырьмя возможными вариантами (pBluescript II SK+/-, pBluescript II KS+/-, где буквами SK и KS обозначены разнонаправленные полилинкеры с граничными сайтами узна-

вания для рестрикционных эндонуклеаз *SacI* и *KpnI* соответственно) [Alting-Mees, Short, 1989]. Данные вектора имеют достаточно большие участки MCS, где, кроме упомянутых выше рестрикционных эндонуклеаз, имеются участки узнавания для следующих ферментов: *SacII*, *NotI*, *XbaI*, *SpeI*, *BamHI*, *SmaI*, *PstI*, *EcoRI*, *EcoRV*, *HindIII*, *ClaI*, *Sall*, *XhoI* и *ApaI*, приведенные здесь в направлении от *SacI* к *KpnI*. Кроме этих векторов линейка фагмидных векторов фирмы "Stratagene" включает ряд сходных векторов pBC+/-, но несущих вместо гена устойчивости к ампициллину ген устойчивости к хлорамфениколу, фагмид pBS+/- с полилинкерами от pUC19 и некоторые другие. Фагмидные вектора семейства pGEM-Zf фирмы "Promega" (pGEM-3Zf, pGEM-5Zf, pGEM-7Zf, pGEM-9Zf, pGEM-11Zf, pGEM-13Zf) существуют в попарных (+/-) комбинациях и отличаются своими полилинкерами, где первый вектор этого семейства pGEM-3Zf(+/-) несет полилинкер от плазмиды pUC19. Однако следует отметить то, что практически все отличия существующих фагмидных векторов друг от друга заключаются в различно организованных полилинкерах и в генах устойчивости к разным антибиотикам.

## 7.7. ШЕСТОЕ ПОКОЛЕНИЕ ПЛАЗМИДНЫХ ВЕКТОРОВ

Первые два поколения плазмидных векторов, предназначенные для клонирования фрагментов ДНК, имели естественное расположение сайтов узнавания как гексануклеотидных, так и тетрануклеотидных рестрикционных эндонуклеаз. Все последующие поколения векторов общего назначения имели, кроме прочих преимуществ, также искусственно созданный полилинкер, в котором были сосредоточены сайты узнавания нескольких гексануклеотидных рестрикционных эндонуклеаз. Для того чтобы появилась возможность расщеплять вектор для целей клонирования и субклонирования по сайту узнавания какой-либо полилинкерной рестрикционной эндонуклеазы, в остальной части вектора такой же сайт (если он имелся) было необходимо удалить. Так, векторы третьего поколения (pUC и им подобные) имели в отличие от своих предшественников уже не естественное расположение сайтов узнавания некоторых ключевых (полилинкерных) гексануклеотидных рестрикционных эндонуклеаз в самой последовательности вектора. Однако все плазмидные, фаговые, фагмидные (а также и космидные) векторы предыдущих поколений все же имели весьма существенный недостаток в виде наличия в векторной последовательности сайтов узнавания для частощепящих рестрикционных эндонуклеаз с тетрануклеотидными участками узнавания, ввиду чего было невозможно расщеплять вставку этими ферментами, поскольку это неминуемо привело бы к разрушению вектора. Поэтому уникальной чертой векторов pSequoia можно считать отсутствие в их нуклеотидной последовательности сайтов узнавания для нескольких частощепящих рестрикционных эндонуклеаз с тетрануклеотидными участками узнавания. Это весьма важное и сулящее значительные удобства и преимущества новшество делает семейства фагмидных векторов pSequoia весьма удобными не только для секвенирования протяженных фрагментов ДНК, но и при проведении раз-

личных молекулярно-биологических экспериментов, требуется работа с частью вставки, как, например, при футпринтинге или исследовании белково-нуклеиновых взаимодействий с помощью метода торможения в геле.

Проведенный анализ наиболее известных фагмидных векторов показал наличие множественных сайтов для большинства тетра-нуклеотидных рестрикционных эндонуклеаз в их последовательности. Так, в векторе pGEM-5Zf имеется 20 сайтов узнавания для рестрикционной эндонуклеазы *HhaI* (CGCG), по 18 для *AluI* (ACGT) и *Tru9I* (TTAA), 14 для *Sau3AI* (GATC) и по 5 для *RmaI* (CTAG) и *TaqI* (TCGA). В то же время в данном векторе имеется всего один сайт тетра-нуклеотидной рестрикционной эндонуклеазы *RsaI*, расположенный в гене  $\beta$ -лактамазы. Для создания вектора с отсутствием в нем сайтов расщепления для тетра-нуклеотидной рестрикционной эндонуклеазы *RsaI* был проведен сайтнаправленный мутагенез с целью замены одного нуклеотида в единственном сайте узнавания тетра-нуклеотидной рестрикционной эндонуклеазы *RsaI* (GTAC) и в совпадающем с ним же сайте гексануклеотидной рестрикционной эндонуклеазы *ScaI* (CGTACG) в гене  $\beta$ -лактамазы, определяющем устойчивость к антибиотику ампициллину, таким образом, что за счет вырожденности генетического кода замены аминокислоты Glu не произошло, но оба эти сайта перестали узнаваться данными ферментами [Чемерис и др., 1996]. В результате такого мутагенеза появилась возможность расщеплять клонированный фрагмент ДНК без опасения расщепить сам вектор с помощью тетра-нуклеотидной рестрикционной эндонуклеазы *RsaI* или ее неизошизомера *Csp6I*, генерирующего в отличие от *RsaI* 5'-выступающий липкий конец 5'-ТА, пригодный для клонирования в некоторые концы, генерируемые гексануклеотидными эндонуклеазами рестрикции и, в частности, в сайт узнавания фермента *NdeI*, имеющийся в полилинкере векторов pSequoia. Созданный таким образом вектор получил наименование pSequoiaT12, стал первым из создаваемого семейства фагмидных векторов pSequoia и открыл новое поколение векторов общего назначения (рис. 7.1).

Другой уникальной чертой вектора pSequoiaT12 является наличие в его полилинкере сайтов узнавания для октануклеотидных рестрикционных эндонуклеаз *SwaI* или ее изошизомера *SmiI* (ATTTAAAT) и *SrfI* (GCCCCGGC). Справедливости ради следует отметить, что в полилинкерах некоторых векторов сайты некоторых октануклеотидных рестрикционных эндонуклеаз имелись и раньше. Так, например, в полилинкерах фагмид семейства pBluescript и pGEMZf имеется сайт для рестрикционной эндонуклеазы *NotI*. Одной из причин отсутствия подобных сайтов для октануклеотидных рестрикционных эндонуклеаз в полилинкерах различных векторов является то, что при конструировании большинства полилинкеров (сохранившихся без изменения до сего дня) был известен только один такой фермент *NotI*, узнающий непрерывную палиндромную октануклеотидную последовательность GCGGCCGC. Однако относительно недавно было обнаружено несколько подобных ферментов, что дало возможность сконструировать новый полилинкер с сайтами узнавания для этих рестрикционных эндонуклеаз. В то же время нельзя не отметить практически полное отсутствие в полилинкерах векторов пятого и предшеству-

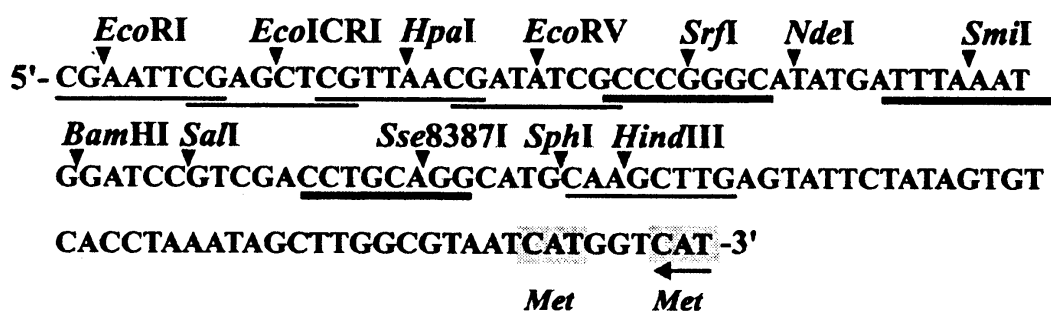
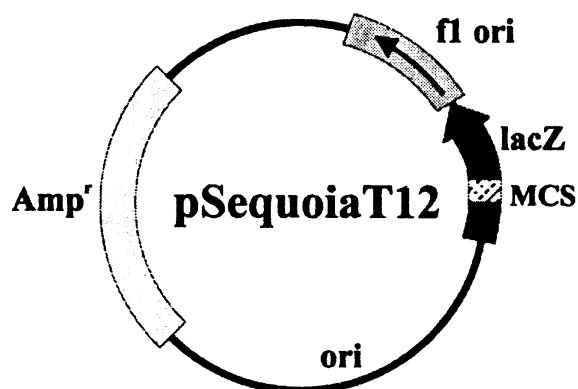


Рис. 7.1. Фагмидный вектор pSequoiaT12 и его полилинкерный участок

ющих поколений полидромных октануклеотидных сайтов узнавания для еще не обнаруженных ферментов, но при этом потенциально существующих. Так, в уже упомянутых фагмидных векторах семейства pBluescript, несмотря на их довольно протяженный полилинкер, имеется всего четыре подобных сайта, включая *NotI*-сайт. Вектор pJRD184 [Heusterspreute et al., 1985] имеет в своей последовательности 43 уникальных сайта рестрикционных эндонуклеаз, но из них в полилинкере локализовано всего 14 и в них только 2 палиндромных октануклеотидных участка для потенциальных рестрикционных эндонуклеаз. Что касается суперполилинкера SL2 вектора pSL301, содержащего все 64 возможных варианта для гексануклеотидных рестрикционных эндонуклеаз, узнающих непрерывные палиндромные последовательности, причем как уже обнаруженных в природе, так и для еще не выявленных [Brosius, 1989], то среди них имеется лишь 9 октануклеотидных палиндромных участка, включая единственный для уже обнаруженного фермента *NotI*. 7 октануклеотидных палиндромных участка содержит в своем полилинкере вектор pNEB193, поставляемый фирмой "New England Biolabs, Inc". (США), из которых 4 доступны для расщепления уже известными ферментами (*AscI*, *PacI*, *PmeI* и *Sse8387I*). Вектор pSequoiaT12, кроме уже упомянутых выше сайтов для октануклеотидных рестрикционных эндонуклеаз *SwaI* (ATTТАААТ) и *SrfI* (GCC-СGGGC), несет еще один подобный сайт для фермента *Sse8387I* (CCTGCAGG), а также еще пять участков для потенциальных октануклеотидных рестрикционных эндонуклеаз (CGAATTCG, CGAGCTCG, CGTTAACG, CGATATCG, CAAGCTTG). В приведенных участках выделены сайты узнавания гексануклеотидных ферментов: *EcoRI*, *SacI*,

*HpaI*, *EcoRV*, *HindIII* соответственно. Конечно, 8 сайтов из потенциальных 256 составляют ничтожную долю, но при составлении новых полилинкеров, видимо, надо принимать в расчет и возможность обнаружения какой-либо октануклеотидной рестрикционной эндонуклеазы. В то же время полилинкер с большим числом сайтов узнавания для потенциальных октануклеотидных рестрикционных эндонуклеаз неприемлем, хотя бы из-за своей длины.

Преимущества, приобретаемые вектором *pSequoiaT12* от удаления в его последовательности сайтов узнавания рестрикционной эндонуклеазы *RsaI* и наличия в полилинкере сайтов узнавания редкощепящих октануклеотидных рестрикционных эндонуклеаз *SwaI* (или ее изошизомера *SmiI*) и *SrfI* детально рассмотрены в разделе, посвященном стратегиям секвенирования протяженных фрагментов ДНК.

## 7.8. РАЗВИТИЕ ВЕКТОРНЫХ СИСТЕМ ДЛЯ КЛОНИРОВАНИЯ И СЕКВЕНИРОВАНИЯ ДНК

На основании изложенных выше черт различных поколений плазмидных векторов для молекулярного клонирования (включая фрагмидные) можно составить весьма упрощенную схему, позволяющую в то же время проследить особенности развития векторных систем общего назначения (рис. 7.2).

Так, векторы для молекулярного клонирования первого поколения представляют собой естественные бактериальные плазмиды так называемого дикого типа, обнаруженные ранее в клетках кишечной палочки *Escherichia coli* как экстрахромосомальные элементы. Плазмида *pSC101*, характеризующаяся большим размером, низкой копийностью и естественным расположением сайтов узнавания всех рестрикционных эндонуклеаз, была одним из таких векторов, который широко использовался в те первые годы молекулярного клонирования. Ее возможности как векторной молекулы были сильно ограничены и поэтому шло конструирование новых плазмид, обладающих целым рядом преимуществ.

Ко второму поколению векторов следует отнести уже искусственно созданные плазмиды, объединившие в себе черты различных природных плазмид дикого типа. Их наиболее ярким представителем явился плазмидный вектор *pBR322*, вобравший в себя такие черты своих предшественников, как репликон *pMB1*, ген устойчивости к тетрациклину  $Tc^R$  и ген устойчивости к ампициллину  $Ap^R$ . За счет ослабленного контроля репликации копийность плазмиды *pBR322* составила уже около 20–50 копий на клетку. Единственное, что осталось неизменным по отношению к первому поколению векторов, как это то же самое естественное расположение сайтов узнавания эндонуклеаз рестрикции. Плазмида *pBR322* оказалась достаточно удобным вектором для молекулярного клонирования и многие ее черты легли в основу третьего поколения плазмидных векторов. Так, от *pBR322* к ним перешли репликон *ColE1* и маркерный ген устойчивости к ампициллину. Размер самой векторной молекулы, впрочем, еще уменьшился, главным образом,



Рис. 7.2. Схема развития плазмидных векторов общего назначения

за счет удаления гена устойчивости к тетрациклину. Копийность на бактериальную клетку возросла и достигла 200 копий. Однако главными новыми чертами этих плазмидных векторов можно считать маркерный ген, кодирующий NH<sub>2</sub>-концевой фрагмент β-галактозидазы *LacZ'*, и встроенный в него искусственно созданный полилинкер, а также то, что в остальной последовательности вектора были удалены сайты узнавания некоторых рестрикционных эндонуклеаз с гексануклеотидными сайтами узнавания.

Следующее поколение плазмидных векторов, первыми представителями которого являются плазмиды pSP64 и pSP65, считать отдельным поколением несколько трудно по причине не очень значительного усовершенствования векторов предыдущего поколения, на основе которых они и были сконструированы. Однако тот факт, что практически все векторы общего назначения следующих поколений также содержат эти небольшие участки промоторных областей бактериофагов SP6, T3 или T7, то их все же можно отнести к четвертому поколению.

Векторы же пятого поколения, сохранив главные черты своих предшественников, весьма сильно отличаются от них, поскольку представляют собой гибридные конструкции, содержащие, кроме плазмидных последовательностей, межгенный участок одноцепочечного нитевидного фага f1, ответственный за репликацию фага. Такие химерные векторы, получившие название фагмид, содержали сразу два участка инициации репликации f1 и ColE1. Одной из первых широко применявшихся фагмид были вектора семейства pEMBL, созданные на основе векторов pUC8/pUC9 и межгенного участка одноцепочечного фага f1.

Шестое поколение векторов, первым представителем которого является вектор pSequoiaT12, отличается от своих предшественников только тем, что в его нуклеотидной последовательности нарушены сайты узнавания для одной или нескольких частощепящих тетра-нуклеотидных рестрикционных эндонуклеаз. Так, первые два поколения плазмидных векторов имели естественное расположение сайтов узнавания как гексануклеотидных, так и тетра-нуклеотидных рестрикционных эндонуклеаз. У всех последующих поколений векторов, начиная с третьего, кроме прочих преимуществ, появился также искусственно созданный полилинкер, в котором были сосредоточены сайты узнавания нескольких гексануклеотидных рестрикционных эндонуклеаз. Однако для его эффективного использования было необходимо нарушить имевшиеся сайты узнавания этих ферментов в остальной последовательности вектора. Поэтому для векторов третьего поколения в отличие от своих предшественников уже не характерно естественное расположение сайтов узнавания некоторых ключевых (полилинкерных) гексануклеотидных рестрикционных эндонуклеаз в самой последовательности вектора. Однако, можно считать, что все ранее созданные векторы предыдущих поколений все же имели весьма существенный недостаток в виде наличия в их векторной последовательности сайтов узнавания для частощепящих рестрикционных эндонуклеаз с тетра-нуклеотидными участками узнавания, ввиду чего было невозможно расщеплять вставку этими ферментами, поскольку это неминуемо привело бы



к разрушению вектора. Уникальной чертой векторов рSequoia является отсутствие в их нуклеотидной последовательности сайтов узнавания для некоторых частощепящих рестрикционных эндонуклеаз с тетра-нуклеотидными участками узнавания. Создание плазмидных векторов с отсутствием в них сайтов тетра-нуклеотидных рестрикционных эндонуклеаз значительно уступает по своей значимости созданию, например, фагидных векторов. Однако то, что данная черта в виде отсутствия сайтов узнавания для отдельных тетра-нуклеотидных рестрикционных эндонуклеаз, несомненно, будет присуща последующим поколениям плазмидных векторов общего назначения, позволяет считать семейство векторов рSequoia новым, шестым поколением плазмидных векторов.

## 7.9. ДРУГИЕ СПЕЦИАЛИЗИРОВАННЫЕ ВЕКТОРЫ ДЛЯ СЕКВЕНИРОВАНИЯ ДНК

В то же время нельзя обойти молчанием ряд специально созданных векторов для секвенирования протяженных фрагментов ДНК, несущих в своем составе различные транспозоны. Так, космидный вектор рAA113X несет транспозон Tn9, позволяющий получать направленные делеции вставки [Ahmed, 1987]. Вектор рAA113M, отличающийся от предыдущего вектора наличием ориджина репликации одноцепочечного фага M13, предназначен для получения одноцепочечных вариантов генерируемых субклонов [Ahmed, 1989]. Недавно был создан весьма совершенный специализированный космидный вектор рAL-F, несущий транспозон IS1, cos-сайты фага лямбда и некоторые другие регуляторные и маркерные участки [Ahmed, Podemski, 1997].

Другой вектор рMM251 также представляет собой весьма сложную конструкцию, в которую встроены регуляторные элементы фага лямбда, инвертированные повторы транспозона Tn3. Механизм получения однонаправленных делеций в данном векторе основан на том, что *in vivo* происходят делеции с фиксированных точек инвертированных повторов до переменных точек соседнего (клонированного) фрагмента ДНК [Sugino, Morita, 1994]. Весьма интересен космидный вектор рDUAL, несущий три гена устойчивости к разным антибиотикам и селективный летальный маркер, а также транспозон Tn1000 [Wang et al., 1993]. Особенностью данного вектора, разработанного для специальной стратегии секвенирования, названной "фабрикой делеций *in vivo*", является возможность одновременного получения двунаправленных делеционных вариантов, отбираемых по устойчивости к разным антибиотикам. Однако все эти векторы имеют два недостатка. Во-первых, их размер достаточно большой, что создает некоторые неудобства и, во-вторых, что гораздо важнее, осуществление этих делеционных процедур требует использования специальных штаммов *E. coli* и весьма сложной схемы их выращивания. Также специального штамма, экспрессирующего фермент рекомбиназу фага лямбда, требует вектор Janus, представляющий собой модифицированный вектор на основе одноцепочечного фага M13 серии mp [Burland et al., 1993]. Отличительной чертой этого вектора является возможность переориентации

вставки *in vivo* и соответственно упаковка разных цепей клонированного фрагмента ДНК. Другой особенностью, связанной с этим вектором Janus™, является то, что его распространением занимается компьютерная фирма “DNASTAR, Inc.” (США), а не какая-нибудь специализирующаяся на продаже молекулярно-биологических продуктов, включая векторные молекулы. Объяснением этому может служить местонахождение лаборатории генетики Висконсинского университета, где этот вектор и был сконструирован, и фирмы “DNASTAR, Inc.”, расположенных в столице штата Висконсин г. Мэдисон.

Таким образом, для использования всех этих векторов с целью получения делеционных вариантов клонированного фрагмента ДНК *in vivo* или упаковки в фаговых частицах противоположной цепи ДНК вставки необходимы специфические штаммы *E. coli*, что служит некоторым препятствием к их широкому использованию.

Весьма интересен целый набор плазмидных векторов plex00-20, разработанных для мультиплексного секвенирования ДНК и представляющих собой 20 самостоятельных векторов, несущих по два различающихся участка для отжига праймеров [Church, Kieffer-Higgins, 1988]. Таким образом, этот набор векторов, используемых одновременно, несет всего 40 вариантов “тэгов”, служащих местами для отжига гибридационных праймеров и для исключения нежелательной вторичной структуры, состоящих из трех нуклеотидов А, С и Т. Впоследствии на основе одноцепочечного вектора M13 и некоторых из этих векторов, последовательность “тэгов” которых не нарушала рамку считывания галактозидазного гена, сконструировано 11 новых M13plex векторов, позволяющих проводить ферментативное мультиплексное секвенирование с одноцепочечными матрицами ДНК [Heller et al., 1991].

Один из ранних векторов pKH47, сконструированный на основе плазмиды pBR322, содержал поли(dA):поли(dT) дуплексный фрагмент протяженностью около 100 пн, что позволило осуществлять успешное разделение цепей клонированной вставки вместе с векторными последовательностями, одна из которых несла поли(dA)- или поли(dT)-участки на колонках с олиго(dT)- и олиго(dA)-целлюлозами соответственно [Hayashi, 1980]. Полученные таким образом отдельные цепи клонированного фрагмента ДНК были пригодны в качестве матриц для их секвенирования ферментативным методом.

Сконструированные производные фага M13 M13-100, M13-101 и M13-102 несли специальный участок гена X, помещенный под промотор фага T7 [Guilfoyle, Smith, 1994; Chen et al., 1996; Johnson et al., 1996]. Сверхпродукция данного белка в штаммах *E. coli*, продуцирующих фермент T7 РНК-полимеразу, приводила к репрессии репликации фага, тогда как инсерция в него клонируемого фрагмента ДНК нарушала рамку считывания и позволяла осуществлять прямой отбор рекомбинантов. Вектор M13-102 отличался от первого вектора данного семейства M13-100 20-нуклеотидным полипуриновым участком для аффинного захвата векторной ДНК и наличием места отжига для универсального секвенирующего праймера, что сделало его более удобным для экспериментаторов [Chen et al., 1996].

## 7.10. ВЕКТОРЫ ДЛЯ КЛОНИРОВАНИЯ ПЦР-ФРАГМЕНТОВ ДНК

Появление ПЦР и широкое использование этого процесса для работки и дальнейшего изучения интересующих исследователей фрагментов ДНК, в том числе, с помощью молекулярного клонирования и секвенирования этих фрагментов, потребовало создания специализированных векторов. Хотя эти векторы и не предназначены непосредственно для секвенирования клонированных продуктов амплификации, но тем не менее они позволяют это сделать и их выделение в отдельный раздел данной главы достаточно оправданно. Обычный подход, рассчитанный на лигирование и клонирование фрагментов ДНК с тупыми концами, показал свою крайне низкую эффективность при работе с ПЦР-фрагментами. Причиной этого является наличие у амплификатов выступающих нематричных нуклеотидов на 3'-концах, образующихся за счет нематричной активности Taq ДНК-полимеразы, присоединяющей по одному дополнительному нуклеотиду, которым преимущественно является остаток аденина [Clark, 1988].

Большинство таких векторов, получивших из-за выступающих тимидинов на своих 3'-концах обозначение Т-векторов и предназначенных для клонирования ПЦР-фрагментов, приобретают свою специализацию в процессе линеаризации обычного вектора путем некоторых дальнейших превращений, но есть и такие, которые содержат специально сконструированные участки в полилинкере. Таким образом, создание Т-векторов достигается различными способами. Например, использование той же нематричной активности Taq ДНК-полимеразы позволило создать специальный вектор с выступающими на 3'-концах дТМФ, пригодный для клонирования в нем таких амплификатов [Marchuk et al., 1991]. В цитируемой работе фагмидный вектор pBluescript [Stratagene, США] расщеплялся рестрикционной эндонуклеазой *EcoRV*, генерирующей тупые концы, и инкубировался с Taq ДНК-полимеразой в присутствии только одного дТТФ, так что у фермента просто не было выбора в использовании того или иного нуклеотида и он был вынужден добавлять в качестве нематричного нуклеотида менее "любимый" им дТМФ. Было описано получение подобных векторов, но уже с ddT-концами с помощью терминальной трансферазы, правда при последующем лигировании вектора и амплификата образовывались "ники" [Holton, Graham, 1991]. Этой же цели – созданию Т-вектора служили специальные Т-касеты, представляющие собой два комплементарных олигонуклеотид, несущих на одном конце "липкий" конец, пригодный для лигирования с вектором, расщепленным соответствующей рестрикционной эндонуклеазой, а на другом выступающие дТ-нуклеотиды, позволяющие осуществить лигирование с фрагментами ДНК, полученными в ходе ПЦР и несущими соответственно выступающий дАМФ [Iwahana et al., 1994].

Другой способ приготовления вектора, несущего на своих 3'-концах выступающие дТМФ, заключался в его расщеплении уникальной рестрикционной эндонуклеазой *XcmI*, два сайта узнавания которой на-

ходились в специально созданном полилинкере [Kovalic et al., 1991; Harrison et al., 1994]. Уникальность этой рестрикционной эндонуклеазы, имеющей сайт узнавания 5'-CCANNNNN<sup>v</sup>NNNNTGG-3', заключается в том, что после расщепления ДНК образуются концы с одним выступающим нуклеотидом, который теоретически может быть любым, поскольку в центральной части сайта расположены незначащие нуклеотиды. Это дало возможность синтезировать участок полилинкера с отличающимися нуклеотидами в двух сайтах узнавания этого фермента таким образом, что вектор после линейаризации становился специализированным T-вектором для клонирования амплификатов.

Целую линейку специальных векторов (pCR-XL-ТОРО, pCR2.1-ТОРО, pTrcHis2-ТОРО, pBAD-ТОРО, pMT/V5-His-ТОРО, pcDNA3.1/CTFP-ТОРО, pcDNA3.1/NTFP-ТОРО) для клонирования амплификатов поставляет фирма "Invitrogen" (США). Выступающие T-нуклеотиды у этих векторов, имеющих, кроме клонирования продуктов ПЦР, каждый свое предназначение, образуются под действием топоизомеразы I [Shuman, 1994]. Реакция "лигирования", протекающая без ДНК-лигазы, завершается всего за 5 мин и в результате трансформации более 90–95% колоний оказываются рекомбинантными.

## 7.11. О НАИМЕНОВАНИЯХ ВЕКТОРОВ

Что касается вектора шестого поколения, названного нами pSequoiaT12, то тут имеет место своеобразная игра слов. Так, *Sequence* это то, что будет определяться с помощью данного вектора, а *Sequoia* это, как известно, самые высокие деревья на Земле и поскольку данный вектор был разработан для секвенирования крупных фрагментов ДНК, то присвоение ему этого имени выглядит вполне логично. В этой связи невозможно не упомянуть об именах некоторых приведенных выше векторов. Существует несколько подходов к формированию названий векторов для молекулярного клонирования. Строчная буква "p", обычно предшествующая имени вектора, указывает на плазмидную принадлежность вектора, однако она продолжает использоваться и для нынешних фагмид. Многие векторы названы по инициалам нашедших или создавших их исследователей. Так, в названии вектора pSC101 зашифрованы инициалы Стенли Коэна (Stanley Cohen). Широко известные плазмиды серии pBR несут имена их создателей (Bolivar F. и Rodriguez R.). Плазмиды семейства pUR названы так создавшим их Ульрихом Рютером (Ulrich Ruther). Можно с уверенностью предположить, что в названии векторов mICE, сконструированных на основе фаговых векторов семейства mp, зашифрованы инициалы их создателя – I.C. Eperon, хотя и здесь присутствует некоторая игра слов. В названии фаговых векторов серии M13mp лежит скорее географический принцип. M13 обозначает коллекционный номер тринадцатого штамма одноцепочечного фага, выявленный в Мюнхене, а аббревиатура mp обозначает институт Макса Планка, в котором проводились работы по созданию этих векторов. В названии векторов семейства pUC также отражено название органи-

зации, где проводились работы по их созданию – University of California. По аналогичному принципу образовано название фрагмидных векторов pEMBL (European Molecular Biology Laboratory). Названия pBluescript и pBluescribe отражают возможность скрининга рекомбинантных клонов по белой или голубой окраске колоний.

### Цитируемая литература

Чемерис А.В., Ахунув Э.Д., Куликов А.М., Вахитов В.А. Прогрессивная тетра/окта стратегия секвенирования протяженных фрагментов ДНК // Междунар. конф. посвящ. памяти акад. А.А. Баева, М., 1996. С. 134.

Ahmed A. A simple and rapid procedure for sequencing long (40-kb) DNA fragments // Gene. 1987. Vol. 61. P. 363–372.

Ahmed A. A vector for sequencing long (40-kb) DNA fragments // Gene. 1989. Vol. 75. P. 315–321.

Ahmed A., Podemski L. Use of ordered deletions in genome sequencing // Gene. 1997. Vol. 197. P. 367–373.

Altling-Mees M.A., Short J.M. pBluescript II: gene mapping vectors // Nucl. Acids Res. 1989. Vol. 17. P. 9194.

Balbas P., Soberon X., Merino E., Zurita M., Lomeli H.H., Valle F., Flores N., Bolivar F. Plasmid vector pBR322 and its special-purpose derivatives – a review // Gene. 1986. Vol. 50. P. 3–40.

Barnes W.M., Bevan M. Kilo-sequencing: An ordered strategy for rapid DNA sequence data acquisition // Nucl. Acids Res. 1983. Vol. 11. P. 349–368.

Barnes W.M., Bevan M., Son P.H. Kilo-sequencing: Creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13 // Meth. Enzymol. 1983. Vol. 101; Recombinant DNA, pt C. P. 98–122.

Bazarel M., Helinski D.R. Characterization of multiple circular DNA forms of colicinogenic factor E-1 from *Proteus mirabilis* // Biochemistry. 1968. Vol. 7. P. 3513–3517.

Benes V., Hostomsky Z., Arnold L., Paces V. M13 and pUC vectors with new unique restriction sites for cloning // Gene. 1993. Vol. 130. P. 151–152.

Bolivar F. Construction and characterization of new vehicles. 3. Derivatives of the plasmid pBR322 carrying unique *EcoRI* sites for selection of *EcoRI* generated recombinant DNA molecules // Gene. 1978. Vol. 4. P. 121–136.

Bolivar F., Rodriguez R., Betlach M., Boyer H.W. Construction and characterization of new vehicles. 1. Ampicillin-resistant derivatives of the plasmid pMB9 // Gene. 1977a. Vol. 2. P. 75–93.

Bolivar F., Rodriguez R., Greene P.J., Betlach M.C., Heineker H.L., Boyer H.W. Construction and characterization of new vehicles. 2. A multipurpose cloning system // Gene. 1977b. Vol. 2. P. 95–113.

Brosius J. Superpolylinkers in cloning and expression vectors // DNA. 1989. Vol. 8. P. 759–777.

Burland V., Daniels D.L., Plunkett G., III, Blattner F.R. Genome sequencing on both strands: the Janus strategy // Nucl. Acids Res. 1993. Vol. 21. P. 3385–3390.

Chambers S.P., Prior S.E., Barstow D.A., Minton N.P. The pMTL *nic*<sup>-</sup> cloning vectors. 1. Improved pUC polyninker regions to facilitate the use of sonicated DNA for nucleotide sequencing // Gene. 1988. Vol. 68. P. 139–149.

Chen D., Johnson A.F., Severin J.M., Rank D.R., Smith L.M., Guilfoyle R.A. M13-102: A vector for facilitating construction and improving quality of M13 shotgun libraries // Gene. 1996. Vol. 172. P. 53–57.

Church G.M., Kieffer-Higgins S. Multiplex DNA sequencing // Science. 1988. Vol. 240. P. 185–188.

Clark J.M. Novel non-template nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases // Nucl. Acids Res. 1988. Vol. 16. P. 9677–9686.

Clewell D.B. Nature of ColE<sub>1</sub> plasmid replication in *Escherichia coli* in the presence of chloramphenicol // J. Bacteriol. 1972. Vol. 110. P. 667–676.

Clewell D.B., Helinski D.R. Supercoiled circular DNA-protein complex in *E. coli*: Purification and induced conversion to an open circular DNA form // Proc. Nat. Acad. Sci. USA. 1969. Vol. 62. P. 1159–1166.

- Cohen S.N., Chang A.C.Y. Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants // *Ibid.* 1973. Vol. 70. P. 1293–1297.
- Cohen S.N., Chang A.C.Y., Boyer H.W., Helling R.B. Construction of biological functional bacterial plasmids in vitro // *Ibid.* 1973. Vol. 70. P. 3240–3244.
- Dente L., Cesareni G., Cortese R. pEMBL: a new family of single stranded plasmids // *Nucl. Acids Res.* 1984. Vol. 11. P. 1645–1655.
- Dotto G.P., Enea V., Zinder N.D. Functional analysis of bacteriophage  $\phi$ 1 intergenic region // *Virology.* 1981. Vol. 114. P. 463–473.
- Eperon I.C. M13 vectors with T7 polymerase promoters: Transcription limited by oligonucleotides // *Nucl. Acids Res.* 1986. Vol. 14. P. 2830.
- Frischauf A.M., Garoff H., Lehrach H. A subcloning strategy for DNA sequence analysis // *Ibid.* 1980. Vol. 8. P. 5541–5549.
- Furuyama H., Okano K., Kambara H. DNA sequencing directly from a mixture using terminal-base-selective primers // *DNA Res.* 1994. Vol. 1. P. 231–237.
- Gronenborn B., Messing J. Methylation of single-stranded DNA in vitro introduces new restriction endonuclease cleavage sites // *Nature.* 1978. Vol. 272. P. 375–377.
- Guilfoyle R.A., Smitn L.M. A direct selection strategy for shotgun cloning and sequencing in the bacteriophage M13 // *Nucl. Acids Res.* Vol. 22. P. 100–107.
- Hayashi K. A cloning vehicle suitable for strand separation // *Gene.* 1980. Vol. 11. P. 109–115.
- Heller C., Radley E., Khurshid F.A., Beck S. M13plex vectors for multiplex DNA sequencing // *Gene.* 1991. Vol. 103. P. 131–132.
- Heusterspreute M., Davison J. A method for the generation of small pre-determined deletions in plasmid DNA: Deletion analysis of the tetR region of vector pBR322 // *Gene.* 1983. Vol. 23. P. 35–40.
- Heusterspreute M., Thi V.H., Emery S., Tournis-Gamble S., Kennedy N., Davison J. Vectors with restriction site banks. 4. pJRD184, a 3793-bp plasmid vector having 43 unique cloning sites // *Gene.* 1985. Vol. 39. P. 299–304.
- Holton T.A., Graham M.W. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors // *Nucl. Acids Res.* 1991. Vol. 19. P. 1156.
- Iwahana H., Yoshimoto K., Tsujisawa T., Itakura M. T-cassette ligation: A method for direct sequencing and cloning of PCR-amplified DNA fragments // *PCR Meth. Appl.* 1994. Vol. 3. P. 219–224.
- Jendrisak J.J., Lewis M.K., Fiandt M.J. Vectors for *in vitro* production of RNA copies of either strand of a cloned DNA sequence. Pat. 4766072. (US). Publ. 1988.
- Johnson A.F., Wang R., Ji H., Chen D. Purification of single-stranded M13 DNA by cooperative triple-helix-mediated affinity capture // *Anal. Biochem.* 1996. Vol. 234. P. 83–95.
- Jupin I., Gronenborn B. Abundant, easy and reproducible production of single-stranded DNA from phagemids using helper phage-infected competent cells // *Nucl. Acids Res.* 1995. Vol. 23. P. 535–536.
- Konings R.N.H., Luiten R.G.M., Reeters B.P.H. Mike, a chimeric filamentous phage designed for the separate production of either DNA strand of pKUN vector plasmids by F<sup>+</sup> cells // *Gene.* 1986. Vol. 46. P. 269–276.
- Kovalic D., Kwak J.-H., Weisblum B. General method for direct cloning of DNA fragments generated by the polymerase chain reaction // *Nucl. Acids Res.* 1991. Vol. 19. P. 4560.
- Marchuk D., Drumm M., Saulino A., Collins F.S. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products // *Ibid.* 1991. Vol. 19. P. 1154.
- Melton D.A., Krieg P.A., Rebagliati M.R., Maniatis T., Zinn K., Green M.R. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter // *Ibid.* 1984. Vol. 12. P. 7035–7056.
- Messing J. New M13 vectors for cloning // *Meth. Enzymol.* 1983. Vol. 101: Recombinant DNA, pt. C. P. 20–78.
- Messing J., Crea R., Seeburg P.H. A system for shotgun DNA sequencing // *Nucl. Acids Res.* 1981. Vol. 9. P. 309–321.
- Messing J., Gronenborn B., Muller-Hill B., Hofschneider P. Filamentous coliphage M13 as a cloning vehicle: Insertion of a *Hind*II fragment of the *lac* regulatory region in M13 replicative form *in vitro* // *Proc. Nat. Acad. Sci. USA.* 1977. Vol. 74. P. 3642–3646.

- Messing J., Viera J.* A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments // *Gene*. 1982. Vol. 19. P. 269–276.
- Norrander J., Kempe T., Messing J.* Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis // *Gene*. 1983. Vol. 26. P. 101–106.
- Okano K., Kambara H.* Fragment walking for long DNA sequencing by using a library as small as 16 primers // *Gene*. 1996. Vol. 176. P. 231–235.
- Peden K.W.C.* Revised sequence of the tetracycline-resistance gene of pBR322 // *Gene*. 1983. Vol. 22. P. 277–280.
- Ruther U.* pUR250 allows rapid chemical sequencing of both DNA strands of its inserts // *Nucl. Acids Res.* 1982. Vol. 10. P. 5765–5772.
- Ruther U., Koenen M., Otto K., Muller-Hill B.* pUR222, a vector for cloning and rapid chemical sequencing of DNA // *Ibid.* 1981. Vol. 9. P. 4087–4098.
- Shuman S.* Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase // *J. Biol. Chem.* 1994. Vol. 269. P. 32678–326784.
- Sugino Y., Morita M.* A new DNA cloning/sequencing vector with a built-in mechanism for generation of nested deletions using transposon Tn3 // *Gene*. 1994. Vol. 148. P. 169–170.
- Sutcliffe J.G.* Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322 // *Cold Spring Harbor Symp. Quant. Biol.* 1978. Vol. 43. P. 77–90.
- Viera J., Messing J.* The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers // *Gene*. 1982. Vol. 19. P. 259–268.
- Wang G., Blakesley R.W., Berg D.E., Berg C.M.* pDUAL: A transposon-based cosmid cloning vector for generating nested deletions and DNA sequencing templates *in vivo* // *Proc. Nat. Acad. Sci. USA*. 1993. Vol. 90. P. 7874–7878.
- Watson N.* A new revision of the sequence of plasmid pBR322 // *Gene*. 1988. Vol. 70. P. 399–403.
- Yanisch-Perron C., Viera J., Messing J.* Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors // *Gene*. 1985. Vol. 33. P. 103–119.

### *Дополнительная литература*

- Balbas P., Soberon X., Merino E., Zurita M., Lomeli H.H., Valle F., Flores N., Bolivar F.* Plasmid vector pBR322 and its special-purpose derivatives – a review // *Gene*. 1986. Vol. 50. P. 3–40.
- Messing J.* New M13 vectors for cloning // *Meth. Enzymol.* 1983. Vol. 101: Recombinant DNA, pt. C. P. 20–78.
- Messing J.* Cloning in M13 phage or how to use biology at its best // *Gene*. 1991. Vol. 100. P. 3–12.

## СТРАТЕГИИ СЕКВЕНИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК

Задачи, стоящие перед исследователями, требовали определения нуклеотидных последовательностей довольно протяженных фрагментов ДНК, которые не могли быть “прочитаны” за одно электрофоретическое разделение продуктов секвенирующих реакций. В первые годы секвенирования ДНК предел “чтения” с радиоавтографа геля редко превышал 250–300 нуклеотидов для одной матрицы. Да и фермент ДНК-полимераза I, ее Кленовский фрагмент, имея низкую процессивность, не мог эффективно строить комплементарную цепь ДНК большой протяженности. Метод химической дегградации вкупе со стандартным гель-электрофорезом того времени также не позволял “читать” более 250–300 нуклеотидов. В связи с этим решение проблемы секвенирования больших вставок, нуклеотидные последовательности которых не могли быть определены в одном эксперименте, заключалось в выработке специальных подходов. С течением времени, благодаря различным новшествам, разрешающая способность методов секвенирования заметно возросла и за счет этого увеличился оптимальный размер вставок до 500 и 1000 пн, позволяющий определить их последовательность в процессе одного электрофоретического разделения. Резко возросли и задачи, предполагающие уже определение нуклеотидных последовательностей еще более крупных фрагментов ДНК или даже целых геномов. Все это привело к разработке большого числа различных стратегий секвенирования протяженных фрагментов ДНК, к рассмотрению которых мы и приступаем.

### 8.1. СТРАТЕГИИ СЕКВЕНИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК МЕТОДОМ ХИМИЧЕСКОЙ ДЕГРАДАЦИИ ПО МАКСАМУ–ГИЛБЕРТУ

Секвенирование больших фрагментов ДНК методом химической дегградации по Максаму–Гилберту было возможно путем построения подробной рестриктазной карты клонированного фрагмента ДНК, разработки на ее основе стратегии секвенирования, заключающейся в выборе тех или иных рестриктазных фрагментов исследуемого клона, полностью перекрывающих всю длину вставки (рис. 8.1). Далее следовал этап препаративного расщепления ДНК соответствующими рестрикционными эндонуклеазами, препаративный гель-электрофорез, концевое мечение элюированных фрагментов и проведение реакций химической модификации и гидролиза. Следует отметить, что данный подход к секвенированию крупных вставок методом Максама–Гилберта вполне может быть применен и в настоящее время, однако разработ-



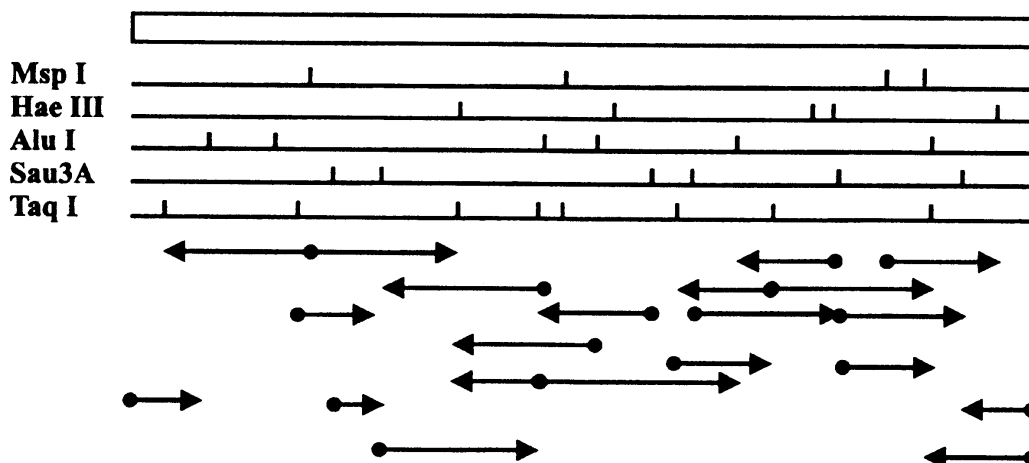


Рис. 8.1. Стратегия секвенирования протяженного фрагмента ДНК методом химической дегградации, основанная на знании детальной рестриктазной карты

ка новых векторов для молекулярного клонирования, подробно рассмотренных в предыдущей главе, оказала заметное влияние на подготовку образцов для секвенирования ДНК этим методом.

Подход с элюцией из геля фрагментов ДНК, предназначенных для секвенирования, все же был не очень производительным и в этой связи предпринимались попытки получения однонаправленных делеционных вариантов клонированного фрагмента ДНК в исходной плазмиде без этапа переклонирования в новом векторе, поскольку данный подход сулил ряд преимуществ. Так, на основе плазмиды рBR322, несущей вставку по сайту рестрикционной эндонуклеазы *Pst*I, после обработки ДНКазой в присутствии ионов  $Mn^{2+}$ , “пришивания” *Eco*RI-линкеров, расщепления рестрикционной эндонуклеазой *Eco*RI и лигирования был получен ряд субклонов с однонаправленным укорочением вставки в направлении против часовой стрелки от сайта *Eco*RI к сайту *Pst*I [Frischauf et al., 1980]. Следует отметить, что те конструкции, у которых разрыв цепей ДНК под действием ДНКазы/ $Mn^{2+}$  произошел в самой векторной последовательности, а не во вставке, оказывались нежизнеспособными и на рис. 8.2., чтобы его не загромождать, не показаны. После повторного расщепления рестрикционной эндонуклеазой *Eco*RI и радиоактивного мечения полученных концов для каждого субклона была получена линеаризованная конструкция с обоими мечеными концами, образованными рестрикционной эндонуклеазой *Eco*RI. Поскольку для секвенирования необходимы фрагменты, меченные по одному концу, второй конец, несущий метку, был удален расщеплением рестрикционными эндонуклеазами *Hind*III или *Cla*I, сайты узнавания которых расположены на расстоянии 29 и 23 пн от сайта рестрикционной эндонуклеазы *Eco*RI соответственно. Таким образом, в результате этих процедур нуклеотидные последовательности этих первых 28 или 22 пн из-за наложения полос могли не всегда достоверно читаться, а сами двадцать девятые или двадцать третьи нуклеотиды не могли быть достоверно прочитанными совсем. Однако главным достижением явилась показанная возможность получения набора последовательно укороченных субклонов.

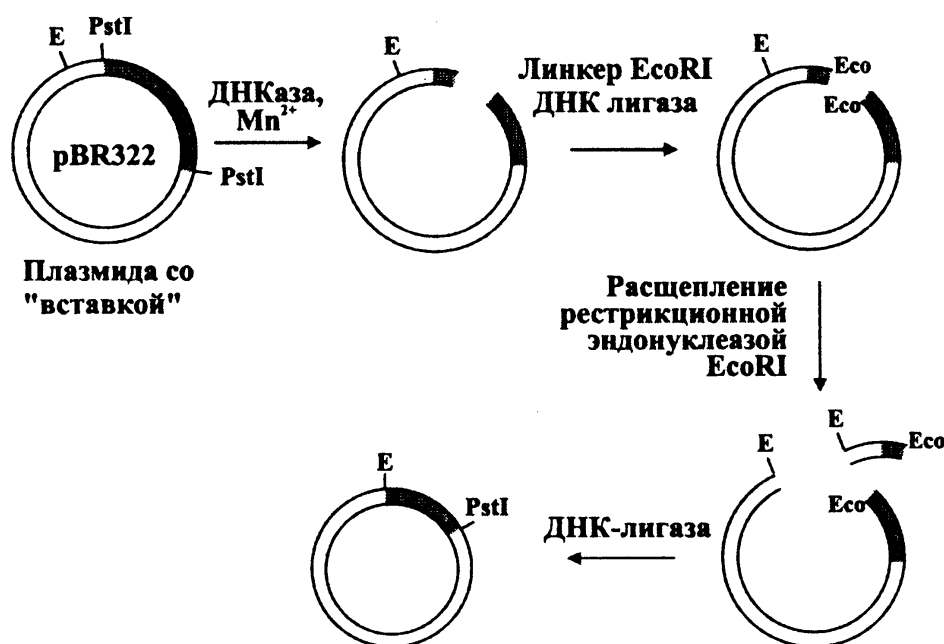


Рис. 8.2. Схема получения однонаправленных делеционных субклонов в векторе pBR322 для их секвенирования методом химической деградации

Подобный подход с гораздо большим успехом стали применять с векторами, имеющими специально созданный полилинкер. Одним из таких векторов является плазмиды рUR222, специально сконструированная для секвенирования обеих цепей ДНК вставки, не требуя ее переклонирования [Ruther et al., 1981]. За счет полилинкера, в котором имелось шесть уникальных сайтов узнавания различных эндонуклеаз рестрикции, в этом векторе появилась расширенная возможность получения коротких субклонов с помощью частощепящих рестрикционных эндонуклеаз. Следующий вектор этого семейства рUR250 имел в своем полилинкере сайты узнавания уже для семи ферментов [Ruther, 1982]. Другое преимущество этих или каких-либо других векторов, несущих полилинкер, заключалось в возможности расщепления вектора в участке полилинкера двумя соответствующими рестрикционными эндонуклеазами, сайты которых расположены по одну и ту же сторону от места клонирования вставки, что приводило к образованию двух рестриктазных фрагментов, сильно отличающихся по размеру. Один из фрагментов, предназначенный для секвенирования, включал в свой состав всю вставку или ее делеционные варианты и почти всю плазмиду, второй же представлял собой небольшой фрагмент полилинкера. После мечения концов данных фрагментов ДНК с помощью Кленовского фрагмента ДНК-полимеразы I они могли сразу быть подвержены реакциям химической деградации без дополнительных стадий очистки (кроме, разве, что, осаждения этанолом).

Эта стратегия секвенирования ДНК методом химической деградации субклонированных фрагментов в векторах, содержащих полилинкерные последовательности, была довольно производительной в сравнении с бесполилинкерным вектором pBR322 и тем более по сравнению с применявшимся ранее подходом с элюцией рестриктазных фрагмен-

тов из геля. Поскольку метод химической дегградации требовал довольно больших количеств ДНК, а элюция из полиакриламидного или агарозного гелей не очень эффективна, то разработка многочисленных векторов, несущих сайты для клонирования по многим рестрикционным эндонуклеазам, позволила, основываясь все на том же рестриктазном картировании, расщеплять вставку подходящими ферментами и получать необходимые субклоны в подобном векторе и затем их секвенировать в составе этого же вектора, получая меченные по одному концу фрагменты ДНК, как описано выше.

Другим преимуществом векторов, несущих полилинкерный участок, была довольно легкая возможность проводить дельта-рестрикционное субклонирование, также основанное на знании детальной рестриктазной карты. Сущность этого метода заключалась в использовании для укорочения вставки полилинкерных редкощепящих рестрикционных эндонуклеаз, сайты которых присутствовали и во вставке. Кроме этого, можно было применять с этой же целью ферменты, чьи сайты имелись во вставке, но отсутствовали в векторе. При этом сам вектор расщеплялся соответствующей полилинкерной рестрикционной эндонуклеазой, желательной образующей у линеаризуемого вектора со вставкой концы, совпадающие с таковыми при расщеплении вставки другим ферментом, что требовалось для облегчения самого этапа лигирования или упрощения подготовки к нему.

Усовершенствование плазмидных векторов для секвенирования ДНК методом химической дегградации позволило создать еще более удобные специальные полилинкеры, содержащие в своей последовательности два участка расщепления рестрикционной эндонуклеазы *Tth1111*, имеющей сайт узнавания GACN<sup>v</sup>NNGTC [Eckert, 1987; Volckaert, 1987]. Удобство этих векторов заключалось в возможности получения с помощью данного фермента фрагментов ДНК, которые потом легко пометить только по одному концу, что как раз требуется для метода секвенирования ДНК по Максаму–Гилберту. Сайты полилинкерных рестрикционных эндонуклеаз, расположенные между фланкирующими сайтами рестрикционной эндонуклеазы *Tth1111*, позволяли, если требовалось, проводить дельта-рестрикционное субклонирование.

Все рассмотренные выше стратегии секвенирования протяженных фрагментов ДНК методом Максама–Гилберта нельзя отнести к стратегиям случайного подхода, так как они основывались, как правило, на знании рестриктазной карты клонированного фрагмента ДНК и были рассчитаны на получение конкретных субфрагментов ДНК или субклонов. В то же время возможность случайного раздробления крупного фрагмента ДНК физическими методами и получение на их основе субклонов для их последующего секвенирования методом химической дегградации принципиально возможны, хотя и не производительны. При этом, однако, необходимо отметить, что совершенствование векторных систем и их использование для определения нуклеотидных последовательностей протяженных фрагментов ДНК химической дегградацией по Максаму–Гилберту не смогло значительно повысить производительность этого метода, разве что, кроме, метода мультиплексного секвенирования.

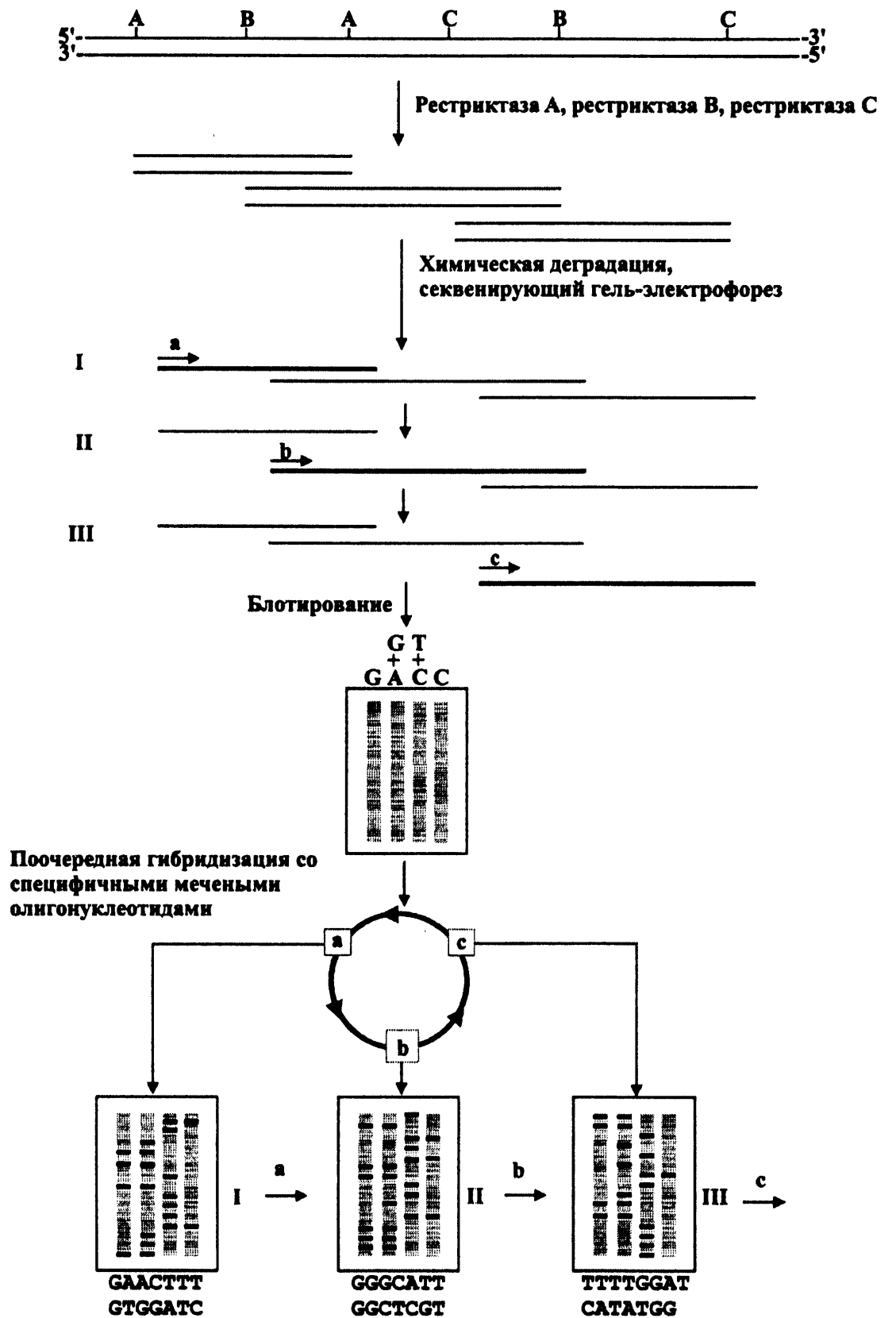


Рис. 8.3. Схема определения нуклеотидной последовательности путем мультиплексной геномной "прогулки"  
а, b, с – специфичные олигонуклеотиды

Первоначально разработанный для секвенирования ДНК по Максиму–Гилберту метод мультиплексного секвенирования ДНК, основанный на одновременном использовании целого набора плазмидных векторов plex00-20, позволял надеяться на резкое увеличение производительности этого метода секвенирования ДНК [Church, Kieffer-Higgins, 1988]. Так, двадцать независимо клонированных фрагментов ДНК, полученных в результате ультразвуковой обработки исходного фрагмента ДНК, в двадцати векторах plex после этапов совместного расщепления соответствующими рестрикционными эндонуклеазами и совместного проведения реакций химической дегградации разделяли в одном секвенирующем геле и переносили на нейлоновый фильтр. Лестницы полос из всех двадцати клонов выявляли поочередной гибридизацией с мечеными праймерами, гомологичными каждый раз только для одного такого вектора из набора. Однако, несмотря на один общий этап химической дегградации, позволяющий экономить время, метод мультиплексного секвенирования достаточно сложен, поскольку для дальнейшего выявления нуклеотидных последовательностей всех клонированных фрагментов ДНК необходим трудоемкий этап блот-гибридизации, удаление связанной метки и следующий этап гибридизации с другим праймером и так последовательно со всеми праймерами, гомологичными соответствующим участкам разных плазмид из набора plex00=20. В то же время нельзя не упомянуть о том, что значительная часть нуклеотидной последовательности полного генома археобактерии *Methanobacterium thermoautotrophicum*, составляющего более 1750 тпн, определена именно с помощью мультиплексного секвенирования ДНК путем химической дегградации [Smith et al., 1997]. Этим же подходом было секвенировано около трети всего генома *Mycoplasma carnicolum*, составляющего более 750 тпн и характеризующегося высоким (75%) содержанием АТ-пар [Dolan et al., 1995].

Ранее схожий метод мультиплексной геномной “прогулки” был применен для определения нуклеотидной последовательности неклонированного гена пируваткиназы *E. coli* [Ohara et al., 1989]. Суть такой “прогулки”, весьма напоминающей праймерную (описанную ниже в разделе 8.4), заключается в расщеплении тотальной ДНК несколькими подходящими рестрикционными эндонуклеазами, разделении полученных фрагментов в денатурирующем секвенирующем гель-электрофореze, их переносе на нейлоновый фильтр и многочисленных гибридизациях с мечеными олигонуклеотидами (рис. 8.3).

Первым таким меченым зондом служит синтезированный олигонуклеотид, комплементарный участку секвенируемой ДНК с известной последовательностью. Следующий, очередной зонд подбирают на основе уже определенной в предыдущем цикле нуклеотидной последовательности. Причем синтез такого олигонуклеотида, комплементарного той или иной цепи, позволяет осуществлять как секвенирование нового участка, так и подтверждающее секвенирование в обратном направлении.

Следует отметить, что метод мультиплексного секвенирования получил дальнейшее развитие при определении нуклеотидных последовательностей ДНК ферментативным методом и поэтому в соответствующем разделе 8.6 данной главы рассмотрен более подробно.

## 8.2. СТРАТЕГИИ СЕКВЕНИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ ПО СЭНГЕРУ С ПОМОЩЬЮ СЛУЧАЙНОГО ПОДХОДА

Что касается ферментативного секвенирования ДНК по Сэнгеру, то этот метод изменился за эти годы весьма сильно и разработанные стратегии секвенирования протяженных фрагментов ДНК оказали заметное влияние на его общую производительность. Большинство существующих стратегий секвенирования протяженных фрагментов ДНК ферментативным методом основаны на получении серии субклонов. Различают два главных подхода к их получению, один из которых считается случайным, а другой направленным. В основе стратегий случайного подхода лежит произвольная фрагментация ранее клонированного фрагмента ДНК и дальнейшее получение субклонов с небольшими вставками в другом векторе (а не в исходном), причем их размер, как правило, позволяет осуществить определение нуклеотидной последовательности всего субклонированного фрагмента за один эксперимент без дополнительного этапа субклонирования. Стратегии направленного подхода предполагают создание серии субклонов обычно в составе исходного вектора, различающихся размером делетированных участков вставки между сайтом какой-либо подходящей полилинкерной рестрикционной эндонуклеазы, предшествующим сайту клонирования (со стороны отжига секвенирующего праймера), и прогрессивно удаляющимися местами вставки. И тот и другой подходы имеют как свои преимущества, так и недостатки, к рассмотрению которых мы и перейдем.

Отсутствие необходимости предварительного рестриктазного картирования протяженной вставки является важным преимуществом стратегий случайного подхода и позволяет экономить много времени и усилий на данном этапе, а также и на этапе лигирования, который проводится только один раз со всем множеством получаемых мелких фрагментов ДНК. Поскольку для получения серии подобных субклонов фрагментация ДНК осуществляется произвольным образом или с помощью ферментов, или физическим воздействием, то знания рестриктазной карты совершенно не требуется. Однако клонирование таких случайных фрагментов приводило, подчас, к многократному секвенированию одних и тех же участков ДНК, что имело, впрочем, некоторое положительное значение, увеличивая достоверность нуклеотидной последовательности ДНК ввиду ее многократного «прочтения», да еще и в обоих направлениях. В то же время отдельные фрагменты иногда никак не удавалось субклонировать и секвенировать.

Проведенные статистические расчеты показали, что для определения нуклеотидной последовательности всего крупного фрагмента ДНК случайным подходом необходимо секвенировать в 7–10 раз больше нуклеотидов, чем сам секвенируемый фрагмент. Причем, чем крупнее фрагмент ДНК, тем труднее завершить проект по его секвенированию этим подходом, поскольку вероятность получения требуемых субкло-

нов на заключительной стадии в этом случае заметно снижается. Поэтому в литературе содержались рекомендации по первоначальному получению двух, трех или большего числа субклонов I порядка (в зависимости от общей протяженности исходного фрагмента ДНК) с использованием подходящих рестрикционных эндонуклеаз. Размер таких субклонов I порядка, как правило, не должен был превышать 5 тпн и их расположение друг относительно друга в исходном полноразмерном фрагменте ДНК должно быть известно. Таким образом, при подобном секвенировании крупных фрагментов ДНК случайным подходом уже прослеживался некий элемент направленности. Следует также отметить, что в целях экономии на определенном этапе выполнения проекта по секвенированию протяженного фрагмента ДНК рекомендовалось проведение реакций только с одним терминирующим нуклеотидом ddT, так называемый T-трейнинг, позволяющий выявить уже ранее секвенированные матрицы сопоставлением полос в T-дорожке и исключить, таким образом, их повторное секвенирование. С этой целью была разработана специальная компьютерная программа TRACTOR, позволяющая выявлять такие клоны и уменьшать за счет этого объем экспериментальной работы на заключительном этапе секвенирования [Niemi, Mantsala, 1994].

Обязательным важным этапом при определении нуклеотидной последовательности ДНК случайным подходом было последующее правильное расположение относительно друг друга секвенированных участков ДНК. Большое количество беспорядочно секвенируемых образцов, которые несли цепи ДНК в противоположных ориентациях, требовало разработки специальных компьютерных программ, позволяющих осуществлять такую стыковку как в ручном, так и в автоматическом режимах. Следует отметить, что особенно серьезные трудности возникали при состыковке секвенированных фрагментов ДНК, содержащих многочисленные повторяющиеся элементы.

Система для секвенирования ДНК методом “дробовика”, основанная на разработанном одноцепочечном векторе, несущем удобный полилинкер, вызвала к жизни способ получения небольших фрагментов ДНК для их последующего клонирования с помощью частощепящих рестрикционных эндонуклеаз с тетра-нуклеотидными сайтами узнавания (рис. 8.4) [Messing et al., 1981]. Преимуществом этого способа можно считать относительную легкость этапа лигирования при условии использования рестрикционных эндонуклеаз, образующих фрагменты ДНК с “липкими” концами. Что касается лигирования рестрикционных фрагментов ДНК, несущих тупые концы, то и в этом случае не требовались дополнительные этапы, предшествующие лигированию, в отличие от других методов получения коротких фрагментов ДНК, рассматриваемых ниже. Главный недостаток данного подхода заключается в том, что сайты используемых рестрикционных эндонуклеаз могут быть расположены не очень равномерно и, таким образом, существует вероятность получить какой-нибудь субклон, несущий вставку довольно крупного размера, не позволяющую осуществить ее “прочтение” за один эксперимент.

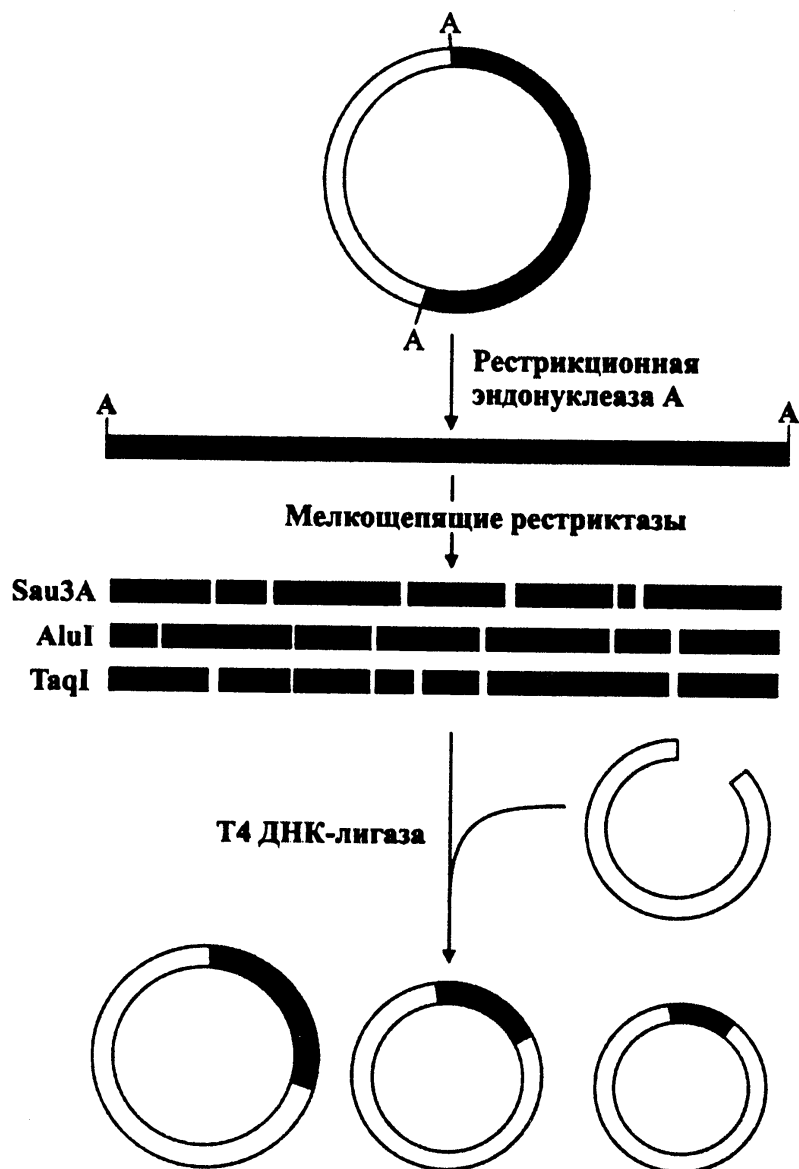


Рис. 8.4. Стратегия получения субклонов протяженного клонированного фрагмента ДНК с помощью частощепящих рестрикционных эндонуклеаз

Другой подход к получению случайных фрагментов ДНК для их последующего клонирования был осуществлен при помощи расщепления вставки ДНКазой I в присутствии ионов  $Mn^{2+}$  (рис. 8.5) [Anderson, 1981]. Известно, что при таких условиях ДНКаза образует разрывы обеих цепей ДНК, причем эти разрывы обычно расположены недалеко друг от друга. Применение этого фермента позволяло получать действительно случайный набор фрагментов ДНК, практически независимый от нуклеотидной последовательности. Таким образом, проведя подбор условий расщепления ДНКазой/ $Mn^{2+}$ , а затем, расфракционировав полученные фрагменты по размеру или ультрацентрифугированием в сахарозном градиенте или же гель-электрофорезом, можно было получить достаточно однородные фрагменты ДНК для создания на их основе субклонов. Однако концы таких фрагментов не были пригодны к лигированию, и требовалось их затупление, осуществляемое обычно с помощью Т4 ДНК-полимеразы.



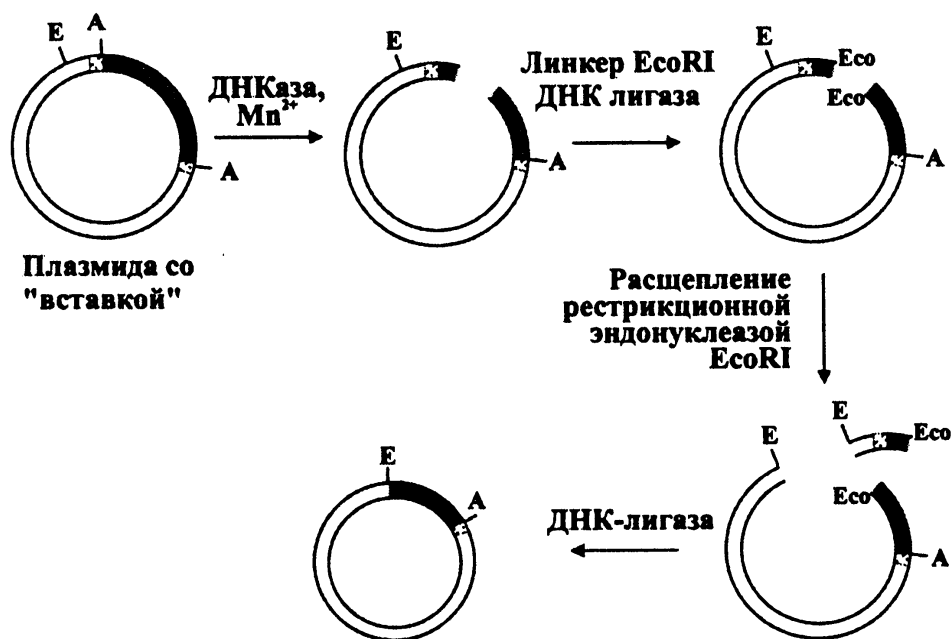


Рис. 8.5. Стратегия получения субклонов протяженного клонированного фрагмента ДНК с помощью расщепления вставки ДНКазой в присутствии ионов  $Mn^{2+}$

Приведенные выше способы получения мелких фрагментов ДНК осуществляются с помощью ферментов, однако применение для этой цели физического воздействия в виде ультразвуковой обработки раствора ДНК или его распыления через соответствующее сопло с помощью газообразного азота (небулизация) также возможны [Deiningger, 1983a, b; Pan et al., 1994]. Механическое воздействие на ДНК крупного размера, приводящее к ее фрагментации, возможно путем продавливания раствора ДНК через тонкую иглу G-30 с диаметром отверстия 0,16 мм [Ansorge et al., 1986]. Для этой же цели применялся специально изготовленный миниатюрный Французский пресс [Schriefer et al., 1990].

Обработка фрагмента ДНК ультразвуком с целью его фрагментации, применяемая, пожалуй, чаще других воздействий, лишь незначительно зависит от особенностей его нуклеотидной последовательности, хотя и считается, что АТ-богатые участки ДНК расщепляются немного чаще, чем GC-богатые. Размер получаемых под действием ультразвуковой обработки фрагментов ДНК зависит от продолжительности озвучивания, его амплитуды. Полученные таким образом фрагменты ДНК также нуждаются в репарации своих концов, протекающей достаточно эффективно, с целью их затупления и пригодности для клонирования в специально приготовленном векторе (рис. 8.6). Эффективное секвенирование субклонированных фрагментов предполагает предварительное обогащение пула "озвученных" молекул ДНК путем ультрацентрифугирования или гель-электрофореза.

Следует отметить, что все описанные выше способы получения субклонов требуют предварительного выделения вставки, так как в противном случае часть субклонов будет нести фрагменты не вставки,

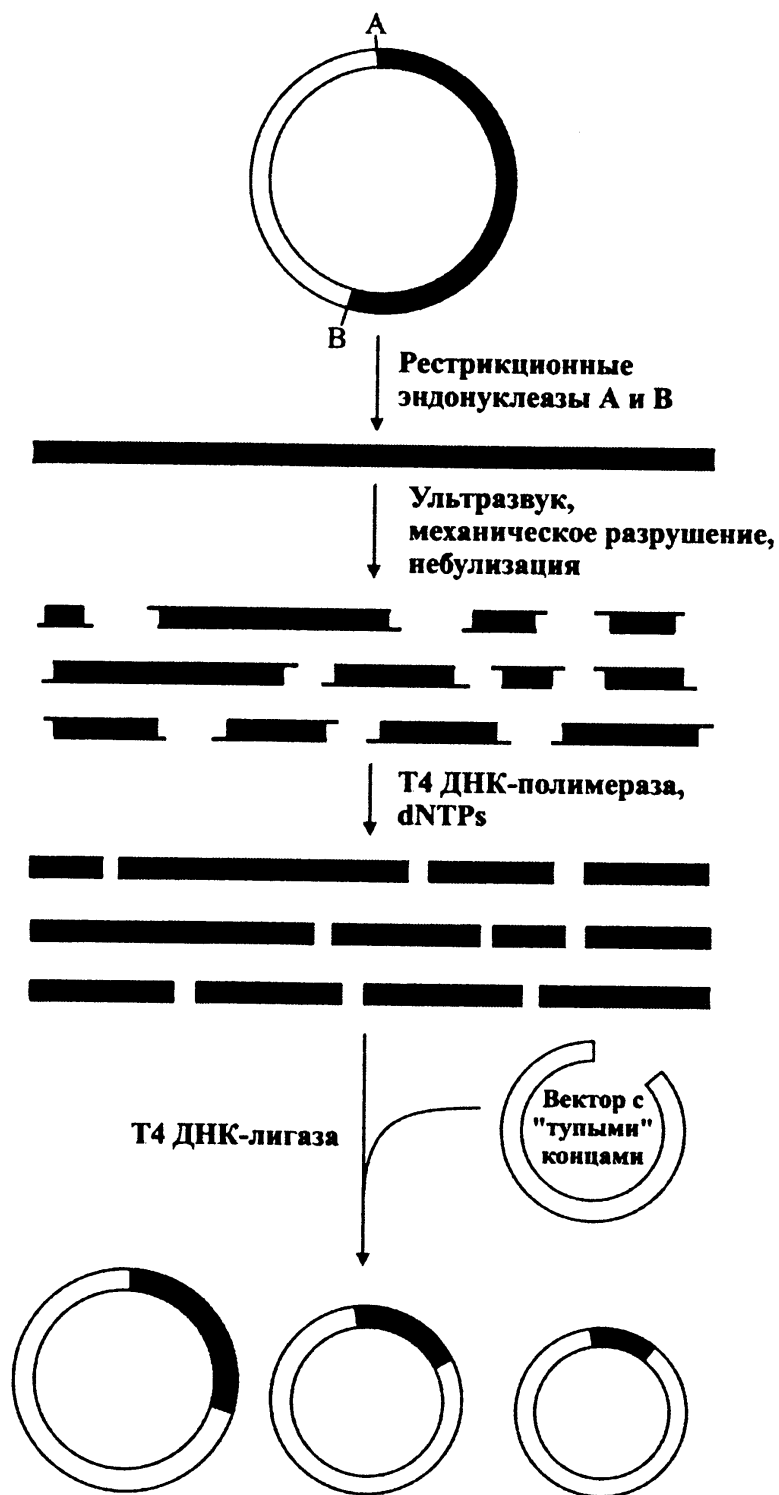


Рис. 8.6. Стратегия получения субклонов протяженного клонированного фрагмента ДНК разрушением вставки с помощью физического или механического воздействия

а вектора. При этом существует возможность поиска среди массы полученных субклонов тех, которые содержат нужные вставки путем молекулярной гибридизации колоний с исходной меченой вставкой. Однако из-за своей трудоемкости этот этап крайне нежелателен и проще изначально работать с очищенной вставкой. Впрочем, при фрагментации ДНК-космид, несущих крупную вставку, сильно превышающую по раз-

меру сам вектор, на первом этапе секвенирования этим можно и пренебречь.

Некоторая опасность при получении субклонов кроется в возможном лигировании между собой мелких фрагментов ДНК в произвольном, неправильном порядке (т.е. не в том, в каком они находятся в исходном фрагменте) и уже в виде такого составного фрагмента лигирования его с вектором. Результатом этого будет неверно определенная последовательность ДНК и дополнительные трудности в состыковке фрагментов при составлении полной нуклеотидной последовательности клонированного фрагмента ДНК. Исключить это можно путем дефосфорилирования 5'-концов фрагментированной ДНК щелочной фосфатазой перед этапом лигирования, что, впрочем, приведет к общему снижению эффективности этого процесса. Другая возможность эффективного субклонирования фрагментов ДНК заключается в применении специальных адапторов. Так, в одной работе [Povinelli, Gibbs, 1993], ввиду низкой эффективности лигирования тупых концов при объединении вектора и вставки, было предложено осуществление на первом этапе лигирования вставки и адаптора, происходящего также по тупым концам, но с большей эффективностью. Далее следовал этап лигирования с вектором, предварительно расщепленным подходящей рестрикционной эндонуклеазой, генерирующей 4 выступающих на 5'-конце нуклеотида, один из которых достраивался с помощью соответствующего дНТФ и Кленовского фрагмента ДНК-полимеразы, что приводило к неспособности вектора к самолигированию (замыканию в кольцо без участия вставки). За счет того, что и вектор, и вставка, точнее, ее адаптор, несли на своих "липких" концах по три комплементарных нуклеотида, достигалась высокая эффективность их лигирования между собой и, как следствие, высокий выход рекомбинантных клонов, превышающий 95%. В своей следующей работе этими авторами был заметно усовершенствован подход к получению подобных субклонов с помощью адапторной технологии [Andersson et al., 1994]. Более высокая эффективность субклонирования с уменьшенным фоном достигалась путем использования адаптора с 11 выступающими нуклеотидами и специального вектора, подготовленного с помощью урацил-ДНК-гликозилазы. В данном случае объединение вектора и вставки происходило без участия Т4 ДНК-лигазы простым отжигом 11 выступающих комплементарных нуклеотидов на концах вектора и вставки. Метод "двойных адапторов" с выступающими 12 нуклеотидами позволил этим же авторам настолько усовершенствовать процедуру субклонирования, что рекомбинантных клонов в результате трансформации оказывалось более 99% [Andersson et al., 1996].

Другой особенностью получения субклонов исследуемой вставки является то, что во избежание так называемого "краевого" эффекта, при котором будут теряться фрагменты ДНК, приходящиеся на концы вставки, желательно фрагментацию клонированной ДНК осуществлять в виде ее не линейного варианта, а кольцевого или конкатамеризованного, получаемого с помощью лигирования Т4 ДНК-лигазой. На этом же принципе получения конкатамерных конструкций основано

крупно-масштабное секвенирование кДНКовых клонов [Andersson et al., 1997]. Данный подход, названный авторами как CCS (concatenation cDNA sequencing), заключается в произвольном объединении с помощью ДНК-лигазы фага Т4 кДНКовых вставок в единый блок, его произвольной дефрагментации и субклонировании в векторах на основе фага М13. Далее проводится стандартное секвенирование и затем, на этапе анализа “прочитанных” нуклеотидных последовательностей, осуществляется восстановление исходных фрагментов кДНК, где в качестве неких маркеров служат сайты рестрикционных эндонуклеаз, характерных для того или иного клона. Как указывают авторы, преимуществом этого подхода является одновременное секвенирование сразу большего числа кДНК-фрагментов, позволяющее повысить эффективность этой процедуры. В своей другой работе они сообщают об успешном секвенировании с помощью CCS-стратегии 100 тпн, принадлежащих 69 индивидуальным кДНКовым клоном, причем точность определения заключительной нуклеотидной последовательности соответствовала установленному стандарту 99,99% [Yu et al., 1997].

Подходы к получению субфрагментов ДНК под действием физических факторов или обработкой ДНКазой, независимые от последовательности нуклеотидов, были несколько предпочтительнее, чем формирование библиотеки субклонов с помощью частощепящих рестрикционных эндонуклеаз с тетра-нуклеотидными сайтами узнавания. Однако, открытие новых ферментов с различными специфичностями действия спустя много лет привело к тому, что этот подход с использованием рестрикционных эндонуклеаз получил новое развитие. Так, благодаря обнаруженной сверхчастощепящей рестрикционной эндонуклеазе *CviII*, узнающей последовательность RGCY, теоретически встречающуюся через каждые 64 нуклеотида, появилась возможность получать необходимые субклоны для секвенирования протяженного фрагмента ДНК, которые бы полностью перекрывали всю вставку без промежутков [Fitzgerald et al., 1992]. В этой работе было также отмечено, что в определенных условиях (в присутствии 1 мМ АТФ и 20 мМ ДТТ) фермент несколько меняет свою специфичность и способен расщеплять не только сайты RGCY, но и RGCR и YGCV, однако не способен расщеплять YGCR. Таким образом, реальная частота встречаемости сайтов этой рестрикционной эндонуклеазы в особых условиях расщепления увеличивается до каждых 16 нуклеотидов. Хотя другими авторами при использовании этой же рестрикционной эндонуклеазы *CviII* с целью получения субклонов клонированного ранее в космиде протяженного фрагмента ДНК было отмечено, во-первых, что расщепление ДНК данным ферментом осуществлялось только по нормальному сайту расщепления RGCY, и, во-вторых, среди 43 тпн (37,8 тпн вставки и 5,2 тпн вектора) был выявлен участок около 1 тпн, не содержащий ни одного такого сайта [Gingrich et al., 1996]. Впрочем, еще одно преимущество данного подхода к получению субклонов по сравнению с упоминавшимися выше вариантами фрагментации протяженного фрагмента ДНК обработкой ультразвуком или другими физическими воздействиями, а также ДНКазой в присутствии ионов марганца, заключается в образо-

вании рестриктазных *Cvi*II-фрагментов с тупыми концами, поскольку данный фермент производит разрыв цепи ДНК между центральными гуанином и цитозином. Тогда как прочие способы фрагментации приводят, главным образом, к образованию участков ДНК не с тупыми концами и поэтому необходимы дополнительные ферментативные этапы по затуплению концов этих коротких фрагментов ДНК.

Нельзя обойти молчанием относительно недавнюю работу французских авторов, посвященную сравнительному анализу эффективности выполнения двух проектов секвенирования фрагмента ДНК протяженностью около 6,2 тпн разными способами [Demolis et al., 1995]. Получение субклонов при выполнении первого проекта осуществлялось с помощью рестрикционных эндонуклеаз с тетрануклеотидными сайтами узнавания *Alu*I, *Sau*3A1 и *Taq*I. Второй проект был выполнен с использованием ДНКазы I/Mn<sup>2+</sup>. Так, в цитируемой работе убедительно показана более высокая производительность подхода, где требуемые субклоны были получены при помощи произвольного расщепления секвенируемого фрагмента ДНК ДНКазой в присутствии ионов марганца.

Недавно для получения субклонов для секвенирования было предложено использовать рестрикционные эндонуклеазы IIС-типа, причем генерирующие фрагменты ДНК с 5 выступающими нуклеотидами на 5'-конце [Rena, Houslay, 1998]. По мнению авторов, данные уникальные последовательности, образуемые этими 5 нуклеотидами, позволят осуществить правильную состыковку секвенируемых субклонов, не прибегая к построению детальной рестриктазной карты всего фрагмента ДНК.

Часто по завершению цикла экспериментов по секвенированию ДНК случайным подходом и после компьютерной состыковки "прочитанных" последовательностей оказывается, что субклоны для отдельных участков среди прочих были не секвенированы. Выявление в созданной клонотеке таких субклонов стандартным способом требует много времени и усилий, поэтому был предложен вариант, в котором с помощью агарозного гель-электрофореза в каждой дорожке разделялось одновременно 10 образцов [Hong, 1988]. Причем, как такового, выделения ДНК не проводилось и электрофоретическому разделению подвергались аликвоты супернатанта после выращивания одноцепочечных фагов с добавлением равного объема буфера для нанесения, содержащего 1%-ный додецилсульфат натрия. Метод очень простой и относительно быстрый, но качество разделения, впрочем, оставляет желать лучшего.

Другой подход к завершению определения нуклеотидной последовательности клонированного фрагмента ДНК заключается в целенаправленном субклонировании конкретного участка исходного клона. Достичь этого можно, построив с помощью компьютера рестриктазную карту секвенированных областей вставки, основываясь на ставшей уже известной последовательности нуклеотидов части секвенируемого фрагмента ДНК, что, возможно, позволит определить необходимые рестрикционные эндонуклеазы для субклонирования нужного фрагмента.

Еще одна стратегия секвенирования протяженных фрагментов ДНК, названная авторами RANDI (RANdom and DIrected), изначально предполагает определение нуклеотидных последовательностей как с помощью методов случайного подхода (на первом этапе), так и оставшихся непокрытыми участков, с помощью направленного подхода на втором этапе [Voss et al., 1995]. Схожий принцип был заложен при разработке стратегии, названной “Sequence scanning”, что можно перевести как “секвенирование сканированием” [Nurminsky, Hartl, 1996]. Авторами данной работы на основе фагового вектора  $\lambda$ ZAP (Stratagene, США) был специально создан вектор LambdaScan, рассчитанный на субклонирование 8–12 тпн вставок и последующее секвенирование их концов. Исходным материалом служили клонированные ранее в бактериофаге P1 фрагменты ДНК с размером около 100 тпн. Необходимым этапом этой стратегии является секвенирование концевых участков (по 350–500 пн) с обеих сторон вставок у достаточного числа полученных случайных субклонов с использованием в качестве мест для отжига праймеров векторных последовательностей, граничащих со вставкой. Проведенные подсчеты показывают, что подобное секвенирование 192 таких субклонов (по числу лунок микротитратора) даст первичную информацию о 135–192 тпн и при этом участки, “прочитанные” по обеим цепям, по одной и еще не секвенированные вовсе, будут соотноситься как 1:2:1. По мнению авторов, компьютерная состыковка должна дать информацию об участках исходного клона, секвенирование которых должно быть продолжено уже стратегией направленного подхода. Несколько иной подход, направленный на уменьшение числа секвенируемых случайных клонов, был предложен другими авторами [Scholler et al., 1995]. Так, для исключения из секвенирования лишнего количества случайных субклонов в фагмидном векторе исходного 11,5 тпн фрагмента ДНК ими была проведена гибридизация ДНК данных субклонов с декамерными олигонуклеотидами. Это позволило распределить полученные субклоны по группам перекрытия и, таким образом, уменьшить количество секвенируемых матриц.

### 8.3. СТРАТЕГИИ СЕКВИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ ПО СЭНГЕРУ С ПОМОЩЬЮ НАПРАВЛЕННОГО ПОДХОДА

При определении последовательности нуклеотидов протяженных фрагментов ДНК направленным подходом получение субклонов носит упорядоченный характер, иногда основанный на знании рестриктазной карты клонированного фрагмента ДНК. В связи с этим состыковка секвенированных фрагментов не представляет каких-либо сложностей, так как заранее известно, какое место вставки подвергается секвенированию. Уменьшается и общее число “прочитанных” нуклеотидов, что позволяет удешевить проект по секвенированию протяженного фрагмента ДНК.

Наиболее оптимальным является получение в составе исходного вектора субклонов, различающихся размером делетированных участков вставки между сайтом какой-либо подходящей полилинкерной рестрикционной эндонуклеазы, предшествующим сайту клонирования (со стороны отжига секвенирующего праймера) и прогрессивно удаляющимися местами вставки. Однако большинство вариантов направленного подхода, предполагающего получение серии таких субклонов, укороченных с одного конца вставки, различаясь значительно при проведении делеционных процедур, требуют первого этапа рестриктазного картирования с целью или построения подробной рестриктазной карты секвенируемого фрагмента ДНК, или обнаружения подходящих рестрикционных эндонуклеаз, отсутствующих во вставке, которые могут быть затем использованы для получения необходимых субклонов.

Описанный в предыдущем разделе метод получения делеционных субклонов с помощью ДНКазы I/Mn<sup>2+</sup> в векторе pBR322 [Frischauf et al., 1980], на самом деле, мало чем отличается от подобных работ, ставящих своей целью получение субклонов для их последующего секвенирования ферментативным методом. Так, в данной статье было просто заявлено, что эти субклоны предназначены для секвенирования по Максаму–Гилберту. Но в то же время другое, более важное отличие заключается в том, что используемый в той работе вектор pBR322 не несет полилинкера и специального участка для отжига универсального праймера. Однако ввиду своей относительной простоты этот метод получения серии делеционных субклонов в составе вектора, несущего полилинкер, с целью их последующего секвенирования ферментативным методом с помощью ДНКазы I в присутствии Mn<sup>2+</sup> применяется весьма часто [Hong, 1982, 1987; Laughon, Scott, 1984; Lin et al., 1985]. Некоторое усовершенствование в виде повторного этапа расщепления лигированных фрагментов ДНК дополнительной рестрикционной эндонуклеазой приводило к обогащению лигазной смеси рекомбинантными плазмидами, несущими делеционные варианты вставок [Lin et al., 1985]. Причиной этому было то, что сайт второй рестрикционной эндонуклеазы расположен между сайтом первой рестрикционной эндонуклеазы и местами разрыва цепи ДНК во вставке, произошедшими под действием ДНКазы. Как следствие, расщеплялись только восстановленные исходные варианты клона, несущие полноразмерную вставку и ненарушенный сайт узнавания для этого дополнительного фермента. Такое же повторное расщепление лигированных фрагментов ДНК, но уже рестрикционной эндонуклеазой, сайт узнавания которой был расположен с противоположной стороны вставки, приводило к тому, что линеаризованными оказывались конструкции, ДНКазное расщепление которых произошло не во вставке, а в противоположной части вектора относительно сайта расщепления первой рестрикционной эндонуклеазы, и не затронуло участки вектора, необходимые для его функционирования [Laughon, Scott, 1984]. Это не позволяет исключить из дальнейшего анализа исходный полноразмерный клон и дает лишь некоторое обогащение пула конечных

субклонов с делеционными вариантами вставок за счет удаления конструкций с векторами, несущими незначительные делеции и сохранившими при этом возможность репликации. Что касается большинства вариантов расщепления вектора, произошедшими под действием ДНКазы в присутствии ионов марганца, то многие из них приводят к потере ими жизнеспособности и такие конструкции из дальнейших экспериментов исключаются.

Данная стратегия с использованием ДНКазы I в сочетании с ПЦР привела к возможности нарабатывать матрицы для секвенирования прямо из трансформированных колоний, несущих делеционные варианты вставки, с последующим анализом амплификатов по размеру в агарозном геле [Burland, Kusukawa, 1997].

Хотя при осуществлении этого подхода с ДНКазным расщеплением и не требуется построения детальной рестриктазной карты клонированного фрагмента ДНК, все же необходимо владеть информацией об отсутствии сайтов расщепления используемых полилинкерных рестрикционных эндонуклеаз во вставке, поскольку в противном случае результатом будет гетерогенная смесь, в которой будет чрезвычайно трудно разобраться.

Был предложен также другой подход к получению делеционных субклонов с использованием ДНКазы I [Varnes, Bevan, 1983]. В этом случае обработка ДНКазой суперскрученной плазмидной ДНК проводилась в присутствии не ионов марганца, а интеркалирующего красителя бромистого этидия (рис. 8.7), что приводило не к разрыву обеих цепей ДНК, а только к формированию “ника”. Следующий этап заключался в расширении с данного созданного “ника” одноцепочечного участка с помощью экзонуклеазы III. Далее в действие вступала нуклеаза Bal31, удалявшая одноцепочечный участок рекомбинантной плазмиды и приводившая, таким образом, к линеаризации вектора со вставкой. Затем с помощью T4 ДНК-лигазы к образовавшимся тупым концам “пришивались” специальные *EcoRI*-линкеры и следовал этап расщепления данной конструкции этой рестрикционной эндонуклеазой. После завершающих этапов лигирования и трансформации проводился отбор фагов, содержащих нужные делеции вставки. Надо отметить, что особенностью используемых в данной работе одноцепочечных векторов серии mWB, сконструированных на основе дикого фага M13, была их способность давать голубое окрашивание на индикаторных чашках Петри (с добавлением *x-gal* и ИПТГ) даже в случае содержания вставки. Однако при нарушении участка *lac*-оператора, помещенного за вставку (что происходило в случае чрезмерного делетирования вставки, приводящего к полному ее удалению), окраска пропадала и, таким образом, подобные фаговые бляшки в дальнейший анализ уже можно было не брать. Другой особенностью всего этого комплексного подхода с участием многочисленных ферментов и названного авторами килосеквенированием, было электрофоретическое разделение не депротенизированной фаговой ДНК, а жизнеспособных фаговых частиц, способных по завершению гель-электрофореза к трансфекции, что дало возможность авторам цитируемых ра-



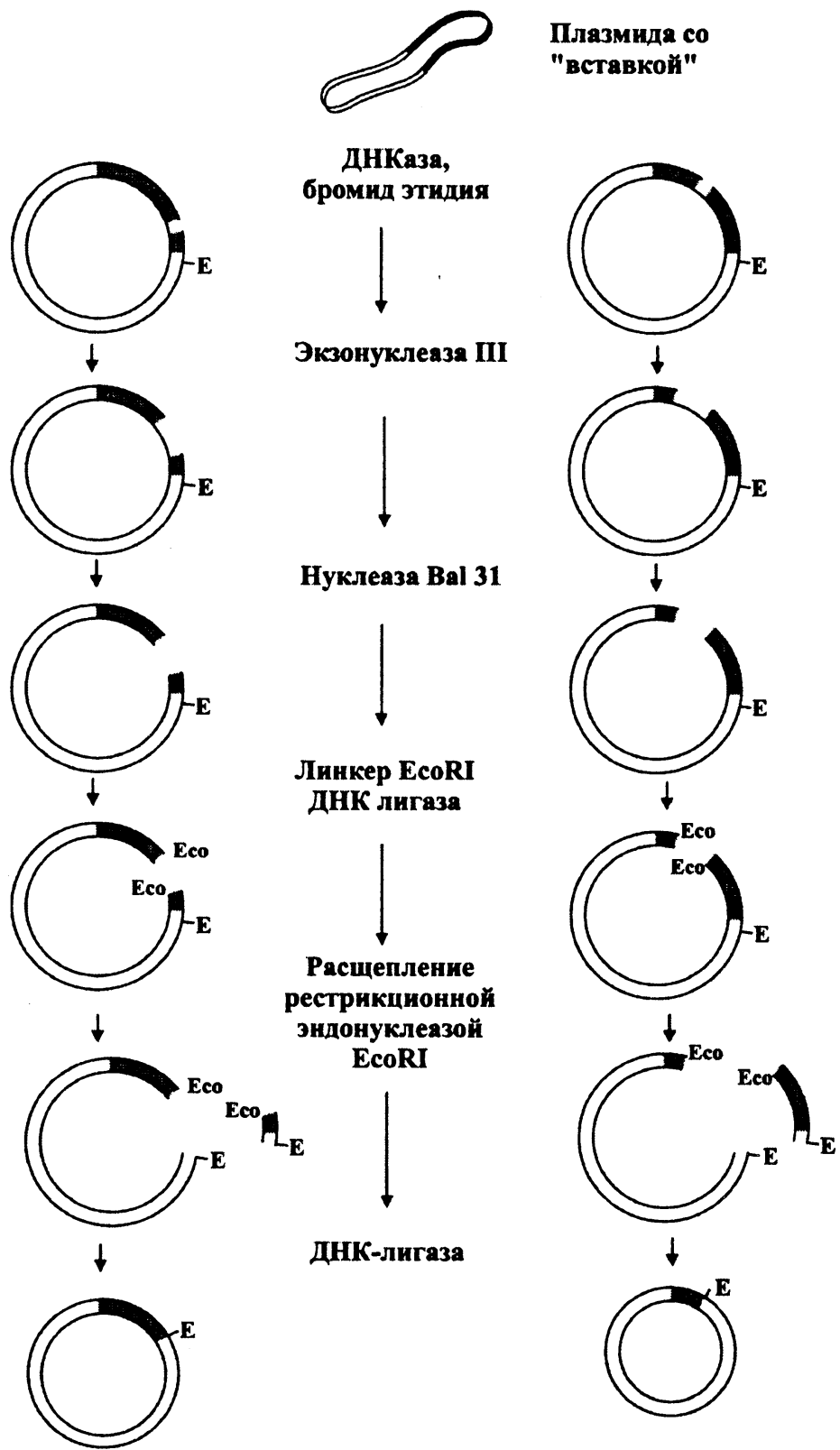


Рис. 8.7. Стратегия получения однонаправленных делеционных субклонов протяженного клонированного фрагмента ДНК с помощью расщепления вставки ДНКазой в присутствии бромистого этидия

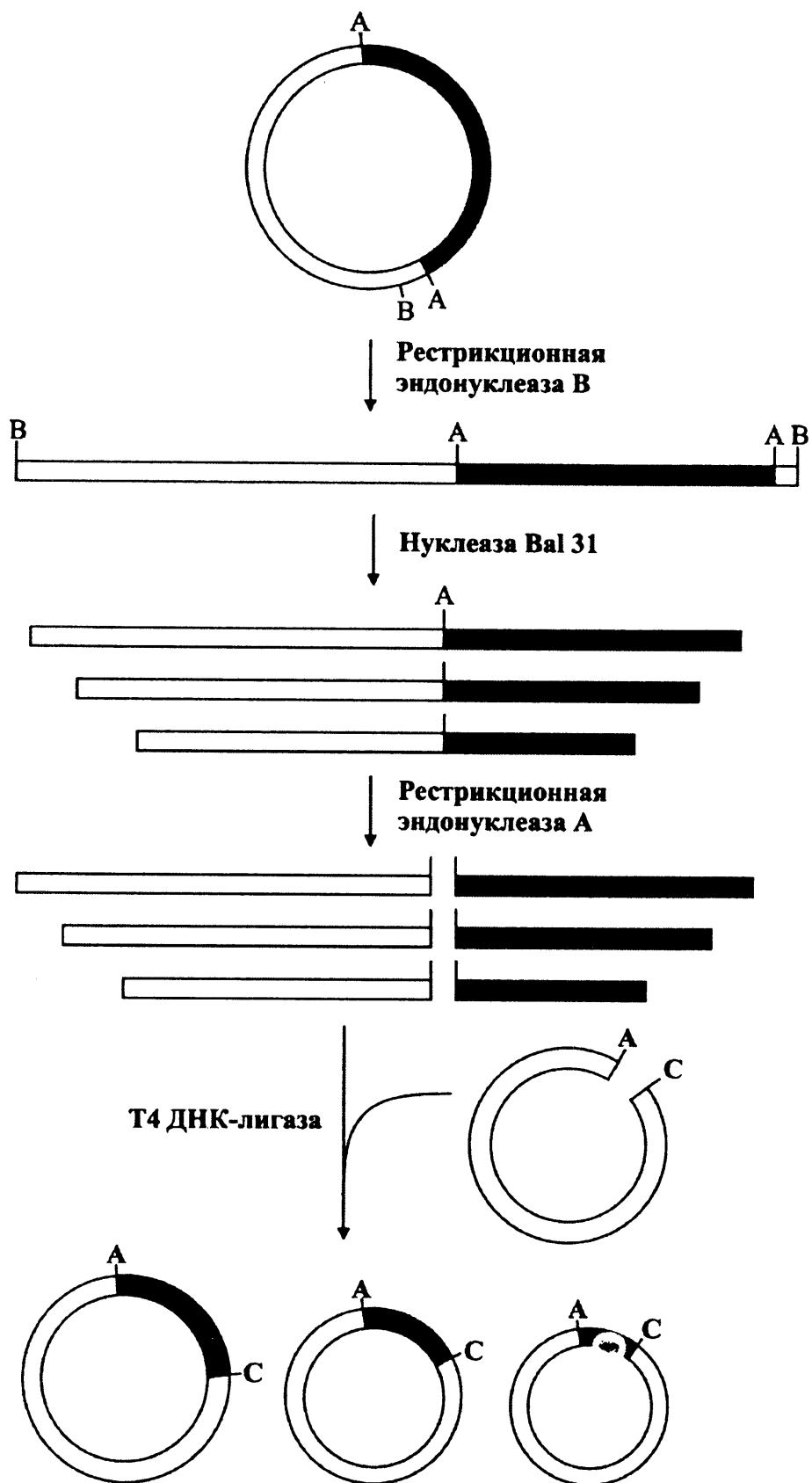


Рис. 8.8. Стратегия получения однонаправленных делеционных субклонов протяженного клонированного фрагмента ДНК с помощью укорочения вставки нуклеазой Bal31

бот значительно увеличить производительность анализа делеционных субклонов.

Также значительное количество различных ферментов для получения делеционных субклонов было использовано в работе других авторов (рис. 8.8) [Guo et al., 1983]. Здесь на первом этапе проводилась линеаризация вектора и вставки одной из полилинкерных рестрикционных эндонуклеаз *Hind*III. Затем под действием нуклеазы *Bal*31 или в другом варианте последовательным расщеплением линейной ДНК экзонуклеазой III и затем S1-нуклеазой, специфичной к односторонней ДНК, происходило двунаправленное укорочение линеаризованного вектора и вставки. Разная продолжительность обработки приводила к появлению образцов с наборами фрагментов ДНК различающейся длины. Расщепление вторым полилинкерным ферментом, расположенным с другой стороны вставки, генерировало смесь фрагментов ДНК, часть из которых представляла собой однонаправленные делеционные варианты вставок. Лигирование этих фрагментов в новом векторе позволяло получить желаемый набор субклонов. Однако был необходим этап анализа полученных клонов, направленный на определение размера содержащейся в них вставки. Следует отметить сложность проведения всех этих последовательных процедур.

Было сообщено об аналогичном использовании нуклеазы *Bal*31 для получения однонаправленных прогрессивных делеций клонированной вставки с последующим переклонированием набора по-разному укороченных вариантов вставки в одноцепочечном векторе M13mp7 [Poncz et al., 1982] и модифицированном M13 векторе mTM010 [Misra, 1985; Misra, 1987]. Причем в последнем случае автор не прибегал к весьма трудоемкой процедуре разделения гетерогенных фрагментов ДНК гель-электрофорезом. В литературе содержатся сведения о быстром несколько модифицированном методе выделения плазмидной ДНК, специально предназначенной для получения серии делеционных субклонов с помощью нуклеазы *Bal*31 [Sharrocks, Hornby, 1987].

Все приведенные выше стратегии секвенирования ДНК предполагали получение однонаправленных делеционных вариантов клонированных фрагментов ДНК и поэтому весьма интересны немногочисленные работы по получению двунаправленных делеционных вариантов вставки [Breault et al., 1995; Ren et al., 1997]. В одной из них под действием нуклеазы *Bal*31 (или комбинацией экзонуклеазы III и нуклеазы золотистой фасоли) проводилось двунаправленное укорочение элюированного из геля клонированного фрагмента ДНК в направлении “снаружи → внутрь” (рис. 8.9) в течение нескольких дискретных промежутков времени [Ren et al., 1997]. После клонирования в подготовленном новом векторе серии укороченных подобным образом фрагментов ДНК появляется возможность секвенирования вставок в обеих ориентациях, как за счет противоположно направленных праймеров, так и за счет вероятности случайного клонирования данных фрагментов в любой ориентации.

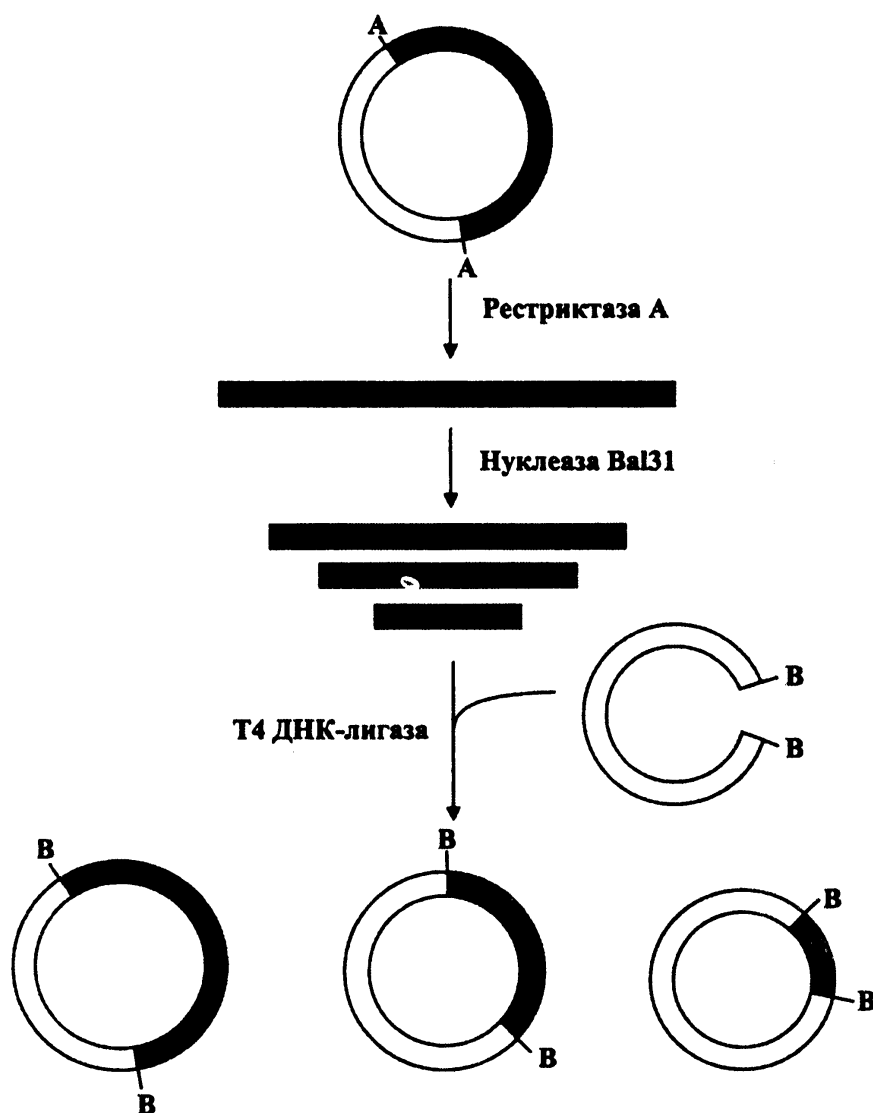


Рис. 8.9. Стратегия получения двунаправленных делеционных субклонов протяженного клонированного фрагмента ДНК с помощью укорочения вставки нуклеазой Bal31 в направлении “снаружи → внутрь”

Также двунаправленное укорочение вставок, но уже в направлении “изнутри → наружу” (рис. 8.10) было проведено с помощью нуклеазы Bal31 в другой работе [Breault et al., 1995]. К преимуществам данного подхода можно отнести то, что укорочение вставки осуществляется без ее элюции из геля и последующее субклонирование проводится в том же исходном векторе. Единственное требование заключается в наличии уникального сайта какой-либо рестрикционной эндонуклеазы, расположенного по возможности ближе к центру исходного клонированного фрагмента ДНК. Расщепление вектора со вставкой данным ферментом приводит к его линейаризации и образованию двух концов, принадлежащих клонированному фрагменту ДНК и доступных для гидролиза нуклеазой Bal31. За счет различного времени протекания реакции происходит соответственно разное укорочение фрагмента ДНК, что позволяет получить набор субклонов, полностью перекрывающих всю последовательность вставки. Некоторой особенностью получения субклонов

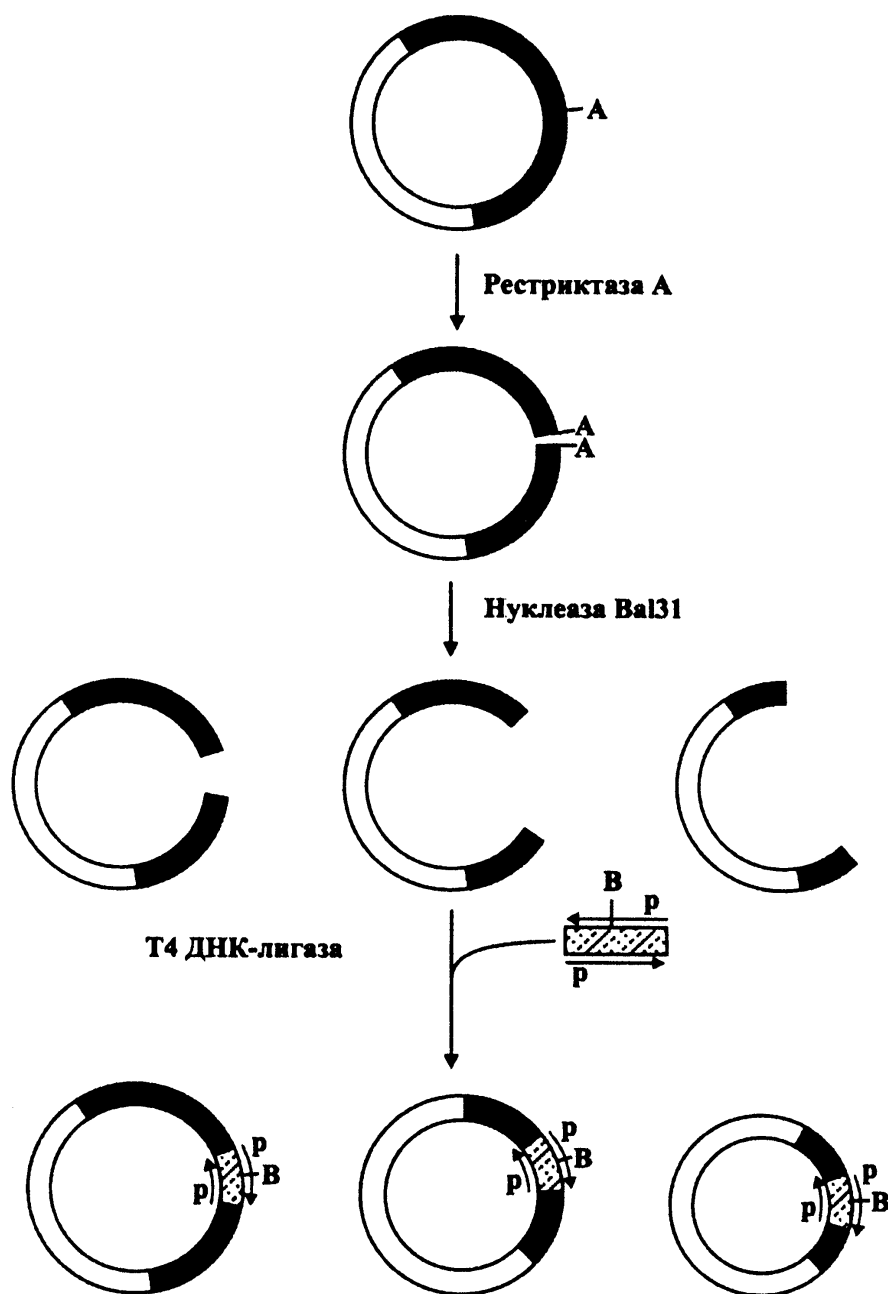


Рис. 8.10. Стратегия получения двунаправленных делеционных субклонов протяженного клонированного фрагмента ДНК с помощью укорочения вставки нуклеазой Val31 в направлении “изнутри → наружу”

и самого проведения этапа лигирования является то, что в образовавшуюся брешь между противоположными концами вставки встраивают синтетический участок ДНК небольшой протяженности, формируемый за счет отжига двух комплементарных олигонуклеотидных праймеров и имеющий тупые концы. Впоследствии обе цепи этого линкера служат, прежде всего, местами отжига этих же олигонуклеотидов, выступающих уже в роли секвенирующих праймеров, что позволяет секвенировать обе цепи клонированного фрагмента ДНК в противоположных направлениях. Другое предназначение данного линкера заключается в отборе с его помощью укороченных делеционных вариантов исходной

вставки. Последнее становится возможным за счет того, что этот синтетический линкер содержит в своей последовательности сайт какой-нибудь редкощепящей рестрикционной эндонуклеазы и те клоны, плазмидная ДНК которых линейаризуется под действием этого фермента, отбираются для их последующего секвенирования, а клоны, расщепляемые первым ферментом, изначально присутствующим во вставке, представляют собой нежелательные варианты и должны отбрасываться. Некоторым недостатком этого метода, как считают отдельные авторы [Rep et al., 1997], является невозможность использования стандартных секвенирующих праймеров, однако, те же олигонуклеотиды, которые и образуют синтетический линкер, служат и секвенирующими праймерами, так что каких-либо дополнительных усилий по синтезу новых праймеров затрачивать не приходится. Другим недостатком, впрочем временным и легко преодолемым, является неизвестность ориентации линкера в каждом конкретном клоне, что приводит к необходимости определения принадлежности секвенированного участка ДНК к той или иной цепи клонированного фрагмента ДНК при его последующем сравнении и состыковке с другими секвенированными участками.

Описанные выше варианты направленной стратегии секвенирования протяженных фрагментов ДНК основаны на особенностях действия используемых ферментов. Так, ДНКаза I в присутствии ионов марганца производит разрыв обеих цепей ДНК с образованием или тупых концов, или с небольшим количеством выступающих нуклеотидов, легко поддающихся репарации. Этот же фермент в присутствии интеркалирующего красителя бромистого этидия делает так называемые ники в виде одиночных разрывов одной цепи ДНК. Специфичная к однонитевой ДНК S1 нуклеаза применяется для разрыва второй цепи ДНК в месте образовавшегося “ника” или целого участка одноцепочечной ДНК, сформировавшегося после “расширения” “ника” с помощью экзонуклеазы III. Для этой же цели используют и другую нуклеазу Bal31, обладающую относительно слабой эндонуклеазной активностью, специфичной для однонитевой ДНК. (Более сильная экзонуклеазная активность этого фермента позволяет деградировать двуцепочечные участки ДНК). Еще один фермент – экзонуклеаза III, используемый вместе с другими ферментами (нуклеазами) в некоторых из упомянутых выше способах для удаления одной цепи ДНК, начиная с “ника” увеличивает в направлении 3' → 5' участок одноцепочечной ДНК.

Однако последний фермент нашел и самостоятельное весьма широкое применение в экспериментах по получению делеционных субклонов, пригодных для секвенирования ДНК (рис. 8.11). Были предложены различные варианты использования экзонуклеазы III для получения делеционных вариантов исходного субклона [Guo, Wu, 1982; Henikoff, 1984, 1987; Yanisch-Perron et al., 1985; Sorge, Blinderman, 1989; Li, Tucker, 1993; Zhu, Clark, 1995]. Общим для всех этих способов является то, что в качестве предварительного этапа необходимо построение рестриктазной карты клонированного фрагмента ДНК. Или точнее, на первом этапе предстоит выяснить отсутствие во вставке сайтов для полилинкерных рестрикционных эндонуклеаз, которые могли бы быть исполь-

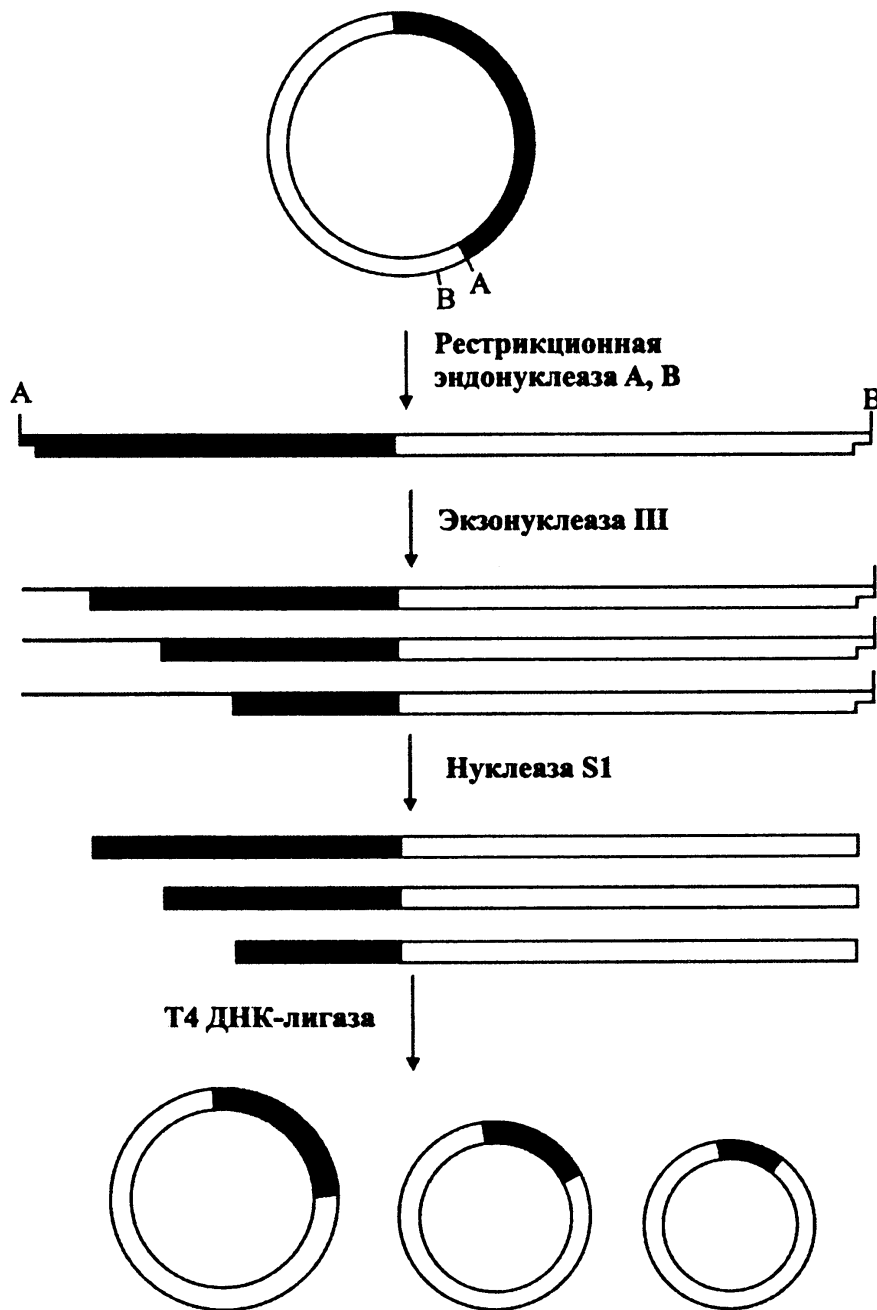


Рис. 8.11. Стратегия получения однонаправленных делеционных вариантов исходного клона с помощью экзонуклеазы III

зованы для однонаправленного укорачивания фрагмента ДНК с помощью экзонуклазы III (рис. 8.11). Причем эти рестрикционные эндонуклеазы должны быть определенным образом ориентированы по отношению ко вставке и генерировать 3'-выступающий (один фермент) и 5'-выступающий или тупой конец (другой фермент). В противном случае при выявлении отсутствующих во вставке полилинкерных ферментов, образующих только 5'-выступающие концы, 3'-конец одного из них необходимо с помощью Кленовского фрагмента ДНК-полимеразы I защитить тиопроизводным соответствующего дНТФ от действия экзонуклеазы III [Putney et al., 1981]. Впрочем, далеко не все фрагменты

ДНК с 3'-выступающими концами защищены от их расщепления экзонуклеазой III. Считается, что менее подвержены экзонуклеазной атаке этим ферментом фрагменты ДНК с 4 выступающими нуклеотидами на 3'-конце. Иной способ защиты одного конца линейаризованного вектора, несущего вставку от действия экзонуклеазы III, был предложен в работе других авторов [Johnson et al., 1990]. Так, ими на основе одноцепочечного вектора M13mp19 был сконструирован специальный вектор, который нес участок *lac*-оператора между полилинкером и сайтом отжига секвенирующего праймера. После расщепления подходящей полилинкерной рестрикционной эндонуклеазой в инкубационную смесь добавлялся белок *lac*-репрессор, прочно связывающийся с *lac*-оператором и защищающий таким образом, данный конец расщепленной репликативной формы вектора от экзонуклеазы III, действие которой как и при стандартном использовании [Henikoff, 1984] продолжалось в течение нескольких дискретных интервалов времени.

Важное значение для этих и других стратегий секвенирования с помощью экзонуклеазы III приобретает получение достаточно дискретных наборов укороченных вставок, отбираемых по времени протекания реакции. Такие наборы должны характеризоваться минимальным разбросом внутри каждого варианта, поскольку в противном случае будет необходим анализ чрезмерного числа клонов и данный подход потеряет свою привлекательность. Так, в литературе содержатся различные рекомендации для уменьшения разброса, заключающиеся в варьировании температуры инкубации или концентрации NaCl в реакционной смеси [Tomb, Varca, 1989].

Владение информацией об отсутствии сайтов узнавания отдельных рестрикционных эндонуклеаз в клонированном фрагменте ДНК позволяет приступить к получению его делеционных вариантов, где первым этапом является линейаризация вектора со вставкой двумя полилинкерными рестрикционными эндонуклеазами, причем сайты узнавания обоих ферментов должны быть расположены между местом отжига секвенирующего праймера и самой вставкой. Вторым условием является то, что расположенный ближе ко вставке фермент должен генерировать фрагменты ДНК с тупыми или 5'-выступающими концами, доступными для их расщепления экзонуклеазой III. Тогда как противоположный конец большого фрагмента ДНК не должен быть подвержен экзонуклеазной атаке. Во время второго этапа экзонуклеазного расщепления вставки в составе линейаризованного вектора через определенные промежутки времени из реакции удаляются аликвоты, что приводит к формированию гетерогенных фрагментов ДНК, у которых образовавшийся участок одноцепочечной ДНК имеет различную дискретную протяженность. Маленький фрагмент ДНК, приходящийся на участок полилинкера и ограниченный этими же использованными рестрикционными эндонуклеазами, настолько мал, что в процессе экзонуклеазной обработки одна его незащищенная цепь перестает существовать и он выключается из дальнейшего хода ферментативных реакций. Полученный в ходе предыдущего этапа набор дискретных фрагментов ДНК с выступающими одноцепочечными участками различной длины на од-



ном из их 5'-концов обрабатывается или S1 нуклеазой, или нуклеазой из золотистой фасоли, или экзонуклеазой VII, специфичными к однонитевым участкам ДНК. В результате этой реакции, а также затупления концов с помощью Кленовского фрагмента ДНК-полимеразы I (его экзонуклеазной активности) образуются полностью двуцепочечные участки ДНК, имеющие разную протяженность делетированной вставки и пригодные для лигирования и последующей трансформации бактериальных клеток.

Несмотря на свою кажущуюся привлекательность и определенное удобство получения однонаправленных делеционных субклонов для секвенирования протяженных фрагментов ДНК, метод с использованием экзонуклеазы III все же весьма трудоемок. Определенные неудачи могут подстергать экспериментатора при использовании в его работе с плазмидной ДНК, содержащей "ники", что приведет к непредсказуемому расщеплению и вектора и вставки, поскольку экзонуклеаза III способна к удалению одной цепи ДНК с таких "ников". Поэтому для проведения подобных экспериментов настоятельно рекомендуется использовать суперскрученную форму плазмид и, желательно, очищенную в градиенте плотности хлористого цезия [Henikoff, 1987]. В то же время, другими авторами в результате клонирования ПЦР-продуктов специально создавались "ники" за счет использования одного фосфорилированного праймера, а другого с отсутствием фосфатной группы на 5'-конце, что позволяло затем с помощью экзонуклеазы III и некоторых других ферментов получать набор по-разному делегированных субклонов [Kitabatake, Inokuchi, 1993].

Также была предложена оригинальная стратегия получения делеционных вариантов субклонов, основанная на применении все той же экзонуклеазы III [Henikoff, 1990]. Суть этой стратегии заключается в образовании сначала двуцепочечной структуры из одноцепочечного фага M13 или ему подобного и уже затем создании делеций. Так, на первом этапе проводится отжиг олигонуклеотидного праймера, расположенного таким образом, чтобы ДНК-полимеразой при удлинении праймера в первую очередь была скопирована сама векторная последовательность. Построение цепи ДНК осуществляется с помощью ДНК-полимеразы фага T4, поскольку этот фермент характеризуется высокой скоростью полимеризации, не несет 5' → 3'-экзонуклеазной активности и не вызывает смещение цепей как некоторые другие ДНК-полимеразы. В результате образуется двуцепочечная молекула ДНК, несущая один "ник" между 5'-концом праймера и 3'-концом вновь синтезированной цепи ДНК. Далее в действие вступает экзонуклеаза III, работающая в обратном направлении и расширяющая "ник" до весьма протяженных участков одноцепочечной ДНК. Отбор алиquot данной реакции по времени расщепления приводит к формированию некоторого пула молекул с одноцепочечными участками различной протяженности. После удаления образовавшихся одноцепочечных участков S1 нуклеазой, репарации концов и их лигирования проводится трансформация компетентных клеток *E.coli* полученными конструкциями. Таким образом, как можно видеть из рис. 8.12, в результате всех этих процедур об-

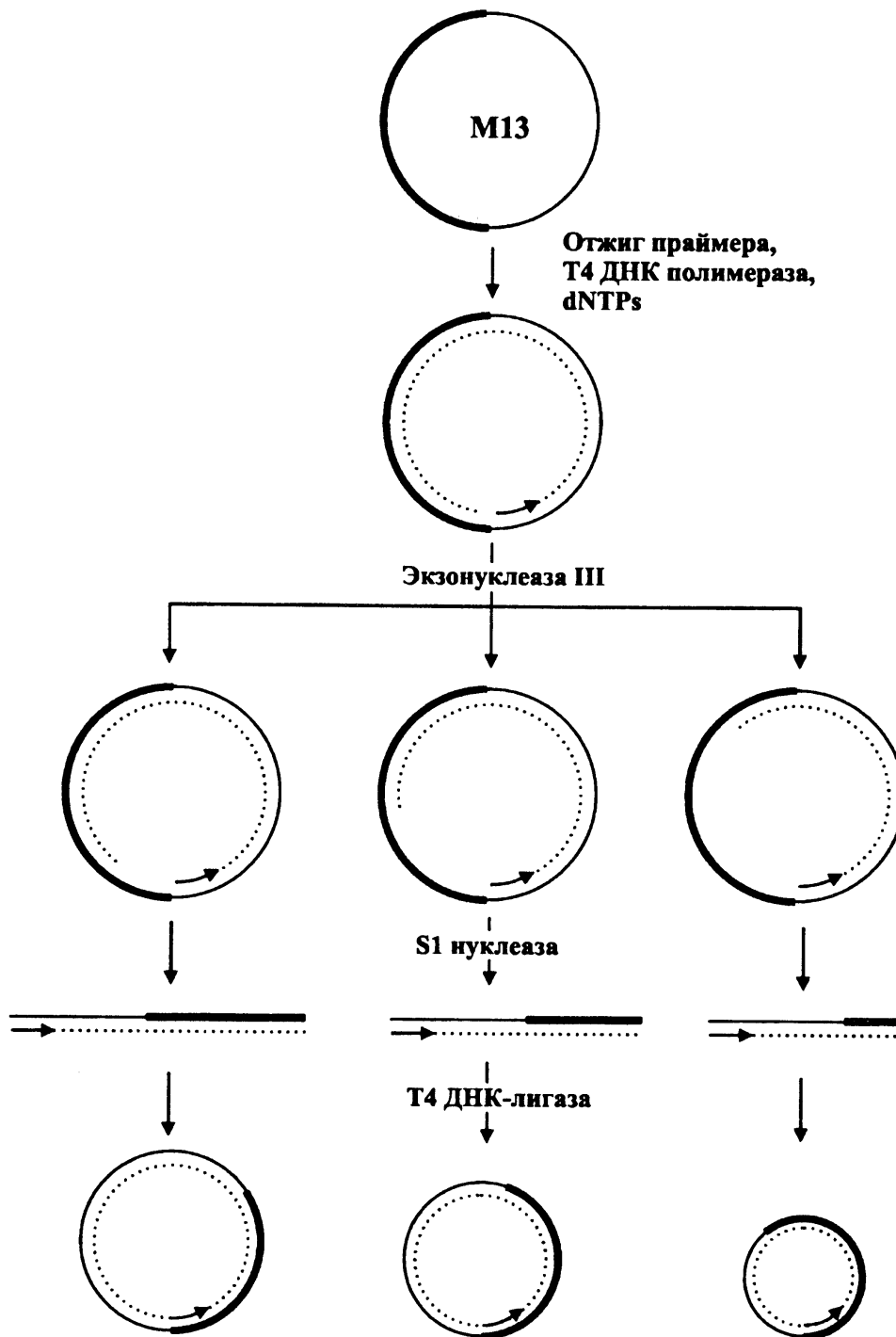


Рис.8.12. Стратегия получения однонаправленных делеционных вариантов вставки с помощью экзонуклеазы III, изначально существующей в одноцепочечном виде в соответствующем векторе

разуется набор рекомбинантных клонов, несущих по разному делетированные вставки. К значительному преимуществу данной стратегии можно отнести ее независимость от наличия или отсутствия сайтов определенных рестрикционных эндонуклеаз, в отличие от большинства других подходов, служащих для получения укороченных вариантов вставки с помощью экзонуклеазы III [Guo, Wu, 1982; Henikoff, 1984; Yanisch-Perron et al., 1985]. Но и недостатки также достаточно серьез-

ные. Пожалуй, главный недостаток заключается в том, что полученные делеционные конструкции после их трансформации и выращивания в виде одноцепочечной фаговой ДНК нельзя секвенировать с помощью универсального праймера. Вернее, секвенировать можно, но при этом будет секвенироваться неподвергнутый делециям участок вставки, расположенный сразу за вектором, поскольку при выращивании того же фага или фагомиды, в которой изначально был клонирован фрагмент ДНК, будет упаковываться та же цепь. Некоторое решение проблемы заключается или в выделении двуцепочечной формы и ее последующем секвенировании, или в использовании специальных фагомид семейства  $\rho$ KUN, имеющих два несколько отличающихся ориджина фаговой репликации, позволяющих нарабатывать ту или иную цепь ДНК в ответ на их инфекцию соответствующими хелперными фагами [Konings et al., 1987].

Другие авторы сообщили об успешном получении подобной серии субклонов в фаговом векторе M13, где однонаправленное укорочение вставок осуществлялось благодаря  $3' \rightarrow 5'$ -экзонуклеазной активности T4 ДНК-полимеразы [Dale et al., 1985]. Впоследствии этот подход был несколько модифицирован и позволил упростить процедуру заключительного лигирования делеционных вариантов вставки с одноцепочечным вектором [Milavetz, 1992]. Схема последовательно протекающих ферментативных реакций при создании делеционных конструкций вставки данным способом выглядит следующим образом (рис. 8.13). На первом этапе происходит отжиг специального олигонуклеотидного праймера  $5'$ -CGACGCCAGTGCCAAGCTTXXXXX- $3'$ , гомологичного участку векторной молекулы (в рассматриваемом примере M13mp18), непосредственно прилегающему к полилинкеру и включающему в свой состав сайт узнавания крайней в данном полилинкере рестрикционной эндонуклеазы *Hind*III (подчеркнуто) и пять переменных нуклеотидов, которые могут быть любыми. Таким образом, данный олигонуклеотид представляет собой набор из 1024 всех возможных вариантов. После его отжига на одноцепочечной матрице ДНК образуется двуцепочечный участок с сайтом узнавания фермента *Hind*III. Чтобы исключить теоретическую вероятность расщепления одноцепочечной последовательности вставки данной рестрикционной эндонуклеазой (как известно, многие ферменты, хотя и с низкой эффективностью, способны расщеплять одноцепочечные молекулы ДНК) в реакционную смесь добавлялся SSB-белок, связывающийся с одноцепочечной ДНК и защищающий ее от расщепления. В результате такой защиты отпала необходимость в предварительном рестриктазном картировании вставки на предмет отсутствия в ней сайта(ов) данного фермента. После линейаризации по небольшому двуцепочечному участку рестрикционной эндонуклеазой *Hind*III одноцепочечной ДНК вектора и вставки, в реакционную среду вносили фермент ДНК-полимеразу фага T4, которая в отсутствие дНТФ проявляет  $3' \rightarrow 5'$ -экзонуклеазную активность. Инкубация с этим ферментом продолжалась разные промежутки времени (от 5 до 30 мин) с отбором аликвот протекающей реакции. В результате ферментативного действия T4 ДНК-полимеразы (ее экзонуклеазной актив-

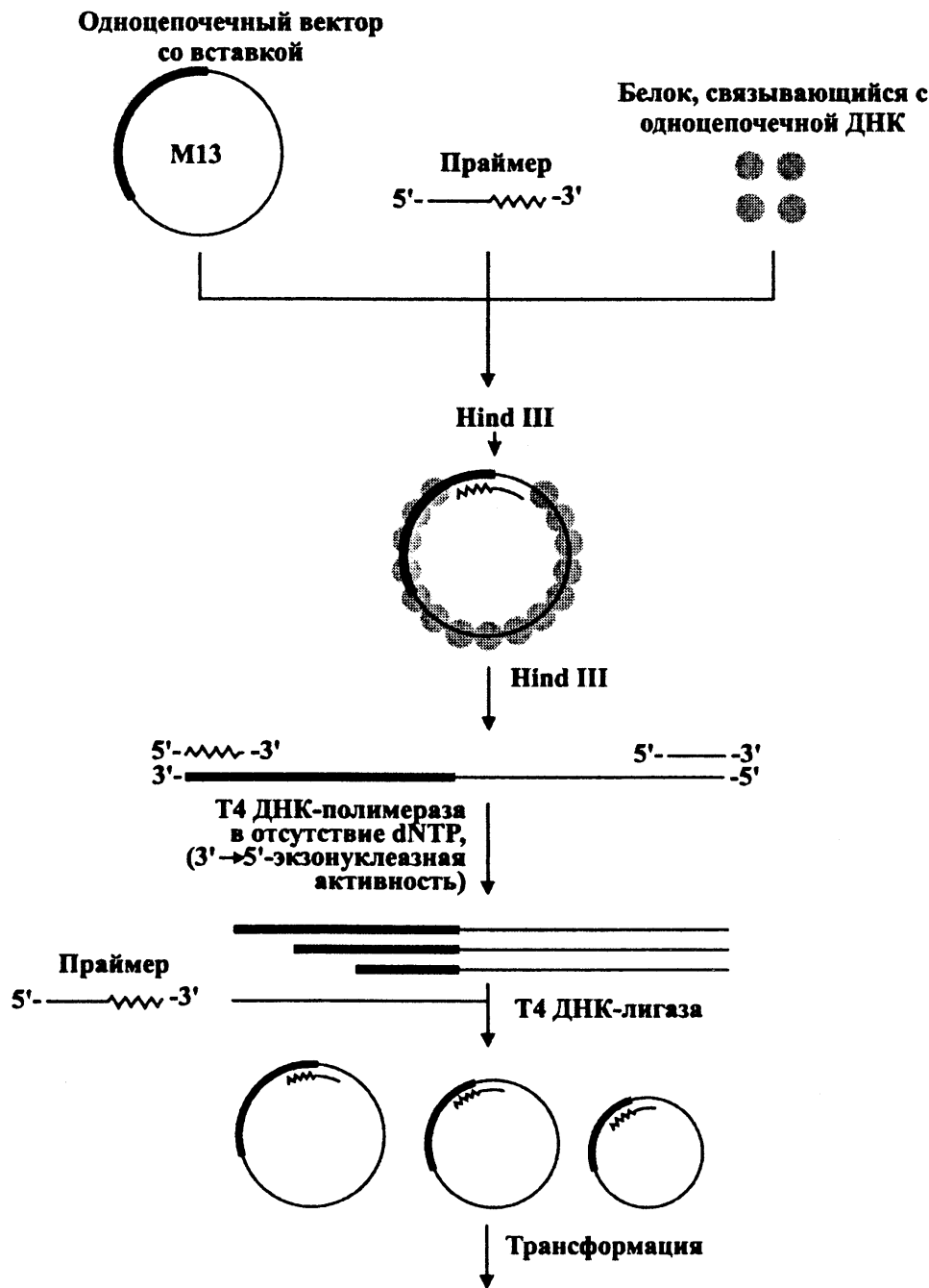


Рис. 8.13. Стратегия получения однонаправленных делеций вставки в составе одноцепочечного вектора M13 с помощью Т4 ДНК-полимеразы

ности) происходило как полное исчезновение небольшого двуцепочечного участка за счет расщепления остатков праймера, начиная с его 3'-конца, так и образование по-разному укороченных фрагментов вставки. После инактивации Т4 ДНК-полимеразы в реакционную смесь добавлялась новая порция указанного выше олигонуклеотидного праймера (названного автором "лигамер") и фермент Т4 ДНК-лигаза. На этой стадии происходил повторный отжиг части данного праймера, гомологичной векторной последовательности с вектором и отжиг какого-либо подходящего варианта (из 1024 возможных) со случайными концами

вставки, образованными под действием T4 ДНК-полимеразы и их лигирование. Лигированная одно/двучепочечная ДНК использовалась для трансформации компетентных клеток *E.coli*.

Ранее был предложен имеющий много общего с описанными выше методами подход к получению делеционных субклонов в одноцепочечном векторе, основанный на построении с помощью Кленовского фрагмента ДНК-полимеразы I комплементарной цепи участков вставки, имеющих разную протяженность [Burton et al., 1988]. В цитируемой работе осуществлялся отжиг олигонуклеотидного праймера, в результате ферментативного удлинения которого происходило увеличение двучепочечного участка вставки, причем производимый по времени отбор аликвот данной реакции обеспечивал получение их разной протяженности (рис. 8.14). Далее остающийся недостроенный одноцепочечный участок вставки и самого вектора удалялся нуклеазой S1. Затем с помощью подходящей рестрикционной эндонуклеазы, сайт которой находился в полилинкере поблизости от места клонирования вставки, “вырезались” делетированные варианты вставки. Их дальнейшее лигирование с заранее приготовленным вектором проводилось с таким расчетом, что происходила реориентация вставки, позволяющая секвенировать фрагменты ДНК со стороны делеций, давая, таким образом, возможность получить перекрывающиеся участки секвенируемой последовательности ДНК.

Интересный метод для создания одноцепочечных матриц с разной протяженностью одноцепочечного участка при секвенировании крупных фрагментов ДНК, не требующий нового этапа клонирования, заключался в линеаризации вектора со вставкой подходящей рестрикционной эндонуклеазой и последующей обработкой обоих концов с помощью экзонуклеазы III [Guo, Wu, 1982]. В этом случае знание рестриктазной карты также являлось необходимым условием, поскольку на первом этапе желательно использовать полилинкерный фермент, отсутствующий во вставке, тогда как после завершения экзонуклеазной обработки, которая обычно длится различные промежутки времени, необходимо проводить расщепление уже самой вставки рестрикционными эндонуклеазами, сайты которых предварительно картированы. Другой особенностью этого способа является ненужность секвенирующего праймера, поскольку в качестве затравки выступает остающийся после экзонуклеазной обработки участок двучепочечной ДНК. Так, на первом этапе проводится расщепление вектора со вставкой. Затем следует обработка экзонуклеазой III в течение дискретных промежутков времени, приводящая к образованию целого спектра одно/двучепочечных фрагментов ДНК. Следующий этап состоит в проведении непосредственно терминирующих реакций с дНТФ и соответствующими ддНТФ. Формирование необходимого для проведения секвенирующего электрофореза общего конца для каждого дискретного класса вновь синтезированных фрагментов ДНК осуществляется выбранными ранее рестрикционными эндонуклеазами, сайты которых локализованы во вставке в различных местах. Главной трудностью этого подхода можно считать подбор условий (времени) расщепления ДНК экзонуклеазой III

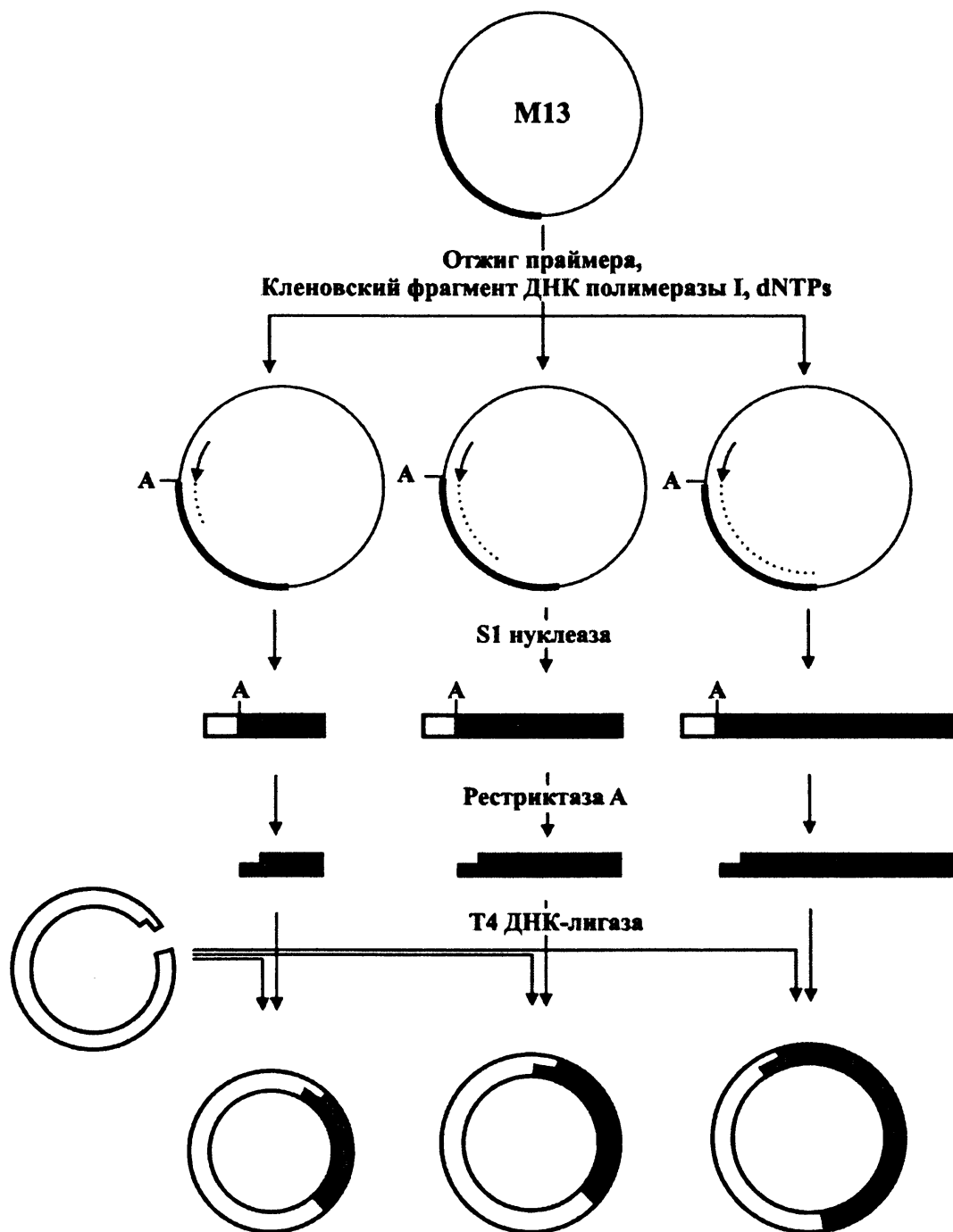


Рис. 8.14. Стратегия получения однонаправленных делеционных субклонов секвенируемой вставки в одноцепочечном векторе путем построения второй цепи клонированного фрагмента ДНК Кленовским фрагментом ДНК-полимеразы I дискретным удлинением праймера

и совмещение мест расщепления рестрикционными эндонуклеазами, чьи сайты расположены во вставке с целью получения набора гетерогенных по 3'-концам и гомогенных по 5'-концу секвенируемых фрагментов ДНК определенного размера, позволяющего осуществить их эффективное разделение с помощью гель-электрофореза.

В результате одновременного электрофоретического разделения продуктов двух секвенирующих реакций, образовавшихся при синтезе

новых цепей ДНК, на противоположных концах линейизованной матрицы со вставкой, возможность “читать” последовательность нуклеотидов вставки существует из-за значительной разницы в размерах построенных цепей, что предопределено выбором для расщепления подходящей рестрикционной эндонуклеазы. Так, при использовании фермента Кленовского фрагмента ДНК-полимеразы I, имеющего низкую процессивность, максимальный размер получаемых фрагментов при проведении терминирующих реакций обычно не превышает 350–500 нуклеотидов, а поскольку новосинтезированная цепь ДНК со стороны вектора за счет асимметричного расщепления второй рестрикционной эндонуклеазой имеет гораздо большую протяженность, то при проведении гель-электрофореза терминированные меченые фрагменты, принадлежащие участку вектора, будут расположены в верхней части геля и не будут, таким образом, интерферировать с секвенируемой областью вставки, что можно видеть из схемы, приведенной на рис. 8.15.

Способ получения однонаправленных делеционных вариантов исходной вставки, названный авторами ExoMeth [Sorge, Blinderman, 1989; Sorge, Huse, 1994], несколько напоминает предыдущий вариант получения одноцепочечных матриц ДНК для их секвенирования [Guo, Wu, 1982]. Общее у них то, что нет необходимости в использовании специальных олигонуклеотидных праймеров, поскольку остающиеся двуцепочечные участки молекулы ДНК способны инициировать построение комплементарной цепи ДНК ДНК-полимеразой. Первые этапы подготовки матриц ДНК этим способом, включая расщепление двумя подходящими рестрикционными эндонуклеазами и однонаправленное удаление цепи ДНК под действием экзонуклеазы III, совпадают с таковыми при стандартном варианте использования экзонуклеазы III [Henikoff, 1984; Yanisch-Petron et al., 1985; Henikoff, 1987]. Главная особенность рассматриваемого способа (рис. 8.16) заключается в том, что в процессе проведения терминирующих реакций вместо дЦТФ в реакционной смеси присутствует 5-метил-дЦТФ. Таким образом, вновь синтезированная цепь ДНК получается метилированной, что имеет важное значение при ее последующем расщеплении рестрикционными эндонуклеазами, чувствительными к метилированию цитозина. В результате использования в условиях недорестрикции какой-нибудь подобной частоты рестабилизующей рестрикционной эндонуклеазы образуется смесь терминированных дидезоксинуклеотидами фрагментов, общим 5'-концом для которых будут служить сайты расщепления этого фермента. А поскольку для одного и того же образца экзонуклеазная реакция длилась различные дискретные промежутки времени, то при этом образовывались соответствующие наборы терминированных фрагментов ДНК. И для каждого такого набора будет иметься свой общий 5'-конец, совпадающий с ближайшим сайтом используемой рестрикционной эндонуклеазы, что позволит провести их электрофоретическое разделение и “прочитать” последовательность нуклеотидов.

Пожалуй, стоит отдельно упомянуть одну работу [Li, Tucker, 1993], в которой авторы предложили метод секвенирования протяженных фрагментов ДНК с помощью экзонуклеазы III, названный ими “экзок-

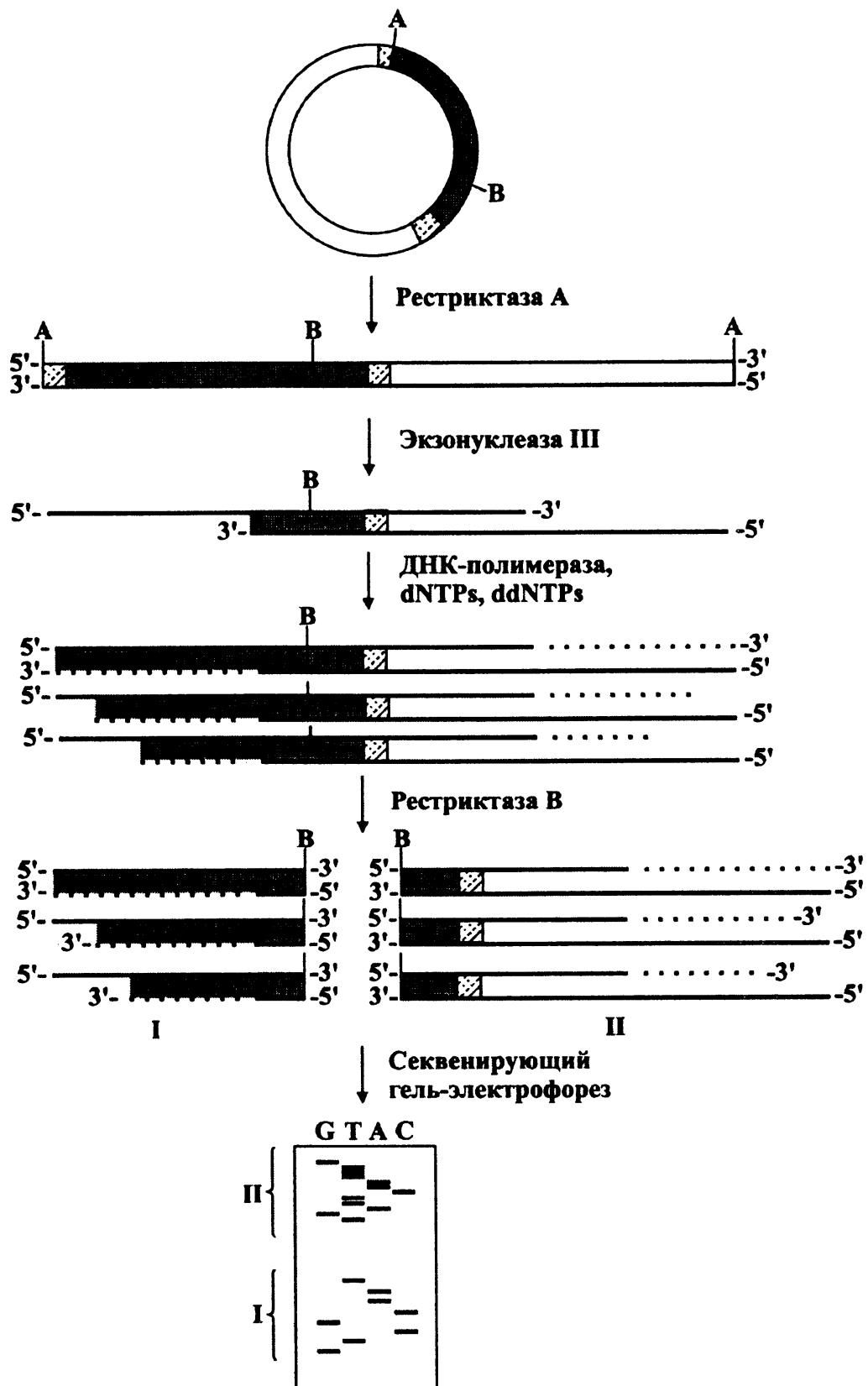


Рис. 8.15. Стратегия получения дискретных матриц ДНК с различной протяженностью одноцепочечного участка с помощью экзонуклеазы III



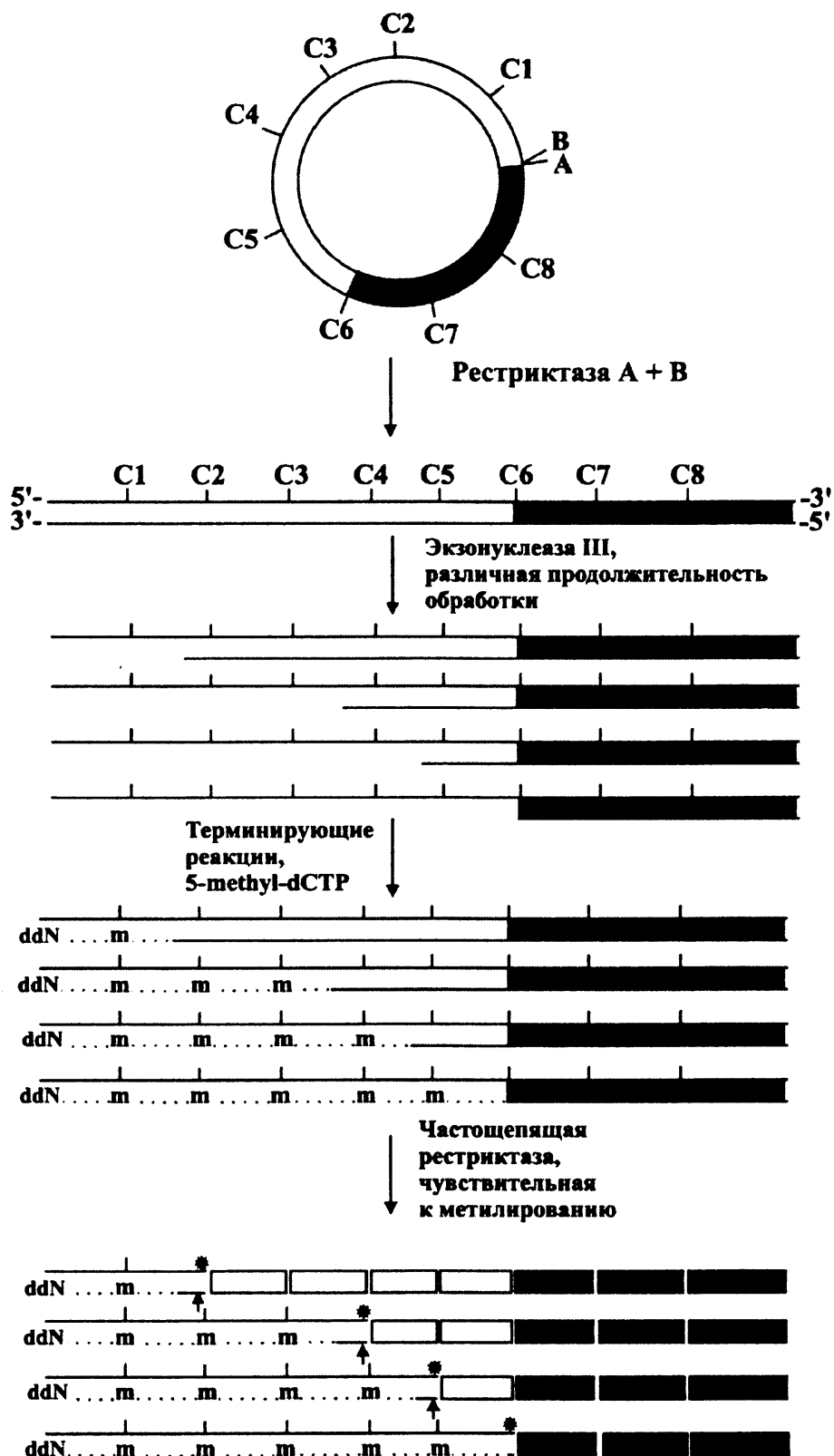


Рис. 8.16. Стратегия определения нуклеотидной последовательности протяженного фрагмента ДНК ферментативным методом в присутствии в реакционной смеси 5-метил-дЦТФ после образования одноцепочечных матриц ДНК с помощью экзонуклеазы III

венс" (exoquence). Данный метод очень сильно напоминает описанную выше ExoMeth стратегию [Sorge, Blinderman, 1989], где затравочными молекулами служат остающиеся двуцепочечные участки ДНК, но в отличие от нее не требует применения весьма дорогого модифицированного 5-метил дЦТФ. Однако предложенная авторами схема такого секвенирования ДНК вызывает некоторое недоверие в возможности получения набора вновь синтезированных молекул ДНК, имеющих уникальный 5'-конец, поскольку для достижения поставленной цели требуется проведение экзонуклеазной обработки в условиях, не дающих большого разброса по протяженности удаленного участка одной цепи ДНК, а также удачное расположение сайтов рестрикционных эндонуклеаз во вставке, которые используют в этом методе для образования гомогенных 5'-концов секвенируемых молекул ДНК после проведения терминирующих реакций. Авторы проанализировали эффективность работы экзонуклеазы III в зависимости от температуры инкубации, стараясь достичь минимального разброса по длине укороченного участка. Однако, на наш взгляд, для осуществления этого подхода необходимо полномасштабное рестриктазное картирование вставки с помощью большого числа редкощепящих рестрикционных эндонуклеаз для выявления наиболее подходящих при формировании гомогенного 5'-конца у терминированных фрагментов ДНК. Тогда как авторы цитируемой работы сообщают о большом числе использованных частощепящих тетра- и пентануклеотидных рестрикционных эндонуклеаз, что в принципе невозможно, поскольку это неминуемо приведет к расщеплению и оставшегося нетронутым двуцепочечного фрагмента ДНК вставки (что не страшно, так как он остается немеченным и не будет виден на заключительном радиоавтографе) и, главным образом, участка ДНК с новосинтезированной в условиях терминации цепью, которая у них не защищена метилированием цитозина, как в ExoMeth стратегии [Sorge, Blinderman, 1989; Sorge, Huse, 1994]. Таким образом, вероятность возникновения гетерогенных по 5'-концу секвенируемых фрагментов ДНК очень высока, что вызовет наложение полос ДНК из разных генераций, произведенных действием частощепящей рестрикционной эндонуклеазы, и не позволит провести однозначное "чтение" нуклеотидной последовательности какого-либо фрагмента.

В связи с увеличением числа обнаруженных рестрикционных эндонуклеаз II типа, имеющих различные сайты узнавания, а также с созданием векторов с различными полилинкерами и даже суперполилинкерами, как, например, в фагмиде pSL301 [Brosius, 1989], рестриктазное картирование вставки стало проводиться с использованием, как правило, гексануклеотидных рестрикционных эндонуклеаз, в первую очередь имеющихся в полилинкере используемого вектора или же отсутствующих во всей последовательности вектора вообще. Выявление таких ферментов (как присутствующих, так и тех, сайты которых отсутствовали во вставке) сделало возможным на основе этой же рекомбинантной плазмиды провести клонирование II порядка или как его еще называют – дельта-рестрикционное клонирование и за счет этого достаточно просто получить серию укороченных вариантов вставки [Ansoerge

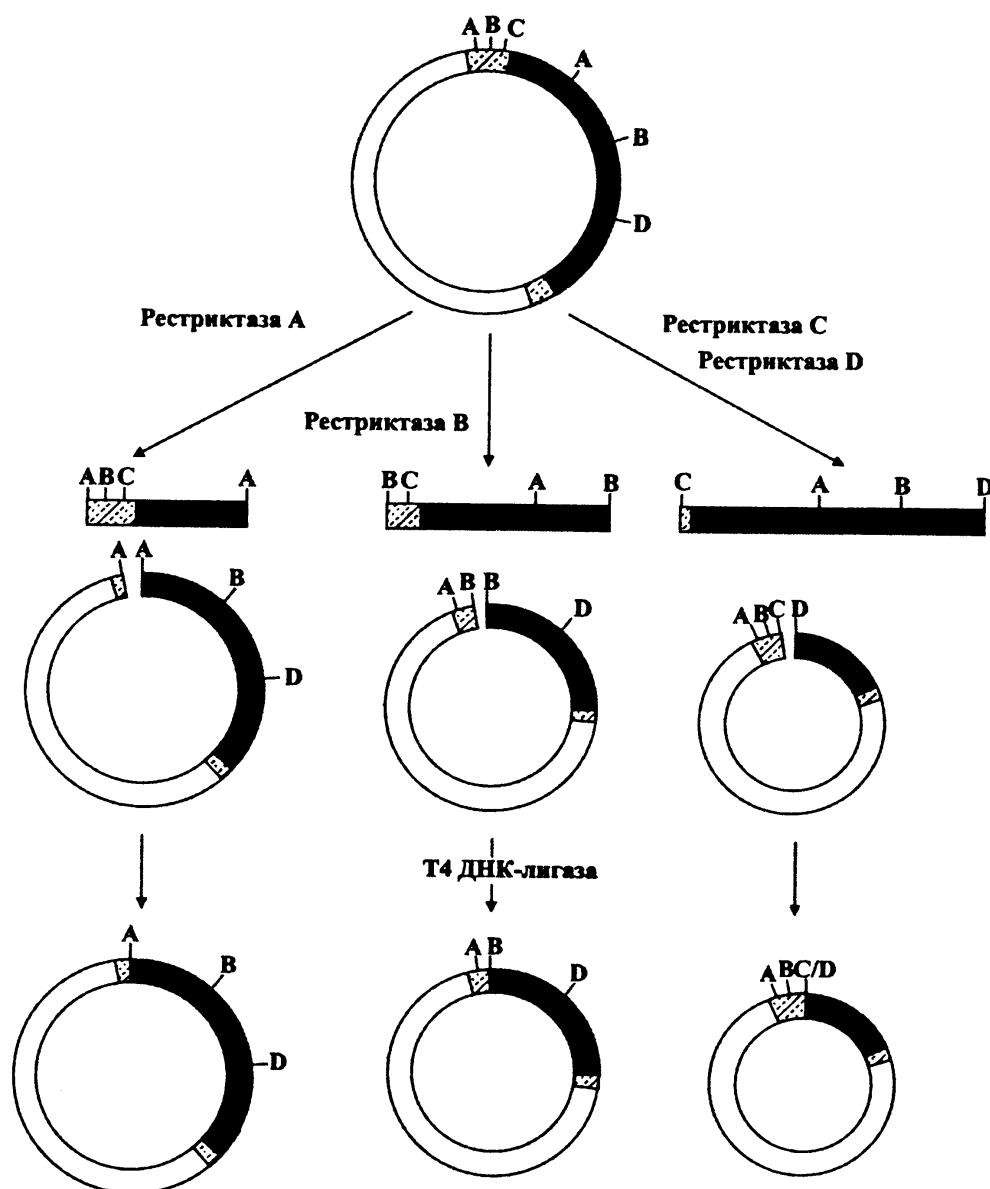


Рис. 8.17. Схема дельта-рестрикционного клонирования для создания однонаправленных делеций вставки и получения необходимых субклонов при секвенировании протяженных фрагментов ДНК

et al., 1996]. Подобная стратегия уже упоминалась выше, поскольку также применима и для метода секвенирования ДНК химической деградацией, хотя с наибольшей эффективностью она используется все же при ферментативном секвенировании ДНК.

Из схематического представления такого способа получения нужных клонов, показанного на рис. 8.17, видно, что дельта-рестрикционное клонирование позволяет удалять из вставки фрагменты ДНК разного размера и за счет этого получать серию однонаправленных делеционных (со стороны места для отжига праймера) субклонов, в отдельных случаях полностью перекрывающую всю длину вставки. Следует отметить, что большее число сайтов удобных рестрикционных эндонуклеаз в полилинкере и одновременно уменьшенное число сайтов других гексануклеотидных рестрикционных эндонуклеаз в остальной последо-

вательности вектора заметно увеличивало возможности этого способа получения необходимых субклонов для завершения проекта по секвенированию ДНК.

Так, идеальными вариантами являются случаи нахождения во вставке сайтов полилинкерных рестрикционных эндонуклеаз (или тех, что образуют фрагменты ДНК с совместимыми с ними концами при условии отсутствия сайтов расщепления этих ферментов в самом векторе), поскольку в этом случае образуются одинаковые концы, принадлежащие как полилинкеру, так и делетированной вставке. Чуть менее удобным вариантом можно считать образование после расщепления полилинкера и вставки разными ферментами фрагментов ДНК с тупыми концами из-за худшей эффективности лигирования последних по сравнению с липкими концами. Еще менее удобными являются случаи несовпадения образуемых под действием соответствующих ферментов концов фрагментов ДНК, что приводит к необходимости проведения дополнительных ферментативных этапов, чтобы сделать их совместимыми между собой. Другое неудобство заключается в том, что почти все полилинкерные рестрикционные эндонуклеазы являются ферментами с гексануклеотидными участками узнавания, что делает вероятность нахождения таких сайтов во вставке не очень большой, поскольку теоретическая частота встречаемости подобных сайтов в произвольной последовательности ДНК составляет один гексануклеотидный сайт на 4096 пн. Таким образом, в отдельных случаях не удастся получить необходимый набор субклонов, позволяющий “прочитать” всю нуклеотидную последовательность клонированного фрагмента ДНК, что приводит к образованию иногда протяженных брешей. Использование же более частощеплящих ферментов сопряжено с разрушением самого вектора, так как в нем также содержатся множественные сайты расщепления таких частощеплящих рестрикционных эндонуклеаз. Следовательно, чтобы получить необходимый набор субклонов, полностью перекрывающий всю последовательность протяженного фрагмента ДНК, нужно для однонаправленного делеционного укорочения вставки использовать частощеплящие рестрикционные эндонуклеазы, не допуская при этом разрушения самого вектора. Подобная стратегия, названная нами тетра/окта стратегия секвенирования, стала возможной благодаря созданию специального вектора и будет рассмотрена чуть ниже. А сейчас, видимо, стоит уделить определенное внимание подходу к получению однонаправленных делеционных субклонов, имеющему нечто общее с дельта-рестрикционным клонированием и с тетра/окта стратегией. Этим общим является получение прогрессивных однонаправленных делеций вставки, а также использование частощеплящей тетрануклеотидной рестрикционной эндонуклеазы для получения необходимых субклонов в исходном векторе.

Так, интересным способом определения нуклеотидной последовательности протяженных фрагментов ДНК является один из вариантов направленного подхода с использованием рестрикционной эндонуклеазы *Bam*HI имеющей гексануклеотидный сайт узнавания  $G^{\vee}GATCC$ , и рестрикционной эндонуклеазы *Sau*3AI с тетрануклеотидным сайтом уз-

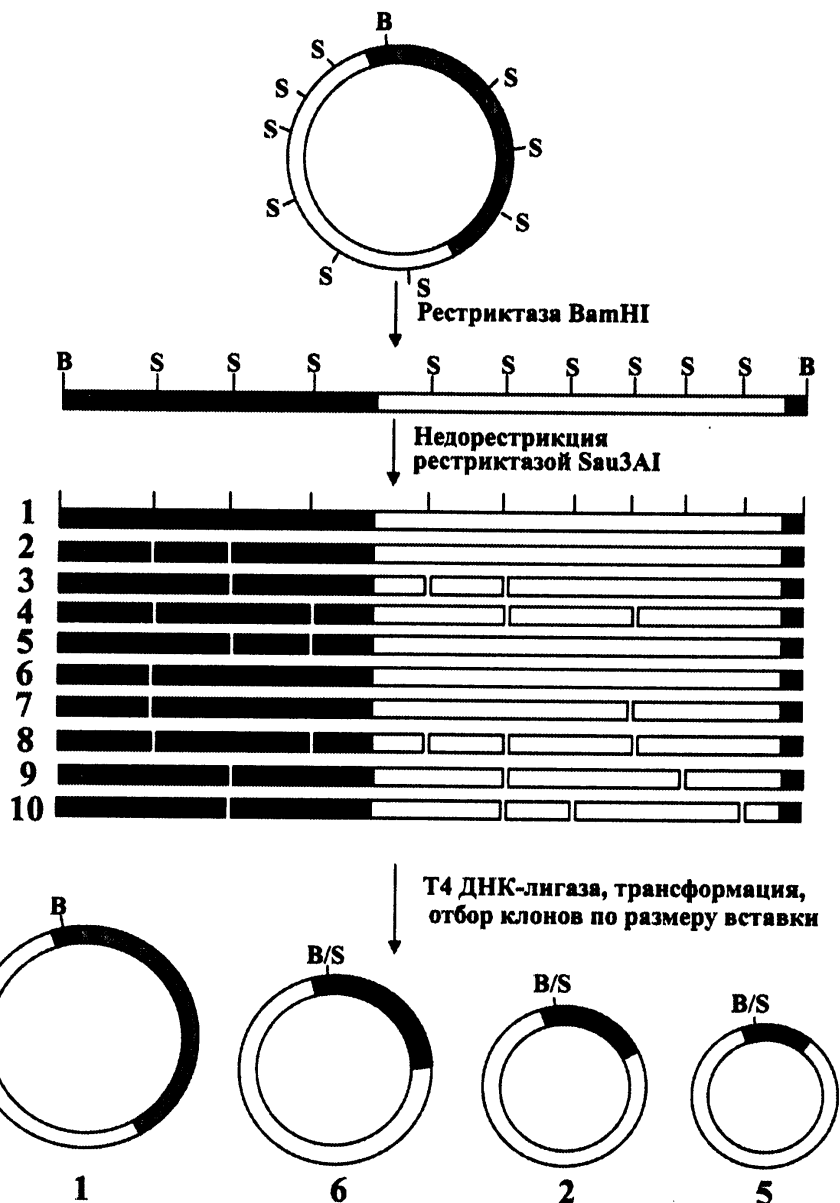


Рис. 8.18. Стратегия получения однонаправленных делеционных субклонов в исходном векторе путем частичного расщепления последнего вместе со вставкой частотщепящей тетра nukлеотидной рестрикционной эндонуклеазой *Sau3AI*

навания  $\nabla$ GATC и генерирующих пригодные для лигирования друг с другом выступающие концы [Lee, Lee, 1988]. Данный способ, не получивший широкого распространения ввиду ряда ограничений, о которых будет сказано ниже, основан на использовании подхода с частичным расщеплением вектора и вставки рестрикционной эндонуклеазой *Sau3AI* (рис. 8.18). При осуществлении этого способа секвенирования протяженных фрагментов ДНК также необходим предварительный этап, во время которого проводится рестриктазное картирование с целью выявления полилинкерных рестрикционных эндонуклеаз, отсутствующих во вставке. Более того, ключевым моментом является необходимость отсутствия в клонированном фрагменте ДНК сайтов узнавания для рестрикционной эндонуклеазы *BamHI*, поскольку в противном случае данный подход становится невозможным. Хотя, справедливости ра-

ди, следует отметить возможность использования других пар гексануклеотидных и тетрануклеотидных рестрикционных эндонуклеаз, как например, *Bgl*III (A<sup>∇</sup>GATCT) и та же *Sau*3AI, *Sph*I (GCATG<sup>∇</sup>C) и *Nla*III (CATG<sup>∇</sup>) и др.

Первый этап данного метода заключается в линеаризации вектора, несущего вставку, рестрикционной эндонуклеазой *Bam*HI. Следующим этапом является расщепление и вектора и вставки в условиях недорестрикции частощепляющим ферментом *Sau*3AI. Так как частота встречаемости какой-либо четверки нуклеотидов довольно высока и приходится в среднем по одному подобному участку на каждые 256 пн, то в любых протяженных вставках теоретически также должно быть несколько сайтов этой тетрануклеотидной рестрикционной эндонуклеазы. Поскольку в используемом авторами векторе на основе фага M13 имеется 6 сайтов узнавания для рестрикционной эндонуклеазы *Sau*3AI, то в результате неполного расщепления образуется набор различных фрагментов, включающих в свой состав укороченные варианты вставки и полноразмерный вектор, но обязательно будут присутствовать и варианты с полной вставкой и укороченным вектором, равно как и различные формы и с укороченными вставкой и вектором одновременно. Во время следующего этапа лигирования T4 ДНК-лигазой происходит рециркулизация данных рестриктазных фрагментов и образуются, в том числе, искомые варианты с полноразмерным вектором и укороченными вставками. Следует отметить, что при инфекции штамма *E.coli* вариантами с укороченным вектором, приводящими к нарушению его нормального функционирования, фаговые бляшки на бактериальном газоне не образуются. В то же время существует высокая вероятность восстановления исходного рекомбинантного клона с полноразмерной вставкой, поскольку расщепление вектора со вставкой рестрикционной эндонуклеазой *Sau*3AI проводится в условиях сильной недорестрикции (из-за опасения необратимо разрушить вектор у большей части клонов), в результате чего многие вставки не успевают подвергнуться расщеплению. Для увеличения выхода искомых укороченных вариантов вставки применяют после лигирования еще один этап расщепления одной из полилинкерных гексануклеотидных рестрикционных эндонуклеаз, отсутствующих во вставке, причем расположение сайта узнавания используемого для этой цели фермента относительно сайта клонирования и сайта рестрикционной эндонуклеазы *Bam*HI должно быть определенным образом ориентировано, в результате чего восстановившиеся и сохранившиеся исходные рекомбинантные конструкции линеаризуются, а, как известно, частота трансформации компетентных клеток *E.coli* линейными молекулами во много раз ниже, чем кольцевыми. Существенным недостатком данного метода является обязательный предварительный этап рестриктазного картирования, имеющий целью выявить, главным образом, отсутствие во вставке сайтов узнавания фермента *Bam*HI, а кроме этого – выяснение отсутствия во вставке сайтов других полилинкерных рестрикционных эндонуклеаз для возможности их последующего использования. Причем с увеличением протяженности вставки вероятность нахождения таких ферментов (и в том числе веро-

ятность отсутствия во вставке сайтов расщепления рестрикционной эндонуклеазы *Bam*HI) пропорционально уменьшается.

Весьма похожая работа по получению однонаправленных делеционных субклонов в фаговом векторе M13 с помощью частичного расщепления вектора и вставки частощеплящими рестрикционными эндонуклеазами после первого этапа линейаризации вектора уникальной гексануклеотидной рестрикционной эндонуклеазой была проведена чуть позже другими авторами [Lamperti, Villa-Komaroff, 1990]. Пожалуй, незначительные отличия данной работы от предыдущей состояли как в дополнительном этапе в виде теста на присутствие места отжига секвенирующего праймера в полученных делеционных субклонах, так и в отсутствии повторного расщепления лигированных конструкций. Поскольку сам вектор, как и вставка, также подвергался расщеплению частощеплящими ферментами, это приводило к образованию нежизнеспособных форм, однако, как справедливо отмечают авторы, не исключены случаи возникновения небольших делеций, не влияющих на жизнеспособность фага. В связи с этим данный тест на сохранность места отжига праймера, проводимый путем отжига праймера и ферментативного включения радиоактивного дНТФ с последующим контролем продуктов реакции с помощью радиоактивного счетчика, выглядит вполне логичным. Введение этого этапа не снимало желательности повторного расщепления лигированных конструкций, поскольку они направлены на достижение разных целей. Так, повторное расщепление определенной полилинкерной рестрикционной эндонуклеазой продуктов лигазной реакции рассчитано на обогащение получаемых субклонов именно делеционными вариантами вставки и его проведение может привести к высокому содержанию среди конечных субклонов вариантов с исходной неукороченной вставкой. Следует отметить, что несмотря на сильное сходство этих двух работ, авторы последней [Lamperti, Villa-Komaroff, 1990] в своем изложении не упоминали работу своих предшественников [Lee Y.-M., Lee S.-C., 1988].

Здесь же можно упомянуть работу еще одних авторов, посвященную получению однонаправленных делеционных субклонов секвенируемой вставки путем ее недорасщепления [Abath, Holder, 1995]. Однако серьезным ограничением предложенного ими варианта являлось то, что для частичного расщепления вставки использовался фермент с гексануклеотидным участком узнавания, а, как уже отмечалось выше, частота встречаемости таких сайтов для подобных экспериментов недостаточна и вероятность получения в этом случае полностью перекрывающихся субклонов крайне низка. Более того, принципиальным моментом является то, что исследователю должно быть заранее известно о наличии в секвенируемой вставке нескольких сайтов конкретного используемого фермента.

Почти все рассмотренные выше стратегии секвенирования протяженных фрагментов ДНК в составе фаговых, плазмидных или фагмидных векторов направленного подхода требуют в той или иной степени предварительного этапа рестриктазного картирования вставки с целью выявления отсутствия в ней сайтов узнавания полилинкерных гексану-

клеотидных рестрикционных эндонуклеаз. Причем, чем больше размер вставки, тем больше вероятность, что такие сайты во вставке все же присутствуют и, таким образом, соответствующие им рестрикционные эндонуклеазы не могут быть использованы для получения однонаправленных делеционных субклонов.

Ликвидация предварительного этапа – рестриктазного картирования вставки возможна за счет использования для расщепления вектора (в участке полилинкера) рестрикционными эндонуклеазами не с гекса-, а октануклеотидными участками узнавания. Теоретическая частота встречаемости последних составляет один сайт на 65536 пн, против одного сайта на 4096 пн для гексануклеотидных рестрикционных эндонуклеаз. Приведенные частоты встречаемости сайтов октануклеотидных рестрикционных эндонуклеаз рассчитаны для ДНК с равным соотношением GC- и AT-пар, но поскольку содержание GC-пар в геномах различных организмов варьирует в широких пределах, то, следовательно, частоты встречаемости участков узнавания для подобных ферментов с GC- или AT-богатыми сайтами узнавания в таких AT- или GC-богатых ДНК будут заметно различаться. Так, для ДНК даже с 60%-ным составом GC-пар будет теоретически приходиться по одному участку АТТГАААТ, узнаваемому октануклеотидной рестрикционной эндонуклеазой *SwaI*, на каждые 390625 пн. Такая же редкая частота встречаемости будет характерна и для рестрикционной эндонуклеазы *SrfI* с октануклеотидным сайтом узнавания GCCCGGGC в ДНК с 60% AT-пар. Для ДНК с 70% AT- или GC-пар теоретически частоты встречаемости будут уже составлять один сайт на почти 4 млн (3901844 пн) для рестрикционных эндонуклеаз *SrfI* и *SwaI* соответственно. Что касается ДНК с приблизительно равным соотношением AT- и GC-пар, для которых теоретическая частота встречаемости этих сайтов такова, что существует вероятность присутствия одного такого участка на 65536 пн, то это, по крайней мере, в 6–7 или большее число раз превышает длину вставок, поскольку размер клонированного фрагмента ДНК в векторах данного типа редко может достигать 10 тпн. Таким образом, благодаря наличию в полилинкере специально сконструированного фагмидного вектора двух сайтов узнавания для редкощеплящих октануклеотидных рестрикционных эндонуклеаз *SwaI* или ее изошизомера *SmiI* (АТТГАААТ) и *SrfI* (GCCCGGGC) появляется возможность использовать их для линеаризации вектора со вставкой, не проводя предварительного широкомасштабного рестриктазного картирования вставки. (Относительно редкие случаи наличия сайтов данных ферментов в предназначенном для секвенирования клонированном фрагменте ДНК и возможные последствия этого достаточно легко преодолимы.)

Во избежание недоразумения считаем важным уточнить, что обычно под GC-составом ДНК какого-то организма понимается соотношение нуклеотидов в обеих цепях ДНК, тогда как при анализе нуклеотидных последовательностей обычно рассматривается только одна цепь ДНК (РНК-подобная, для тех случаев, когда это известно). Так, если какая-либо ДНК содержит 50% GC-пар, то это не означает, что в одной цепи ДНК обязательно будет находиться по 25% аденина, тимидина, гу-



анина и цитидина, хотя это и возможно. Теоретически может оказаться так, что, скажем, аденины и цитидины будут находиться преимущественно в одной цепи, а гуанины и тимины – в другой. При этом общее содержание АТ- и GC-пар для организма остается неизменным, но частота встречаемости сайтов рестрикционных эндонуклеаз будет весьма далекой от той, как если бы все нуклеотиды были распределены по длине одной цепочки ДНК весьма равномерно. Конечно, чем крупнее исследуемый фрагмент ДНК, тем более равномерно в среднем (теоретически) должны быть распределены в нем все характерные для данного организма, которому принадлежит участок ДНК, нуклеотиды, тогда как на ограниченных отрезках ДНК могут быть заметные отклонения, вызванные, например, большими блоками повторов с “простой” последовательностью.

Стратегии секвенирования протяженных фрагментов ДНК в составе фагмидного вектора путем получения субклонов II порядка с использованием гексануклеотидных рестрикционных эндонуклеаз, как уже отмечалось выше, достаточно просты по исполнению, но учитывая весьма редкую встречаемость их сайтов (теоретически один сайт на 4096 пн), часто бывает невозможно получить необходимый набор делеционных вариантов, позволяющих при секвенировании определить полную нуклеотидную последовательность всей вставки без промежутков. Решением проблемы могло бы быть использование для этой цели частощеплящих рестрикционных эндонуклеаз с тетра- и пентануклеотидными сайтами узнавания (теоретическая частота встречаемости которых составляет 1 сайт на 256 пн) в условиях недорестрикции. Однако, использование подобных ферментов для этой цели практически невозможно из-за присутствия множества сайтов этих рестрикционных эндонуклеаз в самой последовательности обычных фаговых, плазмидных или фагмидных векторов, что неминуемо приведет к их разрушению.

Большинства всех этих недостатков лишена новая прогрессивная стратегия секвенирования протяженных фрагментов ДНК на основе специально сконструированного вектора [Чемерис и др., 1996]. Сущность данной тетра/окта стратегии секвенирования заключается в получении серии субклонов с однонаправленными делеционными вариантами вставки, покрывающими всю ее длину, с использованием специализированного фагмидного вектора, в котором отсутствуют сайты расщепления для одной или нескольких рестрикционных эндонуклеаз с тетра- и пентануклеотидными сайтами узнавания и одновременно наличием в полилинкере этого вектора сайтов расщепления для одной или нескольких рестрикционных эндонуклеаз с октануклеотидными сайтами узнавания, путем последовательного расщепления вектора, несущего вставку, одной из этих полилинкерных октануклеотидных рестрикционных эндонуклеаз и затем уже расщепления вставки в составе линейаризованного вектора в условиях недорестрикции тетра- или пентануклеотидной рестрикционной эндонуклеазой, сайт расщепления которой отсутствует в векторной последовательности.

Схематично способ определения нуклеотидной последовательности протяженных фрагментов ДНК в составе фагмидного вектора

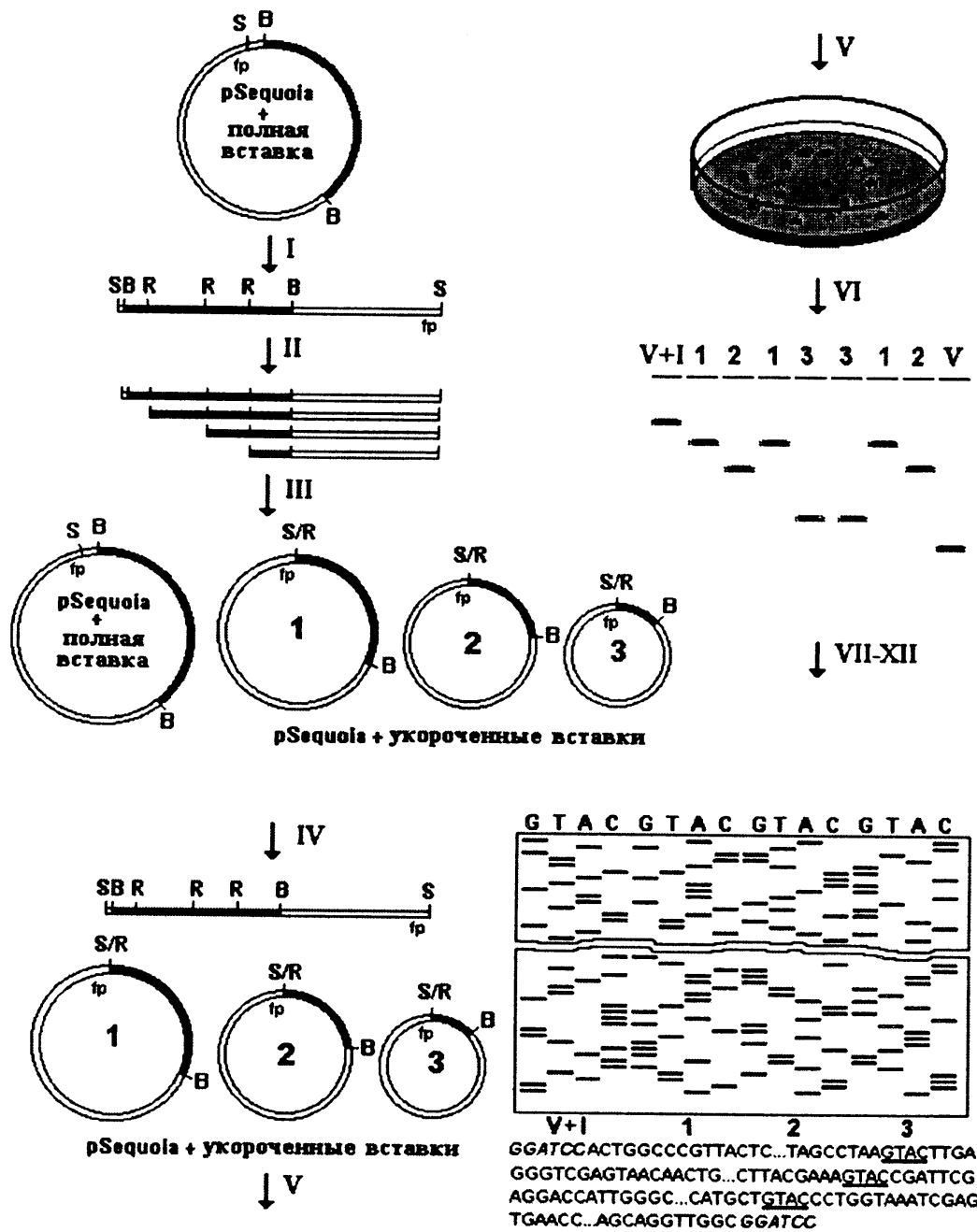


Рис. 8.19. Тетра/окта стратегия секвенирования протяженных фрагментов ДНК в специально сконструированном векторе рSequoiaT12

Пояснения см. текст

рSequoiaT12 с помощью тетра/окта стратегии секвенирования, включая этап повторного расщепления той же октануклеотидной рестрикционной эндонуклеазой, приведен на рис. 8.19, где “В”, “R”, “S” – сайты узнавания эндонуклеаз рестрикции *Bam*HI, *Rsa*I и *Srf*I или *Swa*I соответственно; “V” – вектор, “V + I” – вектор с полной вставкой, изображенной в виде затемненной части кольца; “1”, “2” и “3” – вектор с вариантами укороченной вставки; fp – место отжига праймера.

На I этапе вектор, несущий вставку, расщепляется одной из октануклеотидных рестрикционных эндонуклеаз, т.е. проводится его линейари-

зация. На II этапе осуществляется уже расщепление непосредственно самой вставки в составе линейаризованного вектора тетра nukлеотидным ферментом *RsaI* в условиях недорестрикции, в результате чего получаются все возможные варианты укорочения вставки. Различные *RsaI*-фрагменты клонированного фрагмента ДНК без векторной последовательности автоматически исключаются из дальнейшего обращения во время этапа трансформации и для упрощения схемы не показаны. III этап – лигирование продуктов рестрикции и недорестрикции T4 ДНК-лигазой, в результате которого происходит рециркулизация линейаризованного вектора с различными делеционными вариантами вставки и, в том числе, восстановление исходной рекомбинантной плазмиды с полноразмерной вставкой. В последнем случае восстанавливается сайт октануклеотидной рестрикционной эндонуклеазы, использовавшейся для первоначального расщепления, что очень важно для проведения следующего этапа. Необходимо отметить возможность случайного лигирования отдельных *RsaI*-фрагментов (на схеме не приведены) и данного вектора, несущего варианты вставки, приводящего к неправильному порядку их расположения, несовпадающему с таковыми в исходном клонированном фрагменте ДНК и их последующего клонирования. Однако вероятность этого довольно низка и, как будет показано ниже, не приведет к ошибке при воссоздании полной нуклеотидной последовательности всей вставки. На IV этапе проводится повторное расщепление вектора, лигированного с вариантами вставки, той же октануклеотидной рестрикционной эндонуклеазой, использовавшейся ранее для линейаризации, с целью обогащения выхода клонов с укороченными вставками, поскольку эффективность трансформации линейаризованных форм плазмиды значительно ниже по сравнению с кольцевыми формами, а восстановленная исходная плаزمиды с неукороченным вариантом вставки в этом случае расщепится и превратится в линейную форму. На V этапе осуществляется стандартная трансформация лигазной/рестриктазной смесью компетентных клеток *E.coli*. Следующий, VI этап, необходим для отбора вариантов, наиболее оптимально покрывающих всю длину вставки, путем анализа размера плазмид в различных субклонах. С этой целью проводятся стандартные процедуры по микровыделению рекомбинантных плазмидных ДНК и анализу их размера гель-электрофорезом (на схеме показаны только суперскрученные формы плазмид). VII–XII этапы посвящены приготовлению матриц для проведения терминирующих реакций, проведению самих этих реакций, разделению продуктов реакций в секвенирующем геле, радиоавтографии и чтению нуклеотидной последовательности. На приведенной схеме “прочитанной” нуклеотидной последовательности всей вставки, сайты клонирования (как довольно удобные и часто используемые) по рестрикционной эндонуклеазе *BamH* (GGATCC) показаны курсивом, сайты *RsaI* (GTAC), являющиеся местами состыковки различных субклонов и служащие в качестве удобных ориентиров, подчеркнуты. Потенциальные ошибочные субклоны с неправильным порядком и/или ориентацией *RsaI*-фрагментов, которые могут возникнуть при случайном лигировании, на стадии чтения и состыковки нуклеотид-

ной последовательности будут выявлены за счет несовпадения участков перекрытия и исключены. Для разделения продуктов терминирующих реакций, проведенных с одноцепочечными ДНК-матрицами тех субклонов, которые наиболее оптимально покрывают всю длину вставки, их наносят на секвенирующий гель в порядке уменьшения размера вставок, начиная с исходного клона с полноразмерной вставкой. Нанесение образцов в соседние дорожки геля внутри каждой матрицы в виде четверок реакций GTAC (совпадающих с сайтом узнавания рестрикционной эндонуклеазы *RsaI*, с помощью которой и получались данные субклоны) образует на радиоавтографе в местах нахождения этой последовательности хорошо заметные ровные ступеньки из полос на соседних дорожках GTAC, служащие удобным ориентиром, значительно упрощающим состыковку нуклеотидных последовательностей разных субклонов при восстановлении полной последовательности клонированного фрагмента ДНК.

С использованием тетра/окта стратегии секвенирования возможно также упрощенное получение серии субклонов, несущих укороченные вставки требуемого размера, без трудоемкого анализа многочисленных колоний с целью определения размера вставок. Для этого проводится полупрепаративный электрофорез вектора со вставкой, последовательно расщепленных октануклеотидной и тетрануклеотидной эндонуклеазами рестрикции в условиях полной и частичной рестрикции соответственно, и полосы, содержащие вектор с различными укороченными вариантами вставки, вырезают из геля. Каждый образец ДНК, элюированный из этих кусочков геля, лигируют отдельно и отдельно трансформируют. Таким образом, при трансформации компетентных клеток *E.coli* на той или иной чашке Петри будут присутствовать колонии, несущие плазмиды с определенным, требуемым размером вставки. Отпадает также необходимость в проведении этапа повторного расщепления октануклеотидной рестрикционной эндонуклеазой, поскольку лигирование и трансформация компетентных клеток *E.coli* вектором, несущим исходную полноразмерную вставку, исключены. Практически отсутствует и вероятность случайного лигирования отдельных *RsaI*-фрагментов клонированного фрагмента ДНК и вектора, несущего различные варианты вставки.

Подобное получение серии субклонов с требуемыми размерами укороченной вставки при раздельном лигировании элюированных из геля образцов ДНК и их независимой трансформации возможно только с использованием вектора, в нуклеотидной последовательности которого отсутствуют сайты расщепления для тетрануклеотидных рестрикционных эндонуклеаз, каковым и является фагмидный вектор *pSequoiaT12*, поскольку при наличии в векторной последовательности сайтов узнавания фермента, используемого для недорестрикции, образуется чрезмерно сложная картина электрофоретического разделения, в которой выбрать необходимые полосы не представляется возможным.

Данный упрощенный подход способа определения нуклеотидной последовательности протяженных фрагментов ДНК в виде схемы приведен на рис. 8.20, где “S + R” – продукты рестрикции/недорестрикции

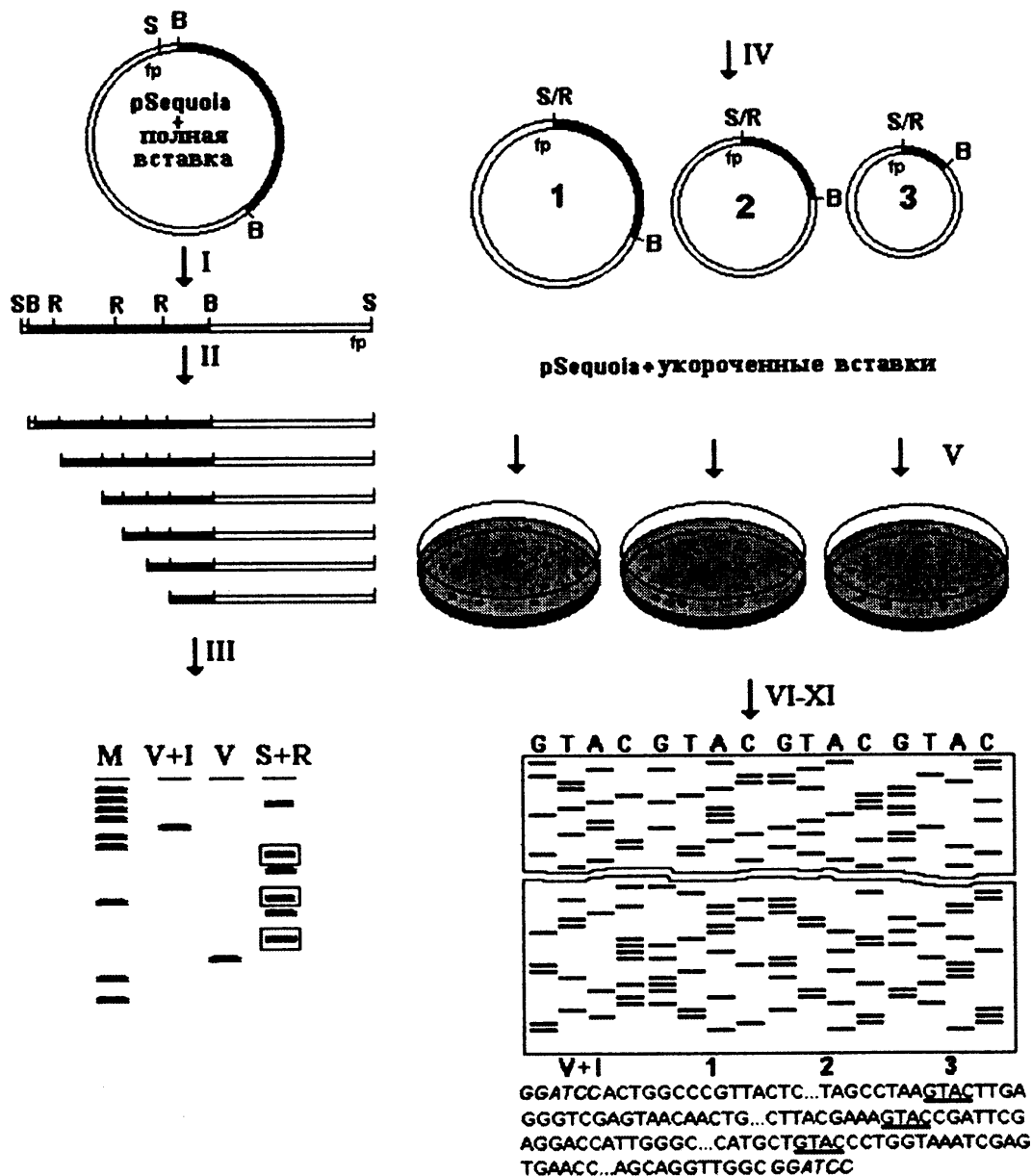


Рис. 8.20. Упрощенный способ тетра/окта стратегии секвенирования протяженных фрагментов ДНК в специально сконструированном векторе pSequoiaT12

исходной рекомбинантной плазмиды октануклеотидной (*SrfI* или *SwaI*) и тетрануклеотидной (*RsaI*) рестриktionными эндонуклеазами; "M" – ДНК фага лямбда, расщепленная ферментом *BstEII*, а все остальные обозначения как в рис. 8.19.

На I и II этапах, как и в описанном выше способе, вектор, несущий вставку, расщепляется одной из октануклеотидных эндонуклеаз рестрикции и затем осуществляется уже расщепление непосредственно самой вставки в составе линейаризованного вектора тетрануклеотидной рестриktionной эндонуклеазой *RsaI* в условиях недорестрикции, в результате чего получают все возможные варианты укорочения вставки. На III этапе проводится полупрепаративное разделение продуктов рестрикции/недорестрикции гель-электрофорезом с целью определить и элюировать из геля фрагменты ДНК, содержащие вектор с требу-

мыми размерами вставок, наиболее оптимально покрывающих всю ее длину. На данной схеме такие полосы ДНК, отобранные для дальнейшего лигирования, трансформации и секвенирования, заключены в рамку. В качестве маркерных молекул используются ДНК фага лямбда, расщепленная рестрикционной эндонуклеазой *Bst*ЕII, а также линейаризованные формы вектора и исходного рекомбинантного клона с полноразмерной вставкой. Различные *Rsa*I-фрагменты клонированного фрагмента ДНК, не превышающие по размеру сам вектор, не мешают анализу рестрикционных фрагментов ДНК и для упрощения схемы не показаны. Что касается присутствия *Rsa*I-фрагментов, не подвергнувшихся полному расщеплению и превышающих по размеру вектор (в случае если общая длина вставки имеет большую протяженность, чем вектор), то они всегда будут представлены в меньшем количестве, чем полосы, содержащие вектор и укороченные варианты вставок. Разная степень интенсивности полос позволяет выбрать необходимые полосы и поэтому *Rsa*I-фрагменты крупного размера также не приведены на схеме для ее упрощения. Раздельное лигирование элюированных из геля фрагментов ДНК ДНК-лигазой фага T4, осуществляемое на IV этапе, приводит к рециркулизации линейаризованного вектора с различными делеционными вариантами вставки. Необходимо отметить отсутствие среди трансформантов исходной рекомбинантной плазмиды с полноразмерной вставкой, а также практически полное исключение случайного лигирования отдельных *Rsa*I-фрагментов и их последующего клонирования. На V этапе проводится стандартная трансформация данными лигазными смесями компетентных клеток *E. coli*. VI–XI этапы посвящены приготовлению матриц для проведения терминирующих реакций, проведению самих этих реакций, разделению продуктов реакций в секвенирующем геле, радиоавтографии и чтению нуклеотидной последовательности аналогично VII–XII этапам на рис. 8.19.

Наиболее важными преимуществами определения нуклеотидной последовательности протяженных фрагментов ДНК с помощью тетра/окта стратегии секвенирования являются:

- 1) не требуется проведения предварительного этапа рестриктазного картирования с целью выявления полилинкерных рестрикционных эндонуклеаз, сайты расщепления которых отсутствуют во вставке;
- 2) использование вектора, в котором:
  - а) отсутствуют сайты расщепления для некоторых рестрикционных эндонуклеаз с тетрануклеотидными участками узнавания;
  - б) в полилинкере присутствуют участки узнавания для октануклеотидных эндонуклеаз рестрикции с АТ-богатым (*Swa*I, АТТТАААТ) и GC-богатым (*Srf*I, GСССGGGC) участками узнавания;
  - в) концы, генерируемые этими октануклеотидными и тетрануклеотидной рестрикционной эндонуклеазой, пригодны для лигирования друг с другом;
- 3) линейаризация данного вектора, несущего вставку одной из полилинкерных октануклеотидных рестрикционных эндонуклеаз, частота встречаемости которых обычно в 15–200 раз реже, чем для гексануклеотидных;

- 4) расщепление протяженной вставки тетрануклеотидной рестрикционной эндонуклеазой *RsaI*, отсутствующей в векторной последовательности, в условиях недорестрикции без опасения расщепить вектор;
- 5) возможность повторного расщепления рециркулированной формы восстановленной исходной рекомбинантной плазмиды той же октануклеотидной рестрикционной эндонуклеазой с целью обогащения выхода искомым укороченным вариантам вставки;
- 6) возможность быстрого получения субклонов, несущих однонаправленные укороченные варианты вставок требуемого размера и наиболее оптимально покрывающих всю длину клонированного фрагмента ДНК без этапа повторного расщепления циркулированных форм вектора, несущего вставки октануклеотидной эндонуклеазой рестрикции, не прибегая к весьма трудоемкому этапу анализа размера вставок у большого числа колоний;
- 7) отсутствие необходимости сложной состыковки прочитанной последовательности с использованием специальных компьютерных программ при разделении в одном секвенирующем геле последовательно друг за другом продуктов терминирующих реакций, проведенных с субклонами с однонаправленными делеционными вариантами вставки и нанесенных на гель в порядке уменьшения размера вставки, начиная с исходного клона с полноразмерной вставкой, поскольку сайты расщепления вставки тетрануклеотидной рестрикционной эндонуклеазой (в случае *RsaI* – GTAC) служат удобными ориентирами (при условии нанесения каждого образца в соседние дорожки геля в соответствующем порядке, для *RsaI* – GTAC), помогающими также исключить вероятность ошибки при определении участков состыковки нуклеотидной последовательности клонированного фрагмента ДНК.

Завершая рассмотрение стратегий секвенирования протяженных фрагментов ДНК направленным подходом, следует отметить, что объединение их в этом разделе диктовалось некоторыми присущими им общими чертами. Большинство этих стратегий предполагает последующее получение субклонов как в новом, так и в исходных векторах и, во вторых, при создании таких субклонов или уже матриц для секвенирования ДНК принимают участие одинаковые или во многом сходные по своему действию ферменты.

В дальнейших разделах этой главы будет говориться о нескольких иных подходах к секвенированию протяженных фрагментов ДНК, хотя практически все они носят направленный характер.

#### 8.4. СТРАТЕГИИ СЕКВИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ ПО СЭНГЕРУ ПРАЙМЕРНОЙ “ПРОГУЛКОЙ”

Как уже отмечалось в главе 2, усовершенствование химического синтеза олигонуклеотидов, сделавшее их более доступными, легло в основу стратегии секвенирования, названной подходом с использова-

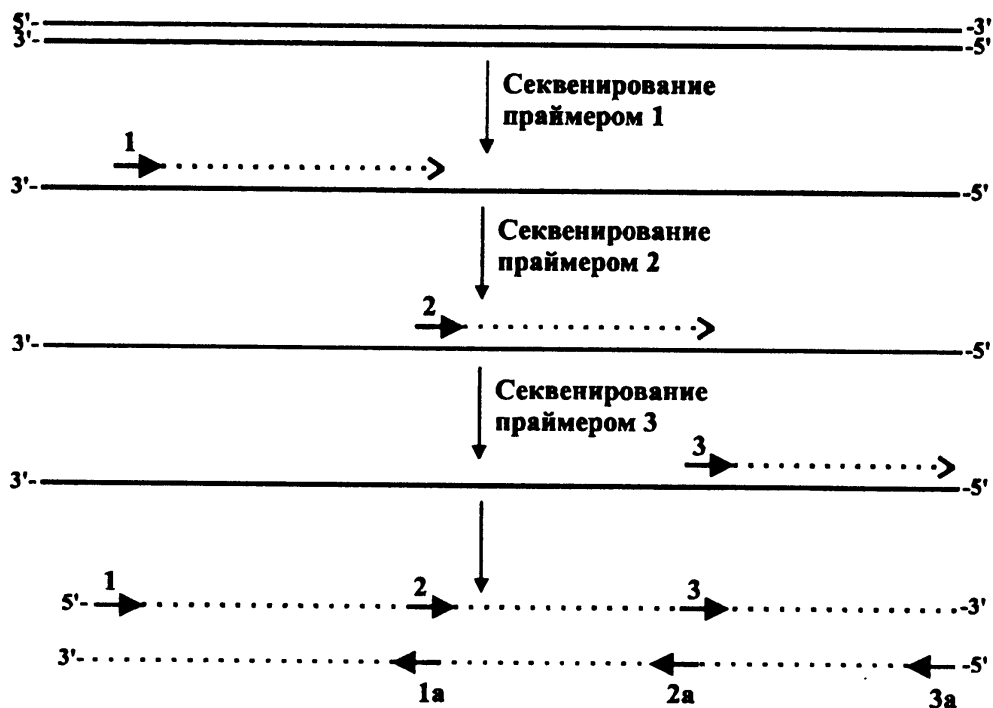


Рис. 8.21. Стратегия секвенирования протяженных фрагментов ДНК праймерной "прогулкой"

нием внутренних или прогрессивных праймеров, или праймерной "прогулкой" [Sanchez-Pescador, Urdea, 1984; Brenner, Shaw, 1985; Strauss et al., 1986]. Суть этого способа заключается в постепенном получении информации о нуклеотидной последовательности протяженного фрагмента ДНК, секвенируя вставку сначала с помощью первого праймера, затем при помощи второго и последующих, синтезированных на основе ставшей известной в первых раундах секвенирования последовательности нуклеотидов клонированного фрагмента ДНК (рис. 8.21). К преимуществу этой стратегии можно отнести ненужность получения субклонов исходного протяженного фрагмента ДНК. Однако существенным недостатком является некоторая задержка с получением информации об очередной "порции" ДНК, связанная со временем синтеза нового секвенирующего праймера. И все же главным недостатком является необходимость постоянного синтеза самих этих праймеров. В разделе 2.3, специально посвященном затравочным молекулам ДНК, в том числе, олигонуклеотидным праймерам, довольно детально рассматриваются рекомендации по выбору таких праймеров для секвенирования ДНК праймерной "прогулкой". Поэтому здесь, видимо, будет целесообразным напомнить лишь то, что для удешевления всего проекта по секвенированию протяженного фрагмента ДНК очередной праймер необходимо подбирать на максимально возможном удалении от предыдущего, но обеспечивающем безошибочное "чтение" нуклеотидной последовательности, что зависит и от типа используемого прибора и от опыта самого экспериментатора и ряда других причин. Стратегия секвенирования ДНК праймерной "прогулкой" с использованием двух типов праймеров, один из кото-



рых предназначается для дальнейшего чтения секвенируемой матрицы, а второй праймер подбирается с таким расчетом, чтобы секвенировать эту же матрицу ДНК в обратном направлении, позволяет выявить возможные неточности в определении нуклеотидной последовательности [Voss et al., 1993]. Таким образом, первый тип праймеров рассчитан на получение новой информации, а второй выступает в качестве подтверждающего праймера.

Одной из главных составляющих стоимости подобного проекта секвенирования ДНК являются сами праймеры. Причем даже самый малый масштаб синтеза олигонуклеотидов позволяет синтезировать такое количество олигонуклеотидов, которое в тысячи раз превышает применяемое в качестве затравки в реакциях секвенирования. Оставшийся праймер, как правило, больше нигде не используется, что крайне неэкономно. Это обстоятельство заставляло искать какие-то новые варианты стратегии секвенирования ДНК праймерной “прогулкой”. По этой причине на протяжении уже многих лет разрабатываются варианты использования или более коротких праймеров, или состоящих из нескольких коротких (5–7 нуклеотидов) модулей. Так, один из подходов заключается в существовании предсинтезированной библиотеки относительно коротких (8–12 звеньев) олигонуклеотидных праймеров, которые могли бы с успехом применяться в разных проектах секвенирования протяженных фрагментов ДНК [Studier, 1989; Siemieniak, Slightom, 1990; Burbelo, Iadarola, 1994; Slightom et al., 1994]. В других работах говорится о библиотеке еще более коротких (4–7 нуклеотидов) блоков-модулей, которые в определенных условиях способны формировать полноценные 12-, 18- или даже большей длины олигонуклеотидные праймеры [Ажикина и др., 1993; Кноппе и др., 1996; Szybalski, 1990; Kieleczawa et al., 1992; Azhikina et al., 1993; Kotler et al., 1993; Kaczorowski, Szybalski, 1994, 1996; Dunn et al., 1995]. Еще один интересный подход к решению проблемы секвенирующих праймеров заключается в ферментативном дифференциальном удлинении короткого праймера с помощью неполного набора дезоксирибонуклеотид трифосфатов [Raja et al., 1997].

Независимо от типа праймерных молекул, представляющих собой или единый олигонуклеотид, или состоящих из отдельных модулей, общая стратегия секвенирования ДНК праймерной “прогулкой”, приведенная на рис. 8.21, остается неизменной. Некоторое отличие реально осуществляемых проектов от того, что приведено на данной схеме, и направленное на ускорение выполнения проекта по определению нуклеотидной последовательности ДНК, заключается в одновременном секвенировании вставки с обоих концов клонированного фрагмента или даже еще с какого-либо произвольного места вставки, что позволяет накапливать данные соответственно в два или три раза быстрее (кратность ускорения зависит от исходного числа точек, с которых начинается секвенирование и праймеров к ним). При этом праймерами для секвенирования краев вставки обычно служат стандартные секвенирующие праймеры, комплементарные соответствующим участкам векторной молекулы.

Так, была предложена стратегия секвенирования космидных клонов, заключающаяся в максимальном увеличении числа точек внутри вставки, которые могли бы служить начальными для секвенирования праймерной “прогулкой” [Siemieniak et al., 1991]. А так как последовательность клонированного фрагмента ДНК неизвестна, то и целенаправленный выбор секвенирующих праймеров невозможен. С этой целью осуществлялось построение рестриктазной карты клонированного фрагмента и секвенирование отдельных рестриктазных фрагментов химической деградацией по Максаму–Гилберту, поскольку этот метод не требует предварительного знания даже части нуклеотидной последовательности. Другой подход заключался в использовании для ферментативного секвенирования специального сплинкера [Kalisch et al., 1986] (подробно описанного ниже) или даже коротких олигонуклеотидных праймеров, отжигающихся на такой матрице достаточно непредсказуемо [Studier, 1989]. Таким образом, на первом этапе секвенирования шло накопление стартовых сведений о нуклеотидной последовательности отдельных участков клонированного в космиде протяженного фрагмента ДНК и уже затем начинался этап широкомасштабного секвенирования праймерной “прогулкой”.

Другой тип олигонуклеотидных молекул при осуществлении праймерной “прогулки” описан в статье Фу и соавт. [Fu et al., 1995]. Ими был применен одно/двухцепочечный праймер, формируемый путем отжига 23-членного и 18-членного олигонуклеотидов, что привело к образованию на одном из его 3'-концов 5 выступающих нуклеотидов. Авторы цитируемой статьи делают заключение, что предсинтезированная библиотека подобных праймеров будет меньше, чем состоящая из более привычных гексамеров, поскольку все возможные варианты 5 неспаренных нуклеотидов приведут к существованию всего 1024 ( $4^5$ ) таких олигонуклеотидов. Принцип же использования подобных праймеров заключается в образовании затравочного комплекса для ДНК-полимеразы путем их отжига на одноцепочечном участке фрагмента ДНК, образуемого подходящей рестрикционной эндонуклеазой (например *AraVI*), и удерживаемых на одноцепочечной матрице ДНК также за счет стэкинг-взаимодействия.

В литературе описан способ определения нуклеотидной последовательности участка ДНК, прилегающего к тому, для которого последовательность уже известна с помощью ПЦР и специальной олиго-кассеты (рис. 8.22) [Rosenthal, Jones, 1990]. Для этого проводится расщепление секвенируемого протяженного фрагмента ДНК какой-либо гексануклеотидной рестрикционной эндонуклеазой. Затем следует этап лигирования полученных рестриктазных фрагментов с олиго-кассетой, представляющей собой двухцепочечный участок ДНК, состоящий из двух отождествленных олигонуклеотидов, с соответствующим “липким” концом. В результате все рестриктазные фрагменты в смеси будут нести по своим концам подобную олиго-кассету. Далее в цитируемой работе авторы проводили 50 циклов линейной ПЦР с одним праймером, несущим биотинилированную метку и отжигающимся на участке ДНК с известной последовательностью. После удаления из реакционной сме-

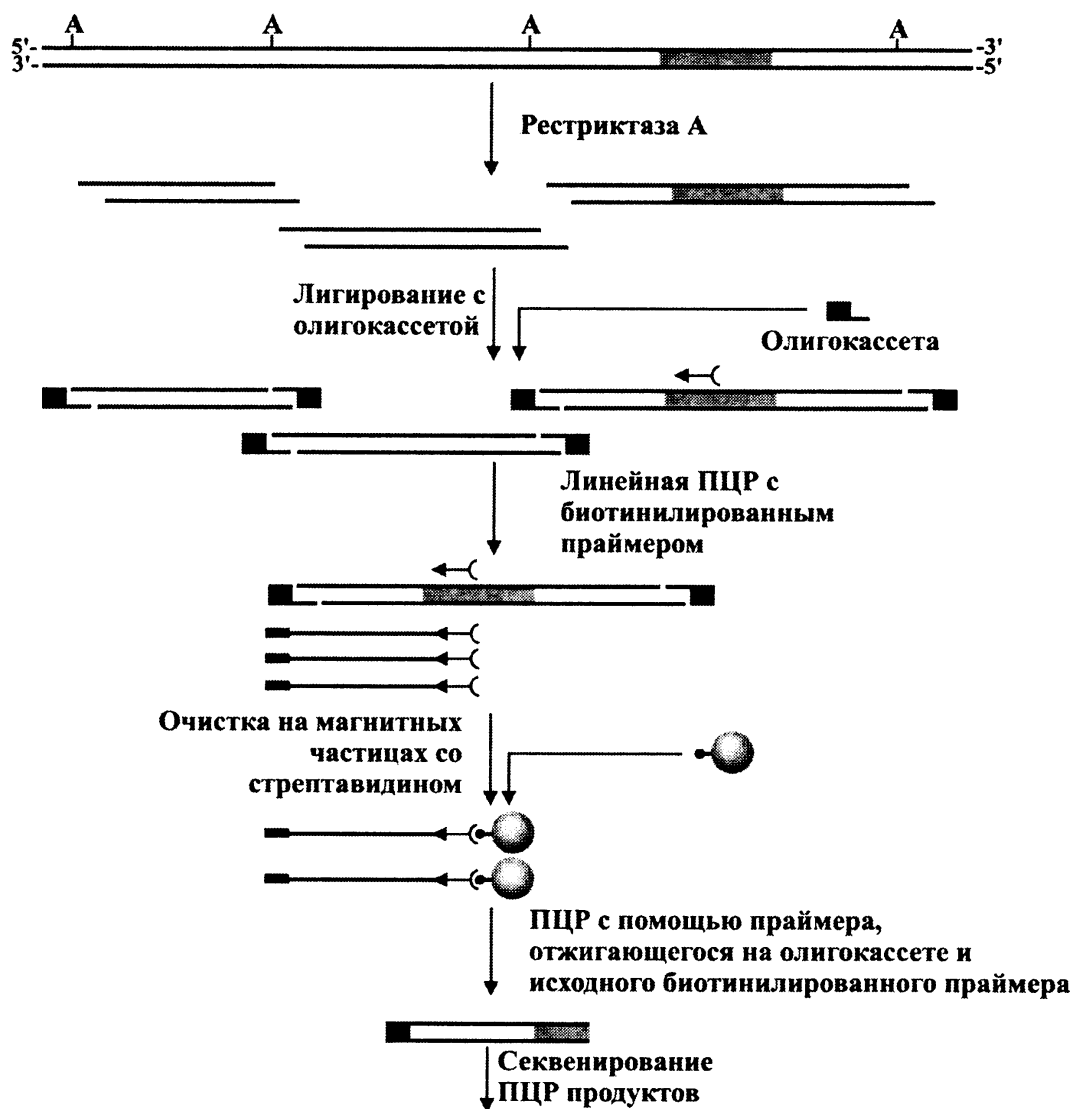


Рис.8.22. Стратегия секвенирования участков ДНК, прилегающих к таковым, с уже известной последовательностью, “геномной прогулкой с олиго-кассетой”

си с помощью магнитных частиц со стрептавидином различных фрагментов ДНК, не несущих молекулы биотина, продукт линейной ПЦР амплифицировался уже экспоненциально при помощи этого же и второго праймеров, причем второй праймер отжигался на олиго-кассете. Таким образом, еще через 35 циклов появлялся двуцепочечный продукт ДНК, часть последовательности которого была известна, а другую часть предстояло определить. Такая схема последовательного определения все новых и новых участков ДНК может действовать и дальше и поэтому авторы назвали свой подход в нашем несколько вольном переводе “геномной прогулкой с олиго-кассетой”.

Другим способом амплификации и последующего секвенирования неизвестных участков ДНК, прилегающих к известным, является подход с проведением ПЦР с пониженной температурой отжига (например, 40 °C вместо 55 °C) в первых циклах с тем, чтобы используемый праймер мог отжечься еще в других участках ДНК [Malo et al., 1994]. Более

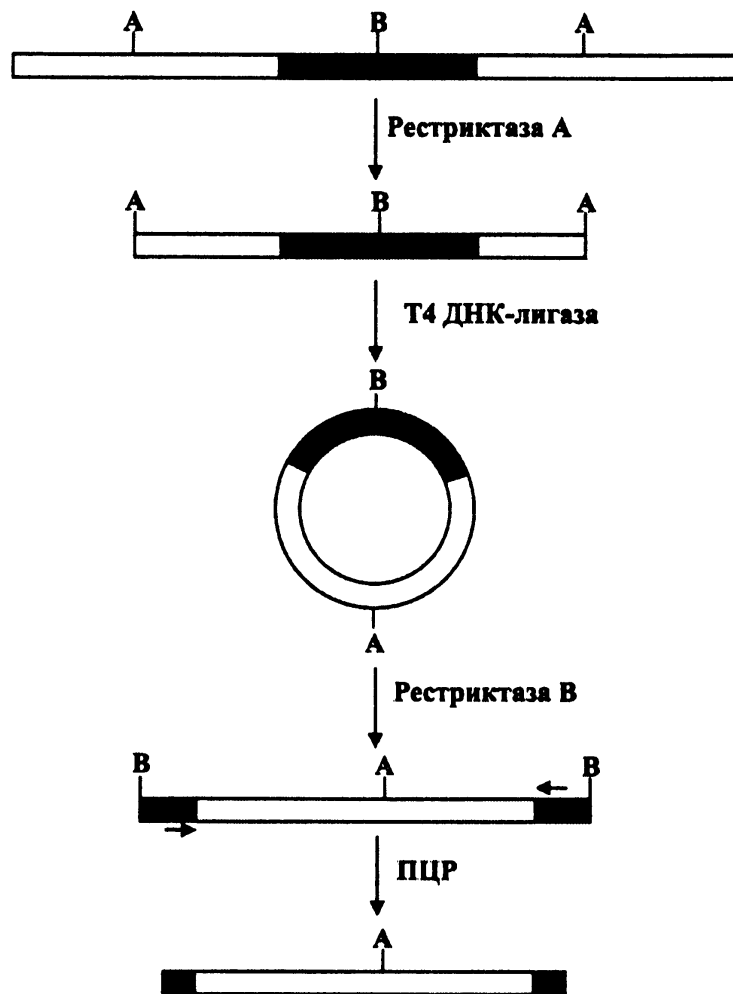


Рис. 8.23. Стратегия амплификации участков ДНК с неизвестной последовательностью, прилегающих к известным, с целью их дальнейшего секвенирования

направленным подходом амплификации подобных участков ДНК является метод обращенной ПЦР, где места отжига праймеров на рестриктазном фрагменте ДНК расположены с таким расчетом, что синтез новых цепей как бы направлен в разные стороны, однако поскольку перед амплификацией производится с помощью T4 ДНК-лигазы циркуляризация рестриктазных фрагментов, то это приведет в конечном итоге к наработке интересующего исследователей участка неизвестной ДНК [Ochman et al., 1988]. Аналогичный подход был предложен другими авторами, единственное отличие которого заключалось в еще одном этапе рестриктазного расщепления по какому-нибудь подходящему сайту в известном фрагменте ДНК, расположенном между местами отжига олигонуклеотидных праймеров [Triglia et al., 1988]. Таким образом, в образующемся линейном фрагменте ДНК происходила реориентация локализации неизвестных участков с краев внутрь (рис. 8.23). В литературе встречается описание ряда других подходов к амплифицированию неизвестных фрагментов ДНК, прилегающих к известным [Arnold, Hodgson, 1991; Parker et al., 1991; Jones, Winistorfer, 1992; Verhasselt et al., 1992; Jones, Winistorfer, 1993; Williamson, Rutherford, 1994; Devon et al.,

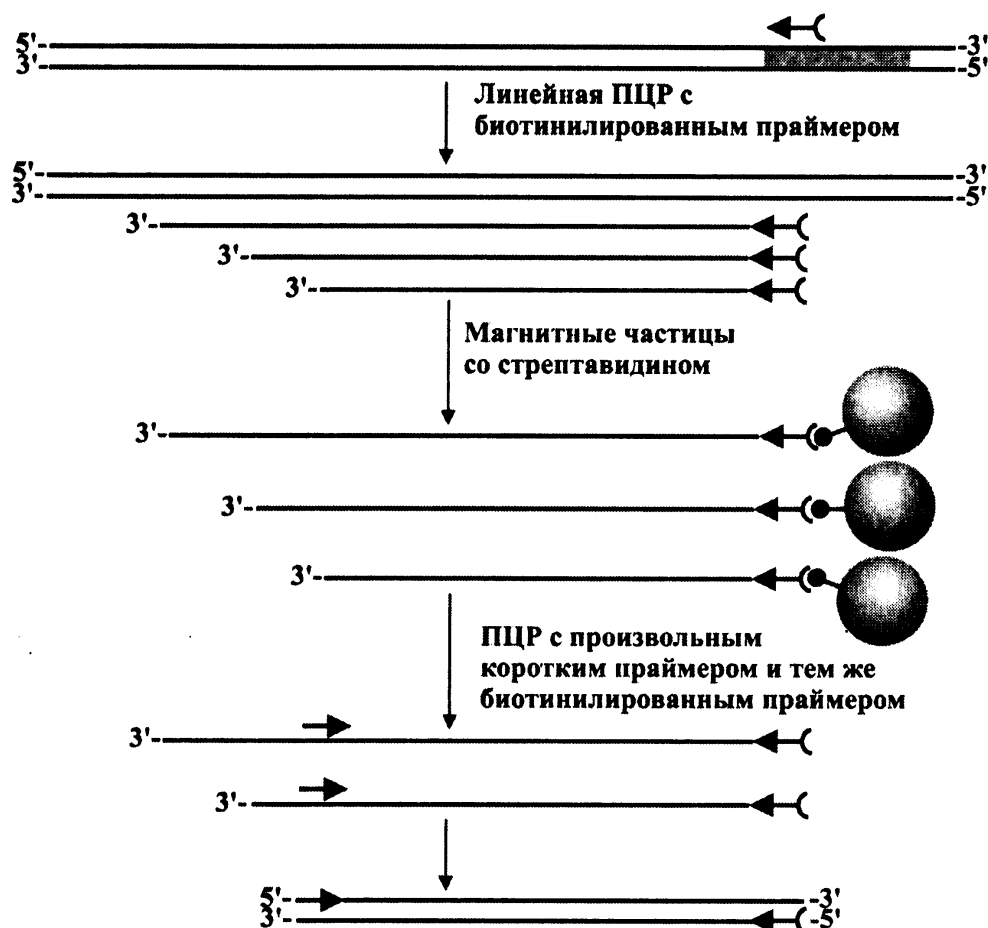
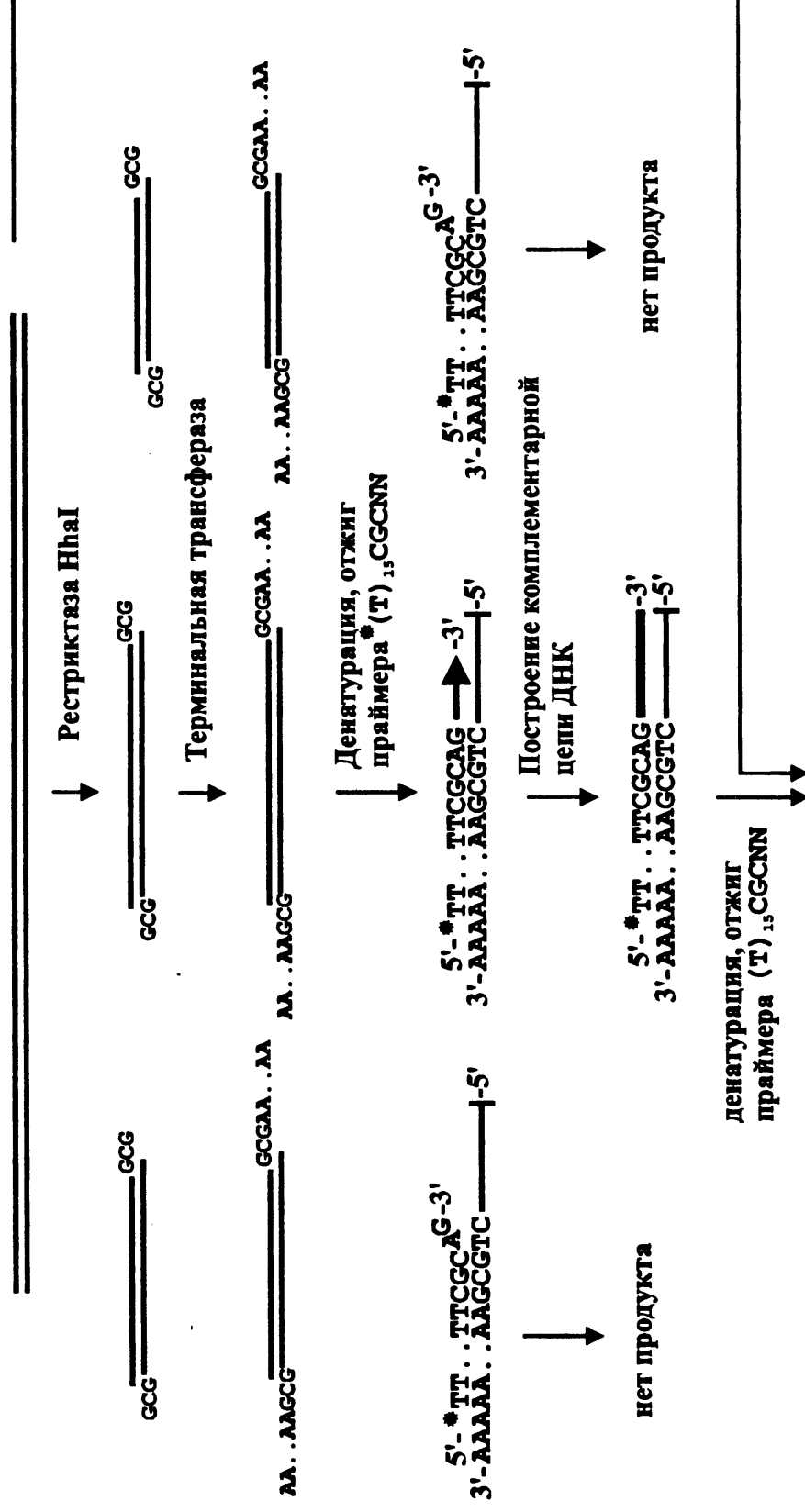


Рис. 8.24. Стратегия амплификации участков ДНК с неизвестной последовательностью, прилегающих к известным, путем использования комплекса биотин/стрептавидин с целью их дальнейшего секвенирования

1995; Dominguez, Lopez-Larrea, 1994; Siebert et al., 1995; Trueba, Johnson, 1996; Sterky et al., 1998; Weber et al., 1998].

Был предложен еще один весьма продуктивный метод секвенирования неизвестных участков геномной ДНК, заключающийся в нескольких циклах линейной амплификации с помощью биотинилированного праймера, отжигающегося на участке ДНК с известной нуклеотидной последовательностью с таким расчетом, что при его ферментативном удлинении рост новосинтезируемой цепи ДНК будет направлен в область с неизвестной последовательностью нуклеотидов, интересующую исследователя [Lagerstrom et al., 1991; Nguyen et al., 1998]. На следующем этапе проводится сорбция этих биотинилированных цепей ДНК на магнитных частицах со стрептавидином и удаление остальной ДНК. Затем с помощью короткого олигонуклеотидного праймера с произвольно выбранной последовательностью и исходного биотинилированного осуществляется второй этап ПЦР, проводимый уже в условиях экспоненциальной амплификации (рис. 8.24). Поскольку короткий произвольный праймер имеет достаточно высокую вероятность отжечься на неизвестном участке ДНК, то в результате 25–30 новых циклов будет наработан фрагмент или фрагменты ДНК, которые можно



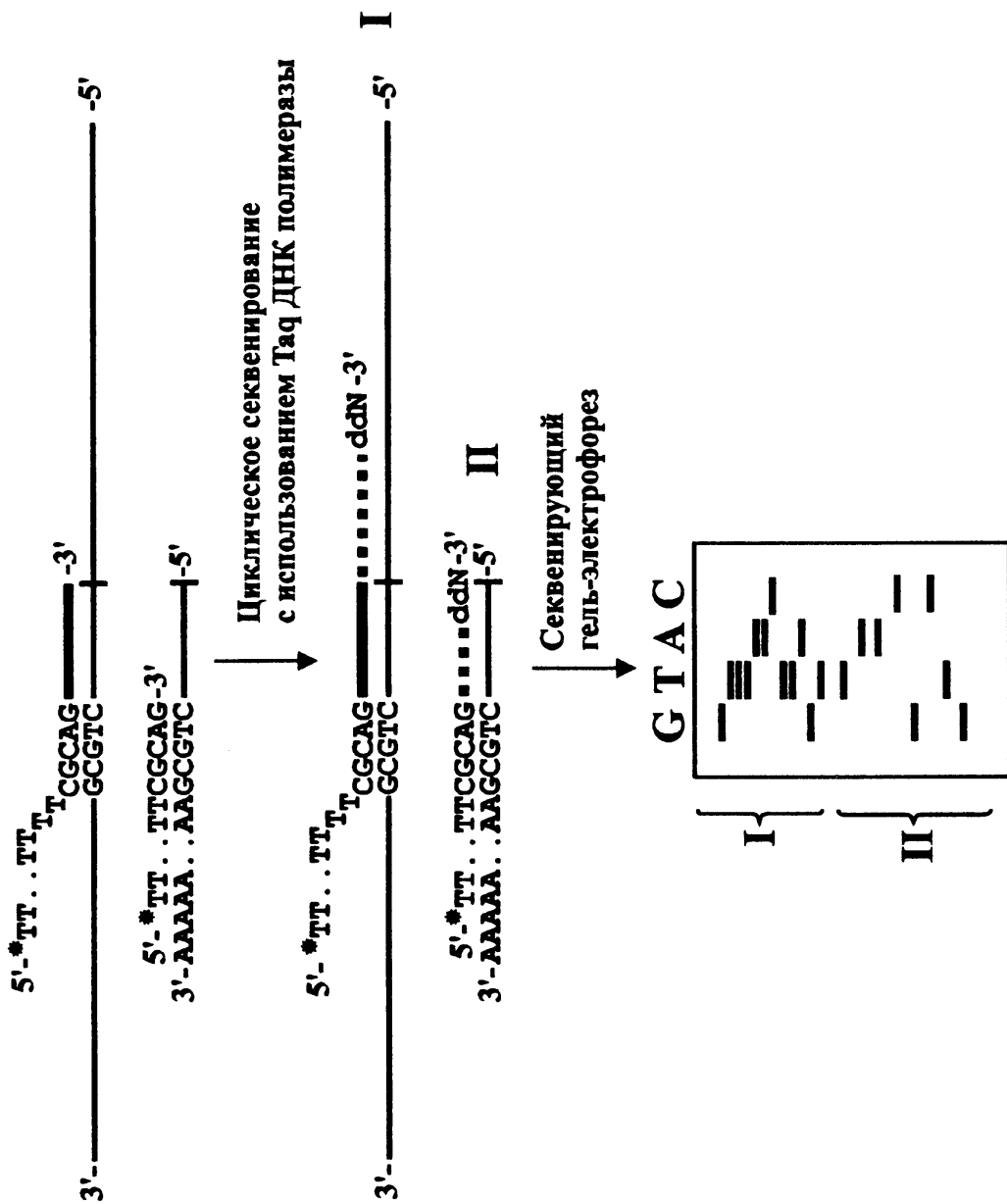


Рис. 8.25. Стратегия секвенирования ДНК "прогулкой" фрагментом

будет отобрать по размеру. Однако, если подобный эксперимент проводить не с изолированными линейными амплификатами, содержащими участок с известной нуклеотидной последовательностью, но и в присутствии всей, не удаленной основной массы тотальной ДНК, то количество амплифицированных случайных фрагментов ДНК будет чрезмерно большим за счет значительного числа мест отжига таких коротких праймеров, что или сильно затруднит, или сделает невозможным поиск среди них целевого фрагмента.

Весьма интересен подход к секвенированию протяженных фрагментов ДНК, названный авторами не праймерной “прогулкой”, а “прогулкой фрагментов” ДНК [Okano, Kambara, 1996]. Назван он так, видимо, потому, что удлиняемые на одном из первых этапов олигонуклеотидные праймеры и используемые затем в качестве затравки на завершающем ввиду их большой длины трудно назвать праймерами. Главной особенностью этого подхода можно считать крайне малую библиотеку предсинтезированных олигонуклеотидных праймеров, состоящую всего из 16 вариантов – 5'-TTTTTTTTTTTTTTGCGNN-3'. 15 нуклеотидов этих праймеров формируют общую часть и представляют собой блок из дезокситимидинов, далее следуют три нуклеотида, комплементарных сайту узнавания (или, вернее, тому, что от него остается после расщепления ДНК) какой-либо рестрикционной эндонуклеазы (в рассматриваемой статье GCG и *HhaI*) и служащих в качестве позиционирующих нуклеотидов. Оставшиеся два нуклеотида (NN) принадлежат к варьирующей части и представляют собой все 16 возможных вариантов динуклеотидов AA, AC, AG, AT, CA, CC и т.д.

На первом этапе клонированный фрагмент ДНК расщепляется соответствующей рестрикционной эндонуклеазой (здесь – *HhaI*), образующей 3'-выступающий конец. Далее с помощью терминальной нуклеотидтрансферазы к полученным рестриктазным фрагментам “пришивается” блок поли(дА). После денатурации ДНК и отжига сразу всех возможных 16 праймеров идет построение цепи ДНК, приводящее к удлинению некоторых праймеров. За счет довольно протяженной общей части этих праймеров (TTTTTTTTTTTTTTGCG) все они “отожгутся” на поли(дА)-последовательности, причем позиционирующие нуклеотиды (GCG) заставят их сделать это вплотную к концу рестриктазного фрагмента ДНК. Однако удлинение праймеров будет возможно только в том случае, если два последних переменных нуклеотида (NN) окажутся комплементарными присутствующим в самой матрице ДНК (при использовании ДНК-полимеразы с отсутствующей или удаленной 3' → 5'-редактирующей активностью). Таким образом, на данном этапе происходит как некоторая селекция праймеров, так и их удлинение. На следующем этапе осуществляется непосредственно само секвенирование ДНК путем построения новых цепей ДНК в условиях терминирования дидезоксинуклеотид трифосфатами. Матрицами служат как повторно денатурированные фрагменты ДНК из предыдущей реакции удлинения праймера, так и вновь добавленный исходный клонированный фрагмент ДНК, не подвергнутый расщеплению рестрикционной эндонуклеазой (здесь – *HhaI*). В результате одновременно протекает две се-



квенирующие реакции. Одна из них использует выбранный какой-то конкретный 5'-TTTTTTTTTTTTTTGCGNN-3' праймер, отжигающий на матрице в виде рестриктазного фрагмента ДНК, другая – удлиненный праймер, или, правильнее сказать, “фрагмент-праймер”, отжигающийся уже своей удлиненной частью на нерасщепленном исходном фрагменте ДНК. Эта стадия проходит в режиме циклического секвенирования и в результате оказывается возможным “чтение” обоих продуктов секвенирующих реакций. Следует отметить, во-первых, некоторую конкуренцию “фрагмента-праймера” и комплементарной ему цепи рестриктазного фрагмента на этапе отжига перед секвенирующими реакциями, и, во-вторых, определенные сложности в выборе того или иного праймера из 16 возможных (описание данного процесса нами просто опущено). Из представленной (упрощенной) схемы на рис. 8.25 можно оценить всю сложность и трудоемкость данного подхода. Однако этот метод возник не неожиданно и явился продолжением предыдущих работ этой группы. Ранее в модельном эксперименте с *HindIII*-фрагментами ДНК фага лямбда эти авторы продемонстрировали успешное применение данного подхода без построения поли(дА)-конца [Furuya et al., 1994].

Стратегия секвенирования рестриктазных фрагментов ДНК, расщепленных ферментами, образующими 3'-выступающие концы, путем “пришивания” к последним поли(дА)блока, приведенная на рис. 8.26, известна давно [Stallard et al., 1987]. В цитируемой работе для проведения терминирующих реакций рестриктазные фрагменты ДНК расщеплялись соответствующими ферментами (здесь – *KpnI* или *PstI*) и после присоединения к ним участка поли(дА) с помощью терминальной трансферазы расщеплялись другой подходящей рестрикционной эндонуклеазой, разделялись электрофорезом и элюировались из геля. 20-звенные праймеры, кроме 16 дТ-остатков, в качестве позиционирующих нуклеотидов содержали элементы сайтов узнавания GTAC и TGCA, использованных для расщепления ферментов *KpnI* или *PstI* соответственно. После отжига *KpnI*- или *PstI*-праймера на элюированной из геля и денатурированной матрице проводились терминирующие реакции и продукты разделялись секвенирующим гель-электрофорезом.

Другой разновидностью праймерной “прогулки” является стратегия секвенирования протяженных фрагментов ДНК ферментативным методом по Сэнгеру с помощью сплинкера – секвенирующего праймер-линкера (рис. 8.27). Одно из основных отличий этого способа секвенирования ДНК от подхода праймерной “прогулкой” заключается в довольно ограниченном числе олигонуклеотидных праймеров, необходимых для его осуществления. Было найдено весьма интересное решение по объединению в одной молекуле праймера и линкера, получившего название спликер (splinker – sequencing primer linker) [Kalisch et al., 1986]. Особенностью такого сплинкера было то, что его последовательность содержала внутренний участок гомологии, что при определенной температуре приводило к формированию вторичной структуры в виде “шпильки” с выступающим или тупым концами. За счет этих выступающих или тупых концов осуществлялось лигирование данного

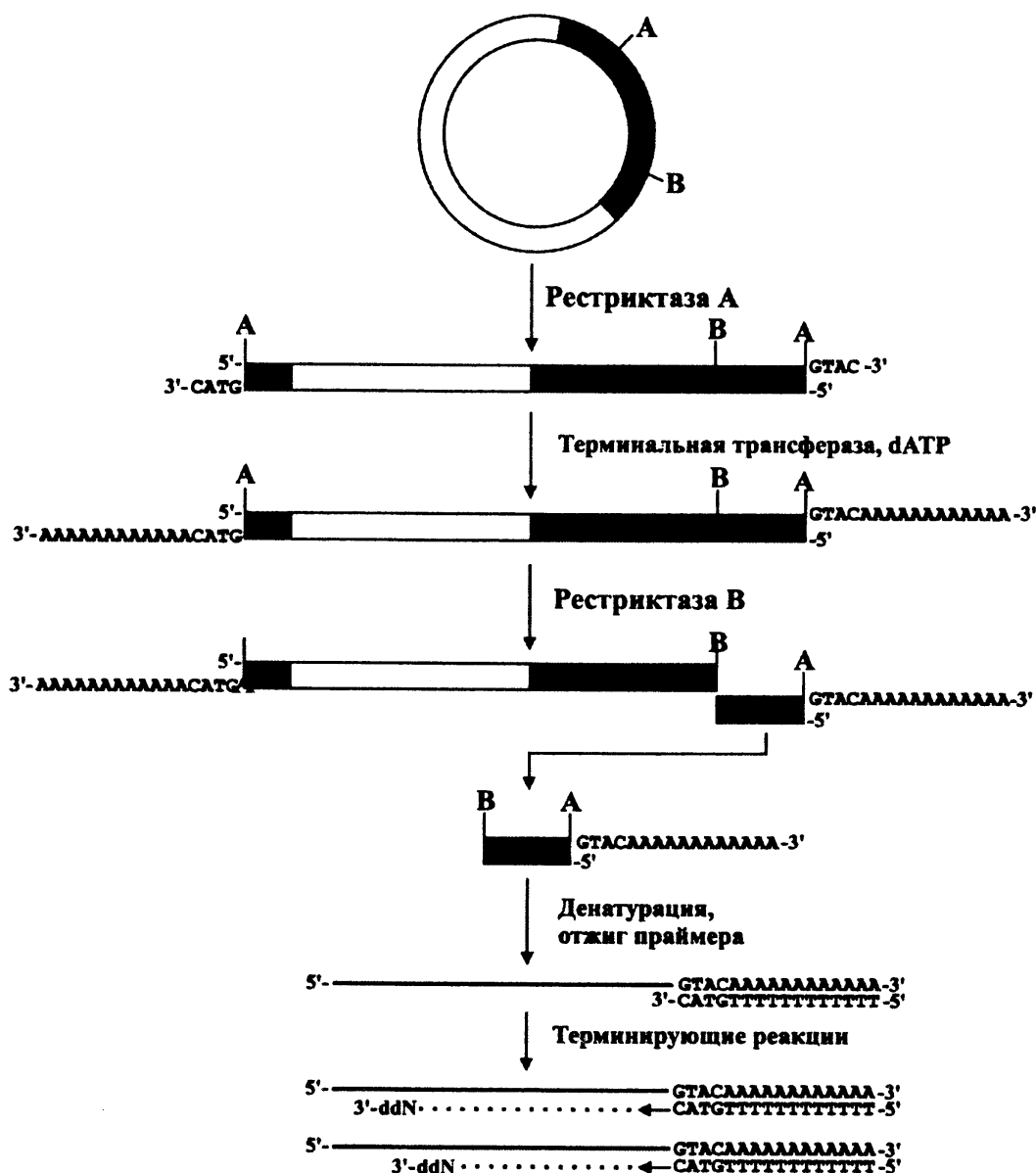


Рис. 8.26. Стратегия секвенирования рестриктазных фрагментов ДНК путем “пришивания” поли(дА) блока и отжига на нем специфического праймера с позиционирующими нуклеотидами

конкретного сплинкера с элюированным из геля рестриктазным фрагментом ДНК, имеющем аналогичные концы. Таким образом, основная последовательность самого сплинкера могла быть практически одинаковой и отличаться только своими концевыми нуклеотидами, выбор вариантов которых довольно ограничен.

Для осуществления этой стратегии секвенирования требуется некоторая предварительная подготовка как самого сплинкера, так и рестриктазного фрагмента ДНК. Для того чтобы сплиinker мог служить в качестве затравочной молекулы для элюированного из геля рестриктазного фрагмента, на его 5'-конец необходимо с помощью полинуклеотидкиназы фага Т4 в присутствии АТФ добавить остаток фосфорной кислоты. А рестриктазный фрагмент ДНК, напротив, надо дефосфори-

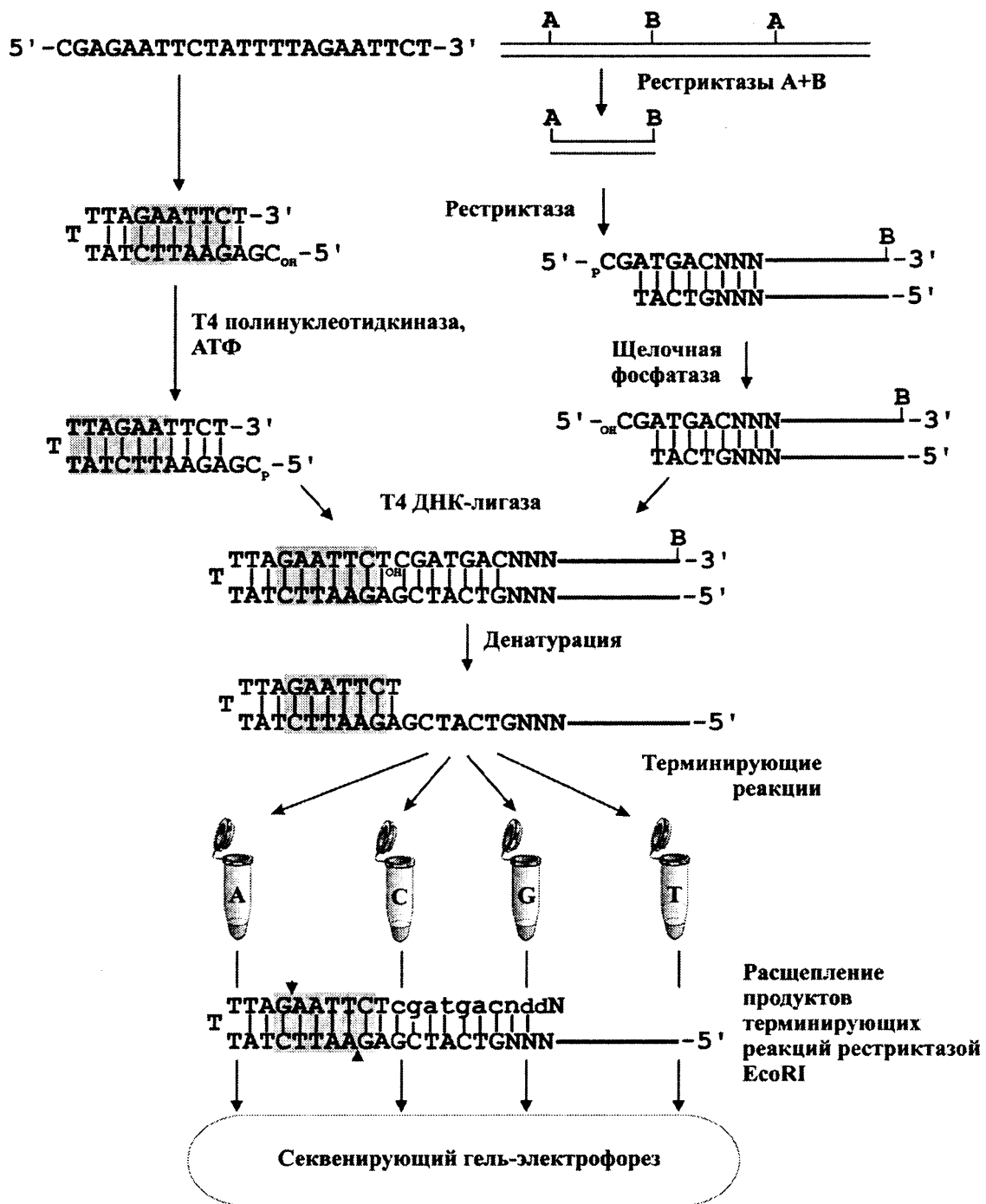


Рис. 8.27. Стратегия секвенирования рестриктазных фрагментов ДНК с помощью сплинкера

лизовать обратной щелочной фосфатазой. В результате этих процедур после лигирования сплинкера с данным рестриктазным фрагментом, как можно видеть из рис. 8.27, образуется “ник”, приводящий после денатурации цепей ДНК к формированию затравки для синтеза новой комплементарной цепи ДНК в условиях терминации с дидезоксинуклеотидтрифосфатами. Перед нанесением продуктов терминирующих реакций на секвенирующий гель-электрофорез необходим этап их расщепления рестрикционной эндонуклеазой, сайт которой имеется в сплин-

кере, поскольку, как отмечалось выше, необходимым элементом сплинкера является наличие в его двуцепочечном участке доступного сайта узнавания относительно редкощепящей рестрикционной эндонуклеазы, расщепление которого позволяет удалить концевую петлю сплинкера. Такое расщепление обеспечивает образование гомогенного 5'-конца для смеси терминированных фрагментов ДНК. Следует отметить, что при этом способе секвенирования ДНК метка не может находиться на 5'-конце праймера и должна включаться в ходе проведения терминирующих реакций.

Еще ряд подходов, описываемых ниже, хотя и не принадлежат к классическим вариантам праймерной "прогулки", все же основаны на использовании олигонуклеотидных праймеров. Вторым важным условием их осуществления является применение ПЦР. Так, был предложен интересный способ получения однонаправленных делеционных вариантов клонированной вставки с помощью ПЦР и биотинилированного праймера [Yohda et al., 1995]. Амплифицированный с использованием фланкирующих его праймеров (один из которых нес биотинилированную метку) фрагмент ДНК был затем расщеплен тетра nukлеотидной рестрикционной эндонуклеазой в условиях недорестрикции. Далее следовал этап лигирования полученных фрагментов со специальным линкером, в результате чего образовывалась смесь разнообразных фрагментов, среди которых присутствовали как ограниченные по обеим концам только линкерами, так и варианты с участием линкера и того или иного праймера. Добавление в реакционную смесь магнитных частиц, покрытых стрептавидином, позволяло сорбировать на них только ограниченные биотинилированным праймером, тогда как прочие исключались из дальнейшего обращения. В результате повторного этапа ПЦР с данным биотинилированным праймером и праймером, гомологичным линкеру, происходила амплификация однонаправленных делеционных вариантов вставки, после разделения гель-электрофорезом которых они становились пригодны для секвенирования. Несколько похожий метод, также основанный на частичном расщеплении амплификатов тетра nukлеотидной рестрикционной эндонуклеазой в условиях недорестрикции, описывается в другой работе [Wright, 1992]. Главное отличие этого подхода состояло в неиспользовании магнитных частиц и биотинилированного праймера. Другими авторами аналогичное частичное расщепление фрагмента ДНК достигалось за счет произвольного включения во время ПЦР метилированных производных цитозина, что защищало фрагмент ДНК от действия ферментов, чувствительных к метилированию цитозиновых остатков [Wong et al., 1997].

Секвенирование ДНК с помощью произвольных праймеров стало возможным благодаря предложенному методу, названному авторами SWAPP (Sequencing With Arbitrary Primer Pairs) [Burt et al., 1994]. Однако подобный подход малоприменим для крупномасштабного секвенирования ДНК и может применяться лишь для определения нуклеотидных последовательностей произвольных фрагментов ДНК у близкородственных организмов с целью выявления полиморфизма каких-то их генов или прочих фрагментов ДНК. Главным условием для успешного

осуществления этого метода является секвенирование амплификатов, ограниченных обязательно разными праймерами. Иначе в противном случае необходимо проводить деление цепей, что весьма трудоемко и сразу снижает ценность метода. Выявление амплификата, пригодного для секвенирования, проводится путем сравнения 3 электрофоретических картин результатов ПЦР, в которых использовались оба праймера вместе и поодиночке. Полосы, видимые только при совместном применении праймеров и совпадающие у разных образцов ДНК по размеру (что косвенно может свидетельствовать об их принадлежности к сходным участкам генома у близкородственных видов), элюировались из геля и брались для дальнейшего секвенирования с использованием в качестве затравки при проведении уже терминирующих реакций одного из амплификационных праймеров.

Неким вариантом секвенирования ДНК с помощью “прогулки” праймерами можно считать “мультипраймерное” секвенирование, где в качестве затравочных молекул служат обе цепи подходящих рестриктазных фрагментов ДНК вставки [Akiyama et al., 1992]. Поскольку построение комплементарных цепей ДНК в данной работе в условиях терминации осуществлялось в присутствии дезапроизводных дГТФ и дАТФ, то у авторов после завершения реакций появилась возможность удаления служащего праймером рестриктазного фрагмента за счет восстановления сайта используемой рестрикционной эндонуклеазы без опасения расщепить новосинтезированную цепь ДНК, несущую модифицированные нуклеотиды, что позволило увеличить число “читаемых” нуклеотидов.

## 8.5. СТРАТЕГИИ СЕКВИРОВАНИЯ ПРОТЯЖЕННЫХ ФРАГМЕНТОВ ДНК ПУТЕМ ПОЛУЧЕНИЯ ДЕЛЕЦИОННЫХ СУБКЛОНОВ IN VIVO

События интрамолекулярной коинтеграции, связанные с наличием в векторе транспозируемых элементов, приводят к образованию дочерних молекул, несущих делеции определенных участков. Такие делеции происходят с фиксированных мест транспозона в варьирующие места прилегающего участка ДНК, который в данном случае представляет собой клонированный фрагмент ДНК, предназначенный к секвенированию. С целью получения субклонов, несущих подобные делеции, были созданы различные специализированные вектора для секвенирования протяженных фрагментов ДНК, несущие в своем составе разные транспозоны. Учитывая то, что транспозиция происходит не с очень высокой частотой, для каждого типа векторов были разработаны соответствующие штаммы *E. coli*, позволяющие в определенных условиях осуществлять поиск нужных субклонов, обычно по признаку летальности, т.е., в этом случае после транспозиции на чашке Петри вырастали только делеционные варианты исходного фрагмента ДНК.

Так, одним из первых было создано целое семейство векторов с транспозоном Tn9, позволяющим получать направленные делеции

вставки, позитивной селекцией для которых служила способность бактериального клона расти на галактозе [Ahmed, 1984, 1985, 1987, 1989]. Первые вектора этого семейства рАА3.7 и рАА3.7Х представляли собой плазмиду, несущую элементы фага лямбда и транспозона Tn9, позволяющие наращивать только двуцепочечную ДНК [Ahmed, 1984, 1985]. Дальнейшее улучшение вектора рАА3.Х состояло в добавлении в его последовательность ориджина репликации из одноцепочечного фага M13, что сделало этот вектор фагмидным и привело к возможности нарабатывать одноцепочечную ДНК [Peng, Wu, 1986, 1987]. Сам автор этого семейства векторов на основе плазмиды рАА3.7Х создал космиду рАА113Х, позволяющую клонировать большие вставки протяженностью около 40 тпн и с помощью транспозона Tn9 получать направленные делеции клонированного фрагмента ДНК [Ahmed, 1987]. Следующий вектор рАА113М, отличающийся от предыдущего вектора наличием ориджина репликации одноцепочечного фага M13, был предназначен для получения уже одноцепочечных вариантов генерируемых субклонов [Ahmed, 1989]. Состыковка нуклеотидных последовательностей фрагментов ДНК, субклоны которых были получены *in vivo* с помощью транспозона имеет одну особенность, которую надо принимать во внимание. Так, в каждом секвенированном участке будет присутствовать блок короткой последовательности, принадлежащий транспозону, но, как наглядно продемонстрировал Ахмед в своей работе, при этом не образуется каких-либо нежелательных aberrаций в месте стыка на этапе субклонирования [Ahmed, 1985]. Транспозон Tn9 характеризуется некоторой избирательностью по отношению к последовательности ДНК, в которую он встраивается, что имеет иногда некоторые отрицательные последствия.

На основе плазмиды рBR322 были сконструированы два вектора рAP13 и рAP18, позволяющие осуществлять делеции вставки *in vivo* за счет встроенного в эти вектора транспозона IS1 и получать необходимые субклоны для секвенирования последовательности ДНК всей вставки [Prentki et al., 1991]. Отбор делеционных вариантов производился на основе устойчивости клеток к стрептомицину. Ферментативное секвенирование ДНК проводилось с помощью олигонуклеотидного праймера, гомологичного IS1 транспозона. Созданный вектор рALF, несущий транспозон IS1, cos-сайты фага лямбда и прочие регуляторные и маркерные элементы, был с успехом применен для получения делеционных вариантов крупной (около 40 тпн) вставки ДНК дрозофилы, вызванных внедрением данного транспозона [Ahmed, Podemski, 1997]. В качестве секвенирующего праймера здесь также, как и в предыдущей работе, служил олигонуклеотид, гомологичный участку IS1 транспозона.

С участием другого транспозона Tn1000 был сконструирован весьма интересный космидный вектор рDUAL, несущий три гена устойчивости к разным антибиотикам [Wang et al., 1993]. Особенностью данного вектора, разработанного для специальной стратегии секвенирования, названной авторами “фабрикой делеций *in vivo*”, является возможность одновременного получения двунаправленных делеционных вари-

антов, отбираемых по устойчивости к двум разным антибиотикам и двум селективным летальным маркерам. Характерной чертой дельта/гамма транспозиций является то, что первые пять нуклеотидов в секвенируемом сегменте ДНК вставки после постоянного участка самого транспозона представляют собой преимущественно А + Т нуклеотиды. В своей предыдущей работе авторы показали, что места внедрения транспозона  $\gamma\delta$  или  $Tn1000$  распределяются достаточно равномерно, хотя даже в GC-богатых участках ДНК внедрение все же чаще отмечается в АТ-блоках (Strausbaugh et al., 1990). Ранее на основе этого же транспозона  $Tn1000$  этими же авторами был создан более простой специализированный вектор и в нем получена серия успешно секвенированных субклонов гена *avtA* кишечной палочки [Liu et al., 1987]. Они же получили конструкцию  $m\gamma\delta-1$ , являющуюся малым дериватом транспозона  $Tn1000$ , и показали ее пригодность для крупномасштабного секвенирования ДНК путем создания направленных делеций вставки *in vivo* [Berg et al., 1991].

Другой вектор рММ251, предназначенный для генерации делеций вставки *in vivo*, также представляет собой весьма сложную конструкцию, в которую встроены регуляторные элементы фага лямбда, инвертированные повторы транспозона  $Tn3$ . Механизм получения однонаправленных делеций в данном векторе основан на том, что *in vivo* происходят делеции с фиксированных точек инвертированных повторов до переменных точек соседнего (клонированного) фрагмента ДНК [Sugino, Morita, 1994]. Преимуществом данной системы является то, что подобные делеции вставки с этим вектором можно осуществлять и *in vitro*, просто добавляя фермент  $Tn3$  транспозазу [Morita et al., 1987; Ichikawa, Ohtsubo, 1990]. В своей следующей работе авторы реализовали эту возможность, добившись более быстрых результатов при получении необходимых субклонов, добавляя в реакционную смесь, содержащую ДНК вектора со вставкой, предварительно гидролизованную соответствующей рестрикционной эндонуклеазой,  $Tn3$  транспозазу, представляющую собой практически неочищенную аммонийсульфатную фракцию клеточного лизата клона *E. coli*, несущего плазмиду с суперпродукцией данного фермента [Morita et al., 1987].

Сообщается о сконструированной космиде, несущей мини-дериват транспозона  $Tn5$ , сконструированной специально для выполнения проекта по секвенированию полного генома кишечной палочки [Kasai et al., 1992]. Ранее с помощью участка этого транспозона  $Tn5$  протяженностью всего 264 пн была создана конструкция на основе фага лямбда, позволяющая проводить систематическое секвенирование, стартуя с по-разному удаленных мест вставки, за счет внедренного транспозона с использованием для отжига праймера его консервативной последовательности [Phadnis et al., 1989]. Применение метода ПЦР позволило в значительной степени усовершенствовать этот процесс такого систематического секвенирования вставки в данном векторе на основе фага  $\lambda$  [Krishnan et al., 1991].

Еще одним транспозоном, применяемым для получения делеций *in vivo*, является транспозон Мю, характеризующийся наибольшей частотой

той транспозиции среди известных транспозонов. На его основе создан специализированный вектор, с помощью которого определена нуклеотидная последовательность довольно протяженного фрагмента гена гиразы *E. coli* [Adachi et al., 1987].

Однако все эти векторы и векторные системы имеют существенный недостаток. Поскольку проведение подобных делеционных процедур требует использования специальных штаммов *E. coli* и весьма сложной схемы их выращивания, это служит некоторым препятствием к их широкому использованию. Другими отрицательными моментами, особенно для векторов, рассчитанных на клонирование крупных фрагментов ДНК, являются некоторая нестабильность вставки, трудности в ее реориентации, вызванной необходимостью “читать” другую цепь ДНК. Также для большинства транспозонов характерным является сосредоточение мест их внедрения у дистального конца вставки по сравнению со значительно меньшей частотой данных событий в центральной и проксимальной частях [Ahmed, Podemski, 1997].

Также специального штамма, экспрессирующего фермент сайт-специфическую рекомбиназу фага лямбда, требует вектор Janus, созданный на основе одноцепочечного фага M13 серии mp и представляющий собой модифицированный вектор, несущий *int* ген фага лямбда [Burland et al., 1993]. Отличительной чертой этого вектора является возможность переориентации вставки *in vivo*, приводящая соответственно к упаковке в фаговые частицы разных цепей клонированного фрагмента ДНК. Произведенный авторами подсчет показал, что для завершения проекта секвенирования ДНК с помощью вектора Janus требуется значительно меньше усилий, чем при секвенировании случайных клонов или при осуществлении праймерной “прогулки”.

## 8.6. МУЛЬТИПЛЕКСНОЕ СЕКВЕНИРОВАНИЕ ДНК

Мультиплексное секвенирование ДНК, основанное на одновременном использовании целого набора плазмидных векторов plex00-20, было разработано первоначально для метода секвенирования ДНК путем химической дегградации с целью повышения его производительности [Church, Kieffer-Higgins, 1988]. Однако наибольшее развитие этот подход получил при секвенировании ДНК ферментативным методом по Сэнгеру. Несмотря на то, что и имеющиеся плазмидные вектора plex00-20 могут быть применимы для секвенирования с помощью дидезокситерминаторов, на основе одноцепочечного вектора M13 и некоторых из этих векторов (последовательность “тэгов” которых не нарушала рамку считывания галактозидазного гена), все же было сконструировано 11 новых M13plex векторов, позволяющих проводить ферментативное мультиплексное секвенирование по нарабатываемым одноцепочечным матрицам ДНК [Heller et al., 1991]. Сообщается об использовании набора из пяти векторов Plex, для которых характерно наличие мест отжига для E и P секвенирующих праймеров [Creasey et al., 1991].



Главный принцип секвенирования ДНК с помощью мультиплексного подхода заключается в одновременном проведении терминирующих реакций со всеми секвенируемыми матрицами ДНК, клонированными изначально в различных специализированных векторах. Затем следует их разделение в секвенирующем полиакриламидном геле, перенос на мембранный фильтр и фиксация перенесенных фрагментов ДНК на этом фильтре. Таким образом, на фильтре оказываются фиксированными “невидимые” лестницы полос всех исследуемых фрагментов ДНК. Для их выявления необходимы этапы молекулярной гибридизации с мечеными олигонуклеотидными зондами, гомологичными каждый раз фрагменту только одного определенного вектора из пула использованных. Так, 20 самостоятельных плазмидных векторов plex00-20 несут по два различающихся участка для отжига праймеров каждый, что составляет 40 вариантов “тэгов”. Аналогичными “тэгами” располагают и одиннадцать M13plex и пять Plex векторов. В результате гибридизации с каким-либо меченым олигонуклеотидным праймером после радиоавтографии на рентгеновской пленке становились видны полосы ДНК, принадлежащие только одной конкретной матрице. Для выявления следующего клона необходим этап полного удаления первого меченого зонда и повторная гибридизация с новым зондом. И так столько раз, сколько матриц в различающихся векторах было первоначально взято в исследование.

Следует отметить, что необходимым условием осуществления мультиплексного секвенирования ДНК является перенос фрагментов ДНК из полиакриламидного секвенирующего геля на мембранный фильтр, поскольку проведение всех этих последовательных процедур гибридизации, окрашивания, отмывки возможно только с мембранным фильтром. Второе условие заключается в прочной сорбции фрагментов ДНК на данном фильтре, выдерживающей многократные инкубации в различных растворах. В противном случае мультиплексное секвенирование будет просто невозможно. Немаловажное значение имеет тип используемой метки. Так, при использовании радиоактивной метки процесс выявления последовательностей всех матриц ДНК довольно продолжителен ввиду требующейся длительной экспозиции фильтра на рентгеновскую пленку, занимающей иногда несколько суток. Гораздо быстрее результаты могут быть получены с помощью хемилюминесценции такого субстрата, как 1,2-диоксетан. Считается, что каждый цикл мультиплексного секвенирования с его использованием требует 1 ч на гибридизацию с меченой пробой и 1,5 ч на все процедуры по выявлению гибридизационных сигналов [Creasey et al., 1991].

Обычно мультиплексное секвенирование проводится с немечеными фрагментами ДНК и уже потом выявление на фильтре секвенируемых полос ДНК зависит от специфической пробы, меченой тем или иным соединением (радиоактивностью, биотином и др.), что делает неизбежным этап блот-гибридизации. Однако было предложено интересное решение по мультиплексному секвенированию нескольких матриц ДНК, не требующее этапа гибридизации со специфическими пробами. Суть этого способа заключается в одновременном разделении в одном геле

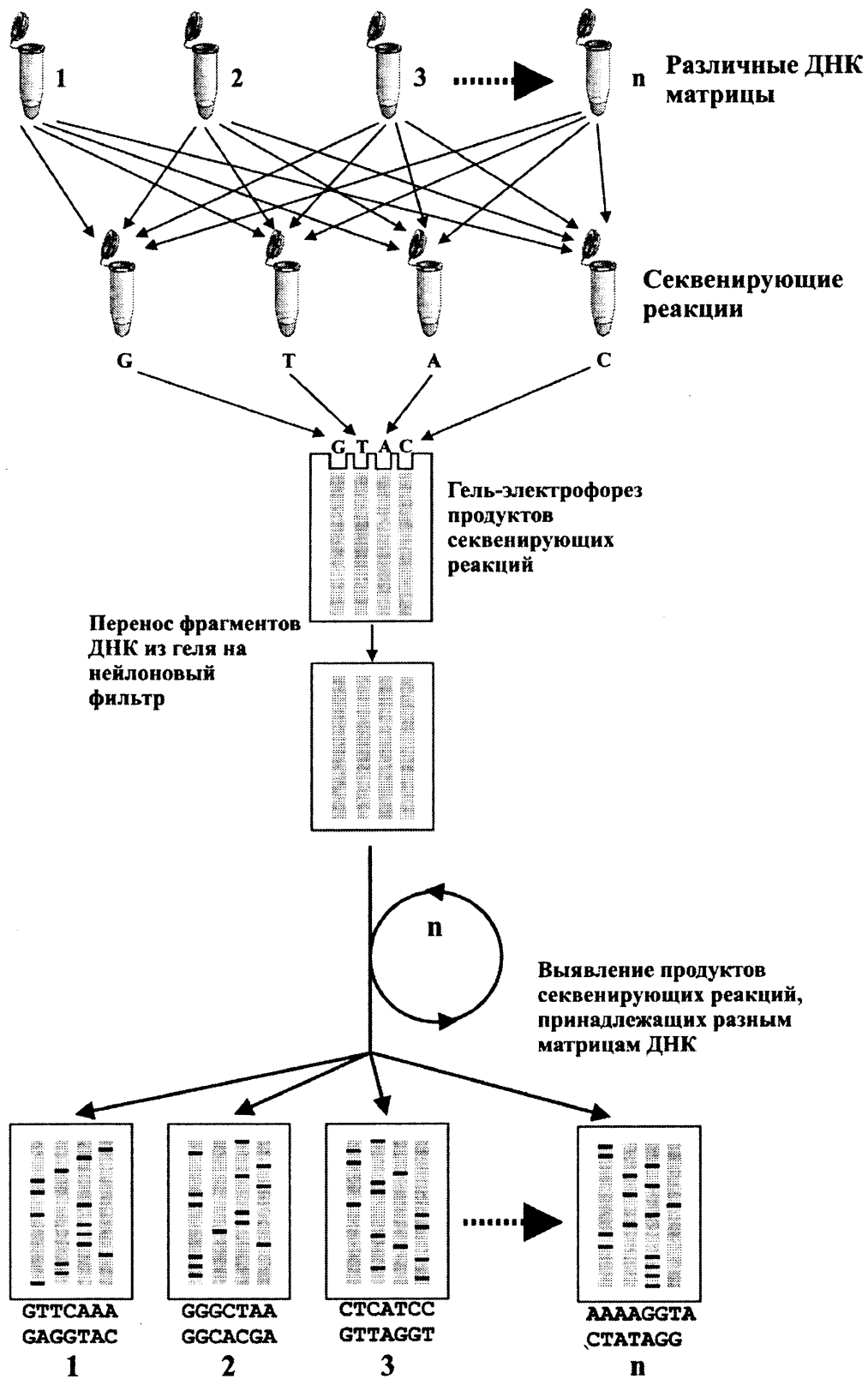


Рис. 8.28. Схема мультиплексного секвенирования фрагментов ДНК ферментативным методом

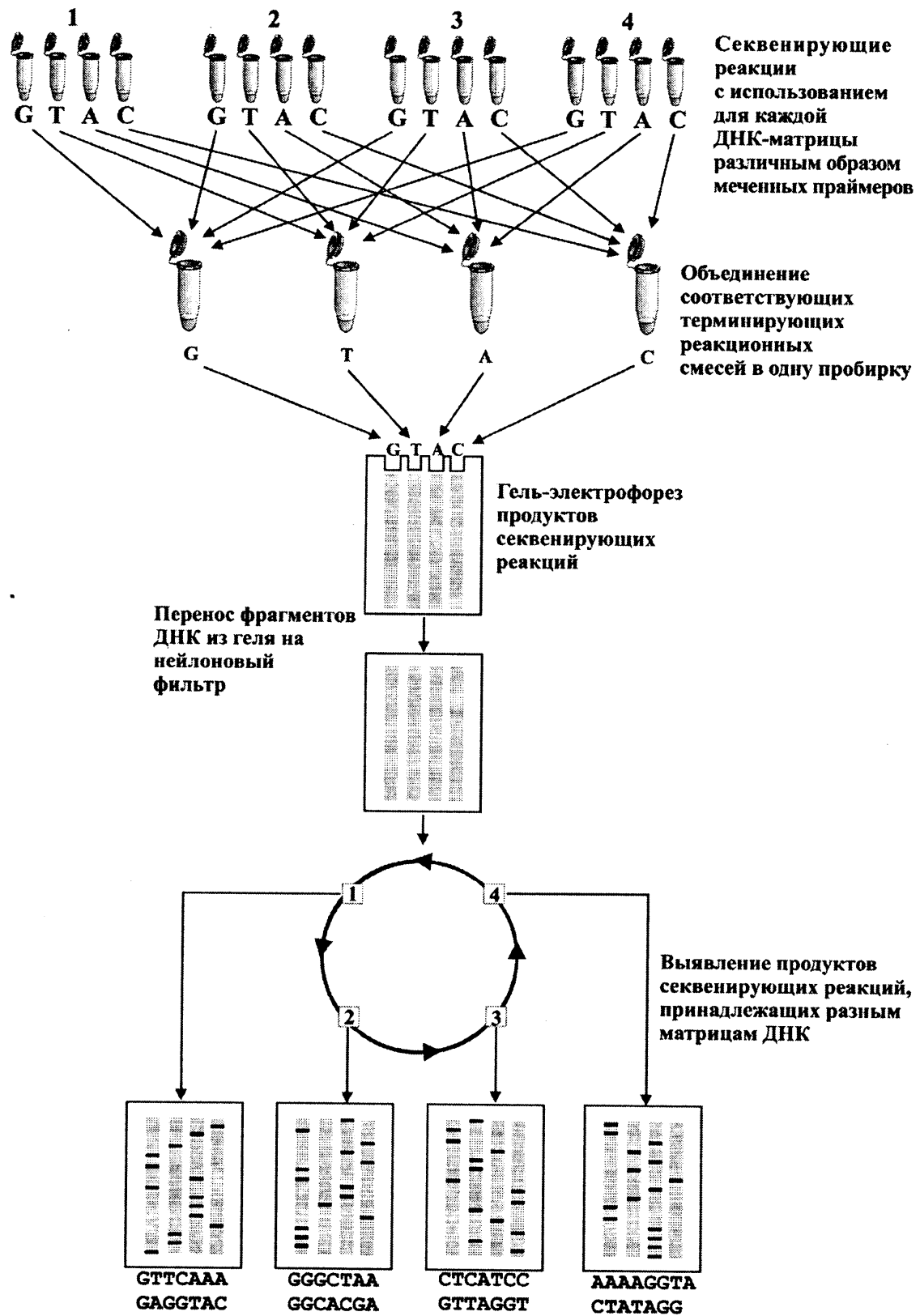


Рис. 8.29. Схема мультиплексного секвенирования ДНК с помощью мультиплексного мечения

по-разному меченных фрагментов ДНК, принадлежащих конкретным матрицам, и поэтому авторы назвали этот метод мультиплексным мечением [Olesen et al., 1993]. Так, ими в качестве праймеров были использованы олигонуклеотиды, меченные по 5'-концу разными гаптенами: биотином, дигоксигенином, 2,4-динитрофенилом и флуоресцеином. В результате терминирующих реакций, секвенирующего электрофореза, переноса фрагментов ДНК из геля на мембранный фильтр, проводимых одновременно со всеми этими матрицами, появилась возможность сразу, без этапа гибридизации, последовательно выявлять на фильтре полосы ДНК, соответствующие определенным матрицам, что в значительной мере способствовало более быстрому получению данных. Одно из главных отличий этого подхода от обычного мультиплексного секвенирования заключается в том, что здесь электрофоретическому разделению подвергаются продукты терминирующих реакций, уже несущие определенные, различные для каждой матрицы метки. Однако для выявления меченных ими фрагментов ДНК, перенесенных на мембранный фильтр, необходимо проведение соответствующих последовательных процедур. Для подобного секвенирования не требуется клонирования или переклонирования вставки в специализированном векторе, поэтому в качестве зондов (в данном случае – секвенирующих праймеров) могут выступать стандартные универсальные праймеры, меченные по 5'-концу различными гаптенами. Таким образом, становятся необходимы только этапы связывания конъюгата антители к соответствующему гаптену и щелочной фосфатазы (или комплекса стрептавидин/щелочная фосфатаза для биотина) и затем с помощью хемилюминесценции 1,2-диоксетанового субстрата выявления полос ДНК на фильтре (рис. 8.29).

Авторы показали, что наибольшую чувствительность имеет система с биотинилированным праймером. В 8 раз худшая чувствительность отмечалась ими при использовании флуоресцеинового гаптена и в 16 – для дигоксигенина и 2,4-динитрофенила. Поскольку при многократных этапах выявления полос ДНК и их последующего удаления с фильтра все же происходит незначительный смыв сорбированных фрагментов ДНК, то в цитируемой работе рекомендовалось последовательное выявление секвенируемой ДНК, несущей дигоксигениновый и/или динитрофенильный гаптены, затем флуоресцеиновый гаптен, и в завершение – биотинилированную метку.

### *Цитируемая литература*

*Ажикина Т.Л., Шевченко Ю.О., Лебедев Ю.Б., Веселовская С.В., Мясников В.А., Потапов В.К., Свердлов Е.Д.* Олигонуклеотиды, образующие высокостабильные дуплексы, их использование в качестве праймеров при секвенировании и в полимеразной цепной реакции // Докл. РАН. 1993. Т. 330. С. 642–645.

*Кнорре Д.Г., Пышный Д.В., Иванова Е.М., Бондарь А.А., Морозов И.В., Мертвцов Н.П., Зарьтова В.Ф.* Тетрануклеотиды и их феназиновые производные в качестве составных праймеров для секвенирования ДНК методом Сэнгера // Там же. 1996. Т. 350. С. 119–120.

*Чемерис А.В., Ахунов Э.Д., Куликов А.М., Вахитов В.А.* Прогрессивная тетра/окта стратегия секвенирования протяженных фрагментов ДНК // Междунар. конф., посвящ. памяти акад. А.А. Баева. М., 1996. С. 134.

- Adachi M., Mizuuchi M., Robinson E.A., Appella E., O'Dea M.H.* DNA sequence of the *E. coli gyrB* gene: Application of a new sequencing strategy // *Nucl. Acids Res.* 1987. Vol. 15. P. 771–784.
- Abath F.G.C., Holder A.A.* A strategy to sequence repetitive DNA based on partial restriction enzyme cleavage // *Appl. Biochem. Biotechnol.* 1995. Vol. 53. P. 195–197.
- Ahmed A.* Use of transposon-promoted deletions in DNA sequence analysis // *J. Mol. Biol.* 1984. Vol. 178. P. 941–948.
- Ahmed A.* A rapid procedure for DNA sequencing using transposon-promoted deletions in *Escherichia coli* // *Gene.* 1985. Vol. 39. P. 305–310.
- Ahmed A.* A simple and rapid procedure for sequencing long (40-kb) DNA fragments // *Gene.* 1987. Vol. 61. P. 363–372.
- Ahmed A.* A vector for sequencing long (40-kb) DNA fragments // *Gene.* 1989. Vol. 75. P. 315–321.
- Ahmed A., Podemski L.* Use of ordered deletions in genome sequencing // *Gene.* 1997. Vol. 197. P. 367–373.
- Akiyama H., Kanai S., Ozawa H.* Multi-priming sequencing: A DNA sequencing method involving restriction enzyme-digested DNA fragments as primers // *J. Biochem.* 1992. Vol. 111. P. 589–593.
- Anderson S.* Shotgun DNA sequencing using cloned DNASE-I generated fragments // *Nucl. Acids Res.* 1981. Vol. 9. P. 3015–3027.
- Andersson B., Lu J., Shen Y., Wentland M.A., Gibbs R.A.* Simultaneous shotgun sequencing of multiple cDNA clones // *DNA Sequenc.* 1997. Vol. 7. P. 63–70.
- Andersson B., Povinelli C.M., Wentland M.A., Shen Y., Muzny D.M., Gibbs R.A.* Adaptor-based uracil DNA glycosylase cloning simplifies shotgun library construction for large-scale sequencing // *Anal. Biochem.* 1994. Vol. 218. P. 300–308.
- Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.* A “double adaptor” method for improved shotgun library construction // *Ibid.* 1996. Vol. 236. P. 107–113.
- Ansorge W., Voss H., Zimmermann J.* (ed.). DNA sequencing strategies. N.Y. Wiley; Spektrum, 1996. 202 p.
- Arnold C., Hodgson I.J.* Vectorette PCR: A novel approach to genomic walking // *PCR Meth. Appl.* 1991. Vol. 1. P. 39–42.
- Azhikina T., Veselovskaya S., Myasnikov V., Potapov V., Ermolayeva O., Sverdlov E.* Strings of contiguous modified pentanucleotides with increased DNA-binding affinity can be used for DNA sequencing by primer walking // *Proc. Nat. Acad. Sci. USA.* 1993. Vol. 90. P. 11460–462.
- Barnes W.M., Bevan M.* Kilo-sequencing: An ordered strategy for rapid DNA sequence data acquisition // *Nucl. Acids Res.* 1983. Vol. 11. P. 349–368.
- Berg C.M., Vartak N.M., Wang G., Xu X., Liu L., MacNeil D.J., Gewain K.M., Wiater L.A., Berg G.E.* The  $\text{my}\delta\text{-1}$  element, a small  $\gamma\delta$  (Tn1000) derivative useful for plasmid mutagenesis, allele replacement and DNA sequencing // *Gene.* 1991. Vol. 113. P. 9–16.
- Breault D.T., Stover M.L., Li M., Lichtler A.C., Rowe D.W.* Novel use for BAL31 nuclease-generated nested deletions: Sequencing from the inside out // *Biotechniques.* 1995. Vol. 18. P. 614–617.
- Brenner D.G., Shaw W.V.* The use of synthetic oligonucleotides with universal templates for rapid DNA sequencing: Results with staphylococcal replicon pC221 // *EMBO. J.* 1985. Vol. 4. P. 561–568.
- Brosius J.* Superpolylinkers in cloning and expression vectors // *DNA.* 1989. Vol. 8. P. 759–777.
- Burbelo P.D., Iadarola M.J.* Rapid plasmid sequencing with multiple octamer primers // *Biotechniques.* 1994. Vol. 16. P. 645–650.
- Burland V., Daniels D.L., Plunkett G., III, Blattner F.R.* Genome sequencing on both strands: the Janus strategy // *Nucl. Acids Res.* 1993. Vol. 21. P. 3385–3390.
- Burland V., Kusakawa N.* Long PCR facilitates concise cloning and sequencing with a minimal tilling set of templates // *Biotechniques.* 1997. Vol. 23. P. 1070–1075.
- Burt A., Carter D.A., White T.J., Taylor J.W.* DNA sequencing with arbitrary primer pairs // *Mol. Ecol.* 1994. Vol. 3. P. 523–525.
- Burton F.H., Loeb D.D., McGraw R.A., Edgell M.H., Hutchinson C.A., III.* A directed nucleotide-sequencing approach for single-stranded vectors based on recloning intermediates of a progressive DNA synthesis reaction // *Gene.* 1988. Vol. 67. P. 159–168.
- Church G.M., Kieffer-Higgins S.* Multiplex DNA sequencing // *Science.* 1988. Vol. 240. P. 47–50.

Creasey A., D'Angio L., Jr., Dunne T.S., Kissinger C., O'Keeffe T., Perry-O'Keeffe H., Moran L.S., Roskey M., Schildkraut I., Sears L.E., Slatko B. Application of a novel chemiluminescence-based DNA detection method to single-vector and multiplex sequencing // *Biotechniques*. 1991. Vol. 11. P. 102–109.

Dale R.M., McClure B.A., Houchins J.P. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing application to sequencing the corn mitochondrial 18S rDNA // *Plasmid*. 1985. Vol. 13. P. 31–40.

Deininger P.L. Approaches to rapid DNA sequence analysis // *Anal. Biochem.* 1983a. Vol. 135. P. 247–263.

Deininger P.L. Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis // *Ibid.* 1983b. Vol. 129. P. 216–223.

Devon R.S., Porteous D.J., Brookes A.J. Splinkerettes – improved vectorettes for greater efficiency in PCR walking // *Nucl. Acids Res.* 1995. Vol. 23. P. 1644–1645.

Dolan M., Ally A., Purzycki M.S., Gilbert W., Gillevet P.M. Large-scale genomic sequencing: Optimization of genomic chemical sequencing reactions // *Biotechniques*. 1995. Vol. 19. P. 264–273.

Dominguez O., Lopez-Larrea C. Gene walking by unpredictably primed PCR // *Nucl. Acids Res.* 1994. Vol. 22. P. 3247–3248.

Dunn J.J., Butler-Loffredo L.L., Studier F.W. Ligation of hexamer on hexamer templates to produce primers for cycle sequencing or the polymerase chain reaction // *Anal. Biochem.* 1995. Vol. 228. P. 91–100.

Eckert R.L. New vectors for rapid sequencing of DNA fragments by chemical degradation // *Gene*. 1987. Vol. 51. P. 247–254.

Fitzgerald M.C., Skowron P., Van Etten J.L., Smith L., Mead D.A. Rapid shotgun cloning utilizing the two base recognition endonuclease *Cvi*II // *Nucl. Acids Res.* 1992. Vol. 20. P. 3753–3762.

Fu D.-J., Broude N.E., Koster H., Smith C.L., Cantor C.R. A DNA sequencing strategy that requires only five bases of known terminal sequence for priming // *Proc. Nat. Acad. Sci. USA*. 1995. Vol. 92. P. 10162–10166.

Furuyama H., Okano K., Kambara H. DNA sequencing directly from a mixture using terminal-base-selective primers // *DNA Res.* 1994. Vol. 1. P. 231–237.

Gingrich J.C., Boehrer D.M., Basu S.B. Partial *Cvi*II digestion as an alternative approach to generate cosmid sublibraries for large-scale sequencing projects // *Biotechniques*. 1996. Vol. 21. P. 99–104.

Guo L.-H., Wu R. New rapid methods for DNA sequencing based on exonuclease III digestion followed by repair synthesis // *Nucl. Acids Res.* 1982. Vol. 10. P. 2065–2084.

Guo L.-H., Yang R.C.A., Wu R. An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid // *Ibid.* 1983. Vol. 11. P. 5521–5540.

Heller C., Radley E., Khurshid F.A., Beck S. M13plex vectors for multiplex DNA sequencing // *Gene*. 1991. Vol. 103. P. 131–132.

Henikoff S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing // *Gene*. 1984. Vol. 28. P. 351–359.

Henikoff S. Unidirectional digestion with exonuclease III in DNA sequence analysis // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 156–165.

Henikoff S. Ordered deletions for DNA sequencing and *in vitro* mutagenesis by polymerase extension and exonuclease III gapping of circular templates // *Nucl. Acids Res.* 1990. Vol. 18. P. 2961–2966.

Hong G.F. The use of DNase I, buffer gradient gel, and <sup>35</sup>S label for DNA sequencing // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 93–110.

Hong G.F. A simple method for identification of gap-bridging subclones in DNA sequencing // *Biosci. Rep.* 1988. Vol. 8. P. 549–551.

Ichikawa H., Ohtsubo E. *In vitro* transposition of the transposon Tn3 // *J. Biol. Chem.* 1990. Vol. 265. P. 18829–18832.

Johnson D.F., Nierlich D.P., Lusic A.J. Use of the *lac* repressor in constructing sequential deletions and a new sequencing vector // *Gene*. 1990. Vol. 94. P. 9–14.

Johnston S.L., Strausbaugh M., Sarkar G., Wettstein P.J. A novel method for sequencing members of multi-gene families // *Nucl. Acids Res.* 1995. Vol. 23. P. 3074–3075.

Jones D.H., Winistorfer S.C. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA // *Ibid.* 1992. Vol. 20. P. 595–600.

- Jones D.H., Winistorfer S.C. Genome walking with 2–4 kb steps using panhandle PCR // PCR Meth. Appl. 1993. Vol. 2. P. 197–203.
- Kaczorowski T., Szybalski W. Assembly of 18-nucleotide primers by ligation of three hexamers: Sequencing of large genomes by primer walking // Anal. Biochem. 1994. Vol. 221. P. 127–135.
- Kaszorowski T., Szybalski W. Co-operativity of hexamer ligation // Gene. 1996. Vol. 179. P. 189–193.
- Kalisch B.W., Krawetz S.A., Schoenwaelder K.-H., Van de Sande J.H. Covalently linked sequencing primer linkers (splinkers) for sequence analysis of restriction fragments // Gene. 1986. Vol. 44. P. 263–270.
- Kasai H., Isono S., Kitakawa M., Mineno J., Akiyama H., Kunit D.M., Berg D.E., Isono K. Efficient large-scale sequencing of the *Escherichia coli* genome: Implementation of a transposon- and PCR-based strategy for the analysis of ordered  $\lambda$  phage clones // Nucl. Acids Res. 1992. Vol. 20. P. 6509–6515.
- Kieleczawa J., Dunn J.J., Studier F.W. DNA sequencing by primer walking with strings of contiguous hexamers // Science. 1992. Vol. 258. P. 1787–1791.
- Kitabatake M., Inokuchi H. A simplified method for generating step-wide deletions using PCR // Gene. 1993. Vol. 123. P. 59–61.
- Konings R.N.H., Verhoeven E.J.M., Peeters B.P.H. pKUN, vectors for separate production of both DNA strands of recombinant plasmids // Meth. Enzymol. 1987. Vol. 153: Recombinant DNA, pt D. P. 12–34.
- Kotler L.E., Zevin-Sonkin D., Sobolev I.A., Beskin A.D., Ulanovsky L. DNA sequencing: Modular primers assembled from a library of hexamers or pentamers // Proc. Nat. Acad. Sci. USA. 1993. Vol. 90. P. 4241–4245.
- Krishnan B.R., Kersulyte D., Brikun I., Berg C.M., Berg D.E. Direct and crossover PCR amplification to facilitate Tn5<sup>supF</sup>-based sequencing of  $\lambda$  phage clones // Nucl. Acids Res. 1991. Vol. 19. P. 6177–6182.
- Lagerstrom M., Parik J., Malmgren H., Stewart J., Petterson U., Landegren U. Capture PCR: Efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA // PCR Meth. Appl. 1991. Vol. 1. P. 111–119.
- Lamperti E.D., Villa-Komaroff L. Generation of deletion subclones for sequencing by partial digestion with restriction endonucleases // Anal. Biochem. 1990. Vol. 185. P. 187–193.
- Laughon A., Scott M.P. Sequence of a *Drosophila* segmentation gene: Protein structure homology with DNA-binding proteins // Nature. 1984. Vol. 310. P. 25–31.
- Lee Y.-M., Lee S.-C. A DNA sequencing strategy // Anal. Biochem. 1988. Vol. 175. P. 521–524.
- Li C., Tucker P.W. Exoquence DNA sequencing // Nucl. Acids Res. 1993. Vol. 21. P. 1239–1244.
- Liu L., Whalen W., Das A., Berg C.M. Rapid sequencing of cloned DNA using a transposon for bidirectional priming: Sequence of the *Escherichia coli* K-12 *avtA* gene // Ibid. 1987. Vol. 15. P. 9461–9469.
- Malo M.S., Srivastava K., Andresen J.M., Chen X.-N., Korenberg J.R., Ingram V.M. Targeted gene walking by low stringency polymerase chain reaction: Assignment of a putative human brain sodium channel gene (*SCN3A*) to chromosome 2q24–31 // Proc. Nat. Acad. Sci. USA. 1994. Vol. 91. P. 2975–2979.
- Messing J., Crea R., Seeburg P.H. A system for shotgun DNA sequencing // Nucl. Acids Res. 1981. Vol. 9. P. 309–321.
- Milavetz B. Preparation of nested deletions in single-strand DNA using oligonucleotides containing partially random base sequences // Ibid. 1992. Vol. 20. P. 3529–3530.
- Misra T.K. A new strategy to create ordered deletions for rapid nucleotide sequencing // Gene. 1985. Vol. 34. P. 263–268.
- Misra T.K. DNA sequencing: A new strategy to create ordered deletions, modified M13 vector, and improved reaction conditions for sequencing by dideoxy chain termination method // Meth. Enzymol. 1987. Vol 155: Recombinant DNA. P. 119–138.
- Morita M., Tsunasawa S., Sugino Y. Overproduction and purification of the Tn3 transposase // J. Biochem. 1987. Vol. 101. P. 1253–1264.
- Morita M., Umemoto A., Li Z.X., Nakazono N., Sugino Y. Nested deletions from a fixed site as an aid to nucleotide sequencing: An *in vitro* system using Tn3 transposase // DNA Res. 1996. Vol. 3. P. 431–433.

- Nguyen T.N., Samuelson P., Sterky F., Merle Poitte C., Robert A., Baussant T., Haeuw J-F., Uhlen M., Binz H., Stahl S. Chromosomal sequencing using a PCR-based biotin-capture method allowed isolation of the complete gene for the outer membrane protein A of *Klebsiella pneumoniae* // *Gene*. 1998. Vol. 210. P. 93–101.
- Niemi J.T., Mantsala P. TRACTOR, a program to locate subclones sequence using only one sequencing reaction // *Comp. Appl. Biosci.* 1994. Vol. 10. P. 563–566.
- Nurminsky D.I., Hartl D.L. Sequence scanning: A method for rapid sequence acquisition from large-fragment DNA clones // *Proc. Nat. Acad. Sci. USA*. 1996. Vol. 93. P. 1694–1698.
- Ochman H., Gerber A.S., Hartl D.L. Genetic applications of an inverse polymerase chain reaction // *Genetics*. 1988. Vol. 120. P. 621–623.
- Ohara O., Dorit R.L., Gilbert W. Direct genomic sequencing of bacterial DNA: The pyruvate kinase I gene of *Escherichia coli* // *Proc. Nat. Acad. Sci. USA*. 1989. Vol. 86. P. 6883–6887.
- Okano K., Kambara H. Fragment walking for long DNA sequencing by using a library as small as 16 primers // *Gene*. 1996. Vol. 176. P. 231–235.
- Olesen S.E., Martin C.S., Bronstein I. Chemiluminescent DNA sequencing with multiplex labelling // *Biotechniques*. 1993. Vol. 15. P. 480–485.
- Pan H.Q., Wang Y.P., Chhissoe S.L., Bodenteich A., Wang Z., Iyer K., Clifton S.W., Crabtree J.S., Roe B.A. The complete nucleotide sequences of the SacBII Kan domain of the P1 pAD10-SacBII cloning vector and three cosmid cloning vectors: pTCF, svPHEP, and LAWRIST16 // *Genet. Anal. Tech. Appl.* 1994. Vol. 11. P. 181–186.
- Parker J.D., Rabinovitch P.S., Burner G.C. Targeted gene walking polymerase chain reaction // *Nucl. Acids Res.* 1991. Vol. 19. P. 3055–3060.
- Peng Z-G., Wu R. A simple and rapid nucleotide sequencing strategy and its application in analyzing a rice histon 3 gene // *Gene*. 1986. Vol. 45. P. 247–252.
- Peng Z-G., Wu R. A new and simple rapid method for sequencing DNA // *Meth. Enzymol.* 1987. Vol. 155: Recombinant DNA. P. 214–238.
- Phadnis S.H., Huang H.V., Berg D.F. Tn5supF, a 264-base-pair transposon derived from Tn5 for insertion mutagenesis and sequencing DNA cloned in phage  $\lambda$  // *Proc. Nat. Acad. Sci. USA*. 1989. Vol. 86. P. 5908–5912.
- Poncz M., Solowiejczyk D., Ballantine M., Scwartz E., Surrey S. “Nonrandom” DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method // *Ibid.* 1982. Vol. 79. P. 4298–4302.
- Povinelli C.M., Gibbs R.A. Large-scale sequencing library production: An adaptor-based strategy // *Anal. Biochem.* 1993. Vol. 210. P. 16–26.
- Prentki P., Binda A., Epstein A. Plasmid vectors for selecting *IsI*-promoted deletions in cloned DNA: Sequence analysis of the omega interposon // *Gene*. 1991. Vol. 103. P. 17–23.
- Putney S.D., Benkovic S.J., Schimmel P.R. A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo* // *Proc. Nat. Acad. Sci. USA*. 1981. Vol. 78. P. 7350–7354.
- Raja M.C., Zevin-Sonkin D., Shwartzburd J., Rozovskaja T.A., Sobolev I.A., Chertkov O., Ramanathan V., Lvovsky L., Ulanovsky E. DNA sequencing using differential extension with nucleotide subsets (DENS) // *Nucl. Acids Res.* 1997. Vol. 25. P. 800–805.
- Ren Z., Nie W., Black L.W. A powerful approach for generating and sequencing DNA deletions: Sequencing from the outside in // *Anal. Biochem.* 1997. Vol. 245. P. 112–114.
- Rena G., Houslay M.D. A simple method for sequencing small DNAs by introducing precise overlapping ends into restriction digestion fragments // *Nucl. Acids Res.* 1998. Vol. 26. P. 3867–3868.
- Rosenthal A., Jones D.S.C. Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction // *Ibid.* 1990. Vol. 18. P. 3095.
- Ruther U. pUR250 allows rapid chemical sequencing of both DNA strands of its inserts // *Ibid.* 1982. Vol. 10. P. 5765–5772.
- Ruther U., Koenen M., Otto K., Muller-Hill B. pUR222, a vector for cloning and rapid chemical sequencing of DNA // *Ibid.* 1981. Vol. 9. P. 4087–4098.
- Sanchez-Pescador R., Urdea M.S. Use of unpurified synthetic deoxynucleotide primers for rapid dideoxynucleotide chain termination sequencing // *DNA*. 1984. Vol. 4. P. 339–343.
- Scholler P., Karger A.E., Meier-Ewert S., Lehrach H., Delius H., Hoheisei J.D. Fine-mapping of shotgun template-libraries: An efficient strategy for the systematic sequencing of genomic DNA // *Nucl. Acids Res.* 1995. Vol. 23. P. 3842–3849.



- Schriefer L.A., Gebauer B.K., Qiu L.Q.Q., Waterston R.H., Wilson R.K.* Low pressure DNA shearing: A method for random DNA sequence analysis // *Ibid.* 1990. Vol. 18. P. 7455–7456.
- Sharrocks A., Hornby D.P.* A rapid method for Ba131 deletion analysis // *Ibid.* 1987. Vol. 15. P. 8564.
- Siebert P.D., Chenchik A., Kellog D.E., Lukyanov K.A., Lukyanov S.A.* An improved PCR method for walking in uncloned genomic DNA // *Ibid.* 1995. Vol. 23. P. 1087–1088.
- Siemieniak D.R., Sieu L.C., Slightom J.L.* Strategy and methods for directly sequencing cosmid clones // *Anal. Biochem.* 1991. Vol. 192. P. 441–448.
- Siemieniak D.R., Slightom J.L.* A library of 3342 useful nonamer primers for genome sequencing // *Gene.* 1990. Vol. 96. P. 121–124.
- Slightom J.L., Bock J.H., Siemieniak D.R., Hurst G.D., Beattie K.L.* Nucleotide sequencing double-stranded plasmids with primers selected from a nonamer library // *Biotechniques.* 1994. Vol. 17. P. 536–537, 540–544.
- Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H., Dubois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K., Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D., Spadafora R., Vicaire R., Wang Y., Wierzbowski J., Gibson R., Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S., McDougall S., Shimer G., Goyal A., Pietrokovski S., Church G.M., Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.* Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics // *J. Bacteriol.* 1997. Vol. 179. P. 7135–7155.
- Sorge J.A., Blinderman L.A.* ExoMeth sequencing of DNA: Eliminating the need for subcloning and oligonucleotide primers // *Proc. Nat. Acad. Sci. USA.* 1989. Vol. 86. P. 9208–9212.
- Sorge J.A., Huse W.D.* Method of DNA sequencing. Pat. 5354656 (US). Publ. 1994.
- Stallard R.L., Certa U., Bannwarth W.* Dideoxy sequencing of double-stranded DNA from poly(A)-extended 3'-ends // *Anal. Biochem.* 1987. Vol. 162. P. 197–201.
- Sterky F., Holmberg A., Alexandersson G., Lundberg J., Uhlen M.* Direct sequencing of bacterial artificial chromosomes (BACs) and prokaryotic genomes by biotin-capture PCR // *J. Biotechnol.* 1998. Vol. 60. P. 119–129.
- Strausbaugh L.D., Bourke M.T., Sommer M.T., Coon M.E., Berg C.M.* Probe mapping to facilitate transposon-based DNA sequencing // *Proc. Nat. Acad. Sci. USA.* 1990. Vol. 87. P. 6213–6217.
- Strauss E.C., Kobori J.A., Siu G., Hood L.E.* Specific-directed DNA sequencing // *Anal. Biochem.* 1986. Vol. 154. P. 353–360.
- Studier F.W.* A strategy for high-volume sequencing of cosmid DNAs: Random and directed priming with a library of oligonucleotides // *Proc. Nat. Acad. Sci. USA.* 1989. Vol. 86. P. 6917–6921.
- Sugino Y., Morita M.* A new DNA cloning/sequencing vector with a built-in mechanism for generation of nested deletions using transposon Tn3 // *Gene.* 1994. Vol. 148. P. 169–170.
- Szybalski W.* Proposal for sequencing DNA using ligation of hexamers to generate sequential elongation primers (SPEL-6) // *Gene.* 1990. Vol. 90. P. 177–178.
- Tomb J.F., Barcak G.J.* Regulating the 3'–5' activity of exonuclease III by varying the sodium chloride concentrations // *Biotechniques.* 1989. Vol. 7. P. 932–933.
- Triglia T., Peterson M.G., Kemp D.J.* A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences // *Nucl. Acids Res.* 1988. Vol. 16. P. 8186.
- Trueba G.A., Johnson R.C.* Random primed gene walking PCR: A simple procedure to retrieve nucleotide fragments adjacent to known DNA sequences // *Biotechniques.* 1996. Vol. 21. P. 20.
- Verhasselt P., Voet M., Volckaert G.* DNA sequencing by a subcloning-walking strategy using a specific and semi-random promoter in the polymerase chain reaction // *DNA Sequenc.* 1992. Vol. 2. P. 281–287.
- Volckaert G.* A systematic approach to chemical DNA sequencing by subcloning in pGV451 and derived vectors // *Meth. Enzymol.* Vol. 155: Recombinant DNA, pt F. P. 231–250.
- Voss H., Schwager C., Wiemann S., Zimmermann J., Stegemann J., Erfle H., Voie A.M., Drzonek H., Ansorge W.* Efficient low redundancy large-scale DNA sequencing at EMBL // *J. Biotechnol.* 1995. Vol. 41. P. 121–129.
- Voss H., Wiemann S., Grothues D., Sensen C., Zimmermann J., Schwager C., Stegemann J., Erfle H., Rupp T., Ansorge W.* Automated low-redundancy large-scale DNA sequencing by primer walking // *Biotechniques.* 1993. Vol. 15. P. 714–721.
- Wang G., Blakesley R.W., Berg D.E., Berg C.M.* pDUAL: A transposon-based cosmid cloning

vector for generating nested deletions and DNA sequencing templates *in vivo* // Proc. Nat. Acad. Sci. USA. 1993. Vol. 90. P. 7874–7878.

Weber K.L., Bolander M.E., Sarkab G. Rapid acquisition of unknown DNA sequence adjacent to a known segment by multiplex restriction site PCR // Biotechniques. 1998. Vol. 25. P. 415–419.

Williamson B.D., Rutherford C.L. Enrichment-mediated PCR amplification of an unknown DNA fragments flanking a known sequence // Ibid. 1994. Vol. 17. P. 670, 672.

Wong K.K., Markillie L.M., Saffer J.D. A novel method for producing partial restriction digestion of DNA fragments by PCR with 5-methyl-CTP // Nucl. Acids Res. 1997. Vol. 25. P. 4169–4171.

Wright W.E. Partial-digest DNA sequencing // Biotechniques. 1992. Vol. 13. P. 772–779.

Zhu K.Y., Clark J.M. Rapid construction of nested deletions of recombinant plasmid DNA for dideoxy sequencing // Ibid. 1995. Vol. 18. P. 222–224.

Yanisch-Perron C., Viera J., Messing J. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors // Gene. 1985. Vol. 33. P. 103–119.

Yohda M., Kato N., Endo I. Solid-phase nested deletion: A new subcloning-less method for generating nested deletions // DNA Res. 1995. Vol. 2. P. 175–181.

Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W., Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A. Large-scale concatenation cDNA sequencing // Genome Res. 1997. Vol. 7. P. 353–358.

## АВТОМАТИЧЕСКОЕ СЕКВЕНИРОВАНИЕ ДНК

Необходимость получения информации о последовательности нуклеотидов в ДНК во всевозрастающих масштабах заставила исследователей заняться автоматизацией всего процесса секвенирования ДНК, поскольку применявшиеся до этого только ручные методы уже не справлялись с такими большими объемами.

Первые успехи в автоматизации секвенирования ДНК, пожалуй, можно связать с полуавтоматическими и автоматическими методами “чтения” последовательности нуклеотидов с радиоавтографа секвенирующего геля или даже еще с более ранней “автоматической” обработкой и анализом последовательности нуклеотидов с помощью компьютеров и специализированных программ. Однако в настоящее время под автоматическим секвенированием ДНК понимается, в первую очередь, электрофоретическое разделение флуоресцентно меченных продуктов терминирующих реакций с помощью специальных приборов – автоматических секвенаторов ДНК. Другой аспект автоматизации процесса секвенирования ДНК заключается в наработке матриц для их последующего ферментативного секвенирования, а также проведении секвенирующих реакций (как в ферментативном методе, так и в методе химической дегградации) с помощью лабораторных биороботов.

### 9.1. МАТРИЦЫ ДЛЯ СЕКВЕНИРОВАНИЯ ДНК АВТОМАТИЧЕСКИМ МЕТОДОМ

Важной составляющей процесса секвенирования ДНК являются одно- или двуцепочечные матрицы. От их количества (как, впрочем, и от качества), во многом зависит производительность всего метода. При крупномасштабном секвенировании ДНК требуется настолько большое число матриц, что их подготовка стандартными способами, без применения автоматизации, становится лимитирующим фактором всего процесса.

При автоматическом секвенировании ДНК в качестве матриц наиболее часто применяются одно- или двуцепочечные молекулы фагов, плазмид или фагмид (кроме ПЦР-продуктов, которые представляют собой отдельный тип матриц для секвенирования). Выделение требуемого большого количества подобных образцов представляет собой некоторую проблему. Так, в работе Эперона сообщалось об изготовлении специальной установки, рассчитанной на одновременное выделение 96 таких образцов с помощью фильтрации [Eperon, 1986]. Выделенные таким образом одноцепочечные препараты ДНК на основе фага M13 показали высокое качество секвенирования, сравнимое с матрицами, получаемыми обычным путем с использованием того же полиэтиленгликоля в качестве осадителя фаговых частиц. Впоследствии этот метод был слегка модифицирован [Eperon et al., 1988]. Предложенные в этих

работах процедуры в дальнейшем подверглись некоторому усовершенствованию путем применения специального лабораторного робота Biomek 1000 (Beckman, США), что позволило выделять до 192 образцов ДНК в день [Smith et al., 1990]. В другой своей работе эти авторы сообщают об успешном выделении 384 образцов фагмидной одноцепочечной ДНК в день ручным способом, который, впрочем, мог быть легко автоматизирован [Rawlinson et al., 1991]. Однако осаждение фаговых частиц с помощью полиэтиленгликоля не совсем удобно, с точки зрения автоматизации этого процесса, из-за довольно высокой вязкости самого раствора, поэтому было предложено осаждать фаг уксусной кислотой [Zimmermann et al., 1989]. В этой работе все процедуры по осаждению фаговых частиц, очистке фаговой ДНК с помощью ее связывания со специальным стеклянным фильтром и последующими отмывками 4М перхлоратом натрия, 70%-ным этанолом и заключительная элюция  $0,1 \times TE$  проводились с использованием лабораторного робота Biomek 1000, что позволило авторам исключить необходимые прежде этапы центрифугирования и автоматизировать весь этот процесс выделения и очистки матриц ДНК полностью. Об очистке плазмидных матриц, пригодных для автоматического секвенирования ДНК с помощью 96-луночных микротитраторных планшетов со стеклофильтром, и сообщается в другой работе [Itoh et al., 1997]. В литературе также приводятся сведения об использовании различных вариантов крупномасштабного выделения одно- и двуцепочечных ДНК матриц для секвенирования, незначительно отличающихся выбором способов осаждения и очистки молекул ДНК [Mardis, Roe, 1989; Kolner et al., 1994; Zollo, Chen, 1994; Andersson et al., 1996; Beg, Holt, 1998]. Впрочем, в отдельных работах встречаются упоминания, что для автоматического секвенирования даже не потребовалось какой-то особой очистки ДНК матриц [Wilson et al., 1990; Trower et al., 1995; Chen et al., 1996]. Так, авторы последней работы сообщают, что они лишь обрабатывали бактериальный осадок протеиназой К, затем нагревали до  $95^{\circ}C$ , центрифугировали и отобранный супернатант уже непосредственно использовали для проведения циклического флуоресцентного секвенирования.

Некоторая экономия в приготовлении матриц на основе фага M13 для автоматического флуоресцентного секвенирования, предложенная японскими авторами [Mita et al., 1994], заключалась в одновременном секвенировании таких одноцепочечных матриц как с помощью прямых, так и обратных праймеров после построения комплементарной цепи, как это было ранее предложено для обычного секвенирования ДНК [Hong, 1981].

Продукты ПЦР, доступные в большом количестве, благодаря различным ДНК амплификаторам, представляют собой другой тип матриц для автоматического секвенирования [Rosenthal et al., 1993]. Если очистка таких ПЦР-продуктов для дальнейшего секвенирования ручным способом могла также проводиться вручную, обеспечивая подготовку требуемого количества матриц, то для автоматического секвенирования с резко возросшим потоком образцов ДНК без применения лабораторных биороботов было уже не справиться. В связи с этим практически

непременным условием автоматизации этого процесса является использование биотинилированного праймера и магнитных частиц со стрептавидином, позволяющих осуществить отделение целевого продукта ПЦР от ставших ненужными ингредиентов этой реакции. Так, в литературе описаны различные модификации лабораторных роботов, позволяющих проводить необходимое пипетирование и разделение фаз с помощью магнита [Alderton et al., 1992; Fry et al., 1992; Wahlberg et al., 1992; Tong, Smith, 1993; Rolfs, Weber, 1994]. Разработанная для этой цели фирмой "DYNAL" (Норвегия) магнитная установка MPC-auto96 позволяет подключать ее к биороботу Biomek и осуществлять необходимые процедуры по выделению и очистке матриц ДНК [Deggerdal et al., 1998].

## 9.2. ПРОВЕДЕНИЕ ТЕРМИНИРУЮЩИХ РЕАКЦИЙ ПРИ АВТОМАТИЧЕСКОМ СЕКВЕНИРОВАНИИ ДНК ФЕРМЕНТАТИВНЫМ МЕТОДОМ

Как уже было описано в главе 2, основными необходимыми ингредиентами при проведении секвенирующих реакций ферментативным методом служат, кроме самих матриц ДНК, затравочные молекулы, ДНК-полимеразы, смеси дНТФ/ддНТФ, меченые молекулы. Что касается затравочных молекул, то в автоматическом секвенировании используются те же олигонуклеотидные праймеры, что и описанные в разделе 2.3, включая как относительно короткие из намеренной библиотеки [Bock, Slinghtom, 1995], так и сегментированные гексамерные праймеры [Hou, Smith, 1994; Johnson et al., 1996; Kaczorowski, Szybalski, 1996; Lodhi, McCombie, 1996; Ruiz-Martinez et al., 1996]. Причем в последней работе описывается применение в качестве затравочных молекул гексамерных модулей вместе с SSB-белком и, как отмечают авторы, это было сделано впервые при разделении продуктов терминирующих реакций капиллярным электрофорезом. Хотя следует отметить, что в некоторых работах подчеркивается, что праймеры, используемые в обычном секвенировании ДНК с помощью радиоактивной метки, не всегда пригодны для автоматического секвенирования с применением флуоресцентной метки и требуют более строгого выбора их нуклеотидной последовательности [De Bellis et al., 1990; 1992].

В автоматическом секвенировании ДНК применяются различные высокопроцессивные ДНК-полимеразы. Выбор конкретной из них во многом зависит от температурных особенностей проводимых терминирующих реакций. Следует отметить, что весьма широкое применение для секвенирующих реакций, проводимых с помощью биоробота, находит фермент Bst ДНК-полимераза из-за ее высокой стабильности при комнатной температуре в отличие от секвеназы. Так, в ряде работ показано высокое качество получаемых результатов при использовании этого фермента для проведения секвенирующих реакций с помощью лабораторного робота [Earley et al., 1993; 1994]. В последней работе авторами было проведено сравнение двух ДНК-полимераз – секвеназы и

Bst ДНК-полимеразы, причем оказалось, что по прошествии уже 2 ч с момента запуска биоробота качество вновь синтезированных продуктов, образующихся в ходе терминирующих реакций с помощью секвенназы, резко ухудшается, тогда как даже после 4-часовой задержки Bst ДНК-полимераза давала практически схожую картину с реакциями, проводимыми сразу после запуска этого робота. Причиной тому является известная нестабильность секвенназы при комнатной температуре. В то же время имеется много работ, где в качестве ДНК-полимеразы при автоматическом секвенировании использовался именно этот фермент [Kristensen et al., 1988; Toneguzzo et al., 1988; Voss et al., 1990]. Был проведен сравнительный анализ эффективности различных ДНК-полимераз в автоматическом флуоресцентном секвенировании [Voss et al., 1993]. Так, было показано, что наилучшие результаты дает секвенназа, а Bst ДНК-полимераза ей несколько уступает, тогда как термостабильные ДНК-полимеразы ( $\Delta$ Taq и Vent(exo<sup>-</sup>)) характеризовались заметно худшей однородностью пиков ДНК. В этой же работе сравнивалось применение данных модифицированных ДНК-полимераз (секвенназа,  $\Delta$ Taq и Vent(exo<sup>-</sup>)) с их нативными формами в автоматическом флуоресцентном секвенировании и заметных различий при этом авторы не обнаружили. Высота пиков в зависимости от используемых ферментов анализировалась также в работах и других авторов (Parker et al., 1995; 1996). Так, ими было показано, что высота пиков, приходящихся на фрагмент ДНК с конкретным ддНМФ на его 3'-конце, зависит от предшествующих ему одного-двух нуклеотидов при построении цепи как с помощью обычной Taq ДНК-полимеразы, так и ее модифицированной версии AmpliTaq FS (F667Y), причем последний фермент все же давал более равномерные по высоте пики.

Достаточно трудоемкой операцией, требующей высокой точности исполнения, является проведение терминирующих реакций, которые при массовом автоматическом секвенировании становятся "узким местом". Использование производимых различными фирмами лабораторных роботов, таких как Biomek 1000, Biomek 2000, Biomek SL, Vistra DNA Labstation 625, ABI Prism CATALYST 800, ABI Prism 877, BioRobot 9600 или им подобных, работающих по принципу XYZ (где X – означает передвижение платформы в горизонтальном направлении, а Y и Z указывают на перемещение подвижного плеча робота в другом горизонтальном, а также вертикальном направлениях), смогло резко повысить производительность этого этапа и освободило экспериментатора от большого объема кропотливой работы [Wilson et al., 1988; Zimmermann et al., 1988, 1990; D'Cunha et al., 1990; Earley et al., 1994]. При этом следует отметить, что применение термостабильных ферментов потребовало использования специальных блоков, поддерживающих необходимую температуру [Ciora et al., 1991; Earley et al., 1994].

Интересным решением проблемы проведения терминирующих реакций одновременно со многими матрицами ДНК, полученными в ходе ПЦР, и содержащими на одной цепи молекулу биотина, явилось приготовление специальной гребенки и соответствующей ей реакционной подставки с лунками [Parik et al., 1993; Lagerkvist et al., 1994]. Особенно-

стью данной гребенки было то, что ее поверхность была покрыта сефарозой, содержащей, в свою очередь, молекулы авидина, что выполняло роль твердой фазы для биотинилированной цепи ДНК при проведении секвенирующих реакций. Еще одно преимущество этой гребенки заключалось в возможности ее использования при одновременном нанесении образцов ДНК на секвенирующий гель, поскольку форма гребенки соответствовала стандартной, используемой при формировании колодцев геля. Схожий принцип одновременного нанесения большого числа образцов ДНК (до 200), удерживаемых на зубцах гребенки ввиду пористости материала, из которого она была изготовлена, был предложен другими авторами [Eggle et al., 1997]. В литературе сообщается об изготовлении специальных пневматических устройств, позволяющих провести быстрое нанесение препаратов ДНК на секвенирующий гель [Panussis et al., 1998].

Подходы к мечению вновь синтезируемых цепей ДНК в условиях терминации при автоматическом секвенировании отличаются от таких в обычном ферментативном секвенировании, пожалуй, лишь типом самой метки, которая представляет собой молекулу какого-либо флуоресцирующего соединения. Но если в случае ручного секвенирования выбор меченых соединений, будь то радионуклиды или какие-либо гаптены для нерадиоактивного секвенирования, довольно ограничен, то для автоматического секвенирования уже синтезировано большое число различных флуоресцентных красителей, а также их производных и этот процесс интенсивно продолжается. К основным требованиям, предъявляемым к подобным соединениям, можно отнести, во-первых, высокий квантовый выход при флуоресценции, обеспечивающий в приборах разного типа детекцию фемто-, атто- или даже пиктомолярных количеств ДНК. Во-вторых, для мечения разных азотистых оснований с их последующей детекцией в одной дорожке секвенирующего геля или в одном капилляре используемые флуорофоры должны иметь неперекрывающиеся или незначительно перекрывающиеся спектры эмиссии. В-третьих, весьма желательно, чтобы флуоресцентные красители, используемые для мечения разных нуклеотидов, не сильно различались по своему влиянию на изменение электрофоретической подвижности меченных ими фрагментов ДНК, поскольку в противном случае будет необходима определенная корректировка получаемых первичных данных и вероятность неверного определения последовательности нуклеотидов при этом несколько возрастает. В этой связи следует отметить, что некоторые из синтезированных флуоресцентных соединений лишь незначительно влияют на электрофоретическую подвижность молекул ДНК или делают это в весьма узком диапазоне. Другой подход выравнивания электрофоретической подвижности фрагментов ДНК, меченных подчас весьма объемными соединениями, заключается в присоединении какой-либо из этих меток путем использования в необходимых случаях специальных линкеров, способных сгладить имеющиеся различия. Так, в литературе описан синтез и использование серии из 4 флуоресцентных красителей, получивших наименование BODIPY (boron dipyrindil) и характеризующихся улучшенными спектральными

характеристиками с меньшей степенью перекрытия между ними [Metzker et al., 1996] по сравнению с используемыми для этих же целей и ставших уже классическими производными флуоресцеина (FAM и JOE) и родамина (TMR и ROX). Другой важной чертой BODIPY красителей, отмечаемой этими авторами, можно считать одинаковую электрофоретическую подвижность фрагментов ДНК равной длины, но несущих разные красители из этой серии, при секвенировании ДНК. В то же время использование различной длины линкеров для присоединения этих красителей к олигонуклеотидным праймерам позволяет изменить их подвижность таким образом, что применяемая компьютерная программа коррекции подвижности для праймеров, меченных стандартными производными флуоресцеина и родамина, может стать пригодной и для BODIPY-олигонуклеотидов. Проведенный анализ подвижности фрагментов ДНК, меченных различными флуоресцентными красителями, при их разделении капиллярным гель-электрофорезом с использованием в качестве матрикса полиэтиленоксида показал, что и в этом случае минимальные отклонения были характерны для BODIPY-олигонуклеотидов [Tan, Yeung, 1997]. Продолжающийся синтез новых более эффективных флуоресцентных красителей приводит к увеличению чувствительности автоматического секвенирования ДНК за счет лучшего квантового выхода этих флуорофоров. Подобное достигается также путем повышения интенсивности флуоресценции за счет передачи энергии от одного красителя к другому, что более подробно рассмотрено ниже.

Несмотря на то, что для автоматического секвенирования ДНК подбираются красители с отличающимися спектрами эмиссии, все же полностью избежать перекрытия не удастся. Один подход к решению данной проблемы заключается в синтезе все новых флуорофоров с улучшенными спектральными характеристиками. Другой же состоит в недавно предложенном способе детекции не спектра эмиссии конкретного соединения, а времени его флуоресценции, измеряемой в наносекундах [Nunnally et al., 1997]. Так, после анализа различных красителей были выбраны наиболее подходящие для этой цели и в модельном эксперименте продемонстрирована их пригодность для автоматического секвенирования ДНК. Авторы отмечают, что такая характеристика флуорофоров, как продолжительность свечения, более дискретна для большинства красителей, нежели их спектры испускания света. Особую актуальность данный подход приобретает в связи с мультиплексным секвенированием ДНК, где необходимо одновременно разделять несколько по-разному меченных образцов [He et al., 1998]. Так, другими авторами за счет детекции времени флуоресценции ряда красителей было осуществлено мультиплексное секвенирование ДНК, показавшее точность свыше 90% для 660 нуклеотидов [Lieberwirth et al., 1998].

Для автоматического секвенирования ДНК, как и для обычного, применяют и внутреннее мечение с помощью несущих флуоресцентную молекулу различных дНТФ, и 5'-концевую метку в виде олигонуклеотидного праймера с флуорофором и 3'-концевую метку, образующуюся за счет терминации синтезируемых цепей ДНК ддНМФ, содер-



жащими какие-либо флуоресцентные красители. Так, внутреннее мечение вновь синтезируемой цепи ДНК с помощью флуоресцентно меченых предшественников в виде дНТФ является относительно дешевой альтернативой использованию праймеров с флуорофором [Ansoerge et al., 1992]. О квазиконцевом мечении вновь синтезируемой цепи ДНК с помощью или дАТФ, или дУТФ, несущих молекулы флуоресцина, а также ограниченного набора остальных дНТФ сообщается в работе Де Беллиса и соавт. [De Bellis et al., 1995]. Другими авторами было обнаружено заметное изменение эффективности мечения ДНК в процессе построения ее новой цепи с использованием аналогичного флуоресцеинового производного дАТФ, связанное с положениями находящихся в исходной матрице ДНК комплементарных ему остатков тимидина [Wiemann et al., 1995a]. Так, максимальная эффективность мечения отмечалась, если первым включаемым вслед за праймером нуклеотидом являлся именно флуоресцеин-15-дАМФ. В других случаях она снижалась очень заметно, причем чем дальше от праймера мог включаться данный меченый нуклеотид, тем с худшей эффективностью это происходило. Таким образом, авторы пришли к выводу о необходимости подбирать праймеры для секвенирования ДНК праймерной “прогулкой” с таким расчетом, чтобы первым включаемым нуклеотидом был тот, который нес флуоресцентную метку [Wiemann et al., 1996]. Основываясь на данном наблюдении, ими был разработан подход, позволяющий осуществить одновременное флуоресцентное мечение обоих концов ДНК разными производными дНТФ [Wiemann et al., 1995b]. Один из используемой пары праймеров подбирался с таким расчетом, что первым включаемым нуклеотидом был флуоресцеин-15-дАМФ, тогда как после второго праймера должен был включаться тexasский красный-5-дЦТФ. Использование автоматического секвенатора с двумя лазерными лучами, возбуждающими тот или иной флуорофор, позволяло осуществлять одновременную детекцию двух типов пиков, принадлежащих разным цепям секвенируемого фрагмента ДНК [Wiemann et al., 1996]. Так, гелий-неоновый лазер возбуждал молекулу тexasского красного, а аргоновый лазер – флуоресцеина.

Другую разновидность внутреннего введения метки можно представить как возможность мечения уже самого праймера по его 3'-концу флуоресцеин-15-дАТФ за счет использования 3' → 5'-экзонуклеазной активности нативной T7 ДНК-полимеразы, способной в отсутствие остальных дНТФ провести обменную реакцию при условии, что концевой нуклеотид праймера будет тот же дАМФ. Добавление различных флуоресцентных красителей к 3'-концу праймера путем его ферментативного удлинения с помощью дУТФ с “пришитыми” флуорофорами показало, что ряд ДНК-полимераз (Кленовский фрагмент ДНК-полимеразы I, обратная транскриптаза и некоторые другие) способны эффективно включать подобные объемные соединения, которые, в свою очередь, не препятствуют дальнейшему синтезу комплементарной цепи ДНК [Александрова и др., 1990]. Упомянутая выше обменная реакция, осуществляемая за счет 3' → 5'-экзонуклеазной активности нативной T7 ДНК-полимеразы, приводила, по существу, к образованию флуорес-

центно меченного по своему 3'-концу олигонуклеотидного праймера. Недавно другими авторами был предложен метод синтеза олигонуклеотидов, несущих различные производные флуоресцеина на своем 3'-конце [Markiewicz et al., 1997]. В этой работе также сообщается, что синтезированные ими праймеры оказались пригодны для проведения секвенирующих реакций и нормально удлинялись секвеназой, что потребовало специальной проверки, поскольку известно, что отдельные флуоресцентные производные дНМФ и ддНМФ различными ДНК полимеразми включаются в синтезируемую цепь ДНК с разной эффективностью и можно было предположить, что и затравочный комплекс с модифицированным нуклеотидом на 3'-конце будет значительно хуже "узнаваться" ДНК-полимеразами. Так, например, известно, что Taq ДНК-полимераза использует нуклеотиды, содержащие молекулы флуоресцеина в качестве субстрата с большей "неохотой", нежели с родамином или другими флуоресцентными группами [Lee et al., 1992]. Отмеченная в другой работе лучшая эффективность включения в комплементарную цепь флуоресцентно меченных дНМФ, несущих объемные группы флуоресцеин изотиоцианата, техасского красного или цианинового красителя Cy5 модифицированной Taq ДНК-полимеразой (delta 280, F667Y), доступной из коммерческого источника под названием Taqenase, позволила авторам сделать вывод, что именно делеция 280 аминокислот этого фермента обеспечивает это свойство [Voss et al., 1997]. Подобного недостатка в виде неэффективного включения модифицированных дНМФ лишено, впрочем, флуоресцентное секвенирование с помощью меченных по 5'-концу праймеров.

О секвенировании ДНК при помощи праймеров, содержащих флуоресцентную метку на 5'-конце, сообщается во многих работах [Smith et al., 1985, 1986; Ansorge et al., 1987; Kaiser et al., 1989]. Этот способ применяется достаточно широко, но одним из сдерживающих моментов является значительная стоимость синтеза подобных праймеров. Так, учитывая необходимость большого числа праймеров при секвенировании ДНК праймерной "прогулкой" и высокую стоимость олигонуклеотидов, содержащих молекулу флуоресцентного красителя, недавно было предложено осуществлять мечение секвенирующих праймеров путем их лигирования со специально синтезированным олигонуклеотидом, несущим подходящий флуорофор [Jang, Steffens, 1997]. Следует отметить, что для этого требовался еще один (третий) олигонуклеотид, содержащий на своем 5'-конце от 1 до 5 вырожденных нуклеотидов и выполняющий роль своеобразного мостика.

Заметное влияние не чувствительность и производительность автоматического секвенирования ДНК оказало создание в 1995 г. специальных несущих флуоресцеин и родамин олигонуклеотидных ET-праймеров (Energy Transfer) с переносом энергии с одного флуорофора на другой [Ju et al., 1995b, 1996b). Следует заметить, что ранее в одной работе был синтезирован универсальный секвенирующий (M13) праймер, содержащий две молекулы флуоресцеинизотиоцианата, у пятого и двенадцатого нуклеотидов, представляющих собой производные урацила и использованных вместо остатков тимидина [Brumbaugh et al., 1988]. Для

ЕТ-праймеров принято сокращенно обозначение типа D-N-A, где D означает молекулу донора, А – акцептора, а N – показывает расстояние между ними в нуклеотидах. Так, например, праймер с 6-карбоксихлорофлуоресцеином (F) на своем 5'-конце в качестве донора и 6-карбоксихлоро-2',7'-диметоксихлорофлуоресцеином (J) модифицированного тимидина в 10-м положении олигонуклеотидной цепочки как акцептор обозначается F10J [Hung et al., 1996b].

Использование ЕТ-праймеров с флуоресцеином и родамином показало 2–14-кратное усиление флуоресценции по сравнению с праймерами, содержащими единичные флуоресцентные группы [Ju et al., 1995a, b]. Проведенный синтез 20 вариантов стандартного универсального M13(–40) праймера с расположенной на его 5'-конце молекулой карбоксихлорофлуоресцеина (FAM или, иначе, просто F), выступающего в качестве донора, и “пришитыми” к по-разному удаленным нуклеотидам этого же (FAM) и другого производного флуоресцеина (JOE или J), и производных родамина (TMR или T и ROX или R) в качестве акцепторных красителей позволил выявить наиболее эффективное их расположение. Так, оказалось, что максимальное увеличение флуоресценции наблюдается в тех случаях, когда акцепторные молекулы применяемых в данном случае флуорофоров располагались через 10 нуклеотидов от донорного красителя, находящегося на 5'-конце олигонуклеотидного праймера. Этими авторами также опробовались несколько иные варианты расположения донорных и акцепторных красителей. Так, вместо F10T и F10R были использованы варианты F3T и F3R соответственно, также приводившие к высокому качеству секвенирования ДНК. В дальнейшем проведенный детальный анализ показал, что максимальное увеличение флуоресценции ЕТ-праймеров проявляется, когда акцепторный краситель отстоит на 7–9 или 10–12 нуклеотидов для разных пар флуорофоров [Glazer, Mathies, 1997]. Целая серия работ этих авторов была посвящена созданию все новых ЕТ-праймеров с различными красителями, которые бы показывали улучшенные спектральные характеристики и незначительные отличия в своем влиянии на электрофоретическую подвижность меченных ими секвенируемых фрагментов ДНК [Hung et al., 1996a, b, 1997, 1998]. Так, использование в качестве донорного флуорофора красителя цианинового ряда, возбуждаемого в широкой области спектра и имеющего несколько меньший квантовый выход, привело к сглаженным различиям в интенсивности сигнала разных акцепторных красителей, что позволило определять до 500 первых нуклеотидов со 100%-ной точностью в 4-цветном секвенировании ДНК [Hung et al., 1996a]. Было проведено исследование спектральных характеристик 56 синтезированных ЕТ-праймеров, отличающихся как донорными и акцепторными группами, так и природой спейсера между ними, который в одних случаях представлял собой цепочку нуклеотидов, а в других был представлен полидезоксирибозофосфатом [Hung et al., 1997]. Авторы остановились на двух наиболее оптимальных наборах, в которых донорами служили красители цианинового ряда, а акцепторами производные родамина.

Таким образом, применение праймеров с усиленной флуоресценцией за счет передачи энергии от одного красителя к другому позволило заметно увеличить чувствительность метода секвенирования ДНК. В случаях ограниченного количества исходной ДНК становилось неизбежным циклическое секвенирование и при этом оказалось достаточно 6 циклов, а для секвенирования всего 10 нг матрицы потребовалось только 15 циклов [Ju et al., 1995a]. Сообщается и об успешном секвенировании вместо обычных 2 мкг всего 0,25 мкг матрицы, для чего с обычными праймерами потребовалось бы уже циклическое секвенирование [Ju et al., 1995b].

С целью снижения затрат на синтез большого числа требуемых ЕТ-праймеров с различающимися последовательностями нуклеотидов, было предложено их мечение с помощью специальной универсальной кассеты, включаемой обычным синтезом на 5'-конец олигонуклеотида [Ju et al., 1996a]. В этой кассете донорный и акцепторный флуорофоры разделены между собой полимерным спейсером, состоящим из остатков 1', 2'-дезоксирибозофосфата. Секвенирование ДНК с такими праймерами показало увеличение сигнала от 2 до 12 раз и позволило "прочитать" с 99%-ной точностью 550 нуклеотидов.

ВODIPY-красители оказались также пригодны для создания праймеров с переносом энергии, образуемой под воздействием соответствующего лазера [Metzker et al., 1996]. При использовании в качестве донора отдельных ВODIPY-красителей акцепторные молекулы других ВODIPY-красителей "пришивались" или к 3-му, или к 6-му нуклеотидам праймера, тогда как дополнительная метильная группа в этих случаях располагалась у 6-го или 3-го нуклеотида соответственно, что в итоге приводило к 180%-му и 360%-му усилению сигнала в том или другом варианте [Metzker et al., 1996]. Также сообщается о синтезе новых флуоресцентных красителей, являющихся производными флуоресцеина и родамина и способных быть донорами и акцепторами в переносе энергии и создании на их основе 4 наборов меченых праймеров для автоматического секвенирования ДНК [Lee et al., 1997].

Автоматическое секвенирование ДНК с помощью флуоресцентно меченных ддНТФ, вызывающих терминирование растущей цепи, было предложено достаточно давно [Prober et al., 1987] и благодаря целому ряду преимуществ данный подход используется весьма активно. Так, во-первых, в этом случае секвенирование ДНК могло осуществляться с использованием в качестве затравки обычных олигонуклеотидных праймеров, не несущих какой-либо другой метки. Во-вторых, имеющее иногда место ложное терминирование цепи ДНК становилось невидимым (поскольку такой фрагмент ДНК не содержал на своем 3'-конце флуоресцентно меченный ддНМФ) и это не вносило ошибки при регистрации нуклеотидной последовательности секвенируемой ДНК. (Попутно необходимо заметить, что прошло около 10 лет, прежде чем ддНТФ, меченные радиоактивным изотопом  $^{33}\text{P}$ , стали использоваться для ручного секвенирования ДНК, о чем уже говорилось в разделе 2.5). Однако из-за неэффективного включения ддНМФ, например Таq ДНК-полимеразой, в заключительном растворе остается довольно

значительное количество таких флуоресцентно меченных ддНТФ, которые интерферируют с разделением специфически терминированных продуктов секвенирующих реакций и ухудшают тем самым разрешающую способность метода. В этой связи желательно провести их удаление, и самым простым способом, хотя и не очень эффективным, является осаждение продуктов терминирующих реакций этанолом. Также был предложен подход к удалению невключившихся ддНТФ с помощью легкого центрифугирования реакционных смесей через сефадекс G-75 в специально подготовленном 96-ячеечном планшете микротитратора [Krakowski et al., 1995]. В другой работе для этой цели был изготовлен специальный блок из органического стекла, вмещавший 24 колонки [Rosenthal, Charnock-Jones, 1992]. При использовании для секвенирования биотинилированных праймеров задача удаления флуоресцентно меченных ддНТФ может быть решена с помощью магнитных частиц M-280 Streptavidin, покрытых стрептавидином, очень просто и быстро, поскольку фирма "DYNAL", являющаяся изготовителем данных частиц, сообщает, что на проведение этой процедуры требуется менее 10 мин [Anonymous, 1998].

Особенностью меченых ддГТФ, ддАТФ, ддЦТФ и ддТТФ, предложенных Пробером и соавт. [Prober et al., 1987; Prober, 1994], было то, что они несли незначительно отличающиеся производные сукцинилфлуоресцеина (SF-505, SF-512, SF-519, SF-526, где цифрами обозначены максимумы эмиссии в нанометрах) с разными спектрами эмиссии, что позволяло детектировать их в одной дорожке секвенирующего геля. Определенным преимуществом данных флуоресцентных соединений было то, что, имея близкие молекулярные массы, они не вызывали серьезных различий в электрофоретической подвижности меченных ими цепей ДНК и, таким образом, не требовалось специальных компьютерных программ для внесения соответствующих поправок, как в случае с некоторыми другими соединениями. Так, использование в автоматическом секвенировании ДНК в качестве флуоресцентных молекул флуоресцеина, тетраметилродамина, NBD (4-хлоро-7-нитробензо-2-окса-1-диазол) и тexasского красного, заметно отличающихся своими размерами, уже требовало определенной коррекции электрофоретической подвижности фрагментов, терминированных различными типами оснований. Другими авторами было проведено детальное сравнение эффективности секвенирования ДНК с помощью указанных выше терминаторов, а также целого ряда синтезированных ими измененных производных этих и некоторых других красителей [Lee et al., 1992]. Приведенные в этой работе результаты показали, что синтезированный набор ддНТФ, меченных различными флуоресцентными красителями (5FAM, 6FAM, DCF-1, DCF-2, EVE, R110-2, BUB1, BUB2, 5ZOE, LOU1, LOU2, NAN2 и др.), приводит к формированию более равномерных пиков флуоресценции при разделении продуктов терминирующих реакций в автоматическом секвенаторе.

Сообщается о синтезе новых, с улучшенными свойствами флуоресцентных терминаторов синтеза цепи ДНК [Rosenblum et al., 1997]. Так, в данной работе было проведено сравнение ддНТФ, несущих производ-

ные родамина, с ддНТФ, мечеными производными дихлорородамина, а также BigDye-терминаторов, представляющих собой комплекс красителей, рассчитанный на передачу энергии от одного красителя к другому. Из 18 синтезированных вариантов BigDye-терминаторов, отличающихся донорными и акцепторными красителями, а также типом линкера, были отобраны наиболее оптимальные варианты, применяемые в дальнейшем исследовании. Так, донорным красителем служило производное флуоресцеина и акцепторным красителем был 5-карбоксихлорородамин. Проведенный анализ результатов секвенирования ДНК с использованием ддНТФ, меченных стандартными производными родамина, дихлорородамина и BigDye-терминаторов показал, что последние характеризовались и более сильным сигналом и повышенной точностью секвенирования. Усредненные показатели для разных матриц ДНК, отличающихся GC-составом, свидетельствуют, что количество “читаемых” с 98%-ной точностью нуклеотидов составляет для ддНТФ с родамином 663 нуклеотида, для ддНТФ, меченных дихлорородамином, – 748 нуклеотидов, и для BigDye-терминаторов – 759 нуклеотидов, но при этом интенсивность флуоресценции у дихлорородамина была значительно ниже, чем у остальных, а максимальной характеризовались, как и следовало ожидать, BigDye-терминаторы. К сходному выводу о предпочтительности использования dRhodamine- и BigDye-терминаторов вместо обычно применявшихся ранее уже привычных терминаторов, несущих различные флуоресцентные группы, пришли и другие авторы [Zakeri et al., 1998]. Преимущества используемых в этих работах терминаторов заключаются в лучшей равномерности пиков и повышенной чувствительности. Сообщается об успешном применении BigDye-терминаторов в широкомасштабном секвенировании геномов микроорганизмов, позволивших “читать” свыше 700 нуклеотидов, что способствовало заполнению имевшихся промежутков на завершающем этапе выполнения этого проекта [Heiner et al., 1998].

Упомянувшиеся выше красители имеют одно общее для всех свойство, а именно, испускание флуоресценции в видимой части спектра. Красители цианинового ряда, флуоресцирующие в области спектра, близкой к инфракрасной, также используются в автоматическом секвенировании ДНК [Shealy et al., 1995; Williams, Soper, 1995; Flanagan et al., 1998; Tu et al., 1998]. К некоторому преимуществу этих красителей можно отнести то, что в этой области спектра отсутствует какая-либо интерференция с биологическими молекулами, что снижает, таким образом, уровень шумов.

### 9.3. РАЗДЕЛЕНИЕ ПРОДУКТОВ ТЕРМИНИРУЮЩИХ РЕАКЦИЙ В АВТОМАТИЧЕСКОМ СЕКВЕНИРОВАНИИ ДНК

Принцип автоматического флуоресцентного секвенирования ДНК заключается в электрофоретическом разделении специфически терминированных продуктов секвенирующих реакций, несущих подходящую

флуоресцентную метку, и их детекции в режиме реального времени в нижней части геля после возбуждения молекул красителя лазерным лучом в момент прохождения фрагментов ДНК, их содержащих, через эту зону. Существуют различные схемы проведения терминирующих реакций и последующего разделения продуктов в секвенирующем геле-электрофорезе. Так, довольно обычным, типичным и для неавтоматического секвенирования, является вариант проведения 4 типов терминирующих реакций с последующим разделением продуктов также в четырех дорожках. Другой вариант (4 реакции/1 дорожка) заключается в проведении также 4 типов реакций с использованием флуоресцентных меток с различными спектрами эмиссии, и поскольку в этом случае их разделение возможно в одной дорожке геля, то это экономит гелевое пространство и, таким образом, сразу в 4 раза повышается производительность данного метода. Более редкими вариантами являются 1 реакция/1 дорожка и 2 реакции/2 дорожки. Эти последние варианты основываются на построении комплементарной цепи ДНК в условиях терминации с помощью секвеназы с использованием какой-то одной флуоресцентной метки, но с различными соотношениями ддНТФ, что позволяет осуществить определение конкретных нуклеотидов по высоте пиков флуоресценции, принадлежащих терминированным этими ддНМФ фрагментам ДНК (Ansorge et al., 1989, 1990; Tabor, Richardson, 1990; Chen et al., 1992).

Первое сообщение об автоматическом секвенировании ДНК с использованием флуоресцентной метки пришло из лаборатории Л. Худа [Smith et al., 1986]. В этой работе после проведения 4 типов реакций (по числу азотистых оснований) с по-разному мечеными праймерами продукты всех реакций объединялись и проводилось их разделение в одной колонке полиакриламидного геля. Таким образом, характер разделения оказался – 4 реакции/1 дорожка. За счет разных длин волн флуоресценции использованных соединений (флуоресцеин, NBD, тетраметилродамин и тexasский красный) и специального вращающегося колеса с 4 фильтрами в виде секторов была определена последовательность нуклеотидов, разделенных в одной дорожке. Созданный в процессе выполнения этой работы прибор был впоследствии усовершенствован и стал поставляться фирмой “Applied Biosystems, Inc.” Несколько отличающийся принцип сбора информации о нуклеотидной последовательности в виде пиков флуоресценции был положен в основу секвенатора ДНК, выпускавшегося в течение ряда лет фирмой “Du Pont” (США). Еще один тип флуоресцентных секвенаторов, разделяющих молекулы ДНК в пластинах полиакриламидного геля и не имеющих подвижных частей, был представлен моделью EMBL (Германия), а также прибором, производимым фирмой “Hitachi” (Япония) и A.L.F. секвенером (“Pharmacia”, Швеция), созданного на основе прибора EMBL. В новом поколении этих приборов (EMBL-2) в дополнении к стандартному аргонному лазеру добавился гелий-неоновый лазер. Высокопроизводительные секвенаторы выпускаются фирмой “Applied Biosystems, Inc.”, ставшей дочерней компанией фирмы “Perkin-Elmer”. Новая модель 377 DNA Sequencer, пришедшая на смену моделям 373 и 370A, с высокой эф-

фективностью позволяет осуществлять одновременное разделение 96 образцов ДНК. Также довольно производительными приборами являются модели 4000LS Long ReadIR и 4000S фирмы “LI-COR” (США), способные разделять до 800 и 550 нуклеотидов соответственно с точностью 99%. Особенностью последних приборов является то, что в отличие от остальных они основаны на детекции молекул ДНК, меченных красителями, флуоресцирующими в области спектра, близкой к инфракрасной.

Перечисленные выше автоматические системы флуоресцентного секвенирования ДНК высокопроизводительны, сложны и весьма дороги. Недавно фирма “Visible Genetics, Inc.” (США) выпустила на рынок новый, менее дорогой прибор OpenGene, не рассчитанный на крупномасштабное секвенирование ДНК, но тем не менее позволяющий осуществить разделение 500 нуклеотидов одного образца за 30 мин с точностью около 98%. Однако всего 16 дорожек в пластине геля, заливаемой и полимеризуемой с помощью специальных устройств всего за 6 мин, не позволяет использовать этот прибор для целей секвенирования полных геномов организмов и поэтому основное предназначение этой системы – это диагностическое секвенирование и анализ полиморфизма генов. В то же время возможность объединения до 4 таких приборов в блок с единым контролем позволяет соответственно увеличить количество секвенируемых матриц.

У российских ученых некоторая надежда на появление отечественного автоматического секвенатора ДНК возникла после объявленного в конце октября 1998 г. тендера по межведомственной научно-технической программе “Научное приборостроение”, где в качестве одного из заданий указана разработка и изготовление прибора для секвенирования<sup>5</sup> и анализа ДНК, причем оговоренный срок исполнения заказа составляет не более двух лет.

Постоянно продолжается работа над усовершенствованием систем для автоматического секвенирования ДНК. Так, недавно предложено использование оптоволоконной техники для подачи возбуждающего луча лазера и сбора сигналов флуоресценции, что позволило исключить применение многочисленных зеркал и линз, требующих весьма частой настройки, причем всей системы сразу [Trost, Guttman, 1998]. В случае же использования оптических волокон настройка каждой части может осуществляться независимо. Более того, в своей статье авторы сообщают, что используемые ими около 6 мес подобные системы не потребовали в этот период какой-либо дополнительной регулировки.

---

<sup>5</sup> Здесь использована орфография, как она была подана в “Перечне заданий” данной Программы. В этой связи следует отметить, что хотя такое произношение и написание этого термина иногда встречается и даже более созвучно с оригиналом, все же в русском языке уже прочно устоялось написание данного слова, как и производных от него, через букву “e”. В качестве дополнительного примера можно привести аналогичное написание другого немного похожего и ставшего недавно печально известным многим нашим согражданам термина “секвестр”, в котором также пишется и произносится буква “e”.



Следует отметить весьма заметное увеличение числа “читаемых” нуклеотидов с одной матрицы ДНК при ее разделении с помощью более совершенных моделей флуоресцентных автоматических секвенаторов. Так, если первые приборы позволяли эффективно разделять лишь 300–400 нуклеотидов [Ansorge et al., 1986], в первой половине 1990-х – 1000 нуклеотидов [Grothues et al., 1993; Zimmermann et al., 1994], то сейчас в руках этих же авторов подобные величины уже не являются пределом. Ими сообщается о секвенировании матрицы в 2000 нуклеотидов в одну реакцию [Voss et al., 1997]. Ранее этой же группой авторов было продемонстрировано эффективное высокоскоростное разделение флуоресцентно меченных фрагментов ДНК с использованием ультратонкого геля (100 мкм), что позволило увеличить градиент напряжения до 80 В на см длины геля и разделить за 1 ч до 1200 нуклеотидов [Stegemann et al., 1991]. Влияние различных факторов на эффективное разделение флуоресцентно меченных фрагментов ДНК было проанализировано в работах японских авторов [Nishikawa, Kambara, 1991, 1992]. Они пришли к выводу, что лучшие разделения фрагментов ДНК отмечаются при оптимальном соотношении градиента напряжения и толщины геля.

Важным аспектом флуоресцентного автоматического секвенирования ДНК является его высокая чувствительность. При осуществлении крупномасштабных проектов с целью уменьшения их стоимости большое значение приобретает экономия различных ингредиентов секвенирующих реакций, да и матрицы ДНК иногда доступны в довольно ограниченных количествах. В этой связи весьма актуальна работа японских авторов, в которой было показано, что всего 1 фмоль матрицы на основе фага M13 и 0,001 единицы ферментативной активности секвеназы 2.0 оказались достаточными для успешного секвенирования приблизительно 400 нуклеотидов клонированного фрагмента ДНК с помощью автоматического флуоресцентного секвенатора [Kawamoto et al., 1994]. Таq ДНК-полимеразе “потребовалось” 0,45 фмоль матрицы и 0,01 единицы фермента, чтобы “прочитать” приблизительно такое же количество нуклеотидов. В другой работе сообщается о секвенировании менее 100 нг двуцепочечной матрицы в автоматическом флуоресцентном секвенаторе, позволившем осуществить “прочтение” более 700 нуклеотидов [Beg, Holt, 1997]. Сообщается также об ином подходе к более экономному автоматическому секвенированию ДНК, заключающемся в повторном (до четырех раз) использовании одной и той же пластины полиакриламидного геля [Swerdlow et al., 1994]. Авторы обнаружили, что гели с формамидом в качестве денатурирующего агента больше подходят для повторного использования по сравнению с содержащими мочевины. Ухудшение качества электрофоретического разделения в таких повторных экспериментах сильно зависело от присутствия значительных количеств остающихся высокомолекулярных матриц ДНК.

Главной целью любого секвенирования ДНК (как ручного, так и автоматического) является определение последовательности нуклеотидов исследуемого фрагмента ДНК. Как это происходит при ручном се-

квенировании, уже было описано в главах 4–6. Что касается флуоресцентного секвенирования, этот процесс здесь также автоматизирован, поскольку автоматические секвенаторы ДНК управляются специально написанными программами. Так, например, приборы фирмы “Applied Biosystems, Inc.” комплектуются двумя программами сбора и анализа данных. После завершения электрофоретического разделения предварительные данные, собранные программой Data Collection, подвергаются анализу специальной программой, нормализующей промежутки между пиками, соответствующими фрагментам ДНК, терминированным тем или иным типом нуклеотидного основания, определяющей относительную высоту таких пиков и ликвидирующей некоторые другие погрешности [Hagemann, Kwan, 1997a, b]. В то же время следует отметить, что в литературе имеются сообщения об аналогичных программах, обеспечивающих даже большую точность в определении конкретных нуклеотидов при сборе и анализе этой информации при разделении фрагментов ДНК в приборах этой же фирмы [Ewing et al., 1998; Ewing, Green, 1998]. Ранее была написана программа Ted (Trace editor), допускающая сбор первичных данных и их анализ с помощью автоматических секвенеров той же фирмы “Applied Biosystems, Inc.”, а также A.L.F. секвенера фирмы “Pharmacia” [Gleeson, Hillier, 1991]. Для автоматических секвенеров этих фирм была написана и другая программа, позволяющая легко обращаться с весьма протяженными последовательностями ДНК при выполнении крупномасштабных проектов [Dear, Staden, 1991]. В другой своей работе этими авторами было уделено серьезное внимание точности определения конкретных нуклеотидов на этапе анализа первичного материала, получаемого с помощью секвенеров 373А и A.L.F. все тех же фирм [Bonfield, Staden, 1995]. Сообщалось, что искажение формы полос разделяемых фрагментов ДНК размером свыше 400 нуклеотидов приводило к нечеткой регистрации соответствующих им пиков и, как следствие, к большему числу ошибочно определяемых нуклеотидов [Sanders et al., 1991]. Разработанная этими авторами компьютерная программа, способная генерировать изображение полос ДНК, позволила повысить точность секвенирования и увеличить длину секвенируемого фрагмента ДНК.

Многие компьютерные программы, рассчитанные на сбор первичных данных при автоматическом секвенировании ДНК, к сожалению, учитывают только информацию о высоте пиков флуоресценции, приходящихся на конкретный нуклеотид. С целью повышения достоверности определяемых этим методом нуклеотидных последовательностей было предложено принимать в расчет не только пики, а также впадины между ними, но и формы кривых, их формирующих [Allex et al., 1996]. Этими авторами и пикам и впадинам, исходя из их характера, были присвоены свои классификации, как сильный(ая), средний(ая) и слабый(ая). Разработанная программа, позволяющая обрабатывать первичный материал в виде пиков флуоресценции секвенируемой ДНК при ее разделении в автоматическом секвенаторе ABI 377, оценивала все эти параметры. Так, было показано, что при “чтении” конкретных нуклеотидов в виде принадлежащих им пиков флуоресценции и с учетом

впадин от других нуклеотидов, совпадающих по месту с этими пиками, заметно повышается точность секвенирования. Другим важным аспектом предложенного разложения записи сигналов флуоресценции секвенируемой ДНК, также способствующего повышению достоверности и производительности данного метода, является возможность более корректного удаления концевых последовательностей перед этапом состыковки различных фрагментов в единый блок. Ряд программ предполагает весьма жесткую установку самим экспериментатором предела чтения, например в 500 нуклеотидов, после которого часто заметно увеличивается процент неверно определяемых нуклеотидов. Другая существующая возможность удаления сомнительных данных предполагает подсчет в специальном окне (составляющем обычно до 40–50 нуклеотидов) неоднозначно определяемых нуклеотидов, и после достижения установленной пороговой величины такие сведения в дальнейший анализ уже не берутся. Однако и во втором, и, особенно, в первом случае есть вероятность как взятия в анализ недостоверных сведений, так и отбрасывания тех, которыми можно было бы воспользоваться. В связи с этим в подобных участках обычно рекомендуется визуальный анализ записи флуоресцентных сигналов, что при большом объеме секвенирования становится малореальным. Разложенные на составляющие кривые флуоресценции позволяют провести более точную дискриминацию данных, пригодных для дальнейшего анализа от неверных. Следует отметить, что все описанные возможности сбора и анализа первичного материала реализованы в модуле Seqman II (более подробно о котором речь пойдет в следующей главе) фирмы “DNASTAR, Inc.” (США), что отличает его от аналогичных программ других фирм [Allex et al., 1997].

Серьезное значение в автоматическом секвенировании ДНК приобретает точность определения нуклеотидной последовательности, поскольку в большинстве случаев экспериментатор не участвует в процессе ее “чтения” и занесения в компьютер. Проводились специальные сравнения числа “ошибок”, производимых различными ДНК-полимеразами. Так, в одной работе секвенировали 94 фрагмента ДНК протяженностью 501 пн с использованием секвеназы и Таq ДНК-полимеразы [Rolf, Weber, 1994]. Авторы сообщают, что при секвенировании отрезка до 300 нуклеотидов уровень ошибок составлял для секвеназы 0,3%, а для Таq ДНК-полимеразы – 1%, однако после 350 нуклеотида количество ошибок резко возрастало для обоих ферментов и варьировало от 6 до 20%. Следует отметить еще одну работу, где была достигнута довольно высокая точность секвенирования, превышающая 99% для нескольких фрагментов ДНК со средней длиной 461 нуклеотид, однако больший интерес к себе она вызывает тем, что в качестве затравочных молекул авторами использовался 31 модульный праймер, составленный каждый из трех различных гексамерных блоков [Lodhi, McCombie, 1996]. С использованием такого же составного (из гексамеров) праймера были секвенированы одноцепочечные матрицы со средней длиной “прочитанного” участка 393 нуклеотида и точностью 99,6% [Hou, Smith, 1994]. Двухцепочечные матрицы характеризовались несколько худшей точностью (99,4%) при средней длине секвенированных фрагментов в 367 пн. Есть

работы, в которых сообщается о секвенировании 700–750 нуклеотидов с точностью более чем 98,5% [Beg, Holt, 1997]. Более высокая точность (99,7%) при секвенировании отрезка в 700 нуклеотидов была достигнута в работе других авторов [Grothues et al., 1993].

Секвенирование ДНК с использованием четырех различных флуоресцентных красителей используется чрезвычайно широко и целая серия статей была посвящена особенностям регистрации их свечения и некоторой коррекции получаемых данных, имеющих целью повысить эффективность и достоверность секвенирования [Yin et al., 1996; Huang et al., 1997, 1998]. В работе других авторов описана многофункциональная компьютерная программа, рассчитанная в том числе на сбор первичной информации при секвенировании ДНК в системе с четырьмя красителями и ее дальнейшую обработку [Wendl et al., 1998].

Некоторую опасность потери информации о секвенируемой матрице ДНК представляет возможный сбой в работе компьютера непосредственно в момент секвенирования. Поскольку при автоматическом флуоресцентном секвенировании занесение предварительных данных в компьютер происходит одновременно с разделением, такой сбой может привести к утере значительной части данных. Для спасения такой информации была разработана специальная программа, восстанавливающая потерянные данные при секвенировании ДНК с помощью секвенаторов моделей 373 и 377 [O'Brien et al., 1997]. Причем авторы сообщают о возможности бесплатного предоставления этой программы, написанной для компьютера Макинтош.

В автоматическом секвенировании ДНК для разделения флуоресцентно меченных продуктов терминирующих реакций, кроме электрофореза в стандартных пластинах полиакриламидного геля, весьма широко используется и капиллярный гель-электрофорез. Причем, для флуоресцентного капиллярного электрофореза характерна как очень высокая чувствительность, так и высокая скорость разделения, являющиеся следствием крайне малого внутреннего диаметра самого капилляра. Так, сравнение разделяемых количеств ДНК в стандартной пластине геля и капилляре путем экстраполяции показывает, что в последнем случае детектируются количества ДНК, которые бы соответствовали аттомолярным, против обычно детектируемых в какой-либо полосе ДНК фемтомолярных [Drossman et al., 1990; Luckey et al., 1990]. К подобному же выводу приходят и другие авторы [Zagursky, McCormick, 1990]. Сообщается об еще более высокой чувствительности этого метода, позволяющего детектировать пептомолярные количества секвенируемых фрагментов ДНК [Swerdlow et al., 1991]. Высокая скорость разделения, присущая капиллярному гель-электрофорезу, является следствием прилагаемого градиента напряжения, достигающего в отдельных случаях до 428 В на 1 см длины капилляра и позволившего разделить 317 нуклеотидов за 31 мин вместо 452 мин, которые требовались для подобного разделения в стандартном геле [Luckey et al., 1990]. Проведенные несколько позже математические расчеты показали, что время разделения продуктов секвенирующих реакций капиллярным электрофорезом может быть даже еще короче, чем обычно затрачиваемое [Aldroubi, Garner, 1992]. Но, как отмеча-

ется в недавнем обзоре Довичи [Dovichì, 1997], сверхвысокая скорость разделения, составившая всего 3,5 мин при градиенте напряжения 1200 В/см, оказалась малоприспособной для геномного секвенирования и может лишь использоваться в экспериментах, связанных с повторным секвенированием (например, в диагностических исследованиях). Более того, количество определяемых нуклеотидов и скорость разделения находятся в обратной зависимости. Таким образом, увеличение производительности метода капиллярного электрофореза было достигнуто не снижением времени разделения, а созданием приборов с многочисленными капиллярами, позволяющими параллельно секвенировать большое число образцов ДНК [Xuàng et al., 1992a]. Хотя следует отметить, что ввиду имеющего место некоторого разброса в скорости разделения фрагментов ДНК от капилляра к капилляру, предпочтительнее использовать подход 4 реакции/1 дорожка [Xuàng et al., 1992b]. Так, сообщается о 4-цветном–1-дорожечном секвенировании в пучке из 25 независимых параллельных капилляров, что обеспечило высокую производительность данной системы [Kheterpal et al., 1996].

Иной подход к увеличению продуктивности метода был продемонстрирован другими авторами [Chen et al., 1992]. Так, проведение терминирующих реакций путем удлинения праймеров, меченных производным флуоресцеина, для ддАТФ и ддЦТФ (взятых в соотношении как 2:1) и меченных производным родамина для ддТТФ и ддГТФ (также взятых в соотношении друг к другу как 2:1) позволило провести их одновременное разделение в одном капилляре и по высоте пиков детектировать конкретный нуклеотид, терминировавший фрагмент ДНК определенной длины и определить, таким образом, последовательность ДНК до 400 нуклеотидов. Дидезокситерминаторы, взятые в реакцию синтеза комплементарной цепи ДНК в соотношении 4:2:1:0 (ddA, ddC, ddG, ddT), позволили разделить терминированные фрагменты в одном капилляре и по высоте пиков флуоресценции, принадлежащим красителю цианинового ряда, “пришитого” к праймеру, определить последовательность около 250 нуклеотидов [Williams, Soper, 1995]. Однако точность секвенирования ДНК при таком подходе составила всего 84%. Более эффективное разделение протяженных фрагментов ДНК было достигнуто с использованием линейных “несшитых” полимеров низкой концентрации в качестве матрикса для капилляров [Ruiz-Martinez et al., 1993], повышением температуры разделения [Kleparnik et al., 1996], применением оптимальных [Lucky, Smith, 1993; Manabe et al., 1994] или ступенчатых электрических полей [Inoue et al., 1998]. Еще один вариант повышения производительности секвенирующего капиллярного электрофореза за счет использования всего 3 флуоресцентных красителей был недавно предложен в качестве способа мультиплексного секвенирования [Kheterpal et al., 1998]. Однако, как отмечают сами авторы, вряд ли предложенная схема сможет заменить стандартное 4-цветное секвенирование ДНК.

Гелевым матриксом в ранних работах по секвенирующему капиллярному электрофорезу служил обычный полиакриламидный гель, однако связанные с ним проблемы в виде его нестабильности, формирова-

ния пузырьков воздуха, видимых при микроскопическом исследовании капилляров, заметно снижали производительность метода [Swerdlow, Gesteland, 1990; Karger et al., 1991; Swerdlow et al., 1992]. Введенный в употребление линейный полиакриламид оказал заметное влияние на последующее развитие этого метода [Ruiz-Martinez et al., 1993]. Кроме уже широко распространенного линейного полиакриламида [Best et al., 1994; Zhang et al., 1995; Wu et al., 1996], в литературе встречаются упоминания об использовании в качестве матрикса мицелл полиэтиленгликоля с фторуглеродными цепочками [Menchen et al., 1996], полиэтиленоксида [Kim, Yeung, 1997], полиметилакриламида [Madabhushi, 1998] и некоторых других [Quesada, 1997; Beale, 1998]. Новый способ полимеризации линейного 2%-ного полиакриламидного геля и его использование в качестве матрикса при заполнении капилляра позволило за 80 мин разделять более 1000 нуклеотидов с точностью выше 97% и 99%-ной точностью для первых 800 нуклеотидов [Goetzing et al., 1998]. Считается, что немаловажную роль в ухудшении разделения играют даже незначительные загрязнения матрицы ДНК, как, впрочем, и сама матрица. Так, удаление ненужной уже матрицы с помощью ультрафильтрации через полиэфирсульфоновую мембрану и снижение концентрации соли до 10 мкМ за счет спин-диализа позволили значительно увеличить эффективность разделения фрагментов ДНК (до 1000 нуклеотидов с точностью 99% для первых 800 нуклеотидов) и продлить срок службы капилляра [Ruiz-Martinez et al., 1998; Salas-Solano et al., 1998].

Особым случаем капиллярного электрофореза можно считать подход, использующий микрокапилляры, изготовленные на плоском стекле с помощью фотолитографии и химического травления и закрытые сверху другим стеклом со специальными микронными отверстиями, проделанными с помощью лазера [Woolley, Mathies, 1994]. После заполнения полученных капилляров линейным полиакриламидным или каким-то другим гелем такие капиллярные “чипы” готовы к электрофоретическому разделению ДНК. В цитируемой выше работе для эффективного разделения рестриктазных фрагментов ДНК с размерами от 70 до 1000 пн на таком чипе с длиной миграции только 3,5 см потребовалось всего 120 с. Следует отметить, что для детекции флуоресценции ДНК в этой работе, кроме всего прочего, потребовался еще и объектив с 40-кратным увеличением. В своей следующей работе эти авторы уже осуществили секвенирование ДНК в таком микрокапиллярном чипе, причем для разделения 433 нуклеотидов в системе с одним красителем (4 реакции/4 дорожки) оказалось достаточно 10 мин [Woolley, Mathies, 1995]. Используя четыре разных красителя (4 реакции/1 дорожка), с 97%-ной точностью за 540 с было секвенировано приблизительно 150 нуклеотидов. В работе других авторов с помощью микрокапиллярного чипа с длиной канала 11,5 см была определена последовательность 400 нуклеотидов за 14 мин [Schmalzing et al., 1998]. Подобные системы секвенирования ДНК в микрокапиллярных электрофоретических чипах имеют целый ряд преимуществ перед другими способами разделения фрагментов ДНК ввиду их чрезвычайно малого размера, сверхвысокой чувствительности, требующей ничтожные количества

ДНК в нанолитровых объемах, высокой скорости разделения, превышающей таковую для стандартного капиллярного электрофореза в 10 раз и для электрофореза в пластинах геля в 100 раз. В связи с подобной миниатюризацией процесса следует отметить, что недавно американской фирмой "Packard Instruments" разработана PiezoTip система, позволяющая дозировать жидкости объемом от 0,2 нл, пригодная наряду с дозированием жидкости для различных анализов в микропланшетах с 384, 864 и даже с 1536 лунками, также и для подобных целей.

Интересный способ "чтения" последовательности нуклеотидов после завершения капиллярного электрофореза был предложен в статье Кима и соавт. [Kim et al., 1996]. Так, проведенное ими последовательное сканирование капилляра содержащимися в нем флуоресцентно мечеными фрагментами ДНК с помощью специального сканера со скоростью около 8 см/с при разных длинах волн возбуждения лазером позволило детектировать до 200 нуклеотидов плазмидной ДНК.

Высокая чувствительность, высокая скорость разделения продуктов секвенирующих реакций методом капиллярного электрофореза, возможность многократного использования капилляров привели к созданию высокопроизводительных, полностью автоматизированных приборов, таких как например, ABI Prism 310 ("Perkin-Elmer", США). Фирмой "Molecular Dynamics" (США), являющейся дочерней компанией концерна "Amersham Pharmacia Biotech", создана система для секвенирования ДНК MegaBACE®1000, позволяющая проводить параллельное электрофоретическое разделение продуктов терминирующих реакций в 96 капиллярах. Данный прибор осуществляет автоматическое внесение препаратов ДНК, регистрацию данных, способен автоматически заменять гелевый матрикс в капиллярах. Высокая скорость разделения (весь цикл от момента внесения препаратов до следующего внесения занимает всего 2 ч) позволяет секвенировать более 500 тпн в течение суток с гарантированным определением более 500 нуклеотидов на один образец ДНК. Капилляры, объединенные в блоки по 16 штук, позволяют проводить легкую смену гелевого матрикса и рассчитаны на их использование в течение, по крайней мере, 100–200 раз. В сентябре 1998 г. было объявлено, что "Molecular Dynamics" совместно с фирмой "Beckman Coulter" (США), также выпускающей высокопроизводительный секвенер SEQ™2000, и известной своими лабораторными работами "Biomek", будет разработан полностью автоматизированный комплекс, который объединит 4 прибора MegaBACE®1000 и увеличит, таким образом, количество секвенируемых последовательностей нуклеотидов до 2 мпн в сутки, причем без участия оператора.

#### 9.4. АВТОМАТИЗАЦИЯ СЕКВИРОВАНИЯ ДНК МЕТОДОМ ХИМИЧЕСКОЙ ДЕГРАДАЦИИ ПО МАКСАМУ–ГИЛБЕРТУ

Все описанное выше в этой главе относилось к автоматическому секвенированию ДНК ферментативным методом по Сэнгеру. Что касается секвенирования путем химической деградации по Максому–Гил-

берту, то этот метод менее пригоден для его автоматизации. Одной из причин этого являются особенности проведения реакций химической модификации оснований и пиперидинового гидролиза, требующих довольно много стадий при их выполнении. Так, была разработана практически единственная модель робота, производимого фирмой "Seiko Instruments" (Япония), для автоматизации проведения химических реакций при секвенировании ДНК по Максаму–Гилберту. Данный микрохимический робот был способен осуществлять все операции по количественному добавлению необходимых реагентов (в диапазоне от 1 мкл до 1 мл), их перемешиванию, центрифугированию, высушиванию, поддержанию заданных температур во время реакций на основе введенной в него программы без непосредственного участия персонала на всех стадиях [Wada, 1984; Wada et al., 1984]. Подобную автоматизацию, впрочем, было трудно считать высокопроизводительной, поскольку на подготовку к секвенирующему электрофорезу 12 образцов уходило около 12 ч [Wada, 1987].

Исключив из схемы проведения реакций химической дегградации этапы центрифугирования, другая группа исследователей смогла автоматизировать этот процесс с помощью ставшего уже к этому времени стандартным лабораторного биоробота Biomek 1000 [Boland et al., 1994]. Вместо трудоемких повторяющихся этапов центрифугирования ими было предложено проводить все химические реакции этого метода секвенирования с препаратами ДНК, сорбированными на стеклянном порошке с завершающей эволюцией уже готовой для электрофоретического разделения ДНК с помощью высокой концентрации NaJ. Определенной автоматизации процесса химической дегградации удалось добиться путем твердофазного секвенирования с использованием в качестве носителя для модифицируемой ДНК CCS-бумаги [Hybond M&G], обладающей высокой механической прочностью [Rosenthal et al., 1985; Ansorge et al., 1988; Voss et al., 1989]. Сообщается также об использовании в качестве твердой фазы магнитных частиц при секвенировании флуоресцентно меченных фрагментов ДНК, что позволило упростить процедуры проведения химических реакций [Ohara et al., 1997].

Об использовании автоматического секвенатора ДНК при разделении флуоресцентно меченных фрагментов ДНК, подвергнутых химической модификации и гидролизу, говорится в работе Розенталя и соавт. [Rosenthal et al., 1990]. Так, в системе 4 реакции/4 дорожки, где образцы ДНК готовились с помощью стандартных реакций модификации по Максаму–Гилберту, было достигнуто разрешение, позволяющее осуществить разделение 250–300 нуклеотидов. Электрофоретическое разделение образцов, подготовленных с помощью гидролиза пиперидином в присутствии NaCl (предварительно обработанных диметисульфатом) с характером расщепления  $G > A > C > T$ , в системе 1 реакция/1 дорожка позволило секвенировать 100–200 нуклеотидов на образце.



## Цитируемая литература

- Александрова Л.А., Лукин М.А., Розовская Т.А., Атражев А.М., Куханова М.К., Краевский А.А.* Флуоресцентные аналоги нуклеозид-5'-трифосфатов для изучения нуклеиновых кислот нерадиоактивными методами // Молекуляр. биология. 1990. Т. 24. С. 1100–1108.
- Alderton R.P., Eccleston L.M., Howe R.P., Read C.A., Reeve M.A., Beck S.* Magnetic bead purification of M13 DNA sequencing templates // *Anal. Biochem.* 1992. Vol. 201. P. 166–169.
- Aldroubi A., Garner M.M.* Minimal electrophoresis time for DNA sequencing // *Biotechniques.* 1992. Vol. 13. P. 620–624.
- Allex C.E., Baldwin S.F., Shavlik J.W., Blattner F.R.* Improving the quality of automatic DNA sequence assembly using fluorescent trace-data classifications // *Proc. Forth Intern. conf. intelligent systems mol. biol.* St. Louis: AAAI press, 1996. P. 3–14.
- Allex C.E., Baldwin S.F., Shavlik J.W., Blattner F.R.* Increasing consensus accuracy in DNA fragment assemblies by incorporating fluorescent trace representations // *Proc. Fifth Intern. conf. intelligent systems mol. biol.* Haikidiki (Greece): AAAI press, 1997. P. 3–14.
- Andersson B., Lu J., Edwards K.E., Muzny D.M., Gibbs R.A.* Method for 96-well M13 DNA template preparations for large-scale sequencing // *Biotechniques.* 1996. Vol. 20. P. 1022–1027.
- Anon.* Dynapure™ dye terminator removal // *DYNALogue.* 1998. N 1. P. 3–4.
- Ansorge W., Rosenthal A., Sproat B.S., Schwager C., Stegemann J., Voss H.* Non-radioactive automated sequencing of oligonucleotides by chemical degradation // *Nucl. Acids Res.* 1988. Vol. 16. P. 2203–2206.
- Ansorge W., Sproat B.S., Stegemann J., Schwager C.* A non-radioactive automated method for DNA sequence determination // *J. Biochem. and Biophys. Meth.* 1986. Vol. 13. P. 315–323.
- Ansorge W., Sproat B.S., Stegemann J., Schwager C., Zenke M.* Automated DNA sequencing: Ultrasensitive detection of fluorescent bands during electrophoresis // *Nucl. Acids Res.* 1987. Vol. 15. P. 4593–4602.
- Ansorge W., Voss H., Wiemann S., Schwager C., Sproat B.S., Zimmermann J., Stegemann J., Erfle H., Hewitt N., Rupp T.* High-throughput automated DNA sequencing facility with fluorescent labels at the European Molecular Biology Laboratory // *Electrophoresis.* 1992. Vol. 13. P. 616–619.
- Ansorge W., Voss H., Wirkner U., Schwager C., Stegemann J., Pepperkok R., Zimmermann J., Erfle H.* Automated Sanger DNA sequencing with one label in less than four lanes on gel // *J. Biochem. and Biophys. Meth.* 1989. Vol. 20. P. 47–52.
- Ansorge W., Voss H., Zimmermann J.* (ed.). DNA sequencing strategies. Wiley; Spektrum, 1996. 202 p.
- Ansorge W., Zimmermann J., Schwager C., Stegemann J., Erfle H., Voss H.* One label, one tube, Sanger DNA sequencing in one and two lanes on a gel // *Nucl. Acids Res.* 1990. Vol. 18. P. 3419–3420.
- Beale S.C.* Capillary electrophoresis // *Anal. Chem.* 1998. Vol. 70. P. 279R–300R.
- Beg O.U., Holt R.G.* A cost-effective plasmid purification protocol suitable for fluorescent automated DNA sequencing // *Mol. Biotechnol.* 1998. Vol. 9. P. 79–83.
- Beg O.U., Holt R.G.* An efficient cost-effective protocol for automated fluorescent-DNA sequencing // *Biotechnol. Appl. Biochem.* 1997. Vol. 26. P. 27–30.
- Best N., Arriaga E., Chen D.Y., Dovichi N.J.* Separation of fragments up to 570 bases in length by use of 6% T non-cross-linked polyacrylamide for DNA sequencing in capillary electrophoresis // *Anal. Chem.* 1994. Vol. 66. P. 4063–4067.
- Bock J.H., Slightom J.L.* Fluorescence-based cycle sequencing with primers selected from a nonamer library // *Biotechniques.* 1995. Vol. 19. P. 60–64.
- Boland E.J., Pillai A., Odom M.W., Jagadeeswaran P.* Automation of the Maxam–Gilbert chemical sequencing reactions // *Ibid.* 1994. Vol. 16. P. 1088–1095.
- Bonfield J.K., Staden R.* The application of numerical estimates of base calling accuracy to DNA sequencing projects // *Nucl. Acids Res.* 1995. Vol. 23. P. 1406–1410.
- Brumbaugh J.A., Middendorf L.R., Grone D.L., Ruth J.L.* Continuous, on-line DNA sequencing using oligonucleotide primers with multiple fluorophores // *Proc. Nat. Acad. Sci. USA.* 1988. Vol. 85. P. 5610–5614.
- Chen D., Harke H.R., Dovichi N.J.* Two-label peak-height encoded DNA sequencing by capillary gel electrophoresis: Three examples // *Nucl. Acids Res.* 1992. Vol. 20. P. 4873–4880.

Chen F.T., Tusak A., Pentoney S., Jr., Konrad K., Lew C., Koh E., Sternberg J. Semiconductor laser-induced fluorescence detection in capillary electrophoresis using a cyanine dye // *J. Chromatogr. A*. 1993. Vol. 652. P. 355–360.

Chen Q., Neville C., MacKenzie A., Korneluk R.G. Automated DNA sequencing requiring no DNA template purification // *Biotechniques*. 1996. Vol. 21. P. 453–457.

Ciora T., Deneffe P., Mayaux J.-P. Rapid one-step automated sequencing reactions for 16 DNA samples using Taq polymerase and fluorescent primers // *Nucl. Acids Res.* 1991. Vol. 19. P. 188.

D' Cunha J., Berson B.J., Brumley R.L., Jr., Wagner P.R., Smith L.M. An automated instrument for the performance of enzymatic DNA sequencing reactions // *Biotechniques*. 1990. Vol. 9. P. 80–90.

Dear S., Staden R. A sequence assembly and editing program for efficient management of large projects // *Nucl. Acids Res.* 1991. Vol. 19. P. 3907–3911.

De Bellis G., Manoni M., Pergolizzi R., Vezzoni P.M., Luzzana M. Primer design in fluorescent DNA sequencing // *Ibid.* 1990. Vol. 18. P. 4951–4952.

De Bellis G., Manoni M., Pergolizzi R., Redolfi M.E., Luzzana M. A more stringent choice of primers can improve the performance of fluorescent automated DNA sequencers // *Biotechniques*. 1992. Vol. 13. P. 892–894.

De Bellis G., Pergolizzi R., Debernardi S., Invernizzi L., Luzzana M. Fluorescence-based automated DNA sequencing by limited primer labeling // *Ibid.* 1995. Vol. 19. P. 66–70.

Deggerdal A., Holmberg A., Bosnes M., Larsen F. Automated DNA isolation method // *DYNALogue*. 1998. N 1. P. 5–7.

Dovichi N.J. DNA sequencing by capillary electrophoresis // *Electrophoresis*. 1997. Vol. 18. P. 2393–2399.

Drossman H., Luckey J.A., Kostichka A.J., D' Cunha J., Smith L.M. High-speed separations of DNA sequencing reactions by capillary electrophoresis // *Anal. Chem.* 1990. Vol. 62. P. 900–903.

Earley J.J., Kuivaniemi H., Prockop D.J., Tromp G. Efficient DNA sequencing on microtiter plates using dried reagents and Bst DNA polymerase // *DNA Sequenc.* 1993. Vol. 4. P. 79–85.

Earley J.J., Kuivaniemi H., Prockop D.J., Tromp G. Robotic automation of dideoxyribonucleotide sequencing reactions // *Biotechniques*. 1994. Vol. 17. P. 156–165.

Eperon I.C. Rapid preparation of bacteriophage DNA for sequence analysis in sets of 96 clones, using filtration // *Anal. Biochem.* 1986. Vol. 156. P. 406–412.

Eperon L.P., Graham I.R., Griffiths A.D., Eperon I.C. Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is a folding limited to a region behind the transcribing RNA polymerase? // *Cell*. 1988. Vol. 54. P. 393–401.

Erfle H., Ventzki R., Voss H., Rechmann S., Stegemann J., Ansorge W. Simultaneous loading of 200 sample lanes for DNA sequencing on vertical and horizontal, standard and ultrathin gels // *Nucl. Acids Res.* 1997. Vol. 25. P. 2229–2230.

Ewing B., Green P. Base-calling of automated sequencer traces using *Phred*. 2. Error probabilities // *Genome Res.* 1998. Vol. 8. P. 186–194.

Ewing B., Hiller L., Wendl M.C., Green P. Base-calling of automated sequencer traces using *Phred*. 1. Accuracy assessment // *Ibid.* 1998. Vol. 8. P. 175–185.

Flanagan J.H., Jr., Owens C.V., Romero S.E., Waddell E., Kahn S.H., Hammer R.P., Soper S.A. Near-infrared heavy-atom-modified fluorescent dyes for base-calling in DNA-sequencing applications using temporal discrimination // *Anal. Chem.* 1998. Vol. 70. P. 2676–2684.

Fry G., Lachenmeier E., Mayrand E., Giusti B., Fisher J., Johnston-Dow L., Cathcart R., Finne E., Kilaas L. A new approach to template purification for sequencing application using paramagnetic particles // *Biotechniques*. 1992. Vol. 13. P. 124–131.

Gleeson T., Hillier L. A trace display and editing program for data from fluorescence based sequencing machines // *Nucl. Acids Res.* 1991. Vol. 19. P. 6481–6483.

Goetzinger W., Kotler L., Carrilho E., Ruiz-Martinez M.C., Salas-Solano O., Karger B.L. Characterization of high molecular mass linear polyacrylamide powder prepared by emulsion polymerization as a replaceable polymer matrix for DNA sequencing by capillary electrophoresis // *Electrophoresis*. 1998. Vol. 19. P. 242–248.

Grothues D., Voss H., Stegemann J., Wiemann S., Sensen C., Zimmermann J., Schwager C., Erfle H., Rupp T., Ansorge W. Separation of up to 1000 bases on a modified A.L.F. DNA sequencer // *Nucl. Acids Res.* 1993. Vol. 21. P. 6042–6044.

Hagemann T.L., Kwan S.-P. ABI analysis: Manipulation of sequence data from the ABI sequencer // *Meth. Mol. Biol.* 1997a. Vol. 70: Sequence data analysis guidebook. P. 39–54.

- Hagemann T.L., Kwan S.-P.* SeqEd: Manipulation of sequence data and chromatograms from the ABI sequencer analysis files // *Ibid.* 1997b. Vol. 70: Sequence data analysis guidebook. P. 55–63.
- He H., Nunnally B.K., Li L.-C., McGown L.B.* On-the-fly fluorescence lifetime detection of dye-labeled DNA primers for multiplex analysis // *Anal. Chem.* 1998. Vol. 70. P. 3413–3418.
- Heiner C.R., Hunkapilleer K.L., Chen S.M., Glass J.I., Chen E.Y.* Sequencing miltimegabase-template DNA with BigDye terminator chemistry // *Genome Res.* 1998. Vol. 8. P. 557–561.
- Hong G.F.* A method for sequencing single-stranded cloned DNA in both directions // *Biosci. Rep.* 1981. Vol. 1. P. 243–252.
- Hou W., Smith L.M.* Fluorescence-based DNA sequencing with hexamer primers // *Nucl. Acids Res.* 1993. Vol. 21. P. 3331–3332; Errata // *Nucl. Acids Res.* 1993. Vol. 21. P. 4158.
- Hou W., Smith L.M.* DNA sequencing with a hexamer string primer and dye-labeled terminators // *Anal. Biochem.* 1994. Vol. 221. P. 136–141.
- Huang X.C., Quesada M.A., Mathies R.A.* Capillary array electrophoresis using laser excited confocal fluorescence detection // *Anal. Chem.* 1992a. Vol. 64. P. 967–972.
- Huang X.C., Quesada M.A., Mathies R.A.* DNA sequencing using capillary array electrophoresis // *Anal. Chem.* 1992b. V. 64. P. 2149–2154.
- Huang W., Fuhrmann D.R., Politte D.G., Thomas L.J., Jr., States D.J.* Filter matrix estimation in automated DNA sequencing // *IEEE Trans. Biomed. Eng.* 1998. Vol. 45. P. 422–428.
- Huang W., Yin Z., Fuhrmann D.R., States D.J., Thomas L.J., Jr.* A method to determine the filter matrix in four-dye fluorescence-based DNA sequencing // *Electrophoresis.* 1997. Vol. 18. P. 23–25.
- Hung S.-C., Ju J., Mathies R.A., Glazer A.N.* Cyanine dyes with high absorption cross section as donor chromophores in energy transfer primers // *Anal. Biochem.* 1996a. Vol. 238. P. 165–170.
- Hung S.-C., Ju J., Mathies R.A., Glazer A.N.* Energy transfer primers with 5- or 6-carboxyrhodamine-6G as acceptor chromophores // *Ibid.* 1996b. Vol. 238. P. 165–170.
- Hung S.-C., Mathies R.A., Glazer A.N.* Optimization of spectroscopic and electrophoretic properties of energy transfer primers // *Ibid.* 1997. Vol. 252. P. 78–88.
- Hung S.-C., Mathies R.A., Glazer A.N.* Comparison of fluorescence energy transfer primers with different donor-acceptor dye combinations // *Ibid.* 1998. Vol. 255. P. 32–38.
- Inoue H., Tsunako M., Baba Y.* Enhanced separation of DNA sequencing products by capillary electrophoresis using a stepwise gradient of electric field strength // *J. Chromatogr. A.* 1998. Vol. 802. P. 179–184.
- Itoh M., Carninci P., Nagaoka S., Sasaki N., Okazaki Y., Ohsumi T., Muramatsu M., Hayashizaki Y.* Simple and rapid preparation of plasmid template by a filtration method using microtiter filter plates // *Nucl. Acids Res.* 1997. Vol. 25. P. 1315–1316.
- Jang G.Y., Steffens D.L.* Ligation mediated fluorescent labeling of DNA sequencing primers // *Ibid.* 1997. Vol. 25. P. 922–923.
- Johnson A.F., Lodhi M.A., McCombie W.R.* Fluorescence-based sequencing of double-stranded DNA by hexamer string priming // *Anal. Biochem.* 1996. Vol. 241. P. 228–237.
- Ju J., Glazer A.N., Mathies R.A.* Cassette labeling for facile construction of energy transfer fluorescent primers // *Nucl. Acids Res.* 1996a. Vol. 24. P. 1144–1148.
- Ju J., Glazer A.N., Mathies R.A.* Energy transfer primers: A new labeling paradigm for DNA sequencing and analysis // *Nat. Med.* 1996b. Vol. 2. P. 246–249.
- Ju J., Kheterpal I., Scherer J.R., Ruan C., Fuller C.W., Glazer A.N., Mathies R.A.* Design and synthesis of fluorescence energy transfer dye-labeled primers and their application for DNA sequencing and analysis // *Anal. Biochem.* 1995a. Vol. 231. P. 131–140.
- Ju J., Ruan C., Fuller C.W., Glazer A.N., Mathies R.A.* Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis // *Proc. Nat. Acad. Sci. USA.* 1995b. Vol. 92. P. 4347–4351.
- Kaczorowski T., Szybalski W.* Automated four-color DNA sequencing using primers assembled by hexamer ligation // *Gene.* 1996. Vol. 179. P. 195–198.
- Kaiser R.J., MacKellar S.L., Vinayak R.S., Sanders J.Z., Saavedra R.A., Hood L.E.* Specific-primer-directed DNA sequencing using automated fluorescence detection // *Nucl. Acids Res.* 1989. Vol. 17. P. 6087–6102.
- Karger A.E., Harris J.M., Gesteland R.F.* Multiwavelength fluorescence detection for DNA sequencing using capillary electrophoresis // *Ibid.* 1991. Vol. 19. P. 4955–4962.
- Kawamoto K., Okano K., Kambara H.* DNA sequencing using ultra small amounts of reagent and template // *DNA Res.* 1994. Vol. 1. P. 297–301.

*Kheterpal I., Scherer J.R., Clark S.M., Radhakrishnan A., Ju, J., Ginther C.L., Sensabaugh G.F., Mathies R.A.* DNA sequencing using a four-color confocal fluorescence capillary array scanner // *Electrophoresis*. 1996. Vol. 17. P. 1852–1859.

*Kheterpal I., Speed T.P., Mathies R.A.* A three-wavelength labeling approach for DNA sequencing using energy transfer primers and capillary electrophoresis // *Electrophoresis*. 1998. Vol. 19. P. 1403–1414.

*Kim S., Yoo H.J., Hahn J.H.* Postelectrophoresis capillary scanning method for DNA sequencing // *Anal. Chem.* 1996. Vol. 68. P. 936–939.

*Kim Y., Yeung E.S.* Separation of DNA sequencing fragments up to 1000 bases using poly(ethyleneoxide)-filled capillary electrophoresis // *J. Chromatogr. A*. 1997. Vol. 781. P. 315–325.

*Kleparnik K., Foret F., Berka J., Goetzinger W., Miller A.W., Karger B.L.* The use of elevated column temperature to extend DNA sequencing read length in capillary electrophoresis with replaceable polymer matrices // *Electrophoresis*. 1996. Vol. 17. P. 1860–1866.

*Kolner D.E., Guilfoyle R.A., Smith L.M.* A high throughput system for the preparation of single stranded templates grown in microculture // *DNA Sequenc.* 1994. Vol. 4. P. 253–257.

*Krakowski K., Bunville J., Seto J., Baskin D., Seto D.* Rapid purification of fluorescent dye-labeled products in a 96-well format for high-throughput automated DNA sequencing // *Nucl. Acids Res.* 1995. Vol. 23. P. 4930–4931.

*Kristensen T., Voss H., Schwager C., Stegemann J., Sproat B., Ansorge W.* T7 DNA polymerase in automated dideoxy sequencing // *Ibid.* 1988. Vol. 16. P. 3487–3496.

*Lagerkvist A., Stewart J., Lagerstrom-Fermer M., Landegren U.* Manifold sequencing: Efficient processing of large sets of sequencing reactions // *Proc. Nat. Acad. Sci. USA*. 1994. Vol. 91. P. 2245–2249.

*Lario A., Gonzplez A., Dorado G.* Automated laser-induced fluorescence DNA sequencing: Equalizing signal-to-noise ratios significantly enhances overall performance // *Anal. Biochem.* 1997. Vol. 247. P. 30–33.

*Lee L.G., Connell C.R., Woo S.L., Cheng R.D., McArdle B.F., Fuller C.W., Halloran N.D., Wilson R.K.* DNA sequencing with dye-labeled terminators and T7 DNA polymerase: Effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments // *Nucl. Acids Res.* 1992. Vol. 20. P. 2471–2483.

*Lee L.G., Spurgeon S.L., Heiner C.R., Benson S.C., Rosenblum B.B., Menchen S.M., Graham R.J., Constantinescu A., Upadhyya K.G., Cassel J.M.* New energy transfer dyes for DNA sequencing // *Ibid.* 1997. Vol. 25. P. 2816–2822.

*Lieberwirth U., Arden-Jacob J., Drexhage K.H., Herten D.P., Muller R., Neumann M., Schulz A., Siebert S., Sagner G., Klingel S., Sauer M., Wolfrum J.* Multiplex dye DNA sequencing in capillary gel electrophoresis by diode laser-based time-resolved fluorescence detection // *Anal. Chem.* 1998. Vol. 70. P. 4771–4779.

*Lodhi M.A., McCombie W.R.* High-quality automated DNA sequencing primed with hexamer strings // *Genome Res.* 1996. Vol. 6. P. 10–18.

*Luckey J.A., Drossman H., Kostichka A.J., Mead D.A., D’Cunha J., Norris T.B., Smith L.M.* High speed DNA sequencing by capillary electrophoresis // *Nucl Acids Res.* 1990. Vol. 18. P. 4417–4421.

*Luckey J.A., Smith L.M.* Optimization of electric field strength for DNA sequencing in capillary gel electrophoresis // *Anal. Chem.* 1993. Vol. 65. P. 2841–2850.

*Madabhushi R.S.* Separation of 4-color DNA sequencing extension products in noncovalently coated capillaries using low viscosity polymer solutions // *Electrophoresis*. 1998. Vol. 19. P. 224–230.

*Manabe T., Chen N., Terabe S., Yohda M., Endo I.* Effects of linear polyacrylamide concentrations and applied voltages on the separation of oligonucleotides and DNA sequencing fragments by capillary electrophoresis // *Anal. Chem.* 1994. Vol. 66. P. 4243–4252.

*Mardis E.R., Roe B.A.* Automated methods for single-stranded DNA isolation and dideoxyribonucleotide DNA sequencing reactions on a robotic workstation // *Biotechniques*. 1989. Vol. 7. P. 840–850.

*Markiewicz W., Groger G., Rosch R., Zebrowska A., Markiewicz M., Klotz M., Hinz M., Godzina P., Seliger H.* A new method of synthesis of fluorescently labelled oligonucleotides and their application in DNA sequencing // *Nucl. Acids Res.* 1997. Vol. 25. P. 3672–3680.

- Menchen S., Johnson B., Winnik M.A., Xu B.* Flowable networks as DNA sequencing media in capillary columns // *Electrophoresis*. 1996. Vol. 17. P. 1451–1459.
- Metzker M.L., Lu J., Gibbs R.A.* Electrophoretically uniform fluorescent dyes for automated DNA sequencing // *Science*. 1996. Vol. 271. P. 1420–1422.
- Mita K., Morimyo M., Hongo E.* One step DNA sequencing of single-stranded DNA with reverse dye-primer // *Nucl. Acids Res.* 1994. Vol. 22. P. 1507–1508.
- Nishikawa T., Kambara H.* Analysis of limiting factors of DNA band separation by a DNA sequencer using a fluorescence detection // *Electrophoresis*. 1991. Vol. 12. P. 623–631.
- Nishikawa T., Kambara H.* High resolution-separation of DNA bands by electrophoresis with a long gel in a fluorescence-detection DNA sequencer // *Ibid.* 1992. Vol. 13. P. 495–499.
- Nunnally B.K., He H., Li L.-C., Tucker S.A., McGown L.B.* Characterization of visible dyes for four-decay fluorescence detection in DNA sequencing // *Anal. Chem.* 1997. Vol. 69. P. 2392–2397.
- O'Brien K.M., Fondon J.W., III, Evans G.A., Garner H.R.* Rescuing corrupted files from model 377 and 373 DNA sequencers // *Biotechniques*. 1997. Vol. 22. P. 1162–1163.
- Ohara R., Tanaka A., Ohara O.* Automated fluorescent DNA sequencing by a simplified solid-phase chemical sequencing method // *Ibid.* 1997. Vol. 22. P. 653–656.
- Panussis D.A., Cook M.W., Rifkin L.L., Snider J.E., Strong J.T., McGrane R.M., Wilson R.K., Mardis E.R.* A pneumatic device for rapid loading of DNA sequencing gels // *Genome Res.* 1998. Vol. 8. P. 543–548.
- Parik J., Kwiatkowski M., Lagerkvist A., Lagerstrom Fermer M., Samiotaki M., Stewart J., Glad G., Mendel-Hartvig M., Landegren U.* A manifold support for molecular genetic reactions // *Anal Biochem.* 1993. Vol. 211. P. 144–150.
- Parker L.T., Deng Q., Zakeri H., Carlson C., Nickerson D.A., Kwok P.Y.* Peak height variations in automated sequencing of PCR products using Taq dye-terminator chemistry // *Biotechniques*. 1995. Vol. 19. P. 116–121.
- Parker L.T., Zakeri H., Deng Q., Spurgeon S., Kwok P.Y., Nickerson D.A.* AmpliTaq DNA polymerase, FS dye-terminator sequencing: analysis of peak height patterns // *Ibid.* 1996. Vol. 21. P. 694–699.
- Prober J.M., Dam R.J., Robertson C.W., Jr., Hobbs F.W., Trainor G.L.* Method systems and reagents for DNA sequencing. Pat. 5306618 (US). Publ. 1994.
- Prober J.M., Trainor G.L., Dam R.J., Hobbs F.W., Robertson C.W., Zagursky R.J., Cocuzza A.J., Jensen M.A., Baumeister K.* A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides // *Science*. 1987. Vol. 238. P. 336–341.
- Quesada M.A.* Replaceable polymers in DNA sequencing by capillary electrophoresis // *Curr. Opin. Biotechnol.* 1997. Vol. 8. P. 82–93.
- Rawlinson W.D., Chee M.S., Smith V., Barrell B.G.* Single stranded rescue from phagemids in microtitre plates // *Nucl. Acids Res.* 1991. Vol. 19. P. 4779.
- Rolfs A., Weber I.* Fully-automated, nonradioactive solid-phase sequencing of genomic DNA obtained from PCR // *Biotechniques*. 1994. Vol. 17. P. 782–787.
- Rosenblum B.B., Lee L.G., Spurgeon S.L., Khan S.H., Menchen S.M., Heiner C.R., Chen S.M.* New dye-labeled terminators for improved DNA sequencing patterns // *Nucl. Acids Res.* 1997. Vol. 25. P. 4500–4504.
- Rosenthal A., Charnock-Jones D.S.* New protocols for DNA sequencing with dye terminators // *DNA Sequenc.* 1992. Vol. 3. P. 61–64.
- Rosenthal A., Coutelle O., Craxton M.* Large-scale production of DNA sequencing template by microtitre format PCR // *Nucl. Acids Res.* 1993. Vol. 21. P. 173–174.
- Rosenthal A., Schwertner S., Hahn V., Hunger H.-D.* Solid-phase methods for sequencing of nucleic acids. 1. Simultaneous sequencing of different oligodeoxyribonucleotides using a new, mechanically stable anion-exchange paper // *Ibid.* 1985. Vol. 13. P. 1173–1184.
- Rosenthal A., Sproat B., Voss H., Stegemann J., Schwager C., Erfle H., Zimmermann J., Coutelle C., Ansorge W.* Automated sequencing of fluorescently labelled DNA by chemical degradation // *DNA Sequenc.* 1990. Vol. 1. P. 63–71.
- Ruiz-Martinez M.C., Berka J., Belenkii A., Foret F., Miller A.W., Karger B.L.* DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection // *Anal. Chem.* 1993. Vol. 65. P. 2851–2858.
- Ruiz-Martinez M.C., Carrilho E., Berka J., Kieleczawa J., Miller A.W., Foret F., Carson S., Karger B.L.* DNA sequencing by capillary electrophoresis using short oligonucleotide primer libraries // *Biotechniques*. 1996. Vol. 20. P. 1058–1069.

Ruiz-Martinez M.C., Salas-Solano O., Carrilho E., Kotler L., Karger B.L. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replacable polymer solutions. A. Development of the cleanup protocol // *Anal. Chem.* 1998. Vol. 70. P. 1516–1527.

Salas-Solano O., Ruiz-Martinez M.C., Carrilho E., Kotler L., Karger B.L. A sample purification method for rugged and high-performance DNA sequencing by capillary electrophoresis using replacable polymer solutions. B. Quantitative determination of the role of sample matrix components on sequencing analysis // *Ibid.* 1998. Vol. 70. P. 1528–1535.

Sanders J.Z., Petterson A.A., Hughes P.J., Connell C.R., Raff M., Menchen S, Hood L.E., Teplow D.B. Imaging as a tool for improving length and accuracy of sequence analysis in automated fluorescence-based DNA sequencing // *Electrophoresis.* 1991. Vol. 12. P. 3–11.

Schmalzing D., Adourian A., Kouthy L., Ziaugra L., Matsudaira P., Ehrlich D. DNA sequencing on microfabricated electrophoretic devices // *Anal. Chem.* 1998. Vol. 70. P. 2303–2310.

Shealy D.B., Lipowska M., Lipowski J., Narayanan N., Sutter S., Strekowski L., Patonay G. Synthesis, chromatographic separation, and characterization of near-infrared-labeled DNA oligomers for use in DNA sequencing // *Ibid.* 1995. Vol. 67. P. 247–251.

Smith L.M., Fung S., Hunkapiller M.W., Humkapiller T.J., Hood L.E. The synthesis of oligonucleotides containing an aliphatic group at the 5' terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis // *Nucl. Acids Res.* 1985. Vol. 13. P. 2399–2412.

Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connel C.R., Heiner C., Kent S.B.H., Hood L.E. Fluorescence detection in automated DNA sequence analysis // *Nature.* 1986. Vol. 321. P. 674–679.

Smith V., Brown C.M., Bankier A.T., Barrell B.G. Semiautomated preparation of DNA templates for large-scale sequencing projects // *DNA Sequenc.* 1990. Vol. 1. P. 73–78.

Stegemann J., Schwager C., Erfle H., Hewitt N., Voss H., Zimmermann J., Ansorge W. High speed on-line DNA sequencing on ultrathin slab gels // *Nucl. Acids Res.* 1991. Vol. 19. P. 675–676.

Swerdlow H., Dew-Jageer K.E., Brady K., Grey R., Dovichi N.J., Gesteland R. Stability of capillary gels for automated sequencing of DNA // *Electrophoresis.* 1992. Vol. 13. P. 475–483.

Swerdlow H., Dew-Jageer K.E., Gesteland R. Reloading and stability of polyacrylamide slab gels for automated DNA sequencing // *Biotechniques.* 1994. Vol. 16. P. 684–693.

Swerdlow H., Gesteland R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing // *Nucl. Acids Res.* 1990. Vol. 18. P. 1415–1419.

Swerdlow H., Zhang J.Z., Chen D.Y., Harke H.R., Grey R., Wu S., Dovichi N.J. Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence // *Anal. Chem.* 1991. Vol. 63. P. 235–284.

Tabor S., Richardson C.C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase // *J. Biol. Chem.* 1990. Vol. 265. P. 8322–8328.

Tan H., Yeung E.S. Characterization of dye-induced mobility shifts affects DNA sequencing in poly(ethylene oxide) sieving matrix // *Electrophoresis.* 1997. Vol. 18. P. 2893–2900.

Toneguzzo F., Glynn S., Levi E., Mjolsness S., Hayday A. Use of a chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA // *Biotechniques.* 1988. Vol. 6. P. 460–469.

Tong X., Smith L.M. Solid phase purification in automated DNA sequencing // *DNA Sequenc.* 1993. Vol. 4. P. 151–162.

Trost P., Guttman A. Fiber bundle based scanning detection system for automated DNA sequencing // *Anal. Chem.* 1998. Vol. 70. P. 3930–3935.

Trower M.K., Burt D., Purvis I.J., Dykes C.W., Christodoulou C. Fluorescent dye-primer cycle sequencing using unpurified PCR products as templates, development of a protocol amenable to high-throughput DNA sequencing // *Nucl. Acids Res.* 1995. Vol. 23. P. 2348–2349.

Tu O., Knott T., Marsh M., Bechtol K., Harris D., Barker D., Bashkin J. The influence of fluorescent dye structure on the electrophoretic mobility of end-labeled DNA // *Ibid.* 1998. Vol. 26. P. 2797–2802.

Voss H., Grothues D., Sensen C., Zimmermann J., Schwager C., Stegemann J., Erfle H., Rupp T., Ansorge W. Automated low-redundancy large-scale DNA sequencing by primer walking // *Biotechniques.* 1993. Vol. 15. P. 714–721.

Voss H., Hentwich U., Wiemann S., Benes V., Zimmermann J., Ansorge W. Automated cycle

sequencing with Taqenase: Protocols for internal labeling, dye primer and "doublex" simultaneous sequencing // *Ibid.* 1997. Vol. 23. P. 312–318.

Voss H., Schwager C., Wirkner U., Sproat B., Zimmermann J., Rosenthal A., Erfle H., Stegemann J., Ansorge W. Direct genomic on-line sequencing and analysis using *in vitro* amplification of DNA // *Nucl. Acids Res.* 1989. Vol. 17. P. 2517–2527.

Voss H., Zimmermann J., Schwager C., Erfle H., Stegemann J., Stucky K., Ansorge W. Automated fluorescent sequencing of cosmid DNA // *Ibid.* 1990. Vol. 18. P. 1066.

Wada A. Automatic DNA sequencing // *Nature.* 1984. Vol. 307. P. 193.

Wada A. Automatic high-speed DNA sequencing // *Ibid.* 1987. Vol. 325. P. 771–772.

Wada A., Yamamoto M., Soeda E. Automatic DNA sequencer: Computer programmed microchemical manipulator for the Maxam–Gilbert sequencing method // *Rev. Sci. Instrum.* 1984. Vol. 54. P. 1569–1572.

Wahlberg J., Holmgren A., Begh S., Hultman T., Uhlen M. Automated magnetic preparation of DNA templates for solid phase sequencing // *Electrophoresis.* 1992. Vol. 13. P. 547–551.

Wendl M.C., Dear S., Hodgson D., Hillier L. Automated sequence preprocessing in a large-scale sequencing environment // *Genome Res.* 1998. Vol. 8. P. 975–984.

Wiemann S., Schilke A., Rechmann S., Zimmermann J., Voss H., Ansorge W. Reducing "double-sequences" in automated DNA sequencing with T7 DNA polymerase and internal labeling // *Biotechniques.* 1996a. Vol. 20. P. 791–792.

Wiemann S., Rupp T., Zimmermann J., Voss H., Schwager C., Ansorge W. Primer design for automated DNA sequencing utilizing T7 DNA polymerase and internal labeling with fluorescein-15-dATP // *Ibid.* 1995a. Vol. 18. P. 688–697.

Wiemann S., Stegemann J., Grothues D., Bosch A., Estivill X., Schwager C., Zimmermann J., Voss H., Ansorge W. Simultaneous on-line sequencing on both strands with two fluorescent dyes // *Anal. Biochem.* 1995b. Vol. 224. P. 117–121.

Wiemann S., Stegemann J., Zimmermann J., Voss H., Benes V., Ansorge W. "Doublex" fluorescent DNA sequencing: Two independent sequences obtained simultaneously in one reaction with internal labeling and unlabeled primers // *Ibid.* 1996. Vol. 234. P. 166–174.

Williams D.C., Soper S.A. Ultrasensitive Near-IR fluorescence detection for capillary gel electrophoresis and DNA sequencing applications // *Anal. Chem.* 1995. Vol. 67. P. 3427–3432.

Wilson R.K., Chen C., Avdalovic N., Burns J., Hood L. Development of an automated procedure for fluorescent DNA sequencing // *Genomics.* 1990. Vol. 6. P. 626–634.

Woolley A.T., Mathies R.A. Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips // *Proc. Nat. Acad. Sci. USA.* 1994. Vol. 91. P. 11348–11352.

Woolley A.T., Mathies R.A. Ultra-high-speed DNA sequencing using capillary electrophoresis chips // *Anal. Chem.* 1995. Vol. 67. P. 3676–3680.

Wu C., Quesada M.A., Schneider D.K., Farinato R., Studier F.W., Chu B. Polyacrylamide solutions for DNA sequencing by capillary electrophoresis: mesh sizes, separation and dispersion // *Electrophoresis.* 1996. Vol. 17. P. 1103–1109.

Yin Z., Severin J., Giddings M.C., Huang W.A., Westphall M.S., Smith L.M. Automatic matrix determination in four dye fluorescence-based DNA sequencing // *Ibid.* 1996. Vol. 17. P. 1143–1150.

Zagursky R.J., McCormick R.M. DNA sequencing separations in capillary gels on a modified commercial DNA sequencing instrument // *Biotechniques.* 1990. Vol. 9. P. 74–79.

Zakeri H., Amparo G., Chen S.M., Spurgeon S., Kwok P.Y. Peak height pattern in dichlororhodamine and energy transfer dye terminator sequencing // *Ibid.* 1998. Vol. 25. P. 406–414.

Zhang J., Fang Y., Hou J.Y., Ren J., Jiang R., Roos P., Dovichi N.J. Use of non-cross-linked polyacrylamide for four-color DNA sequencing by capillary electrophoresis separation of fragments up to 640 bases in length in two hours // *Anal. Chem.* 1995. Vol. 67. P. 4589–4593.

Zimmermann J., Voss H., Kristensen T., Schwager C., Stegemann J., Erfle H., Ansorge W. Automated preparation and purification of M13 templates for DNA sequencing // *Meth. Mol. and Cell. Biol.* 1989. Vol. 1. P. 29–34.

Zimmermann J., Voss H., Schwager C., Stegemann J., Ansorge W. Automated Sanger dideoxy sequencing reaction protocol // *FEBS Lett.* 1988. Vol. 233. P. 432–436.

Zimmermann J., Voss H., Schwager C., Stegemann J., Erfle H., Stucky K., Kristensen T., Ansorge W. A simplified protocol for fast plasmid DNA sequencing // *Nucl. Acids Res.* 1990. Vol. 18. P. 1067.

*Zimmermann J., Wieman S., Vos H., Schwager C., Ansorge W.* Improved fluorescent cycle sequencing protocol allows reading nearly 1000 bases // *Biotechniques*. 1994. Vol. 17. P. 302–307.

*Zollo M., Chen E.Y.* A manual high-throughput M13 DNA preparation // *Ibid.* 1994. Vol. 16. P. 370–372.

### *Дополнительная литература*

*Haugland R.P.* Molecular probes: Handbook of fluorescent probes and research chemicals. 1992–1994. 390 p.

*Hunkapiller T., Kaiser R.J., Koop B.F., Hood L.* Large-scale and automated DNA sequence determination // *Science*. 1991. Vol. 254. P. 59–67.

*Trainor G.L.* DNA sequencing, automation, and the human genome // *Anal. Chem.* 1990. Vol. 62. P. 418–426.



## КОМПЬЮТЕРНЫЙ АНАЛИЗ И БАЗЫ ДАННЫХ

Появление быстрых методов секвенирования ДНК путем химической дегградации по Максаму–Гилберту и ферментативным построением новой цепи ДНК в условиях терминации по Сэнгеру привело к резкому увеличению числа секвенированных фрагментов ДНК. Определяемые в ходе секвенирования нуклеотидные последовательности, в виде так называемых ДНКовых текстов, повлекли за собой разработку специализированных компьютерных программ по их анализу, поскольку обработка таких больших массивов данных без помощи компьютеров стала просто невозможна. Однако первые появившиеся компьютерные программы обращения с нуклеотидными последовательностями и анализа ДНК [Kom et al., 1977; Staden, 1977, 1978; Schroeder, Blattner, 1978] характеризовались минимальным набором сервисных функций. Стремительный рост числа разнообразных программ, рассчитанных на проведение тех или иных операций с нуклеотидными последовательностями ДНК, даже потребовал выделения отдельных номеров журнала “Nucleic Acids Research”, целиком посвященных данной проблеме [1982. Vol. 10. N1; 1984. Vol. 12. N1. Part 1, 2; 1986; Vol. 14. N1].

Многие компьютерные программы тех лет представляли собой небольшие программки для решения конкретных задач, вроде поиска сайтов рестрикционных эндонуклеаз, определения размеров получающихся фрагментов после расщепления ими молекул ДНК или определения молекулярной массы таких фрагментов, их нуклеотидного состава. С целью некоторого упорядочения информации обо всех этих программах и лучшей ориентации исследователей был подготовлен специальный указатель, вобравший в себя максимально возможное число известных к тому времени компьютерных программ и дающий краткое описание их возможностей [Rawling, 1986]. Однако становилась очевидной насущная потребность создания так называемых пакетов прикладных программ, которые бы позволяли проводить целый ряд необходимых операций по всевозможному анализу секвенированных фрагментов ДНК, начиная от занесения последовательности нуклеотидов в компьютер до выявления особенностей кодируемых ими белков.

Другим аспектом компьютерного анализа секвенированных молекул ДНК стал вопрос хранения полученных данных и необходимость обеспечения широкого доступа ученых к уже известным нуклеотидным последовательностям. Это привело к образованию специализированных банков данных, сначала одного, потом нескольких и уже затем многочисленных. В настоящее время, кроме трех основных, так называемых первичных банков данных (GenBank, EMBL, DDBJ), главной целью ставящих сбор нуклеотидных последовательностей, существует еще множество различных баз данных, преследующих какую-

либо конкретную цель. Те из них, которые в той или иной степени имеют отношение к секвенированию ДНК, будут кратко рассмотрены в данной главе наряду с уже упомянутыми базами данных GenBank, EMBL, DDBJ.

### 10.1. ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ ПОЛНОГО АНАЛИЗА НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СЕКВЕНИРОВАННЫХ ФРАГМЕНТОВ ДНК

Чтобы выявить какие-то особенности или характерные черты исследуемого гена или фрагмента ДНК, необходимо провести анализ его нуклеотидной последовательности, ставшей известной в результате секвенирования. Причем, зачастую требуется всесторонний анализ, который более удобно осуществить с помощью интегрированного пакета специализированных программ. В настоящее время имеется широкий выбор различных пакетов таких программ, отличающихся требованиями, предъявляемыми ими к самой компьютерной технике, к операционным системам. В силу особенностей компьютерного парка нашей страны, представленного подавляющим количеством IBM-совместимых компьютеров, программы, написанные для других типов машин, здесь, за редкими исключениями, упоминаться не будут.

К наиболее популярным пакетам можно отнести такие программы, как, например, Staden Plus, Pustell, PC/GENE, DNASIS, DNASTAR, доступные из коммерческих источников, представленных различными фирмами. В настоящее время многие из них имеют современные версии, работающие в среде Windows. Общей чертой всех этих программ является их довольно высокая, особенно для отечественных ученых, стоимость. В то же время отсутствие современной компьютерной программы может свести на нет все усилия, затраченные на получение в результате секвенирования ДНКовых текстов, которые могут остаться не до конца расшифрованными или даже просто разрозненными кусками. Чтобы облегчить для исследователей выбор той или иной программы анализа ДНК, наиболее соответствующей их запросам, компьютерные фирмы предоставляют возможность поближе познакомиться с их "детищем". Причем различные фирмы делают это несколько по-разному. Так, кроме привычных рекламных материалов, исследователям предоставляются для испытаний сами программы, обычно являющиеся в подобных случаях демо-версиями. Некоторые из таких демо-версий позволяют провести всесторонний анализ только какой-то одной нуклеотидной последовательности конкретного фрагмента ДНК или их небольшого числа, предложенных поставщиками. Другие предоставляют возможность работать с любыми нуклеотидными последовательностями, выбираемыми самими экспериментаторами, однако такие опции, как сохранение результатов анализа или их распечатка, в этих демо-версиях исключены. Ряд фирм после подписания специального соглашения, оговаривающего условия испытательного периода, запрещает

ющего копирование и использование программы в последующем, позволяют тестировать полностью функциональный пакет своих программ (обычно не самую последнюю версию) в течение 30 дней, после чего установочные материалы (дискеты, компакт-диски) должны быть возвращены владельцу, их предоставившему. Выбор программ последним способом позволяет в достаточно полной мере оценить их возможности и принять соответствующее решение. Так, в качестве примера приобретения пакета компьютерных программ можно привести опыт нашей лаборатории, когда было осуществлено тестирование ряда программ, как в виде различных демо-версий, так и предоставленной на 30-дневный испытательный срок программы Lasergene фирмы "DNASTAR, Inc." (США), по результатам которого выбор был остановлен именно на последней программе.

Фирма "DNASTAR, Inc." была основана еще в 1976 г. молекулярным биологом Ф.Р. Блаттнером (под руководством которого недавно завершено секвенирование полного генома *Escherichia coli* – см. следующую главу) и компьютерщиком Дж.Д. Шредером. На протяжении этих лет компьютерные программы, разрабатываемые специалистами данной фирмы, зарекомендовали себя как весьма удобные в работе и предоставляющие экспериментатору широкий выбор тех или иных способов анализа каких-то черт нуклеотидной или аминокислотной последовательностей исследуемых генов и кодируемых ими белков. Можно сказать, что эта фирма всегда "шла в ногу со временем", поставляя комплект специализированных программ, необходимый для всестороннего анализа ДНК, постоянно адаптируя его к реалиям текущего момента. Так, в настоящее время в полный пакет компьютерных программ Lasergene данной фирмы входят 8 модулей, один из которых (GeneQuest) добавился только в 1998 г. Следует отметить, что каждая программа из пакета Lasergene может поставляться и использоваться отдельно, что весьма удобно в тех случаях, когда пользователь может ограничиться лишь необходимым ему модулем и таким образом сэкономить деньги, однако модуль EditSeq, хотя и может работать в "одиночестве", является начальным уровнем всех остальных и требует за счет этого обязательного приобретения.

Любой компьютерный анализ последовательности нуклеотидов ДНК становится возможным только после занесения в память компьютера соответствующего ДНКового текста, что может осуществляться разными способами, как в ручном, так и автоматическом режимах, рассмотренных выше в главах 6 и 9. В комплекте программ Lasergene имеется специальный модуль EditSeq, который позволяет осуществлять это как с помощью клавиатуры (причем с возможностью контроля заносимой последовательности с помощью синтезатора речи), так и посредством специального дигитайзера. Кроме этого, возможен импорт каких-либо нуклеотидной или аминокислотной последовательностей из других молекулярно-биологических программ или баз данных. Последовательности из базы данных, предоставляемой самой фирмой "DNASTAR" в виде модуля GeneMap, могут быть взяты в анализ и редакцию непосредственно, не прибегая к процедуре импорта. Также EditSeq способен воспринять файл с текстом в виде ASCII кодов.

Выбор исследователем следующего модуля из пакета Lasergene для анализа секвенированной им последовательности ДНК зависит от информации, которой он уже владеет. Так, в случае секвенирования известного гена на данном этапе нет особой необходимости проводить поиск по всему банку данных на предмет выявления каких-либо гомологичных последовательностей и здесь очередной программой анализа может стать модуль MapDraw. Другое дело, когда секвенируются достаточно произвольные участки ДНК. Здесь следующим этапом после занесения нуклеотидных последовательностей в компьютер может стать состыковка всех секвенированных мелких фрагментов в один цельный с помощью модуля SeqMan или SeqMan II и уже затем поиск гомологии по базе данных Lasergene, способной пролить свет на то, что же все-таки секвенировано. Осуществить такой поиск поможет модуль GeneMap.

Так, при выполнении крупномасштабного проекта по секвенированию ДНК с помощью случайного подхода генерируется большое количество данных в виде последовательностей нуклеотидов относительно коротких фрагментов, принадлежащих произвольным участкам какого-либо протяженного фрагмента ДНК. Ручная состыковка таких коротких фрагментов в единую последовательность является весьма трудоемкой процедурой. Использование специализированных программ, таких как, например, модуль SeqMan II позволяет резко повысить эффективность этого процесса. Главное отличие новой версии этой программы заключается в том, что ранее SeqMan позволял обращаться с последовательностями ДНК, ограничиваемыми уровнем в тысячи пар нуклеотидов, тогда как новая версия SeqMan II способна осуществлять состыковку фрагментов ДНК, измеряющихся уже миллионами пар нуклеотидов. Высокое качество конечной последовательности ДНК обеспечивается в том числе отдельным этапом подготовки последовательности субклонов к компьютерной состыковке. На этом этапе происходит автоматическая идентификация и удаление загрязняющих векторных последовательностей. Еще один тип контаминации представляют собой случайно клонированные участки ДНК *E. coli*, также автоматически удаляемые на данном этапе. Кроме этого, увеличить достоверность заключительной последовательности и повысить эффективность процесса состыковки позволяет исключение из анализа 3'- и 5'-концевых участков, как правило, характеризующихся несколько большим числом ошибочно "прочитанных" нуклеотидов. Выбор отдельных важных параметров при проведении самой состыковки таких предварительно "улучшенных" последовательностей позволяет регулировать процесс формирования заключительной последовательности. Имеющаяся на данном этапе возможность осуществления редакторских функций в особо трудных и проблематичных регионах также способствует повышению качества заключительной последовательности. Также следует отметить, что SeqMan II способен проводить сбор первичных данных, получаемых в процессе автоматического секвенирования флуоресцентно меченных фрагментов ДНК, при их разделении гель-электрофорезом в приборе ABI 377, обеспечивая выверенную последовательность нуклеотидов за счет соответствующей обработки [Allex et al., 1997].

Еще один уже упоминавшийся выше модуль MapDraw из пакета программ Lasergene рассчитан на поиск в секвенированной последовательности сайтов различных рестрикционных эндонуклеаз, поиск открытых рамок считывания и их трансляцию в белковые продукты. Построение как линейных, так и кольцевых карт сопровождается возможностью добавления большого числа поясняющих символов различных элементов. Важной чертой этого модуля является создание рисунков, пригодных для публикации.

Широкие возможности анализа белков заключены в модуле Protean. Так, данная программа позволяет выявлять в исследуемых белках участки с  $\alpha$ -спиральной структурой и  $\beta$ -складками и прочие элементы их структурной организации, рассчитывать гидрофильные и гидрофобные области белков, предсказывать физико-химические и электрофоретические свойства анализируемых с помощью компьютера белков.

Важным элементом анализа секвенированных последовательностей ДНК является их сравнение друг с другом или так называемое множественное выравнивание. Ранее большинство компьютерных программ позволяло одновременно сравнивать друг с другом только две последовательности, что снижало ценность получаемых результатов. В настоящее время уже многие программы анализа ДНК и белков рассчитаны на одновременный анализ большого числа родственных последовательностей. Такая же возможность реализована в модуле MegAlign пакета программ Lasergene. Причем данный пакет позволяет проводить множественное выравнивание как последовательностей ДНК, так и белков. Повышению достоверности проведенного анализа способствует возможность ручного редактирования и окончательного “доведения” таких данных, поскольку случается так, что компьютер в некоторых случаях дает определенные систематические ошибки. Затем на основе уже выровненных последовательностей можно реконструировать филогенетические деревья, с вычисленным процентом сходства и генетическими расстояниями. Полученные результаты возможно сохранить в виде специальных файлов, что весьма удобно.

Если для широкомасштабного анализа множественным выравниванием недостаточно своих собственных данных, полученных в результате секвенирования каких-либо родственных последовательностей ДНК, то можно привлечь аналогичные сведения, хранящиеся в банках данных и полученные ранее другими авторами. Помочь этому может специальный модуль GeneMap. Он позволяет проводить интерактивный поиск по ряду объединенных генетических баз данных DDBJ/EMBL/GenBank и представленных на специальных компакт-дисках фирмы “DNASTAR”, обновляющихся каждые 2 мес. Гипотетические белки, полученные в результате трансляции секвенированного фрагмента ДНК, могут быть сравнены с белковыми базами данных SWISS-PROT, PIR/NBRF, GBTrans. Следует отметить, что поиск может проводиться с использованием Булевых операторов (AND, OR и NOT), варьируя различные критерии. Весьма удобным оказывается то, что можно сохранять результаты поиска в виде отдельного файла.

Новый модуль Gene Quest наряду с обычными чертами последовательностей ДНК и белков, выдаваемых и другими модулями, предназначен для выявления самих генов, их регуляторных элементов, причем не только путем обнаружения открытых рамок считывания, а и более сложным анализом частот встречаемости определенных олигонуклеотидных блоков. Кроме этого, Gene Quest способен сделать компьютерную имитацию результатов агарозного гель-электрофореза рестриктазных фрагментов ДНК, предсказать вторичную структуру молекул РНК. При условии подключения компьютера к сети Интернет данный модуль позволяет организовать поиск гомологичных последовательностей в ГенБанке.

Учитывая широкое применение олигонуклеотидных праймеров как в ПЦР, секвенировании ДНК, так и в других экспериментах, в пакет Lasergene входит специальный модуль PrimerSelect, предназначенный для анализа фрагментов ДНК с известной последовательностью нуклеотидов с целью подбора участков, наиболее оптимально соответствующих таким праймерам. Большой выбор задаваемых параметров позволяет подобрать олигонуклеотидные праймеры с минимальными различиями в их температуре отжига на матрице ДНК, что заметно повышает эффективность процесса ПЦР. Аналогичным образом можно подобрать оптимальные праймеры для секвенирования ДНК праймерной “прогулкой”.

Завершая описание пакета компьютерных программ Lasergene фирмы “DNASTAR” можно сказать, что это довольно стандартный набор необходимых программ для молекулярно-биологических исследований, включая секвенирование ДНК. Однако, несмотря на то что практически все его модули способны готовить данные, пригодные для публикации с использованием специального шрифта GeneFont, все же существуют специально написанные для этого программы, характеризующиеся уже соответственно большими возможностями. Объявленный пакет Lasergene99 вдобавок к стандартным функциям будет характеризоваться возможностью состыковки до 64000 первичных последовательностей фрагментов ДНК при секвенировании методом “дробовика”, BLAST-поиска из любого модуля (при условии подключения компьютера к сети Интернет), расширенными опциями импорта и экспорта нуклеотидных последовательностей, а также некоторыми другими улучшенными чертами.

Пожалуй, первая специализированная программа DNADRAW, готовящая информацию о нуклеотидной последовательности какого-либо фрагмента ДНК для ее опубликования, была разработана в 1986 г. [Shapiro, Senarathy, 1986]. Это не потеряло своей актуальности и в дальнейшем, поскольку как обычные текстовые редакторы, так и большинство программ анализа ДНК не позволяют придавать описываемой последовательности ДНК, готовящейся к опубликованию в виде рисунка, большое число характерных черт. Так, новая версия этой программы DNAdraw способна отображать последовательность ДНК или белков, используя для выделения отдельных мотивов и элементов до 8 уровней затенения, подчеркнутый, курсивный, жирный или обра-

щенный шрифты с фиксированными интервалами между знаками, а также еще целый ряд других возможностей [Shapiro, 1995]. К программам отображения последовательностей ДНК нового поколения можно отнести программу DNAView, которая легко работает с фрагментами ДНК до 100 тпн [Singh, Krawetz, 1995]. DNAView, используя цветовое кодирование, позволяет на одной странице компактно разместить информацию о 100 тпн, причем такое изображение может быть сохранено в виде отдельного файла в стандартном графическом формате TIFF. Интересное решение по использованию интенсивности цветового изображения было предложено этими авторами, чтобы показать сколько раз “читался” тот или иной нуклеотид при выполнении проекта по секвенированию ДНК.

## 10.2. БАЗЫ ДАННЫХ ПО НУКЛЕОТИДНЫМ И АМИНОКИСЛОТНЫМ ПОСЛЕДОВАТЕЛЬНОСТЯМ

Нарастающее число секвенированных нуклеотидных последовательностей, принадлежащих разным организмам, заставило задуматься об упорядочении их хранения, что привело к созданию “коллекций” таких последовательностей. Первая компьютеризованная база данных “The Los Alamos DNA Database”, содержащая информацию о секвенированных последовательностях ДНК, ставших известными с 1967 г., была организована У. Гоадом и его коллегами в Лос-Аламосской национальной лаборатории (LANL) в США. К лету 1981 г. в ней уже содержалась информация о 280 опубликованных нуклеотидных последовательностях общей протяженностью около 370 тпн [Kanehisa, 1982]. В 1982 г. на ее основе под патронажем Национального института здоровья (NIH) был организован новый банк данных генетических последовательностей GenBank. Сотрудники группы теоретической биологии и биофизики LANL продолжали для этой новой “старой” базы данных собирать нуклеотидные последовательности, давать им аннотации, формировать на их основе ГенБанк [Bilofsky et al., 1986]. На протяжении ряда лет распространением собранных сведений в виде данных на магнитных носителях занимался специальный отдел фирмы “Bolt Beranek and Newman, Inc.” С 1 октября 1987 г. эти функции были переданы компьютерной фирме “IntelliGenetics, Inc.”, занимавшейся этим до 1992 г. Произшедшая в 1992 г. реорганизация ГенБанка и его переподчинение Национальному центру биотехнологической информации (NCBI), созданному при NIH, переместила ГенБанк с западного на восточное побережье США. Хотя, начиная с 1982 г. главная финансовая поддержка ГенБанка исходила от NIH, Лос-Аламосская команда играла центральную роль в сборе секвенированных нуклеотидных последовательностей. Делавшая это на протяжении почти 20 лет в результате подобного решения она была отодвинута на второй план, поскольку новые сведения о нуклеотидных последовательностях теперь исследователи должны были уже посылать в NCBI.

GenBank хотя и содержит информацию, собранную со всего мира, является американской базой данных, поскольку финансируется правительством США. Созданная в 1980 г. европейская аналогичная база данных The EMBL Data Library, явилась первым международным хранилищем информации о последовательностях ДНК. Основными целями, преследовавшимися при ее организации, были: обеспечение бесплатного доступа к коллекции опубликованных нуклеотидных последовательностей, выработка определенных стандартов и развитие информационного и компьютерного обеспечения проводимых молекулярно-биологических исследований [Hamn, Cameron, 1986]. Свой первый релиз, охватывающий информацию о 568 последовательностях общей протяженностью 585433 нуклеотида, The EMBL Data Library выпустила в апреле 1982 г. Первое время очередные релизы, распространяемые на магнитных лентах, предоставлялись бесплатно любому желающему по его запросу. Однако рост как самой базы, так и числа пользователей привел к тому, что дальнейшие выпуски стали распространяться по подписке, как, например, 12 выпуск (июль 1987 г.), вместивший в себя информацию о почти 14 миллионах пар нуклеотидов [Cameron, 1988]. В 1994 г. база данных EMBL Data Library трансформировалась в EMBL Nucleotide Sequence Database и стала курироваться Европейским институтом биоинформатики (European Bioinformatics Institute или EBI) [Emmert et al., 1994]. В настоящее время ее адрес в ИНТЕРНЕТ – <http://www.ebi.ac.uk> .

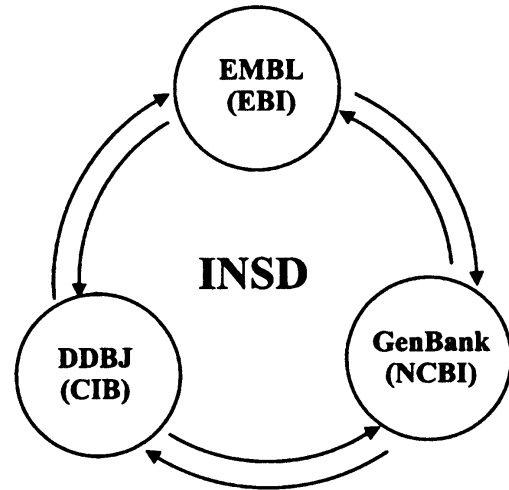
Третий из группы основных – банк данных DDBJ (DNA Data Bank of Japan) начал функционировать в Японии с 1986 г. В 1995 г. в Японии был создан Центр информационной биологии (CIB) и DDBJ, организованный ранее по инициативе Национального института генетики, перешел к нему в подчинение. Ныне его адрес во всемирной сети Интернет – <http://www.ddbj.nig.ac.jp> .

В настоящее время действует некий альянс этих трех основных банков по нуклеотидным последовательностям в виде The International Nucleotide Sequence Database (INSDB). INSDB есть результат трехстороннего сотрудничества GenBank (NCBI), EMBL (EBI) и DDBJ (CIB), главными составляющими которого являются единые требования, предъявляемые к авторам при занесении ими нуклеотидных последовательностей, присвоение единых для всех этих трех банков инвентарных номеров, регулярный ежедневный обмен информацией о новых поступивших нуклеотидных последовательностях (рис. 10.1).

В феврале 1998 г. в DDBJ/EMBL/GenBank содержалась информация о более чем 2 000 000 последовательностях ДНК, общая длина которых перевалила за 1 млрд нуклеотидов. Свыше 30 000 видов и подвидов различных организмов представлены в этих первичных банках и каждый месяц добавляется около 600 новых [Benson et al., 1998]. Если ранее для удвоения общего числа нуклеотидов, хранимых в этих банках последовательностей, требовалось 18 мес, потом 1 год, то в 1998 г. темпы еще более возросли и уже в декабре в этих банках данных стала содержаться информация в виде более чем 2 млрд нуклеотидных оснований для 3 044 000 различных последовательностей. Считается что около 43%



Рис. 10.1. Международная база данных по нуклеотидным последовательностям INSD, существующая благодаря сотрудничеству трех основных банков данных DDBJ/EMBL/GenBank



хранящихся последовательностей принадлежат ДНК человека (причем около половины из них являются EST), далее следует мышь (10%) и нематода (9%). Следует отметить, что около 70% от всех хранящихся в банках данных нуклеотидных последовательностей представлены в виде участков EST.

Вся информация, хранящаяся в первичных банках данных DDBJ/EMBL/GenBank, организована по таксономическому признаку, за исключением нескольких частных случаев, где она сгруппирована на основе специфической природы этой информации (например, раздел HTG, посвященный полным геномам организмов). Такой подход оказался достаточно удобным, и по мере необходимости происходит дробление каких-либо все же условных групп. Так, в настоящее время в базе данных EMBL все последовательности распределены по 17 разделам, перечисленным здесь с указанным в скобках их трехбуквенным кодом: Bacteriophage (PHG), ESTs (EST), Fungi (FUN), High throughput genome (HTG), Genome survey sequences (GSS), Human (HUM), Invertebrates (INV), Organelles (ORG), Other mammals (MAM), Other vertebrates (VRT), Plants (PLN), Prokaryotes (PRO), Rodents (ROD), STSs (STS), Synthetic (SYN), Unclassified (UNC), Viruses (VIR). Даже по названиям разделов легко видеть, что выделение какой-либо группы последовательностей в отдельный раздел напрямую связано с количеством последовательностей, к ней относящихся. В ГенБанке к ранее выделенным в виде отдельных разделов dbSTS и dbEST в сентябре 1998 г. добавился новый – dbSNP (Single-Nucleotide-Polymorphism), предназначенный для сбора информации о полиморфизме человеческого генома в виде единичных замен нуклеотидов.

В первые годы существования первичных баз данных (“The Los Alamos DNA Database”, GenBank, EMBL, DDBJ) их пополнение происходило путем поиска штатными сотрудниками статей в научных журналах и ручное занесение найденных в них нуклеотидных последовательностей, сопровождаемое аннотацией, составленной также на основе сведений, содержащихся в статье, в компьютер. Для исключения ошибок этот процесс повторялся даже на разных компьютерах и данные потом сличались. Кроме этого, бумажная копия с компьютерным вариантом последовательности и ее аннотацией посылалась авторам для подтверждения. Такой порядок действовал чуть ли не до конца 80-х годов, до тех пор, пока большинство научных журналов в качестве обязательного условия опубликования статьи, содержащей информацию о последовательности ДНК, стали требовать предварительного занесения данной

последовательности в какой-либо из трех основных банков данных и присвоения ей специального инвентарного номера (Accession number). В результате такого решения была разработана специальная форма, сопровождающая каждую регистрируемую последовательность фрагмента ДНК, причем в первое время базы данных, кроме электронных вариантов ДНК-овых текстов, в исключительных случаях продолжали принимать нуклеотидные последовательности в виде бумажных копий. Следует отметить, что, например, штат EMBL продолжает просматривать европейские журналы, но уже на предмет поиска библиографических данных с целью уточнения и пополнения сведений, сопровождающих уже существующие занесения нуклеотидных последовательностей в базу данных [Stoesser et al., 1998].

Развитие электронных средств связи сделало возможным занесение нуклеотидной последовательности и сведений о ней в банки данных посредством электронной почты. С целью упрощения процедуры представления данных и их регистрации сотрудниками ГенБанка совместно с фирмой "IntelliGenetics" в 1991 г. была разработана специальная программа Authorin, рассчитанная на единообразное оформление подаваемых сведений о секвенированном фрагменте ДНК и самой нуклеотидной последовательности. Весьма важной чертой этой программы явилась возможность импорта ею файлов с последовательностями нуклеотидов из других компьютерных программ анализа ДНК. Программа Authorin, умещавшаяся на трех дискетах емкостью 360 Кб, вместе с весьма объемным руководством бесплатно высылалась после получения запроса сначала фирмой "IntelliGenetics", а затем и штатом NCBI. По мере развития всемирной компьютерной сети программа Authorin стала доступна посредством FTP. Несмотря на значительное удобство регистрации новых последовательностей ДНК в банках данных, эта программа не была лишена некоторых недостатков, и на смену ей в 1994 г. пришла программа BankIt, с помощью которой можно произвести занесение нуклеотидной последовательности в ГенБанк посредством ИНТЕРНЕТА после того, как GenBank открыл свой сервер в сети WWW (<http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html> ). Сообщается, что ныне свыше 65% всех занесений нуклеотидных последовательностей в ГенБанк осуществляется с помощью программы BankIt [Benson et al., 1998]. Другая программа занесения Sequin, разработанная сотрудниками ГенБанка, рассчитана, в первую очередь, на обращение с длинными и сегментированными последовательностями, а также с целыми наборами сходных последовательностей и используется несколько реже более простой программы BankIt. Что касается программы Authorin, то в настоящее время в Генбанке с ее помощью регистрируется менее 1% всех последовательностей, а начиная с 1 января 1999 г. подобные занесения просто перестают приниматься этой базой данных.

С помощью современной программы BankIt авторы вносят всю информацию о последовательности ДНК непосредственно в специальную форму, причем сама нуклеотидная последовательность может быть перенесена в нее из буфера обмена. Затем посредством сети ИНТЕРНЕТ заполненная форма направляется в ГенБанк, где создается черновик

этого занесения и у автора, таким образом, имеется возможность еще раз провести ревизию регистрируемых данных. Следует отметить, что присваиваемый перед заполнением формы идентификационный номер (BankIt ID) не является инвентарным номером (Accession number) и используется, например, для вызова свежезаполненной формы с целью каких-либо изменений в ней, когда окончательный инвентарный номер еще не присвоен. Accession number (регистрационный номер) нуклеотидной последовательности служит подтверждением того, что последовательность секвенированного фрагмента ДНК помещена в банк данных и отныне может использоваться как авторами, так и другими исследователями при ссылках на нее в публикуемых материалах.

ГенБанк обычно присваивает инвентарный номер какой-либо нуклеотидной последовательности в течение 1 рабочего дня с момента получения. Это время уходит на проверку “загрязненности” данной последовательности участками вектора, на правильность транслирования при условии, что фрагмент ДНК содержит кодирующие регионы и некоторые другие процедуры. По крайней мере, 4 сотрудника ГенБанка вручную обрабатывают представленную последовательность ДНК вместе с ее аннотацией, после чего автору обычно по электронной почте сообщается надлежащая информация. Исключение составляют EST, STS и некоторые другие типы данных, которые обрабатываются в полуавтоматическом режиме. Однако только по прошествии 5 дней сама последовательность вместе с аннотацией помещается в базу данных и становится доступной всем пользователям, кроме тех случаев, когда автором в процессе занесения было указано, что данная нуклеотидная последовательность должна сохраняться конфиденциальной до момента опубликования связанной с ней статьи.

Аналогичная BankIt программа занесения Webin разработана штатом EMBL. Начиная с июня 1997 г., она стала наиболее популярной при регистрации нуклеотидных последовательностей секвенированных фрагментов ДНК в этой базе данных. Кроме Webin, другой используемой программой для банка данных EMBL, остается Sequin, доступная также по компьютерным сетям из сервера EBI через WWW browser (<http://www.ebi.ac.uk/subs/allsubs.html> ) или FTP (<ftp://ftp.ebi.ac.uk/pub/software/sequin/>). Для банка DDBJ, помимо той же программы Sequin, действует еще и Sakura, аналогичная BankIt и Webin.

Штат EMBL извещает пользователей о том, что в случае занесения ими большого числа (более 25) схожих последовательностей, авторам рекомендуется заранее связаться со служащими банка данных и указать количество таких фрагментов ДНК с тем, чтобы в дальнейшем присвоить им очередные порядковые инвентарные номера, поскольку в среднем каждую минуту кем-то заносится новая последовательность ДНК. Другим положительным моментом такого массового занесения является облегчение и ускорение анализа этих последовательностей и их аннотаций за счет присущих им общих черт отвечающих за это кураторами [Stoesser et al., 1998].

Что касается самих инвентарных номеров, то присваивающиеся ранее состояли из одной впереди стоящей буквы и пяти цифр (напри-

мер, X66106), однако бурный рост числа секвенированных фрагментов ДНК вызвал к жизни необходимость перейти к новой системе, состоящей уже из двух букв и 6 цифр (например, AF047363). Следует отметить, что прежние номера остаются шестисимвольными и не будет проводиться их трансформация в восьмизначные. Кроме этого, при аннотации нуклеотидной последовательности появилось еще одно новшество в виде дополнительного номера GI (GenBank Identifier). Эта информация является своеобразным идентификатором нуклеотидной последовательности и будет меняться при возможных последующих изменениях самой последовательности, что позволит проследить этот процесс во времени, тогда как ее инвентарный номер будет оставаться неизменным. Так, сотрудниками EMBL отмечается, что в отдельные месяцы они получают до 1500 извещений о частичном изменении или самой последовательности, или аннотации к ней. Около 900 изменений из этих 1500 связаны с публикациями, поскольку большинство занесений нуклеотидных последовательностей в банк данных осуществляется обычно еще до того, как статьи приняты к опубликованию. В связи с этим штат EMBL обращается к авторам с просьбой информировать их о каких-либо изменениях данного статуса, что должно сделать банк данных более совершенным. Выявляемые иногда неточности в секвенированной последовательности ДНК уже после опубликования статьи крайне редко служат поводом для написания новой. В этой связи внести исправления в конкретную последовательность, депонированную в банке данных, гораздо проще, однако сделать такие изменения могут только сами авторы. Весьма удобным инструментом для этого является, например, тот же BankIt, его Update функция, причем не важно, с помощью какой программы ранее была занесена данная последовательность. Стремясь сделать базу данных максимально достоверной, ГенБанк, например, приглашает всех пользователей, обнаруживших какие-либо неточности и ошибки, сообщать об этом посредством email [Venson et al., 1998].

Второй важной составляющей сбора и хранения информации о секвенированных нуклеотидных последовательностях в базах данных является ее доступность для исследователей. Этому аспекту деятельности банков данных всегда уделялось должное внимание и выше уже упоминались рассылаемые ими постоянно увеличивающиеся в объеме релизы. Так, например, в 1985 г. информация о более чем 4 000 000 пн, принадлежащих почти 6300 зарегистрированным последовательностям ДНК, разбитым по таксономическим признакам на 12 групп, поставлялась на 14 дискетах емкостью 360 Кб каждая вместе со специальным 56-страничным руководством (GenBank Release 28.1 и EMBL Release 4.0). Бурный рост секвенируемых последовательностей ДНК и стремительное развитие компьютерной техники привели к выпуску в 1992 г. подобной базы данных уже на лазерном компакт-диске емкостью 680 Мб. В декабре 1997 г. для ГенБанковского релиза 104 вместе с базой данных Entrez потребовалось уже 12 подобных компакт-дисков. В связи с высокой стоимостью такого комплекта и неудобства обращения с ним было принято решение о прекращении выпуска баз данных в подобном виде

с апреля 1998 г. Хотя подобные намерения существовали уже давно, когда только хватало еще “всего” 6 компакт-дисков. Так, штатом ГенБанка был проведен специальный опрос пользователей, результаты которого показали, что более 70% американских и около 50% остальных исследователей предпочитают интернетовскую версию, кроме большего удобства, также и по причине ее ежедневного обновления (Update секции), тогда как релизы на компакт-дисках выходят только ежеквартально. Отмечалось также, что максимальное число подписчиков базы данных Entrez с 2200 в прошлом значительно снизилось (до 1400 в 1996 г.). Здесь, видимо, необходимо отметить, что весьма важной вехой в развитии сервисных услуг Генбанка оказалось его начавшееся в 1988 г. взаимодействие с другой известной базой данных по библиографии MEDLINE, приведшее к формированию перекрестных ссылок как на опубликованные статьи, связанные с секвенированием ДНК, так и на регистрационные номера нуклеотидных последовательностей [Bicknell et al., 1988]. Свое дальнейшее развитие такое взаимодействие получило при формировании новой системы PubMed, предназначенной для поиска библиографических данных в области молекулярной биологии, что, однако, сильно увеличило объем данных и потребовало большого числа компакт-дисков. Что касается базы данных EMBL, то ее штат продолжает осуществлять ежеквартальный выпуск релизов на нескольких CD-ROM, который, кроме самих последовательностей ДНК, еще содержит другие поддерживаемые и распространяемые EBI базы данных общим количеством свыше 50.

Следует отметить, что вслед за отказом многих журналов публиковать в оригинальных статьях последовательности нуклеотидов, ограничиваясь только их инвентарными номерами, еще более возросла роль банков данных как источников важной информации уже не доступной из статей. У отдельных ученых появилась потребность иногда даже не в самой базе со всеми ее нуклеотидными последовательностями, а только в конкретном фрагменте ДНК более не доступным ему из статьи. Развитие сети электронных коммуникаций позволило EMBL организовать новые возможности обслуживания клиентов на своем сервере с помощью или email, или FTP [Stoehr, Omond, 1989; Fuchs et al., 1990]. Так, получив по электронной почте следующий запрос GET NUC: X66105 (где X66105 представлял собой найденный в статье Accession number интересующей экспериментатора последовательности нуклеотидов) сервер автоматически высылал запрашиваемый файл. Развитие всемирной сети Интернет позволило осуществить прямой поиск интересующей исследователя последовательности на серверах каждого из этих трех первичных банков данных. Другой предоставляемой ГенБанком возможностью является поиск гомологий с помощью целого семейства программ BLAST (Basic Local Alignment Search Tool) [Benson et al., 1998]. Новой мощной программой этого семейства является PowerBLAST, поддерживающая возможность анализа ДНК на уровне “сравнительной геномики” за счет обращения с последовательностями ДНК любой длины, поскольку она способна их разбивать и затем сращивать в автоматическом режиме [Zhang, Madden, 1997].

Кроме самого EBI-сервера, в Европе с 1988 г. действует широкоразвитая сеть EMBnet, охватывающая большинство стран с центральным узлом, расположенным в Голландии (<http://www.embnet.org> ). В декабре 1996 г. к этой сети присоединилась и Россия. Российский узел EMBnet, находящийся в НИИ физико-химической биологии им. А.Н. Белозерского МГУ (<http://www.genebee.msu.su/> ) предоставляет зарегистрированным пользователям широкий круг возможностей анализа ДНК. Следует отметить, что компьютерный центр НИИ физико-химической биологии им. А.Н. Белозерского МГУ после соответствующей регистрации и лицензирования предоставляет возможность анализа ДНК с помощью пакета программ GCG. The Genetic Computer Group при лаборатории генетики Висконсинского университета существует с 1980 г. и за это время ею подготовлен весьма мощный пакет программ, требующий, однако, соответствующего компьютерного обеспечения, доступного далеко не каждой лаборатории. В связи с этим возможность использовать пакет программ GCG с помощью данного сервера является удобной альтернативой.

Еще одной существующей с 1981 г. фактографической базой данных, в которой содержатся сведения о более чем 300000 аннотированных нуклеотидных и аминокислотных последовательностях, является GENESEQ английской фирмы "Derwent" (<http://www.derwent.com> ). Особенность этой базы данных заключается в том, что сами последовательности и информация о них собрана из патентной литературы до 1982 г. (имеется в виду от настоящего времени) и большей частью отсутствует в остальных базах данных.

База данных GSDB (Genome Sequence DataBase), существующая в США при Национальном центре геномных ресурсов (NCGR), собирает сведения о полных геномах различных организмов с целью сделать более удобной сравнительную и функциональную геномику, в первую очередь, прокариот, ведущую к пониманию функции генов, а также биохимии и физиологии бактерий [Harger et al., 1998]. Причем GSDB, доступная также через ИНТЕРНЕТ (<http://www.ncgr.org> ), хранит нуклеотидные последовательности геномов, отдельных хромосом в виде единых файлов независимо от их размера.

Придавая серьезное значение достоверности имеющихся сведений, штат этой базы данных в качестве одной из своих целей ставит перед собой идентификацию и удаление нуклеотидных последовательностей, принадлежащих векторам и занесенных в банк данных по недосмотру авторов. Так, в результате обследования почти 1 млн последовательностей было обнаружено, что 0,36% из них загрязнены векторными последовательностями (представляющими собой обычно область полилинкера), причем подавляющее большинство таких участков располагалось или на 5'-, или на 3'-конце заносимых в базы данных фрагментов ДНК и лишь немногие несли векторные участки на обоих концах. Еще меньшее число несло подобные участки внутри секвенированного фрагмента ДНК, являющиеся результатом или каких-то перестроек, или образования химерных конструкций на этапе клонирования. После обнаружения такого векторного загрязнения де-

лалось его подробное описание, однако удаление этих участков из зарегистрированных последовательностей проводилось только после консультаций с авторами. Проведенный анализ показал, что более 50% обнаруженных векторных “загрязнений” приходится на последовательности EST и STS.

К проблеме “загрязнения” нуклеотидных последовательностей в банках данных участками векторов обращались и ранее [Hodgson, 1990; Lopez et al., 1992; Kristensen et al., 1993; Reynolds, 1994]. В одной работе проведен детальный анализ и все случаи присутствия последовательностей, принадлежащих и плазмидным, фаговым и фагмидным векторам, разделены на 5 категорий, исходя из причин их попадания в базы данных вместе с основными последовательностями [Lamperti, 1992]. К другим “загрязнениям” банков данных случайными последовательностями следует отнести присутствие участков рибосомных РНК и их генов, особенно среди EST и STS [Gonzalez, Sylvester, 1997]. Главной причиной этого служит повышенное содержание этих элементов генома при создании геномных и особенно кДНКовых библиотек.

Базой данных, специализирующейся на информации, получаемой в рамках проекта “Геном человека”, является Genome Database (GDB). К основным типам данных, хранящихся в ней, относится информация о генетических картах, о генах, их полиморфизме, клонах и литературных ссылках [Letovsky et al., 1998]. Кроме основного адреса в Интернете (<http://gdb.org>), расположенного в США, эта база данных имеет дополнительные международные сайты в ряде стран.

Среди большого числа баз данных, хранящих сведения об аминокислотных последовательностях, следует отметить, пожалуй, лишь некоторые, в большей степени связанные с банками нуклеотидных последовательностей, по причине получения оттуда части информации в виде транслированных последовательностей. Одной из таких баз данных является SWISS-PROT, существующая с 1986 г. и с 1987 г. поддерживаемая совместно Женевским университетом и EMBL Data Library (в настоящее время – EMBL Outstation и Европейским институтом биоинформатики (EBI)). База данных SWISS-PROT вместе с приложением TrEMBL, представляющим собой кодирующие последовательности, выведенные из нуклеотидных последовательностей EMBL, распространяется EBI на лазерных компакт-дисках с периодичностью 1 раз в три месяца, а также доступна в WWW (<http://www.expasy.ch> или через сервер EBI <http://www.ebi.ac.uk/>). Важной чертой аминокислотных последовательностей, хранящихся в этой базе данных, является их подробное описание [Bairoch, Arweiler, 1998]. Другая аналогичная база данных PIR (Protein Information Resource) под патронажем Национального фонда биомедицинских исследований США (NBRF) существует даже с 1984 г. С 1988 г. она превратилась в международную PIR – International путем кооперации с базами данных JPID (Япония) и MIPS (Германия). Основная web-страница этой базы данных в сети ИНТЕРНЕТ может быть найдена по адресу: <http://www-nbrf.georgetown.edu/pir/>. Кроме такого доступа, PIR распространяется с периодичностью 4 раза

в год на компакт-дисках (Barker et al., 1998). База данных MIPS кроме последовательностей белков поддерживает также информацию о полных геномах различных организмов [Mewes et al., 1998]. Так, ее сотрудниками подготовлен специальный компакт-диск с интерактивным доступом к информации о всех 16 хромосомах дрожжей *Saccharomyces cerevisiae*.

Пожалуй, особым типом баз данных являются собранные воедино сведения о нуклеотидных и аминокислотных последовательностях разных первичных банков данных и поставляемые вместе с пакетами коммерческих компьютерных программ, таких как, например, DNASIS, Lasergene и др. Их особенность заключается в том, что они позволяют осуществлять поиск только с помощью соответствующей программы из “своего” пакета. Другой особенностью является достаточно частое обновление и выпуск новых CD-ROM релизов каждые два месяца (для Lasergene), что все же чаще, чем это делает база данных EMBL или аналогичные.

Среди других баз данных, имеющих определенное отношение к секвенированию ДНК, можно выделить существующую уже на протяжении многих лет базу данных по рестрикционным эндонуклеазам и метилазам REBASE, поддерживаемую известной фирмой-производителем этих ферментов “New England Biolabs, Inc.” (США). По состоянию на ноябрь 1997 г. в ней находилась информация о 2900 рестрикционных эндонуклеазах, включая изошизомеры [Roberts, Macelis, 1998]. Большинство представленных в ней ферментов относится ко II типу, число прототипов которых составляет уже 212. Кроме сведений о ферментах, в REBASE содержится информация о более чем 5000 библиографических ссылках. Обновление этой базы данных происходит ежедневно и раз в месяц обновленные сведения по электронной почте рассылаются ученым, чьи адреса находятся в листе рассылки, после того, как ими был сделан соответствующий запрос. Кроме этого, доступ к базе данных REBASE возможен также через ИНТЕРНЕТ. Адрес REBASE в WWW—<http://www.neb.com/rebase> .

### Цитируемая литература

*Allex C.E., Baldwin S.F., Shavlik J.W., Blattner F.R.* Improving the quality of automatic DNA sequence assembly using fluorescent trace-data classifications // Proc. Forth Intern. conf. intelligent systems mol. biol. AAAI press, 1996. P. 3–14.

*Allex C.E., Baldwin S.F., Shavlik J.W., Blattner F.R.* Increasing consensus accuracy in DNA fragment assemblies by incorporating fluorescent trace representations // Proc. Fifth Intern. conf. intelligent systems mol. biol. AAAI press, 1997. P. 3–14.

*Bairoch A., Apweiler R.* The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998 // Nucl. Acids Res. 1998. Vol. 26. P. 38–42.

*Barker W.C., Garavelli J.S., Haft D.H., Hunt L.T., Marzec C.R., Orcutt B.C., Srinivasarao G.Y., Yeh L.-S.L., Ledley R.S., Mewes H.W., Pfeiffer F., Tsugita A.* The PIR-International protein sequence database // Ibid. 1998. Vol. 26. P. 27–32.

*Benson D.A., Boguski M., Lipman D.J., Ostell J., Ouellette B.F.F.* GenBank // Ibid. 1998. Vol. 26. P. 1–7.

*Bicknell E.J., Rada R., Davidson S., Stander R.* Mapping from GenBank to MEDLINE // Ibid. 1988. Vol. 16. P. 1667–1680.



- Bilofsky H.S., Burks C., Fickett J.W., Goad W.B., Lewitter F.I., Rindone W.P., Swindell C.D., Tung C.-S.* The GenBank® genetic sequence databank // *Ibid.* 1986. Vol. 14. P. 1–4.
- Bishop M.J.* Computing // DNA sequencing: A practical approach / Ed. C.T. Howe, E.S. Ward. Oxford: IRL press, 1989. P. 185–219.
- Cameron G.N.* The EMBL data library // *Nucl. Acids Res.* 1988. Vol. 16. P. 1865–1867.
- Emmert D.B., Stoehr P.J., Stoesser G., Cameron G.N.* The European Bioinformatics Institute (EBI) databases // *Ibid.* 1994. Vol. 22. P. 3445–3449.
- Gonzalez I.L., Sylvester J.E.* Incognito rRNA and rDNA in databases and libraries // *Genome Res.* 1997. Vol. 7. P. 65–70.
- Fuchs R., Stoehr P., Rice P., Omond R., Cameron G.* New services of the EMBL data library // *Nucl. Acids Res.* 1990. Vol. 18. P. 4319–4323.
- Hamm G.H., Cameron G.N.* The EMBL data library // *Ibid.* 1986. Vol. 14. P. 5–9.
- Harger C., Skupski M., Bingham J., Farmer A., Hoisie S., Harber P., Kiphart D., Krakowski L., McLeod M., Schwertfeger J., Seluja G., Siepel A., Singh G., Stamper D., Steadman P., Thayer N., Thompson R., Wargo P., Waugh M., Zhuang J.J., Schad P.A.* The genome sequence database (GSDB): Improving data quality and data access // *Ibid.* 1998. Vol. 26. P. 21–26.
- Hodgson C.P.* Cloning vector artifacts in the DNA database // *Biotechniques.* 1990. Vol. 9. P. 54–55.
- Kanehisa M.* Los Alamos sequence analysis package for nucleic acids and proteins // *Nucl. Acids Res.* 1982. Vol. 10. P. 183–196.
- Korn L.J., Queen C.L., Wegman M.N.* Computer analysis of nucleic acid regulatory sequences // *Proc. Nat. Acad. Sci. USA.* 1977. Vol. 74. P. 4401–4405.
- Kristensen T., Lopez R., Prydz H.* An estimate of the sequencing error frequency in the DNA sequence databases // *DNA Sequenc.* 1992. Vol. 2. P. 343–346; Erratum // *DNA Sequenc.* 1993. Vol. 3. P. 337.
- Lamperti E.D., Kittelberger J.M., Smith T.F., Villa-Komaroff L.* Corruption of genomic databases with anomalous sequence // *Nucl. Acids Res.* 1992. Vol. 20. P. 2741–2747.
- Letovsky S.I., Cottingham R.W., Porter J., Li P.W.D.* GDB: the human genome database // *Ibid.* 1998. Vol. 26. P. 94–99.
- Lopez R., Kristensen T., Prydz H.* Database contamination // *Nature.* 1992. Vol. 355. P. 211.
- Mewes H.W., Hani J., Pfeiffer F., Frishman D.* MIPS: a database for protein sequences and complete genomes // *Nucl. Acids Res.* 1998. Vol. 26. P. 33–37.
- Rawling C.J.* Software directory for molecular biologists: A complete guide to the selection of computer software for management and analysis of molecular sequences. N.Y.: Stockton press, 1986. 412 p.
- Reynolds T.L.* Vector DNA artifacts in the nucleotide sequence database // *Biotechniques.* 1994. Vol. 16. P. 1124–1125.
- Roberts R.J., Macelis D.* REBASE – restriction enzymes and methylases // *Nucl. Acids Res.* 1998. Vol. 26. P. 338–350.
- Schroeder J.L., Blattner F.R.* Least-squares method for restriction mapping // *Gene.* 1978. Vol. 4. P. 167–174.
- Shapiro M.* DNAdraw: a program for preparing DNA and protein data for publication // *Biotechniques.* 1995. Vol. 18. P. 1064–1067.
- Shapiro M., Senapathy P.* Automated preparation of DNA sequences for publication // *Nucl. Acids Res.* 1986. Vol. 14. P. 65–73.
- Singh G.B., Krawetz S.A.* DNAView: a quality assessment tool for the visualization of large sequenced regions // *Comp. Appl. Biosci.* 1995. Vol. 11. P. 317–319.
- Staden R.* Sequence data handling by computer // *Nucl. Acids Res.* 1977. Vol. 4. P. 4037–4051.
- Staden R.* Further procedures for sequence analysis by computer // *Ibid.* 1978. Vol. 5. P. 1013–1015.
- Stoehr P.J., Omond R.A.* The EMBL network file server // *Ibid.* 1989. Vol. 17. P. 6763–6764.
- Stoesser G., Moseley M.A., Sleep J., McGowran M., Garcia-Pastor M., Sterk P.* The EMBL nucleotide sequence database // *Ibid.* 1998. Vol. 26. P. 8–15.
- Tateno Y., Fukami-Kobayashi K., Miyazaki S., Sugawara H., Gojobori T.* DNA data bank of Japan at work on genome sequence data // *Ibid.* 1998. Vol. 26. P. 16–20.
- Zhang J., Madden T.L.* PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and automation // *Genome Res.* 1997. Vol. 7. P. 649–656.

## Дополнительная литература

Гельфанд М.С. Компьютерный анализ последовательностей ДНК // Молекуляр. биология. 1998. Т. 32. С. 103–120.

Компьютерный анализ генетических текстов / А.А. Александров, Н.Н. Александров, М.Ю. Бородавский, Ю.А. Каламбет, А.Э. Кистер, А.А. Миронов, П.А. Певзнер, П.А. Шепелев. М.: Наука, 1990. 267 с.

Компьютерный анализ структуры, функции и эволюции генетических макромолекул // Ред. Н.А. Колчанов. Новосибирск, 1989. 282 с.

*Alphey L.* DNA sequencing. from experimental methods to bioinformatics. Berlin etc.: BIOS sci. publ., 1997. 224 p.

*Ansoerge W., Voss H., Zimmermann J.* DNA sequencing strategies: Automated and advanced approaches. N.Y.: Wiley; Spektrum, 1996.

NCBI news, 1992–1998. (Издаваемые на регулярной основе новости о работе GenBank.)

Nucleic Acids Research. 1982. Vol. 10, N 1 (special issue).

Nucleic Acids Research. 1984. Vol. 12, N 1, pt 1, 2 (special issues).

Nucleic Acids Research. 1986. Vol. 14. N 1 (special issue).

*Swindle S.R.* (ed.). Sequence data analysis guidebook. Totowa: Humana press, 1997. 324 p. (Meth Mol. Biol.; Vol. 70).

## СЕКВЕНИРОВАНИЕ ЦЕЛЫХ ГЕНОМОВ

Пожалуй, первыми из природных нуклеотидных последовательностей, ставших известными, были участки бактериофага лямбда протяженностью 12 пн, формирующие “липкие” концы или так называемые *coc*-сайты этого фага [Wu, Taylor, 1971]. Следующей вехой можно считать 50 нуклеотидов одноцепочечного фага *f1* [Sanger et al., 1973]. Разработка быстрых методов секвенирования ДНК сделало возможным определение уже довольно крупных молекул ДНК, какой является одноцепочечный бактериофаг *φX174*, имеющий размер 5386 нуклеотидов [Sanger et al., 1977]. Следующий рубеж был преодолен при определении нуклеотидной последовательности митохондриальной ДНК человека (16 569 пн) в 1981 г. [Anderson et al., 1981]. Серьезным успехом явилось определение полной нуклеотидной последовательности бактериофага *λ*, состоящей уже из 48 502 пн [Sanger et al., 1982]. Дальнейшее развитие методов секвенирования ДНК привело к появлению высокопроизводительных подходов, сделавших возможным определение нуклеотидных последовательностей еще более крупных фрагментов ДНК и даже целых геномов. Значительная роль в этом принадлежит автоматизация всего процесса секвенирования, начиная от приготовления матриц и заканчивая занесением определенных последовательностей ДНК в компьютер без непосредственного участия оператора.

До 1995 г. наиболее крупными геномами с известной последовательностью нуклеотидов были геном цитомегаловируса (229 тпн) [Bankier et al., 1991], геном вируса осповакцины (192 тпн) [Goebel et al., 1990], геном натуральной оспы (186 тпн) [Shchelkunov et al., 1993; Massung et al., 1993], митохондриальный (187 тпн) и хлоропластный (121 тпн) геномы мха *Marchantia polymorpha* [Oda et al., 1992], хлоропластный геном эвглены *Euglena gracilis* (143 тпн) [Hallick et al., 1993]. Однако все эти геномы принадлежат несвободноживущим организмам. Первый геном самостоятельно существующего организма (хотя и ведущего паразитический образ жизни), полная нуклеотидная последовательность которого стала известна, принадлежит бактерии *Haemophilus influenzae* и состоит из 1 830 137 пн [Fleischmann et al., 1995]. К настоящему времени определены нуклеотидные последовательности полных геномов еще 16 микроорганизмов, включая кишечную палочку *E. coli*, часть из которых кратко охарактеризована в ходе дальнейшего изложения. Еще более пяти десятков проектов по определению нуклеотидных последовательностей прокариотических и эукариотических геномов близки к завершению или активно выполняются. Кроме этого, стала известна нуклеотидная последовательность всего генома первого эукариотического организма, которым оказались дрожжи *Saccharomyces cerevisiae* [Goffeau et al., 1996, 1997]. Предполагается, что в декабре 1998 г. будет сообщено о завершении проекта по секвенированию гено-

ма нематоды *Caenorhabditis elegans*. Считается, что и выполнение проекта “Геном человека” перешло в свою решающую стадию [Collins et al., 1998]. Таким образом, можно сказать, что молекулярная биология через крупномасштабное секвенирование ДНК вступила в так называемую геномную эру.

### 11.1. ГЕНОМЫ ПРОКАРИОТИЧЕСКИХ ОРГАНИЗМОВ

В июле 1995 г. в американском журнале “Science” появилась статья, сообщающая об определении нуклеотидной последовательности полного генома самостоятельно существующего организма – бактерии *Haemophilus influenzae*, состоящего из 1 830 137 пн [Fleischmann et al., 1995]. В выполнении этого проекта приняло участие 40 ученых из 4 исследовательских центров США. Но основная нагрузка выпала на долю сотрудников Института геномных исследований (TIGR – The Institute for Genomic Research) и его директора Дж.К. Вентера. Весьма примечательным можно считать то, что второй руководитель проекта Г. Смит, являющийся лауреатом Нобелевской премии и получивший ее за открытие рестрикционных эндонуклеаз, обнаружил эти ферменты четверть века назад именно у данного вида микроорганизмов [Smith, Wilcox, 1970]. Так что бактерия *Haemophilus influenzae*, уже дважды открывшая для молекулярной биологии новые перспективы, может этим “гордиться”.

Основной стратегией определения нуклеотидной последовательности при выполнении проекта по секвенированию генома бактерии *Haemophilus influenzae*, RD штамма KW20 можно считать случайный подход с субклонированием фрагментов ДНК, разрушенных механическим путем. На разных стадиях выполнения проекта осуществлялось и секвенирование вставок, клонированных в  $\lambda$ -векторах. Заключительная ликвидация промежутков проводилась с помощью праймерной “прогулки”. В результате было достигнуто 6-кратное покрытие всего генома.

Проведенная компьютерная состыковка секвенированных участков позволила выявить, что полный геном бактерии состоит из 1 830 137 пн и характеризуется относительно низким содержанием GC-пар (38%), причем найдено 7 протяженных участков с более высоким (около 50%) содержанием GC-пар. Дальнейший анализ нуклеотидной последовательности привел к обнаружению предположительного ориджина репликации, состоящего из 280 пн, 6 оперонов рРНК, 54 генов тРНК для всех 20 аминокислот. На основании полученных данных авторами была составлена кольцевая карта хромосомы *Haemophilus influenzae*, состоящая из нескольких концентрических колец, несущих различную информацию о данном геноме. Из 1743 предполагаемых кодирующих регионов для 736 не удалось выявить их функции. 389 предсказанных кодирующих участков не показали значимой гомологии с известными на тот момент сведениями, содержащимися в ГенБанке и прочих базах данных. Таким образом, около 78% открытых рамок считы-

вания *Haemophilus influenzae* имели гомологию с обнаруженными ранее и представленными в базах данных подобными последовательностями других организмов, причем для 20% из этих 78% функциональная роль кодируемых ими гипотетических белков неизвестна.

В заключительном разделе цитируемой статьи авторы говорят о важных уроках, которые они извлекли при выполнении данного проекта и выражают уверенность, что это поможет в дальнейшем при секвенировании геномов других микроорганизмов. Доказательством этому явилось сообщение уже в октябрьском номере "Science" о секвенировании практически тем же коллективом авторов полного генома еще одной бактерии *Mycoplasma genitalum* [Fraser et al., 1995], причем для выполнения этого проекта потребовалось всего 4 мес.

Полная нуклеотидная последовательность генома *Mycoplasma genitalum* оказалась равной "всего" 580 070 пн. Изучение такого небольшого генома свободноживущего, хотя и ведущего паразитический образ жизни организма, представляло интерес в связи с возможностью выявления чуть ли не минимального набора генов, необходимых для самостоятельного существования. Генеральная стратегия секвенирования ДНК этого объекта принципиально не отличалась от таковой при секвенировании генома *Haemophilus influenzae*. Было определено в общей сложности 3 806 280 нуклеотидов в виде первичных данных, что составило приблизительно 6,5-кратное покрытие полного генома.

Дальнейший анализ показал, что нуклеотидный состав генома *Mycoplasma genitalum* варьирует от 27 до 37% GC-пар со средней величиной, равной 34%. Самое низкое содержание GC-пар, как и у *Haemophilus influenzae*, отмечено в области предполагаемого ориджина репликации, а самое высокое характерно для генов рРНК (44%) и тРНК (52%), что легко объясняется присущей этим молекулам сильной вторичной структурой. Выявление открытых рамок считывания позволило обнаружить 470 регионов, которые могли бы кодировать белки. 96 регионов не обнаружили при этом заметной гомологии с другими последовательностями, находящимися в базах данных, что не позволило приписать им какую-либо функцию. Оставшиеся 374 оказались возможным идентифицировать, исходя из имеющейся гомологии и разделить на условные ферментативные группы, отвечающие за те или иные процессы в жизнедеятельности клетки.

Сравнительная геномика этих двух микроорганизмов – *Haemophilus influenzae* и *Mycoplasma genitalum* – показала некоторую разницу между ними в доле тех или иных генов, кодирующих белки с определенной биологической ролью, от всего генома или, правильнее сказать, от общего числа обнаруженных открытых рамок считывания. Так, для *Mycoplasma genitalum* гены с биологической ролью в виде репликации, транскрипции и трансляции составляют несколько большую долю по сравнению с таковыми у *Haemophilus influenzae*. Что касается таких функциональных категорий, как биосинтез аминокислот, метаболизм жирных кислот и фосфолипидов, биосинтез кофакторов и некоторых других, то можно видеть, что гены, относящиеся к этим группам, в геноме *Mycoplasma genitalum* представлены очень скупо. Причиной этому

является как раз довольно малый размер генома и вынужденная “экономия” своей ДНК. Это же объяснение может быть распространено и на то, что доля потенциальных белковых продуктов с невыясненной ролью у *Haemophilus influenzae* составляет 43% от всех теоретически выявленных, а у *Mycoplasma genitalum* – только 32%.

Очередным геномом, нуклеотидная последовательность которого стала известна благодаря, в основном, той же группе авторов, стал геном археобактерии *Methanococcus jannaschii* [Bult et al., 1996]. Стратегия его секвенирования также принципиально не отличалась от задействованных при секвенировании двух предыдущих геномов *Haemophilus influenzae* и *Mycoplasma genitalum* и лишь покрытие генома было приблизительно 10-кратным. Особенностью данного генома можно считать то, что он состоит из трех частей, кольцевой хромосомы (1 664 976 пн), большого экстрахромосомного элемента (58 407 пн) и малого экстрахромосомного элемента (16 550 пн). Главная составная часть генома в виде кольцевой хромосомы имеет средний GC-состав около 31% и содержит 1682 региона, для которых обнаружены открытые рамки считывания. GC-состав экстрахромосомных элементов незначительно различается и составляет 28,2% для большого и 28,8% для малого. В большом экстрахромосомном элементе найдено 44 потенциальных региона, кодирующих белки, тогда как для малого их выявлено всего 12. Дальнейший анализ всех этих потенциальных белков показал, что только для 38% из них в базах данных обнаруживаются гомологичные последовательности с известной функцией и еще для 6% имеется заметная гомология с некими гипотетическими белками. Следует отметить, что для ранее секвенированных геномов *Haemophilus influenzae* и *Mycoplasma genitalum* характерны другие соотношения, причем 83% потенциальных белков *Mycoplasma genitalum* имело своих гомологов в геноме *Haemophilus influenzae*. Для предсказанных белков *Methanococcus jannaschii* характерно лишь 11%-ное совпадение с белками *Haemophilus influenzae* и 17%-ное с *Mycoplasma genitalum*. Таким образом, для большинства генов и потенциально кодируемых ими белков *Methanococcus jannaschii* не выявлена их функция, исходя из гомологии последовательностей. В геноме *Methanococcus jannaschii* обнаружено два оперона рРНК и 37 генов тРНК, причем некоторой особенностью последних является то, что почти все аминокислоты, кодируемые парами триплетов, у данной бактерии представлены только одним типом тРНК для каждой из них. Исключение составляет лишь глутаминовая кислота, для которой в геноме присутствуют оба типа тРНК.

Однако следует особо подчеркнуть, что у археобактерии *Methanococcus jannaschii* для части генов с известными функциями, как, например, отвечающими за клеточное деление, метаболизм, продукцию энергии, более высокая гомология обнаруживается с прокариотическими организмами. Тогда как гены, вовлеченные в процессы репликации, транскрипции и трансляции, более подобны таковым эукариот. Это может лишним раз свидетельствовать, что археобактерии имеют с эукариотами общую эволюционную траекторию, независимую от эволюции эубактерий.

В связи с обнаруженными в результате этих работ серьезными различиями между эубактериями и архебактериями теперь и на геномном уровне особое значение приобрело определение полной нуклеотидной последовательности другого представителя археобактерий – *Methanobacterium thermoautotrophicum* [Smith et al., 1997]. В выполнении этого проекта принимала участие большая группа ученых из четырех исследовательских центров США, возглавляемая Д.Р. Смитом.

Стратегия секвенирования полного генома этой бактерии *Methanobacterium thermoautotrophicum* ДН заметно отличалась от применяемых сотрудниками Дж.К. Вентера и Г. Смита. Так, нуклеотидная последовательность значительной части генома была определена с помощью мультиплексного секвенирования, причем как ферментативным методом, так и методом химической дегградации. В общей сложности было секвенировано около 14,8 мпн, что соответствовало 8,5-кратному покрытию всего генома, который в итоге составил 1 751 377 пн и содержал около 49,5% GC-пар. Проведенный анализ этой последовательности нуклеотидов позволил выявить 1855 открытых рамок считывания. Для 844 из них (46%) была идентифицирована предполагаемая функция на основе аналогичных последовательностей, хранящихся в различных базах данных. Кроме этого, 514 (28%) регионов соответствовали гипотетическим белкам с неизвестной функцией, но представленным в банках данных. Оставшиеся 496 (27%) или вообще не имели какой-либо гомологии, или характеризовались ее низким уровнем с известными последовательностями, что также не позволило приписать им какую-либо функциональную роль. В геноме *Methanobacterium thermoautotrophicum* обнаружено 2 оперона рРНК и 39 генов тРНК, причем для этой бактерии, как и для *Methanococcus jannaschii*, характерно существование только по одному типу тРНК для аминокислот, кодируемых парами триплетов. Исключение также составляет лишь глутаминовая кислота, для которой в геноме *Methanobacterium thermoautotrophicum* присутствуют оба типа тРНК.

Проведенный сравнительный анализ предсказанных белков *Methanobacterium thermoautotrophicum* с известными последовательностями в специальных базах данных по эубактериям, архебактериям и эукариотическим организмам показал, что 54% предположительных генных продуктов этой архебактерии показывают все же более высокую гомологию именно с архебактериями. Это было бы совсем не удивительно, если бы не факт довольно низкой гомологии между двумя представителями метаносбраживающих археобактерий *Methanobacterium thermoautotrophicum* и *Methanococcus jannaschii*. Так, только 352 (19%) предсказанных белковых продукта *Methanobacterium thermoautotrophicum* обнаруживаются и у *Methanococcus jannaschii*. Анализ предсказанных белков *Methanobacterium thermoautotrophicum* показал, что 42% из них все же ближе к эубактериальным и только 13% имеют более высокую гомологию с эукариотическими. Причем, как и в случае *Methanococcus jannaschii*, к последним относятся ферменты и прочие белки, вовлеченные в метаболизм ДНК, транскрипцию и трансля-

цию. К бактериальному же типу принадлежат белковые продукты, принимающие участие в общем метаболизме клеток.

Пожалуй, наиболее долгожданным из прокариотических геномов, чья полная нуклеотидная последовательность была определена, явился геном кишечной палочки *Escherichia coli*. Поскольку этот микроорганизм был наиболее широко используемым и исследуемым объектом в молекулярной генетике и биологии, то можно было предполагать, что уж его-то геном будет секвенирован в первую очередь. Более того, к моменту только начала выполнения проектов по секвенированию геномов некоторых бактерий для *E. coli* было уже известно около 800 тысяч нуклеотидных пар. Однако весьма значительный размер генома данной бактерии (около 4,7 млн) позволил ей стать только шестым микроорганизмом с полностью известным геномом, определение нуклеотидной последовательности которого в рамках специального проекта заняло около 6 лет [Blattner et al., 1997].

4 639 221 пн генома *E. coli* K-12 штамма MG1655 была определена с помощью различных стратегий секвенирования, главной из которых было клонирование в специализированном векторе M13 Janus, созданном в лаборатории генетики Висконсинского университета, где преимущественно и выполнялся данный проект [Blattner et al., 1997]. Кроме этого, еще три исследовательских центра США и один Мексики приняли участие в осуществлении этого проекта. Так, на первом этапе было достигнуто 4–5-кратное покрытие генома за счет секвенирования случайных субклонов в фаговом векторе M13 Janus. После компьютерной состыковки “прочитанных” последовательностей применялось секвенирование недостающих клонов в фаге лямбда, а также ограниченная праймерная “прогулка”.

Содержание GC-пар в геноме *E. coli* составило 50,8%. Проведенный компьютерный анализ выявил 4288 открытых рамок считывания, принадлежащих как действительно существующим белоккодирующим генам, так и только гипотетическим. Дальнейший анализ показал, что для 38% из них пока не может быть определена функциональная роль. Для реальных и потенциальных белков стартовыми кодонами являются ATG (3542 случая), GTG (612), TTG (130). Возможно еще есть один CTC и один CTG кодоны в качестве стартовых. Частоты встречаемости терминирующих кодонов также заметно различаются: TAA (2705 раз), TGA (1257), TAG (326). 405 генов показывают перекрытие стартового и терминирующего кодонов: ATGA (224 случая), TAATG (98), TGATG (48), GTGA (28), TAGTG (4), TTGA (3). Было проведено сравнение белков (открытых рамок считывания) *E. coli* с аналогичными *Haemophilus influenzae*, причем в расчет принималось только совпадение по крайней мере 60% всего белка с 30%-ной идентичностью аминокислот. В результате оказалось, что 1130 потенциальных белков *Haemophilus influenzae* из 1743 возможных совпадают с таковыми кишечной палочки. Привлеченные в анализ полные геномы других бактерий *Synechocystis* sp. и *Mycoplasma genitalum* показали несколько худшее совпадение. Следует отметить, что 111 генов были присущи всем сравниваемым организмам. В то же время геномы организмов, относящих-



ся к другим группам археобактерий и эукариот (*Methanococcus jannaschii* и *Saccharomyces cerevisiae* соответственно), имели с *E. coli* 231 и 254 сходных белка. Однако для всех этих геномов, принадлежащих разным организмам (*E. coli*, *Haemophilus influenzae*, *Mycoplasma genitalum*, *Synechocystis* sp., *Methanococcus jannaschii*, *Saccharomyces cerevisiae*), характерно только 16 общих белков с заданным при данном сравнении уровнем гомологии.

Выявленный ориджин репликации и место его терминации разделили геном *E. coli* на две реплицирующиеся в противоположных направлениях половины, названных авторами реплихорами. Оказалось, что около 55% генов расположены в направлении, совпадающем с репликацией. Так, все семь оперонов рРНК и 53 из 86 генов тРНК транскрибируются в направлении репликации. Давно известна чрезвычайно редкая встречаемость в ДНК *E. coli* сайтов узнавания рестрикционной эндонуклеазы *MaeI* (СТАГ), причины которой так до конца и не ясны, хотя в литературе высказывались различные предположения. Полная нуклеотидная последовательность генома *E. coli* позволила подтвердить это с помощью компьютера и оказалось, что количество таких сайтов составляет всего 886 против 18101 теоретически ожидаемого. Как отмечают авторы цитируемой статьи, еще предстоит тщательный анализ нуклеотидной последовательности потенциальных генов и кодируемых ими белков *E. coli* и сопоставление этих результатов с известными биохимическими данными.

О начале секвенирования полного генома другого штамма *Escherichia coli* K-12 W3110 и об объединении для этой цели нескольких групп исследователей из Японии и США было объявлено в конце 1992 г. [Kasai et al., 1992]. В настоящее время японскими учеными полностью завершено секвенирование участка генома 0–68,8 мин. Совместно с банком данных DDBJ на основе секвенированного ими генома *Escherichia coli* разрабатывается информационная интерактивная система, охватывающая различные аспекты геномики этой бактерии и доступная посредством Интернета из web-сайта – <http://mol.genes.nig.ac.jp/ecoli/> [Tateno et al., 1998]. При этом отмечается, что данная система может послужить прообразом для представления в сети WWW прочих полных геномов организмов.

Секвенирование полного генома грамположительного микроорганизма – почвенной бактерии *Bacillus subtilis*, штамма 168 – было осуществлено усилиями международного консорциума, в который входили 25 европейских, 7 японских и 1 лаборатория из Южной Кореи [Kunst et al., 1997]. Каждая исследовательская группа была вольна в выборе той или иной стратегии секвенирования, однако общим требованием было секвенирование обеих цепей ДНК. Размер генома *Bacillus subtilis* оказался довольно большим и составил 4 214 810 пн. Усредненный GC-состав ДНК генома равен 43,5%, однако, если принимать в расчет кодирующие последовательности ДНК по одной цепи, то соотношения нуклеотидов будут весьма различаться между собой (G – 24%, A – 30%, C – 20%, T – 26%). При анализе генома были обнаружены открытые рамки считывания для приблизительно 4100 потенциальных белков, причем

следует отметить, что около четверти из них организованы в генные семейства. Самым большим таким семейством является группа из 77 генов, кодирующих белки-транспортёры, что вполне объяснимо, учитывая особенности организации клеточной стенки у грамположительных бактерий. Еще одно крупное семейство генов кодирует белки, секретирующиеся в культуральную среду, что также характерно для данной группы бактерий. Стартовым кодоном для 78% генов служит ATG, для 13% – TTG и для 9% – GTG, что несколько отличается от обнаруженных у кишечной палочки. У бациллы выявлены также и необычные стартовые кодоны для нескольких белков – ATT и CTG.

Значительный интерес представляло сравнение геномов двух бактерий *Bacillus subtilis* и *E. coli* даже по причине сходства их размеров. Так, в цитируемой работе было определено, что около 1000 генов бациллы имеют своих ортологов в геноме кишечной палочки, причем некоторые из них характеризуются иным порядком генов в едином опероне. Так, например, арабинозный оперон в геноме бациллы организован как *araABD*, тогда как у *E. coli* он представлен в виде *araBAD*. Сравнительный анализ большого генома *Bacillus subtilis* и “крохотного” генома *Mycoplasma genitalum* позволил сделать интересные заключения, поскольку считается, что микоплазмы произошли от грамположительных бактерий. Действительно, 300 генных продуктов микоплазмы совпадают с таковыми у бациллы. Из оставшихся 146 открытых рамок считывания у микоплазмы 3 совпадают с белками других видов бацилл, еще 9 – с белками других грамположительных бактерий, 25 – подобны белкам разных грамотрицательных бактерий, 19 – с белками других видов микоплазм. Таким образом, можно предполагать, что только 90 генов у *Mycoplasma genitalum* кодируют белки, вовлеченные во взаимодействие микроорганизма и его хозяина. Как видно из приведенного примера, сравнительная геномика уже сейчас позволяет выявлять определенные гены какого-либо организма (пока микроорганизма), кодирующие белки, ответственные за те или иные особенности его жизнедеятельности.

Еще одним грамположительным организмом, геном которого секвенирован, стала бактерия *Mycobacterium tuberculosis* [Cole et al., 1998]. Геном *Mycobacterium tuberculosis* довольно крупный (4 411 529 пн) и превосходит по размеру геном *Bacillus subtilis* и лишь немного уступает геному кишечной палочки. К особенностям данного генома можно отнести довольно высокое содержание GC-пар, равное 65,6%. Кроме этого, в геноме этой бактерии обнаружена повторяющаяся ДНК, мультигенные семейства, дублицированные гены. Число открытых рамок считывания насчитывает 3924. В 35% случаев стартовым кодоном является GTG, что объясняется высоким GC-составом ДНК этой бактерии.

Из других бактериальных геномов, полные нуклеотидные последовательности которых известны, можно отметить небольшой геном (816 394 пн) *Mycoplasma pneumoniae*, содержащий информацию о 677 открытых рамках считывания, причем для 75,9% из них обнаруживаются гомологичные последовательности в банках данных [Himmelreich et al., 1996]. Геном несколько большего размера характерен для патогенной бактерии *Borrelia burgdorferi* B31, содержащей линейную хромосому

(910 725 пн), а также серию плазмид общей протяженностью около 533 тпн [Fraser et al., 1997]. Определена полная нуклеотидная последовательность генома одноклеточной цианобактерии *Synechocystis* sp. штамма PCC6803, составившая 3 573 470 пн [Kaneko et al., 1996]. Компьютерным анализом найдены потенциальные открытые рамки считывания для 3168 белков, из которых 145 были идентичны известным ранее генам. Более 50% белковых продуктов показали определенную гомологию с последовательностями белков, занесенными в банки данных, как с известными функциями (1257), так и с гипотетическими (340) соответственно. Оставшиеся 45% не имели значимой гомологии с последовательностями в банках данных.

Секвенирован геном бактерии *Helicobacter pylori* штамма 26 695, состоящий из 1 667 867 пн [Tomb et al., 1997]. К некоторым особенностям белков, кодируемых этим микроорганизмом, ответственного, как считается, за хронический гастрит, можно отнести повышенное содержание основных аминокислот (аргинина, лизина), что, видимо, является следствием его приспособления к кислотности желудочного сока. Определена нуклеотидная последовательность всего генома еще одного опасного для человека микроорганизма *Treponema pallidum*, являющегося возбудителем сифилиса [Fraser et al., 1998]. Его размер оказался равен 1 138 006 пн с довольно высоким GC-составом (52,8%). Обнаружена 1041 открытая рамка считывания и для 287 из них (28%) не обнаружено гомологичных белковых продуктов в банках данных. Следует отметить наличие в относительно небольшом геноме бледной спирохеты *Treponema pallidum* 61 типа (всех возможных) генов тРНК. Недавно сообщено о завершении секвенирования полного генома другого опасного микроорганизма *Rickettsia prowazekii*, являющегося внутриклеточным паразитом и вызывающего сыпной тиф [Andersson et al., 1998]. Его геном состоит из 1 111 523 пн и несет информацию о 834 белковых продуктах. Своеобразной чертой этого генома можно считать высокую долю некодирующих последовательностей (около 24%), что не характерно для прочих микроорганизмов. Так, для генома *Haemophilus influenzae* этот показатель составляет 13%, а для генома *Aquifex aeolicum* [Deckert et al., 1998] – он еще меньше (6%). Другой интересной особенностью генома *Rickettsia prowazekii* является высокая гомология значительной части его генов и кодируемых ими белков с митохондриальными. Секвенированный геном гипертермофильной бактерии *Aquifex aeolicum*, являющейся облигатным хемолитоавтотрофом, состоит из 1 551 335 пн с GC-составом 43,4% [Deckert et al., 1998]. 93% генома кодирует какие-либо белки. Так, 849 открытых рамок считывания гомологичны белкам с известной функцией, 256 – белкам с неизвестной функцией и для оставшихся 407 в банках данных не обнаружено гомологичных последовательностей. Кроме главной хромосомы, у данной бактерии имеется еще экстрахромосомный элемент протяженностью 39 456 пн с более низким GC-составом (36,4%). Он кодирует только 1 белок с известной функцией, 4 белка с неизвестными функциями, но представленными в банках данных и 27 гипотетических белков, не выявленных у других организмов ранее.

Небольшая плазмида, имеющая размер 7493 пн, характерна и для генома другой бактерии *Chlamydia trachomatis*, состоящего из 1 042 5119 пн и содержащего 41,3% GC-пар [Stephens et al., 1998]. Геном этой бактерии был секвенирован с помощью клонирования случайных фрагментов в векторе на основе фага M13, что привело приблизительно к 10-кратному покрытию. Анализ открытых рамок считывания позволил выявить белковые продукты, потенциально связанные с вирулентностью. 255 потенциальных белков (28%) не обнаружили гомологии с ранее секвенированными последовательностями, представленными в банках данных. Особенностью генома этой бактерии является наличие ряда генов, показывающих заметную гомологию с эукариотическими, причем наиболее удивительным является то, что они оказались более близки генам представителей царства растений и только затем – крысе и человеку. Подобные гены бактерий *E. coli*, *Haemophilus influenzae* и *Helicobacter pylori* оказались еще более удаленными. В то же время у *Chlamydia trachomatis* выявлено 58 семейств паралогичных генов, что сравнимо с количеством таковых у прочих микроорганизмов с большими геномами.

Большинство микроорганизмов, геномы которых оказались секвенированы, характеризуются относительно низким содержанием GC-пар, достигающим у *Rickettsia prowazekii* даже до 29,1%. Некоторым исключением является спирохета *Treponema pallidum*, геном которой содержит более 50% GC-пар (52,8%). Геном *E. coli* несет приблизительно равное соотношение GC-пар и AT-пар – 50,8%, однако авторы отмечают заметный уровень компрессии, мешавший эффективному секвенированию ДНК этой бактерии. Самый высокий GC-состав (65,6%) среди всех секвенированных геномов характерен для грамположительной бактерии *Mycobacterium tuberculosis*, что вызывало высокий уровень компрессии при секвенировании ДНК. Также большие трудности поджидали экспериментаторов при секвенировании симбиотической мегаплазмиды клубеньковой азотфиксирующей бактерии *Rhizobium* sp., штамма NGR234 [Freiberg et al., 1996]. Причиной тому было высокое содержание GC-пар у этого микроорганизма, варьирующее от 59 до 65% и приводящее к значительной компрессии (до 400 случаев на среднюю космидную вставку). После завершения секвенирования ДНК этой крупной плазмиды (536 165 пн), лишь немного уступающей целому геному *Mycoplasma genitalium* (580 070 пн), оказалось, что усредненное содержание GC-пар в данной плазмиде составляет 58,5%, тогда как полный геном этой бактерии *Rhizobium* sp. характеризуется еще более высоким GC-составом (62,2%), определенным пока только косвенными методами [Freiberg et al., 1997].

Секвенирован геном еще одной археобактерии *Archaeoglobus fulgidus*, размер которого составил 2 178 400 пн [Klenk et al., 1997]. В его последовательности найдено 2436 открытых рамок считывания, четвертая часть из которых приходится на не охарактеризованные еще белки и, в свою очередь, две трети которых совпадают с также не охарактеризованными гипотетическими белками *Methanococcus jannaschii*. Другая археобактерия *Puugococcus horikoshii* штамм OT3 имеет несколько

меньший размер генома, состоящий из 1 738 505 пн с GC-составом 42,0% [Kawarabayasi et al., 1998]. Проведенный поиск выявил 2002 открытые рамки считывания протяженностью более 100 аминокислотных остатков, 44,4% которых кодировали гипотетические белки, не обнаруженные ранее у других организмов.

В настоящее время около 50 проектов по секвенированию полных геномов прокариотических организмов находятся или на стадии завершения или активно выполняются. Следует отметить, что значительная часть бактерий, полная нуклеотидная последовательность геномов которых уже определена (*Chlamydia trachomatis*, *Borrelia burgdorferi*, *Helicobacter pylori*, *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae*, *Mycoplasma genitalum*, *Rickettsia prowazekii*, *Treponema pallidum*), а также и тех, что еще только секвенируются (*Campylobacter jejuni*, *Chlamydia pneumoniae*, *Clostridium acetobutylicum*, *Desulfovibrio vulgaris*, *Enterococcus faecalis*, *Francisella tularensis*, *Lactobacillus acidophilus*, *Legionella pneumophila*, *Mycobacterium leprae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, *Salmonella typhimurium*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Treponema denticola*, *Ureaplasma urealyticum*, *Vibrio cholerae*, *Yersinia pestis*), относится к возбудителям опасных заболеваний человека. Так, знание нуклеотидной последовательности всего генома этих микроорганизмов может оказать большую помощь в борьбе с данными инфекциями. В процессе выполнения находятся и проекты, направленные на выяснение полных геномов археобактерий: *Halobacterium salinarium*, *Methanosarcina mazei*, *Pyrococcus furiosus*, *Sulfolobus solfataricus*, *Thermoplasma acidophilum* и некоторых других.

Значительное внимание, уделенное в данной главе именно геномам прокариотических организмов, объясняется завершенностью большого числа этих проектов, их изученностью и возможностью сравнительного анализа. Другой причиной служит то серьезное влияние, которое оказала расшифровка первого полного бактериального генома *Haemophilus influenzae* не только на молекулярную биологию, но и на микробиологию и, как отмечается в одной статье, привела к новому пониманию прокариотического мира [Tang et al., 1997].

## 11.2. ГЕНОМЫ ЭУКАРИОТИЧЕСКИХ ОРГАНИЗМОВ

Первым геномом эукариотического организма, полная последовательность которого определена, стал геном дрожжей *Saccharomyces cerevisiae* [Goffeau et al., 1996]. Это явилось результатом усилий огромного числа ученых и лаборантов, работающих в лабораториях и крупных исследовательских центрах стран Западной Европы, а также в США и Японии. При этом следует отметить, что это был первый, стартовавший еще в 1989 г., проект по секвенированию полного генома свободноживущего организма, опередивший в этом отношении многие

другие аналогичные проекты [Thomas, 1996]. В результате выполнения этого проекта было определено 12 млн 68 тпн, хотя весь полный геном дрожжей *Saccharomyces cerevisiae*, сосредоточенный по 16 хромосомам, оценивается в 13 389 тпн. Таким образом, более 1 млн, приходящихся на повторяющиеся элементы в нескольких хромосомах и, главным образом, на повторы рДНК, расположенные в виде отдельного кластера на XII хромосоме, остались не секвенированными. Однако это не служит каким-то препятствием для дальнейшего анализа устройства и функционирования всего генома дрожжей и даже позволяет считать, что в результате выполнения этого проекта стала известна все-таки практически полная его последовательность. Более того, оказалось возможным определить число всех этих несеквенированных повторов. Так, например, было подсчитано, что кластер генов рРНК состоит приблизительно из 140 повторов и локализован на XII хромосоме. Гены малых ядерных РНК, напротив, разбросаны по многим хромосомам. Аналогичное распределение по разным хромосомам характерно и для 275 генов тРНК, относящихся к 43 семействам. К другим особенностям организации хромосом дрожжей можно отнести участки размером по 31 тпн, расположенные по краям самой малой хромосомы I и практически не содержащие генных последовательностей. Для многих хромосом характерно наличие протяженных или GC-богатых, или, наоборот, GC-обедненных участков, коррелирующих в некоторой степени с распределением генных последовательностей.

Проведенный анализ нуклеотидной последовательности дрожжей *Saccharomyces cerevisiae* позволил выявить 6275 открытых рамок считывания, но, как считают авторы цитируемой работы, не похоже, что 390 из них действительно транслируются в какие-либо белки. Таким образом, число открытых рамок считывания для генома дрожжей принято равным 5885. Относительно небольшой размер генома дрожжей *Saccharomyces cerevisiae* в некоторой степени можно объяснить как малым размером самих интронов, так и их незначительным числом, поскольку интроны содержат только около 4% генов. Следует отметить, что больший размер генома (около 15 млн) других дрожжей *Schizosaccharomyces pombe*, секвенирование которого также осуществляется в рамках специального проекта, обусловлен в значительной степени наличием протяженных интронов в 40% генов этого организма. Для 50% предполагаемых белковых продуктов из всего “протеома” были выявлены их принадлежности к той или иной функциональной группе белков. (Использованный авторами цитируемой работы термин “протеом”, имеющий определенную аналогию с “геномом”, и обозначающий всю совокупность продуцируемых организмом белков, достаточно емко и лаконично). Так, 11% белков протеома дрожжей отвечают за метаболизм клетки, 3% – за репликацию, репарацию и рекомбинацию ДНК, 7% и 6% – за транскрипцию и трансляцию соответственно. Как подчеркивают авторы, необходимы дальнейшие усилия по выяснению точной роли белковых продуктов с привлечением в том числе методов биохимии и генетики. Уже стартовали проекты, посвященные функциональному анализу дрожжевого генома.

Из других проектов по секвенированию эукариотических геномов необходимо отметить завершающееся в декабре 1998 г. секвенирование 100-мегабайтного генома нематоды *Caenorhabditis elegans*. Секвенирование генома другого организма, ожидаемого не меньше *E. coli*, плодовой мушки *Drosophila melanogaster* с ее геномом, равным приблизительно 130 млн, находится сейчас на начальных стадиях и его ориентировочное завершение предполагается, исходя из реалий сегодняшнего дня, в 2002 г. Фронтальное секвенирование генома мыши (с предположительным завершением к 2005 г.) еще не начато и идет лишь подготовительная работа. Уже выполняются проекты по секвенированию геномов довольно просто устроенных организмов: *Candida albicans*, *Dictyostelium discoideum*, *Giardia lamblia*, *Plasmodium falciparum*, *Trypanosoma rhodesiense* и некоторых других. Кроме этого, проводится секвенирование генома рыб, собаки и некоторых других представителей царства животных.

Среди объектов растительного царства ведется секвенирование двух модельных растений из класса однодольных и двудольных – риса (*Oryza sativa*) и резуховидки Таля, или арабидопсиса, (*Arabidopsis thaliana*) соответственно, а также некоторых других растений. Выбор этих модельных растений (рис и арабидопсис) связан, главным образом, с относительно малыми размерами их геномов [Gale, Devos, 1998]. Так, недавно было сообщено о секвенировании двух протяженных участков хромосомы 5 (1,6 млн) и хромосомы 4 (1,9 млн) арабидопсиса *Arabidopsis thaliana* [Sato et al., 1997; Bevan et al., 1998]. По состоянию на конец октября 1998 г. в общей сложности секвенировано и сопровождено аннотацией около 30 млн арабидопсисного генома [Meinke et al., 1998]. Вообще, предполагается, что геном арабидопсиса содержит около 120 млн, кодирующих приблизительно 20000 генов, распределенных по 5 хромосомам. В выполнении данного проекта принимают участие 25 исследовательских групп (17 из Европы, 7 – из США и 1 – из Японии). Его завершение запланировано на конец 2000 г. [Meinke et al., 1998]. Геном риса значительно крупнее и составляет около 430 млн, однако, учитывая важность этого злака для человечества, предполагается, что усилиями исследовательских групп из Европы, Китая, Японии, Южной Кореи, США и других стран его секвенирование будет завершено к 2005 г. При этом следует отметить, что геном риса уже картирован с высоким разрешением [Hong et al., 1997; Kurata et al., 1997].

В литературе появилось сообщение о принятом решении объединить усилия ученых нескольких институтов из Канады, Австралии и Германии для секвенирования мельчайших эукариотических геномов (380 тпн и 600 тпн), каковыми являются нуклеоморфы из двух групп водорослей [McFadden et al., 1997]. Нуклеоморфы представляют собой остатки ядер прежних свободноживущих эукариотических водорослей, поглощенные другими водорослями и существующие как бы полуавтономно. В то же время они сохранили все основные черты организации генетического материала, присущие эукариотам, и определение полных нуклеотидных последовательностей этих двух геномов может оказать значительную пользу при выявлении функциональной роли протеома более высокоорганизованных эукариот из царства растений.

### 11.3. ГЕНОМ ЧЕЛОВЕКА

Первые достаточно серьезные разговоры о необходимости определения нуклеотидной последовательности всех 24 хромосом (22 аутосом и половых X и Y хромосом) человека стали вестись в США в 1984–1986 гг. [Cantor, 1990; Watson, 1990]. И в 1988 г. был создан Национальный центр по исследованию генома человека, который возглавил Нобелевский лауреат Дж. Уотсон. В 1990 г. Национальный институт здоровья и Департамент энергии объединили свои финансовые возможности и разработали первый пятилетний план совместных исследований по проекту “Геном человека”, однако значительный прогресс в исследовании ДНК и изменившаяся за счет этого ситуация заставили в 1993 г. его пересмотреть и принять второй пятилетний план, выполнение которого завершилось в сентябре 1998 г. [Collins, Galas, 1993; Collins et al., 1998]. Новый пятилетний план на период 1998–2003 гг. ставит задачей завершение секвенирования генома человека к концу 2003 г. [Collins et al., 1998]. При этом отмечается, что задачи, стоявшие перед исследователями в 1993–1998 гг., полностью решены. Так, построена генетическая карта с разрешением 1 сантиморган вместо предполагавшихся 2–5 сантиморган. Вместо планировавшихся 80 млн к концу 1998 г. по всем проектам, связанным с секвенированием геномов, уже только для генома человека определено около 180 млн и еще 111 млн прочих организмов. Подобные успехи позволили предположить, что треть генома человека будет известна уже к концу 2001 г. и к этому же сроку будет подготовлен черновой вариант (с точностью, по крайней мере, 99%) оставшейся части с тем расчетом, чтобы к 2003 г. (вместо планировавшегося когда-то 2005 г.) геном будет полностью уточнен. 2003 г. станет юбилейным годом для молекулярной биологии, поскольку к этому времени пройдет 50 лет с момента исторического открытия Дж. Уотсоном и Ф. Криком двойной спирали ДНК [Watson, Crick, 1953a, b], и лучшего подарка, чем последовательность ДНК всего генома человека, для молекулярных биологов трудно вообразить, что и является дополнительной причиной установки этой даты [Collins et al., 1998]. Хотя это является весьма трудной задачей, поскольку в октябре 1998 г. определено только около 6% от всего генома. Достигнутая сейчас производительность секвенирования генома человека составляет около 90 млн в год, а в течение последующих пяти лет она должна быть доведена до 500 млн/год за счет улучшения существующих и разработки новых технологий секвенирования. Предполагается, что американскими участниками проекта к моменту завершения будет сделан вклад в пределах 60–70% от всего генома, тогда как на долю Сэнгеровского центра в Англии и других иностранных партнеров придется остальная часть. Здесь следует отметить, что проект, начавшись как американский, практически сразу стал международным, поскольку многие страны, включая СССР, изъявили желание принять участие в его выполнении. Причем Советский Союз стал второй страной после США, где в 1988 г. по инициативе выдающегося ученого академика А.А. Баева была сформирована Государственная программа “Геном человека”, воз-



главляемая им вплоть до самой его кончины (1994 г.). В настоящее время Научный совет РАН по проблеме “Геном человека” возглавляет известный ученый член-корреспондент РАН, профессор Л.Л. Киселев, являющийся также одним из вице-президентов международной организации по изучению генома человека HUGO (HUMAN Genome Organisation). HUGO, в состав Совета которого входят около двадцати крупнейших ученых всего мира, играет важную роль, координируя научные усилия различных групп ученых, работающих по этой проблеме, и разрабатывая общую стратегию исследований. HUGO также выпускает специальный бюллетень “Genome Digest” и организует международные конференции по геному человека, ставшие в последнее время ежегодными.

Важными составными частями проекта “Геном Человека” является создание баз данных по STS (sequence-tagged site) и EST (expressed sequence tag). Первые необходимы для картирования человеческого генома [Olson et al., 1989], и по состоянию на октябрь 1998 г. их уже выявлено около 52000 [Collins et al., 1998]. Вторые используются для идентификации генов и представляют собой короткие (обычно 150–300 нуклеотидов) участки кДНК, секвенированные не с очень высокой точностью [Adams et al., 1991]. Их уже известно около 30000 [Collins et al., 1998]. Предполагается, что к моменту завершения определения геномной последовательности ДНК человека будут секвенированы полноразмерные молекулы кДНК. К новым целям, которые планируется достигнуть в рамках проекта “Геном человека” и не ставящимся ранее на период 1993–1998 гг., можно отнести функциональный анализ генома и изучение разнообразия человеческих геномов, что уже называют вторым проектом “Геном человека” [Brower, 1998]. Ясно, что существует заметный полиморфизм геномов разных людей, так называемый SNP (Single-Nucleotide Polymorphism), или однонуклеотидный полиморфизм, и считается, что единичные замены нуклеотидов имеются в среднем в каждой тысяче пар нуклеотидов. К другим типам вариаций, исследования которых также являются одной из целей принятого пятилетнего плана, относятся делеции, инсерции, дубликации и прочие перестройки генома с низкой частотой встречаемости.

Нельзя не отметить серьезное отличие данного проекта от прочих проектов, также направленных на секвенирование полных геномов каких-либо организмов, связанное с возможными социальными последствиями. Обеспокоенность всего мирового сообщества вызвана определенными опасениями из-за существующей возможности использования полученной в ходе выполнения этого проекта информации для ущемления прав отдельного человека или целых рас, а также для человеконенавистнических целей. Серьезное опасение также представляет собой возможность направленного вмешательства в геном человека. Все это заставило Организацию Объединенных Наций в ноябре 1997 г. провозгласить ряд основополагающих принципов, регламентирующих этические аспекты проекта “Геном человека”, изложенные во “Всеобщей декларации о геноме человека и правах человека” (Universal Declaration on the Human Genome and Human Rights. Paris, 11.11.1997).

Кроме этой декларации существует множество других международных документов, также связанных с этико-правовыми аспектами данного проекта, достаточно широкая подборка которых проведена коллективом авторов [Иванов, Юдин и др., 1998].

#### 11.4. СТРАТЕГИИ СЕКВЕНИРОВАНИЯ ПОЛНЫХ ГЕНОМОВ

При рассмотрении работ, в которых сообщается об определении нуклеотидной последовательности целого генома какого-либо организма, обращает на себя внимание очень большое число соавторов. Впрочем, это не удивительно, если представить объем выполненной работы. Так, например, секвенирование генома дрожжей *Saccharomyces cerevisiae* было осуществлено усилиями около 600 ученых разных стран. Причем, при выполнении этого проекта было задействовано два глобальных подхода или даже, можно сказать, стратегии. Они заключались в том, что в выполнении этого проекта наряду с 5 научно-исследовательскими центрами крупного и среднего размеров, специализирующимися на секвенировании ДНК (названных в статье А. Гоффи и соавт. [Goffeau et al., 1996] – “фабриками”) принимало участие и множество обычных лабораторий из разных стран. Так вот, эта сеть из 92 лабораторий внесла решающий вклад в осуществление данного проекта (как отмечают авторы – к удивлению некоторых), определив 55% всей последовательности генома, что наглядно продемонстрировало возможность международной кооперации очень большого числа ученых при выполнении крупномасштабных проектов по секвенированию ДНК целых геномов различных организмов.

Что касается выбора стратегий непосредственного получения данных по секвенированию ДНК “на рабочем месте”, то следует отметить, что в цитируемых выше работах, в которых сообщалось о завершении секвенирования какого-либо генома, не использовалось чего-то особенного, поскольку специальной стратегии секвенирования геномов как бы нет и используются те или иные подходы, описанные выше в гл. 8. Причем разными исследовательскими группами на этапе получения первичной информации применялись отличающиеся подходы. Так, секвенирование геномов микроорганизмов *Haemophilus influenzae* [Fleischmann et al., 1995], *Mycoplasma genitalum* [Fraser et al., 1995] и *Methanococcus jannaschii* [Bult et al., 1996] и некоторых других, осуществленное одними и теми же группами ученых, было проведено с помощью вариантов случайного подхода путем механического расщепления их ДНК с последующим секвенированием случайно клонированных последовательностей. Секвенирование генома *Methanobacterium thermoautotrophicum* [Smith et al., 1997] было проведено, главным образом, с помощью мультиплексной технологии как ферментативным методом (38 мембранных фильтров), так и методом химической деградации (31 мембранный фильтр). Сообщается, что около трети генома микоплазмы *Mycoplasma carnicolum* (250 тпн из 750 тпн) секвенировано методом химической деградации с помощью мультиплексной геномной

“прогулки” [Dolan et al., 1995]. Однако, как правило, завершающие этапы этих и других проектов проводятся с использованием ПЦР и праймерной “прогулки”.

Рассматриваемые в гл. 8 стратегии секвенирования протяженных фрагментов ДНК ориентированы, в основном, на определение последовательности нуклеотидов в клонированных вставках длиной от 2 тпн до 45 тпн, что соответствует клонам в обычных плазмидных или космидных векторах. Однако простое переложение этих стратегий для секвенирования полных геномов невозможно, поскольку здесь должен действовать комплексный подход. Более того, учитывая большой объем предстоящей работы при секвенировании целых геномов организмов, должен соблюдаться некий баланс между производительностью выбранного метода, включая избыточность, получаемой с его помощью информации и стоимостью всего проекта. Это вместе с большими размерами геномов как бактерий, так и тем более высших организмов диктует свои особенности. Для последних считается, что усилия, прилагаемые на этапе картирования генома, должны соотноситься с тем количеством нуклеотидов, которые необходимо определить [Roach et al., 1995].

Так, общая стратегия, специально предложенная для секвенирования геномов микроорганизмов [Fraser, Fleischmann, 1997], выглядит следующим образом: фрагментация ДНК бактерии случайным образом; субклонирование этих фрагментов; их секвенирование; компьютерная состыковка ставших известными нуклеотидных последовательностей клонированных фрагментов ДНК; выявление не полностью перекрытых участков; выбор дальнейшей стратегии секвенирования таких участков или с помощью ПЦР или праймерной “прогулкой”; заключительный анализ секвенированной последовательности; составление кольцевой хромосомы.

Первый этап этой общей стратегии состоит в получении полностью случайного набора фрагментов ДНК бактерии и их клонировании в плазмидных, фаговых или фагмидных векторах. Расщепление ДНК частощеплящими рестрикционными эндонуклеазами зависит от распределения их сайтов узнавания в последовательности ДНК, которое не бывает равномерным, и это может привести к образованию фрагментов ДНК различной длины и с большим разбросом. Результатом этого будет то, что последовательность нуклеотидов более крупных субклонов не сможет быть “прочитана” в одном эксперименте и, следовательно, при состыковке субклонов это приведет к появлению несеквенированных промежутков. Во избежание этого предлагается разрушать ДНК с помощью ультразвукового или механического воздействий, практически независимых от конкретной последовательности нуклеотидов в цепи ДНК и генерирующих фрагменты ДНК с небольшими вариациями по длине.

Затем, перед началом выполнения крупномасштабного проекта по секвенированию всего генома какого-либо микроорганизма, необходимо убедиться действительно в случайном характере полученных субклонов. С этой целью проводится секвенирование от 500 до 1000 таких

Табл. 11.1

*Характер накопления данных при секвенировании случайным подходом  
условного генома протяженностью 1 млн  
с помощью клонотеки ДНК с размерами вставок 500 пн*

Количество клонов	Общее количество секвенированных нуклеотидов	Покрывтие генома (кратность)	Несеквенированный геном (пн)	Несеквенированный геном (%)
5000	2500000	2,5	82100	8,21
10000	5000000	5,0	6700	0,67
15000	7500000	7,5	600	0,06
20000	10000000	10,0	40	0,004
25000	12500000	12,5	3	0,0003

субклонов и последующий компьютерный анализ секвенированных нуклеотидных последовательностей позволяет определить нет ли в данной клонотеке каких-либо фрагментов, представленных в большей пропорции. При обнаружении подобного факта необходимо создать такую клонотеку заново, не допуская подобной избирательности, имевшую место для первой клонотеки. В противном случае значительные усилия и время будут тратиться на секвенирование одних и тех же участков генома, что приведет к его большому покрытию, но не позволит определить полную последовательность. Таким образом, от качества созданной клонотеки в значительной мере зависит успешное выполнение всего проекта по секвенированию полных геномов и в связи с этим данному аспекту необходимо уделять самое пристальное внимание.

Но и работа даже с такой полностью случайной клонотекой неизбежно приведет к повторному секвенированию одних и тех же участков ДНК, незначительно сдвигающихся от клона к клону. Не очень сложные математические расчеты показывают, чтобы полностью секвенировать почти весь геном, необходимо определить количество нуклеотидов, более чем в 10 раз превышающее необходимое. Табл. 11.1 позволяет сопоставить общее количество “прочитанных” нуклеотидов и достигаемое при этом покрытие генома.

Из таблицы видно, что в рассматриваемом частном случае определения нуклеотидной последовательности условного генома, состоящего из 1 млн, секвенирование первых 5000 клонов даст свыше 90% нуклеотидной последовательности всего генома, тогда как следующие 5000 клонов прибавят лишь около 8% новых последовательностей. И в дальнейшем эффективность получения новых данных при продолжении секвенирования ДНК случайным подходом будет снижаться. Достаточно оптимальным является достижение 6-кратного покрытия генома и завершение всего проекта с помощью каких-либо других стратегий направленного подхода [Vouk et al., 1998]. В этой же работе говорится, что 3–4-кратное покрытие генома приведет к формированию знаний,

которые могут оказаться полезными при первичном анализе кодирующих областей, тогда как 2-кратное покрытие может дать предварительную информацию об экзонах и позволит вовлечь в анализ EST-последовательности. После компьютерной состыковки “прочитанных” последовательностей и выявления незавершенных промежутков приступают к их секвенированию или с помощью ПЦР, или праймерной “прогулкой”. Разработанная специальная программа BULLET, позволяющая проводить компьютерную симуляцию нарастания данных при секвенировании ДНК с помощью случайного подхода, показала, что когда становятся известны 98% от всей последовательности секвенируемого фрагмента ДНК (по одной цепи), это должно соответствовать приблизительно его 80%-ному покрытию по обеим цепям [Smith, Bernstein, 1995]. Другая компьютерная программа PRIMO была способна после состыковки фрагментов, секвенированных случайным подходом, автоматически подбирать олигонуклеотидные праймеры для завершения всего проекта по секвенированию с помощью праймерной “прогулки” [Li et al., 1997]. Несколько ранее был предложен подход к секвенированию больших геномов только с использованием праймерной “прогулки” [Szybalski, 1993], в основе которой лежал принцип формирования затравочных молекул из библиотеки предсинтезированных 6-мерных олигонуклеотидов путем их лигирования Т4 ДНК-лигазой непосредственно на самой матрице ДНК [Szybalski, 1990], что теоретически позволяет снизить избыточность первичной информации и секвенировать геномы со значительно меньшим покрытием.

Предложенная относительно недавно стратегия геномного секвенирования более крупных геномов, включая геном человека [Venter et al., 1996], несколько отличается от таковых, рассчитанных на секвенирование бактериальных хромосом [Fraser, Fleischmann, 1997] и от применявшихся ранее для крупных геномов, где считалось, что каждый регион должен быть прежде картирован. В ее основе лежит принцип последовательного получения случайных субклонов с уменьшающимися размерами вставок в соответствующих векторах, начиная с вектора на основе искусственной бактериальной хромосомы. Так, ВАС-вектор на основе бактериальной искусственной хромосомы (Bacterial Artificial Chromosome) способен нести вставки размером до 350 тпн и, таким образом, весь человеческий геном с 15-кратным покрытием может “поместиться” в клонотеке из 300 тысяч таких клонов, принимая средний размер вставки равным приблизительно 150 тпн. Важным преимуществом ВАС-вектора перед YAC-векторами на основе дрожжевой искусственной хромосомы (Yeast Artificial Chromosome) является его большая стабильность [Shizuya et al., 1992], поскольку считается, что до 50% YAC-клонов могут быть химерами, что внесет сильную путаницу при состыковке секвенированных участков и заключительной реконструкции полного генома. Полученные клоны в ВАС-векторе распределяются по микротитраторным планшетам, хранятся соответствующим образом и берутся для исследования по мере надобности. Следующим этапом является получение очередной клонотеки уже вставки из ВАС-вектора в фаговом, плазмидном или фагмидном векторе. Так, 3000 субклонов со

вставками длиной около 1000 пн в этих типах векторов будут достаточны для субклонирования каждого такого клона из ВАС-вектора с условным двойным покрытием. Секвенирование концевых 500 нуклеотидов вставки каждого ВАС-клона (с каждого из концов, граничащих с векторными последовательностями) сделает известной около 10% ( $3 \cdot 10^8$  пн) всей последовательности человеческого генома ( $3 \cdot 10^9$  пн). Теоретически секвенированные таким образом участки будут встречаться во всем геноме в среднем через 5 тпн, что позволит потом каждый ВАС-клон соотнести с 30 другими, в которых будут находиться эти же последовательности нуклеотидов ( $150 \text{ тпн} / 5 \text{ тпн}$ ). Авторы цитируемой работы обозначили такие нуклеотидные последовательности концов вставок в ВАС-клонах как STC (sequence-tagged connector). В дальнейшем, после завершения секвенирования каждого набора субклонов, соответствующего одному ВАС-клону, проводится этап состыковки известных последовательностей и с помощью STC выявляются участки перекрытия вставок из ВАС-клонов. При этом может оказаться, что для секвенирования всего генома человека, имеющего размер ( $3 \cdot 10^9$  пн), или любого другого, ему подобного, может оказаться достаточным всего 20 тыс клонов с ВАС-векторами. Для организмов с еще меньшими геномами потребуется соответственно и меньшее число ВАС-клонов.

Считается, что современная техника секвенирования ДНК позволяет завершить проект определения нуклеотидной последовательности какого-либо прокариотического организма с геномом до 2 мпн за 6 мес, если задействовать требуемое число сотрудников и получить должное финансирование. Секвенирование геномов с размерами от 2 до 6 мпн уже требует значительно большего времени, значительных усилий и средств. Так, считается, что на сегодняшний день стоимость одного определенного нуклеотида составляет около 50 центов и предполагается, что к моменту завершения секвенирования генома человека за счет продолжающегося усовершенствования методов секвенирования ДНК она будет составлять не более 25 центов [Collins et al., 1998].

## 11.5. ДОСТОВЕРНОСТЬ ДАННЫХ ПРИ СЕКВЕНИРОВАНИИ ДНК

Немаловажное значение имеет достоверность получаемой информации в виде последовательности нуклеотидов, чему всегда уделялось серьезное внимание [Chen, Hunkapiller, 1992]. Разработаны специальные компьютерные программы, направленные на выявление неточностей при анализе первичных данных, а также контролирующие весь ход выполнения проекта по крупномасштабному секвенированию ДНК. Особенно важным это становится в случаях когда в одной лаборатории выполняется более одного крупномасштабного проекта и поток данных чрезвычайно велик. Осуществлять контроль за такими экспериментами призвана программа Kaleidaseq [Dedhia, McCombie, 1998]. Авторы сообщают, что программа доступна для некоммерческих организаций через Интернет (<http://www.cshl.org/kaleidaseq> ) на бесплатной основе. На этапе сбора пер-

вичной информации процент ошибок сильно зависит от применяемого метода секвенирования, типа метки, используемого фермента, автоматического или ручного секвенатора и варьирует в зависимости от участка секвенируемой ДНК и опыта экспериментатора. В связи с точностью секвенирования особое значение приобретает и цель, стоящая перед исследователем. Так, например, при секвенировании коротких участков молекул кДНК, называемых ESTs и служащих идентификаторами генов, достоверности подобных данных серьезного значения не придается и, по некоторым оценкам, они могут содержать более 1% ошибок [Gibbs, 1997], что совершенно недопустимо в других случаях.

Неточности в определении того или иного нуклеотида на этапе получения первичных данных, как правило, сосредоточены в регионах, приходящихся на начало и конец матрицы. Так, обычно определенное количество ошибок характерно для первых двух десятков нуклеотидов матрицы и оно заметно увеличивается после 400–500 нуклеотидов. Было проведено специальное исследование точности определения нуклеотидной последовательности с помощью разных типов флуоресцентных меток 80 разными лабораториями США, которым были розданы одинаковые образцы ДНК [Naeve et al., 1995]. Последующий анализ показал, что лаборатории по точности секвенирования могут быть разделены на три группы. Первая группа была способна секвенировать ДНК с почти 100%-ной точностью в диапазоне 100–500 нуклеотидов, с 500 по 600 нуклеотидов точность определения заметно снижалась. Для второй группы почти 100%-ная достоверность была доступна только в диапазоне 100–400, тогда как третья группа 92–100%-ную точность смогла обеспечить только в интервале 100–300 нуклеотидов. Однако на завершающем этапе точность секвенирования ДНК должна составлять, по крайней мере, 99,95%, что соответствует 5 ошибкам на 10000 нуклеотидов, впрочем в последнее время вполне достижимым стандартом считается 99,99% или 1 ошибка на 10000 нуклеотидов [Marshall, 1995]. Используемые в геномных проектах стратегии случайного подхода, и достигаемое за счет этого в среднем 7–8-кратное покрытие генома, кроме избыточных данных, дают существенное преимущество в виде многократного “чтения” одних и тех же участков (причем, по обеим цепям) ДНК, что приводит к необходимой достоверности получаемых результатов. Так, например, при секвенировании генома *Mycoplasma genitalium* был достигнут требуемый стандарт – 99,99% [Fraser et al., 1995], а для генома *Haemophilus influenzae* точность определения нуклеотидной последовательности оказалась еще выше и составила менее одной ошибки на 19000 нуклеотидов [Fleischmann et al., 1995]. Данный подсчет в некоторой степени условен и основан, в первую очередь, на определении частоты сдвига рамки считывания [Fichant, Quentin, 1995]. Другими авторами была разработана программа, способная благодаря специальному алгоритму выявлять и исправлять в кодирующих регионах определенный тип ошибок, получивших общее обозначение “индел” (*indel* – *insertion, deletion*), в виде инсерций и делеций нуклеотидов, нарушающих рамку считывания [Xu et al., 1995]. Недавно созданная программа PHRED [Ewing et al., 1998; Ewing, Green, 1998] позволяет осуществлять предсказание специфических “ошибок” в “прочтении”

определенных нуклеотидов на этапе сбора первичной информации, что делает ее весьма удобной при выполнении крупномасштабных проектов. Так, с помощью данной программы был проведен анализ первичных данных 6 американских лабораторий, выполняющих проекты по секвенированию геномов [Richterich, 1998]. В результате оказалось, что предсказанные данной программой “ошибки” в определении тех или иных нуклеотидов во многом совпали с действительными, хотя справедливости ради следует отметить, что последних было все же несколько меньше. Ранее похожая программа, направленная на предсказание потенциальных ошибок в первичном материале секвенирования ДНК, была предложена другими авторами [Lawrence, Solovyeu, 1994].

Возвращаясь к вопросу о достоверности последовательности ДНК генома *Haemophilus influenzae*, надо отметить, что сами авторы подчеркивают, что выявленные ими отдельные несовпадения (замены нуклеотидов, делеции, инсерции) при сравнении с аналогичными генами *Haemophilus influenzae*, определенными ранее другими авторами и помещенными в ГенБанк, вовсе не свидетельствуют, что “их” последовательность не верна [Fleischmann et al., 1995].

Что касается, например, генома дрожжей *Saccharomyces cerevisiae*, то авторы сообщают, что в заключительной последовательности может содержаться до 3 ошибок на 10000 нуклеотидов [Goffeau et al., 1996], что соответствует точности секвенирования 99,97%.

Как уже отмечалось в предыдущей главе, дополнительная коррекция полученных результатов возможна на этапе занесения нуклеотидных последовательностей в компьютер, стыковки “прочитанных” участков и последующего анализа.

## 11.6. СТРАТЕГИИ АНАЛИЗА ПОЛНЫХ ГЕНОМОВ

Главной задачей, стоящей перед исследователями, и решению которой служит расшифровка всей последовательности ДНК, является выяснение всех аспектов жизнедеятельности того или иного организма, геном которого оказался полностью секвенированным. Решение этой задачи в полном объеме дело будущего, причем, наверное, отдаленного. Однако приближение к такому полному знанию лежит через постепенное выяснение основных принципов и особенностей функционирования геномов. Необходимым этапом служит обнаружение *in silico* (т.е. на уровне ДНК) всех потенциальных белковых продуктов, составляющих протеом исследуемого организма. Далее необходимо каким-то образом определить функции этих еще гипотетических белков. Весьма заметную роль в таком познании выполняет сравнительный анализ этих белковых продуктов с уже известными.

После завершения составления нуклеотидной последовательности полного генома, исключаяющей какие-либо неоднозначности, с помощью специальной компьютерной программы, например *Consed* [Gordon et al., 1998] или GRM [Lawrence et al., 1994], наступает этап анализа всего генома через последовательность его ДНК. Кроме становящихся ав-



томатически известными общего количества нуклеотидов в геноме и GC-состава ДНК, следуют стандартные процедуры по определению количества рибосомных оперонов, генов тРНК, повторяющихся элементов, транспозонов, обнаружения ориджина репликации и др. [Fraser, Fleischmann, 1997]. Пожалуй, самым важным элементом анализа, как уже отмечалось выше, является обнаружение всех потенциальных открытых рамок считывания, имеющих достаточную протяженность и выявление, таким образом, всего протеома. Идентификация среди потенциальных продуктов протеома тех, что обнаруживают заметную гомологию с белками с известной функцией, ранее занесенными в банки данных, позволяет приписать им конкретную роль. Однако выявленные подобным образом гомологичные гены могут быть как ортологичными, так и паралогичными, что, естественно, скажется на функции, выполняемой их белковыми продуктами. Проблема заключается в том, что паралогичные гены, возникающие в результате дупликации, часто за счет дивергенции накапливают определенные замены, приводящие к тому, что кодируемые ими белки могут уже выполнять отличающуюся функцию. Но при этом уровень гомологии таких генов может оказаться довольно значительным, что в итоге приведет к ошибке при определении роли какого-либо нового белка из протеома исследуемого организма. Кроме орто- и паралогичной, принято различать еще несколько типов гомологии генов, характерные примеры которых приведены ниже:

Гомологичные Ортологичные	Гены, имеющие общего предка, например все глобины Гомологичные гены, дивергировавшие после событий видообразования (например, $\beta$ -глобин человека и $\beta$ -глобин шимпанзе)
Паралогичные	Гомологичные гены, дивергировавшие после событий дупликации (например, $\beta$ - и $\gamma$ -глобины)
Ксенологичные	Гомологичные гены, дивергировавшие после событий горизонтального переноса генов (например, гены устойчивости к антибиотикам у бактерий)
Позиционная гомология	Специфические позиции аминокислот или нуклеотидов в разных белках или кодирующих их генах, совпадающие с таковыми у их общих предковых форм

В то же время нельзя забывать, что такой эволюционный процесс, как конвергенция, может привести к образованию подобных по своей последовательности белков и их генов, которые при этом не будут гомологичными.

Большое значение в анализе протеома приобретает сравнительная геномика [Koonin et al., 1996], ставшая возможной после завершения секвенирования нескольких геномов прокариот. Причем сравнение одного полного генома микроорганизмов даже с незавершенным геномом другого позволило выявить важные черты обоих [Tatusov et al., 1996]. Сравнение полных геномов микроорганизмов на предмет их сходства и эволюционных взаимоотношений вызвало к жизни термин “филогеномика” [Eisen, 1998]. В этой работе автор проводит анализ подходов при выявлении эволюционных семейств белков и предостерегает от ошибочного выбора. Проведенное сравнение белковых продуктов, кодиру-

емых 6 прокариотическими геномами, включая 1 археобактериальный, и одним эукариотическим, позволило уже другим авторам выявить так называемые кластеры ортологичных групп (Clusters of Orthologous Groups или COGs) [Tatusov et al., 1997]. В результате оказалось идентифицировано 720 таких групп, каждая из которых представлена как минимум 3 белками, причем для большинства этих групп уже известна функциональная роль белков, входящих в данный COG. Поскольку ортологичные гены обычно кодируют белки с одинаковыми функциями, то, основываясь на этих выявленных кластерах, становится значительно легче приписывать “новым” белкам конкретную роль.

Кодирующие и не кодирующие участки генома могут быть также обнаружены и путем определения частот встречаемости отдельных олигонуклеотидов с помощью подхода, получившего название “внутреннего” [Borodovsky et al., 1994]. “Внешний” подход, также использованный этими авторами, заключался в обнаружении значительного подобия продуктов потенциальных открытых рамок считывания, представленных в геноме данного организма с уже известными белковыми последовательностями. Был предложен метод анализа протяженных последовательностей ДНК, применимый и к полным геномам, основанный на определении частот встречаемости октануклеотидных блоков и рассчитанный на выявление нетипичных для данного организма регионов, возможно являющихся следствием горизонтального переноса генов или каких-то других событий [Scherer et al., 1994]. Интересный подход для анализа полных бактериальных геномов и просто устроенных эукариотических был предложен другими авторами [Lashkari et al., 1997]. Его суть заключалась в амплификации всех потенциальных открытых рамок считывания и получения соответствующих фрагментов ДНК, пригодных затем для проведения с ними различных экспериментов. Причем авторы отмечают, что синтез такого большого количества олигонуклеотидных праймеров возможен с помощью созданного ими ранее высокопроизводительного мультиплексного ДНК-синтезатора [Lashkari et al., 1995], способного после некоторой проведенной модификации синтезировать за 8 ч 288 25-мерных праймеров (три 96-луночных микротитраторных планшета).

Для анализа эукариотических, более сложно устроенных, геномов, кроме некоторых из этих подходов, также используют и другие, связанные, в первую очередь, с существованием у этих организмов мРНК, несущих поли(А)-последовательность. Так, например, выявление кодирующих последовательностей возможно путем сравнения ДНК генома и 3'-участков мРНК в виде последовательностей, известных как EST и уже упоминавшихся выше при описании генома человека. Специфичность функционирования отдельных генов может выясняться с помощью процедуры, называемой Different Display. К более сложному анализу можно отнести недавно предложенную “интегрированную процедуру идентификации генов”, которая объединяет многие из независимо существующих методов [Wang, Rowley, 1998]. Большие возможности в идентификации генов заключены в использовании приготовленных с помощью фотолитографии ДНК-чипов.

## 11.7. БУДУЩЕЕ СЕКВЕНИРОВАНИЯ ПОЛНЫХ ГЕНОМОВ

В 1995 г. стали известны нуклеотидные последовательности двух прокариотических геномов. 1996 г. увеличил число таких геномов до 5 и добавил еще и один эукариотический геном. В 1997 г. было завершено секвенирование сразу 6 бактериальных геномов. Столько же новых геномов с определенной последовательностью нуклеотидов уже принес 1998 г. и в декабре предполагается завершение секвенирования генома еще одного эукариотического организма – нематоды *Caenorhabditis elegans*. Можно ожидать, что в 1999 г. темпы расшифровки последовательности нуклеотидов полных геномов различных организмов еще увеличатся. Также не вызывает сомнений, что секвенирование полных геномов различных организмов станет важной задачей молекулярных биологов следующего столетия и продолжающееся бурными темпами развитие необходимой для этого техники (включая разработку новых подходов к секвенированию ДНК, создание соответствующих приборов, компьютерное обеспечение) сделает со временем подобные задачи весьма обычным делом, хотя и дорогостоящим. Пока же секвенирование геномов (особенно крупных) представляет собой трудную задачу, справиться с которой в обозримые сроки можно только при условии объединения усилий сразу нескольких исследовательских групп. В то же время, помимо непосредственного секвенирования полных геномов, все большее значение приобретает функциональная геномика, направленная на выяснение механизмов функционирования отдельных генов и их взаимодействия в составе целого организма. Именно в этом случае секвенирование геномов дает те сведения, которые от него ожидают, поскольку отдельные новые гены, которые становятся известны в результате секвенирования какого-либо полного генома, могли бы быть клонированы и секвенированы обычным путем и не они представляют собой главную цель подобных проектов. Но, учитывая огромный объем уже сейчас известной информации в виде последовательности нуклеотидов, можно представить, что пройдет еще много времени, пока станут детально известны особенности функционирования хоть какого-нибудь небольшого генома.

Так, в настоящее время, после анализа какого-либо ставшего известным бактериального генома становится очевидным, что в его протеоме более половины открытых рамок считывания принадлежит потенциальным генам, функции которых не известны. Более того, для части из них даже не обнаруживается заметной гомологии с уже имеющимися в банках данных последовательностями. И это для не очень больших геномов с довольно ограниченным набором генов. После завершения секвенирования генома человека с его 100 тыс. генов или около того (причем считается, что 30 тыс. из них экспрессируется в мозге) надо думать, что пропорция будет совсем иной. В то же время нельзя не учитывать постоянное пополнение банков данных нуклеотидными последовательностями, принадлежащими генам с известной функцией, что, несомненно, окажет влияние на соотношение известных и неизвестных открытых рамок считывания в протеоме человека.

Сравнительный анализ генов или кодируемых ими белков является мощным инструментом познания, позволяющим выявить определенные различия и особенности как их структуры, так и функционирования. В этой связи секвенирование геномов родственных организмов может оказать незаменимую помощь при анализе, например, генома человека. С этой целью уже стартовал проект по секвенированию генома другого млекопитающего, а именно мыши, поскольку она является очень хорошо изученным объектом на биохимическом и молекулярном уровнях. В то же время нельзя исключать вероятность секвенирования в будущем геномов и более родственных человеку видов животных. Так, уже сейчас ученые задаются вопросом “какие гены делают нас человеком?”, поскольку известно, что ДНК человека и шимпанзе гомологичны друг другу на 98,5% (Gibbons, 1998).

Завершая данный раздел, посвященный перспективам секвенирования полных геномов различных организмов, нельзя не отметить некое знаменательное событие, произошедшее на выставке ANALYTICA 98, проходившей в Мюнхене в апреле 1998 г., где было объявлено об объединении усилий пяти немецких биотехнологических компаний с целью предоставить заказчикам полный спектр услуг, начиная от создания геномных библиотек до детального анализа секвенированного генома. В этот генный альянс (Gene Alliance) вошли AGOWA GmbH (специализирующаяся на крупно-масштабном заказном секвенировании ДНК), Biomax Informatics GmbH (анализ генетической информации), GATC GmbH (создание геномных клонотек, заказное секвенирование), MediGenomix GmbH (заказное секвенирование и диагностика) и Qiagen GmbH (поставщик реактивов для выделения и очистки ДНК, пригодной для секвенирования). Так, заказное секвенирование каких-либо последовательностей ДНК существует уже на протяжении многих лет и стало обычным явлением и вот теперь желающие могут заказать секвенирование уже целого генома.

### *Цитируемая литература*

Этико-правовые аспекты проекта “Геном человека”: (Международные документы и аналитические материалы) / Ред.-сост. Б.Г. Юдин, В.И. Иванов. М.: РНКБ РАН, 1998. 190 с.

Adams M.D., Kelly J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., Kerlavage A.R., McCombie W.R., Venter J.C. Complementary DNA sequencing: Expressed sequence tags and Human Genome Project // *Science*. 1991. Vol. 252. P. 1651–1656.

Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H.L., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J.H., Staden R., Young I.G. Sequence and organization of the human mitochondrial genome // *Nature*. 1981. Vol. 290. P. 457–465.

Andersson S.G.E., Zomorodipour A., Andersson J.O., Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naslund A.K., Eriksson A.S., Winkler H.H., Kurland C.G. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria // *Nature*. 1998. Vol. 396. P. 133–140.

Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R., Chee M.S., Hutchinson C.A., III, Kouzarides T., Martignetti J.A., Preddie E. The DNA sequence of the human cytomegalovirus genome // *DNA Sequenc.* 1991. Vol. 2. P. 1–12.

Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C., Bergkamp R., Dirkse W., Van Staveren M., Stiekema W., Drost L., Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P., Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terry N., Gielen J., Villarroe R., De Clerck R., Van Montagu M., Lecharny A., Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P., Entian K.-D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S., Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B., Hilbert H., Dusterhoft A., Moores T., Jones J.D.G., Eneva T., Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S., Schueller C., Chalwatzis N. Analysis of 1.9 Mb of continuous sequence from chromosome 4 of *Arabidopsis thaliana* // *Nature*. 1998. Vol 391. P. 485–488.

Blattner F.R., Plunket M.G., Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y. The complete genome sequence of *Escherichia coli* K-12 // *Science*. 1997. Vol. 277. P. 1453–1462.

Borodovsky M., Koonin E.V., Ruud K.E. New genes in old sequence: A strategy for finding genes in the bacterial genome // *Trends Biochem. Sci.* 1994. Vol. 19. P. 309–313.

Bouck J., Miller W., Gorrell J.H., Muzny D., Gibbs R.A. Analysis of the quality and utility of random shotgun sequencing at low redundancies // *Genome Res.* 1998. Vol. 8. P. 1074–1084.

Brower V. Genome II: The next frontier // *Nature Biotechnol.* 1998. Vol. 16. P. 1104.

Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I., Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D., Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C. Complete genome sequence of the methanogenic archaeon, *Methanococcus janaschii* // *Science*. 1996. Vol. 273. P. 1058–1073.

Chen W.Q., Hunkapiller T. Sequence accuracy of large DNA sequencing projects // *DNA Sequenc.* 1992. Vol. 2. P. 335–342.

Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Takaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Fetwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver K., Osborne J., Quail M.A., Rajandream M.-A., Rogers J., Rutter S., Seeger K., Skelton R., Squares S., Sulston J.E., Taylor K., Whithead S., Barrell B.G. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence // *Nature*. 1988. Vol. 393. P. 537–44; Errata *Nature*. 1998. Vol. 396. P. 190–198.

Collins F., Galas D. A new five-year plan for the U.S. Human genome project // *Science*. 1993. Vol. 262. P. 43–6.

Collins F.S., Patrinos A., Jordan E., Chakravarti A., Gesteland R., Walters L. New goals for the U.S. Human genome project: 1998–2003 // *Ibid.* 1998. Vol. 282. P. 682–689.

Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R., Feldman R.A., Short J.M., Olsen G.J., Swanson R.V. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus* // *Nature*. 1998. Vol. 392. P. 353–358.

Dedhia N.N., McCombie W.R. Kaleidaseq: A Web-based tool to monitor data flow in a high throughput sequencing facility // *Genome Res.* 1998. Vol. 8. P. 313–318.

Dolan M., Ally A., Pyrzycki M.S., Gilbert W., Gillevet P.M. Large-scale genomic sequencing: Optimization of genomic chemical sequencing reactions // *Biotechniques*. 1995. Vol. 19. P. 264–273.

Eisen J.A. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis // *Genome Res.* 1998. Vol. 8. P. 163–167.

Ewing B., Green P. Base-calling of automated sequencer traces using *Phred*. 2. Error probabilities // *Ibid.* 1998. Vol. 8. P. 186–194.

Ewing B., Hiller L., Wendl M.C., Green P. Base-calling of automated sequencer traces using *Phred*. 1. Accuracy assessment // *Ibid.*, 1998. Vol. 8. P. 175–85.

Fichant G.A., Quentin Y. A frameshift error detection algorithm for DNA sequencing projects // *Nucl. Acids Res.* 1995. Vol. 23. P. 2900–2908.

Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., FitzHugh W., Fields C., Gocayne J.D., Scott J., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd // *Science*. 1995. Vol. 269. P. 496–512.

Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R., Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M., Dougherty B., Tomb J.F., Fleischmann R.D., Richardson D., Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J., Venter J.C. // Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi* // *Nature*. 1997. Vol. 390. P. 580–586.

Fraser C.M., Fleischmann R.D. Strategies for whole microbial genome sequencing and analysis // *Electrophoresis*. 1997. Vol. 18. P. 1207–1216.

Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J., Nguyen D., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchinson C.A., III, Venter J.C. The minimal gene complement of *Mycoplasma genitalium* // *Science*. 1995. Vol. 270. P. 397–403.

Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G., Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T., McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete // *Science*. 1998. Vol. 281. P. 375–388.

Freiberg C., Fellay R., Bairoch A., Broughton W.J., Rosental A., Perret X. Molecular basis of symbiosis between *Rhizobium* and legumes // *Nature*. 1997. Vol. 387. P. 394–401.

Freiberg C., Perret X., Broughton W.J., Rosenthal A. Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable “sequenase”: A beginning // *Genome Res*. 1996. Vol. 6. P. 590–600.

Gale M.D., Devos K.M. Plant comparative genetics after 10 years // *Science*. 1998. Vol. 282. P. 656–659.

Gibbons A. Which of our genes make us human? // *Ibid*. 1998. Vol. 281. P. 1432–1434.

Gibbs R.A. Hares and tortoises in the race to sequence the human genome: Expectation and realities // *Trends Genet*. 1997. Vol. 13. P. 381–383.

Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P., Paoletti E. // The complete DNA sequence of vaccinia virus // *Virology*. 1990. Vol. 179. P. 247–266.

Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldman H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., Louis E.J., Mewes H.W., Murakami Y., Philippsen P., Tettelin H., Oliver S.G. Life with 6000 genes // *Science*. 1996. Vol. 274. P. 546–567.

Goffeau A., Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W., Benes V., Bruckner M., Delius H., Dubois F., Dusterhoft A., Entian K.D., Floeth M., Hebling U., Heumann K., Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter P., Louis E.J., Messenguy F., Mewes H.W., Hoheisel J.D. The yeast genome directory // *Nature*. 1997. Vol. 387, suppl. P. 5–105; Errata // *Nature*. 1997. Vol. 387. P. 737; *Nature*. 1997. Vol. 389. P. 412.

Gordon D., Abajian C., Green P. *Consed*: A graphical tool for sequence finishing // *Genome Res*. 1998. Vol. 8. P. 195–202.

Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A., Orsat B., Spielmann A., Stuitz E. Complete sequence of *Euglena gracilis* chloroplast DNA // *Nucl. Acids. Res*. 1993. Vol. 21. P. 3537–3544.

Himmelreich R., Hilbert H., Plages H., Pirkl E., Li B.C., Herrmann R. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae* // *Ibid*. 1996. Vol. 24. P. 4420–4449.

Hong G., Qian Y., Yu S., Hu X., Zhu J., Tao W., Li W., Su C., Zhao H., Qiu L., Yu D., Liu X., Wu B., Zhang X., Zhao W. A 120 kilobase resolution contig map of the rice genome // *DNA Sequenc*. 1997. Vol 7. P. 319–335.

Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A.,

Nakazaki N., Naruo K., Okumura S., Shimpo S., Takeuchi C., Wasa T., Watanabe A., Yamada M., Yasuda M., Tabata S. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. 2. Sequence determination of the entire genome and assignment of potential protein-coding regions // DNA Res. 1996. Vol. 3. P. 109–136.

Kasai H., Isono S., Kitakawa M., Mineno J., Akiyama H., Kunit D.M., Berg D.E., Isono K. Efficient large-scale sequencing of the *Escherichia coli* genome: Implementation of a transposon- and PCR-based strategy for the analysis of ordered  $\lambda$  phage clones // Nucl. Acids Res. 1992. Vol. 20. P. 6509–6515.

Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K., Kicuchi H. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3 // DNA Res. 1998. Vol. 5. P. 147–155.

Klenk H.P., Clayton R.A., Tomb J.F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou I., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus* // Nature. Vol. 390. P. 364–370; Erratum Nature. 1998. Vol. 394. P. 101.

Koonin E.V., Mushegian A.R., Rudd K.E. Sequencing and analysis of bacterial genomes // Curr. Biol. 1996. Vol. 6. P. 404–416.

Kunst F., Ogasarawa N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero V., Bessiere P., Bolotin A., Borchert S., Borriss R., Brousier R., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Brouchi C.V., Caldwell B., Capuano V., Carter N.M., Choi S-L., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S-Y., Glaser P., Gopffeau A., Golightly E.J., Grandi G., Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hullo M-F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S-M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S-H., Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B-S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H-F., Zumstein E., Yoshikawa H., Danchin A. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis* // Nature. 1997. Vol. 390. P. 249–256.

Kurata N., Umehara Y., Tanoue H., Sasaki T. Physical mapping of the rice genome with YAC clones // Plant Mol. Biol. 1997. Vol. 35. P. 101–113.

Lashkari D.A., Hunicke-Smith S.P., Norgren R., Davis R.W., Brenmann T. An automated multiplex oligonucleotide synthesizer: Development of high-throughput, low-cost DNA synthesis // Proc. Nat. Acad. Sci. USA. 1995. Vol. 92. P. 7912–7915.

Lashkari D.A., McCusker J.H., Davis R.W. Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR // Ibid. 1997. Vol. 94. P. 8945–8947.

Lawrence C.B., Honda S., Parrott N.W., Flood T.C., Zhang L., Jain M., Larson S., Myers E.W. The genome reconstruction manager: A software environment for supporting high-throughput DNA sequencing // Genomics. 1994. Vol. 23. P. 192–201.

Lawrence C.B., Solovyev V.V. Assignment of position-specific error probability to primary DNA sequence data // Nucl. Acids Res. 1994. Vol. 22. P. 1272–1280.

- Li P., Kupfer K.C., Davies C.J., Burbee D., Evans G.A., Garner H.R. PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing // *Genomics*. 1997. Vol. 40. P. 476–485.
- Marshall E. Emphasis turns from mapping to large-scale sequencing // *Science*. 1995. Vol. 268. P. 1270–1271.
- Massung R.F., Esposito J.J., Liu J.-I., Jin Q., Utterback T.R., Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N., Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J., Venter J.C. Potential virulence determinants in terminal regions of variola smallpox virus genome // *Nature*. 1993. Vol. 366. P. 748–751.
- McFadden G.I., Gilson P.R., Douglas S.E., Cavalier-Smith T., Hofmann C.J.B., Maier U.G. Bonsai genomics: Sequencing the smallest eukaryotic genomes // *Trends Genet.* 1997. Vol. 13. P. 46–49.
- Meinke D.W., Cherry J.M., Dean C., Rounsley S.D., Koornneef M., *Arabidopsis thaliana*: A model plant for genome analysis // *Science*. 1998. Vol. 282. P. 662–682.
- Naeve C.W., Buck G.A., Niece R.L., Pon R.T., Robertson M., Smith A.J. Accuracy of automated DNA sequencing: A multi-laboratory comparison of sequencing results // *Biotechniques*. 1995. Vol. 19. P. 448–453.
- Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N., Akashi K., Kanegae T., Ogura Y., Kohchi T. Gene organization deduced from the complete sequence of livermoss *Marchantia polymorpha* mitochondrial DNA: A primitive form of plant mitochondrial genome // *J. Mol. Biol.* 1992. Vol. 223. P. 1–7.
- Olson M., Hood L., Cantor C., Botstein D. A common language for physical mapping of the human genome // *Science*. 1989. Vol. 245. P. 143–145.
- Richterich P. Estimation of errors in “raw” DNA sequences: A validation study // *Genome Res.* 1998. Vol. 8. P. 251–259.
- Roach J.C., Boysen C., Wang K., Hood L. Pairwise end sequencing: A unified approach to genomic mapping and sequencing // *Genomics*. 1995. Vol. 26. P. 345–353.
- Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fidders J.C., Hutchinson C.A., III, Slocombe P.M., Smith M. Nucleotide sequence of bacteriophage  $\phi$ X174 DNA // *Nature*. 1977. Vol. 265. P. 687–695.
- Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B. Nucleotide sequence of bacteriophage  $\lambda$  DNA // *J. Mol. Biol.* 1982. Vol. 162. P. 729–773.
- Sanger F., Donelson J.E., Coulson A.R., Kossel H., Ficher D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA // *Proc. Nat. Acad. Sci. USA*. 1973. Vol. 70. P. 1209–1213.
- Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M., Miyajima N., Tabata S. Structural analysis of *Arabidopsis thaliana* chromosome 5. 1. Sequence feature of 1.6 Mb region covered by twenty physically assigned P1 clones // *DNA Res.* 1997. Vol. 4. P. 215–230.
- Scherer S., McPeck M.S., Speed T.P. A typical regions in large genomic DNA sequences // *Proc. Nat. Acad. Sci. USA*. 1994. Vol. 91. P. 7134–7138.
- Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S. Genes of variola and vaccinia viruses necessary to overcome the host protective mechanisms // *FEBS Lett.* 1993. Vol. 319. P. 80–83.
- Shizuya H., Birren B., Kim U.-J., Mancino V., Slepak T., Tachiiri Y., Simon M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector // *Proc. Nat. Acad. Sci. USA*. 1992. Vol. 89. P. 8794–8797.
- Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H., Dubois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K., Harrison D., Hoang L., Leagle P., Lumm W., Pothier B., Qiu D., Spadafora R., Vicaire R., Wang Y., Wiersbowski J., Gibson R., Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S., McDougall S., Shimer G., Goyal A., Petrokovski S., Church G.M., Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N. Complete genome sequence of *Methanobacterium thermoautotrophicum*  $\Delta$ H: Functional analysis and comparative genomics // *J. Bacteriol.* 1997. Vol. 179. P. 7135–7155.
- Smith G.D., Bernstein K.E. BULLET: a computer simulation of shotgun DNA sequencing // *Comput. Appl. Biosci.* 1995. Vol. 11. P. 155–157.
- Smith H.O., Wilcox K.W. A restriction enzyme from *Haemophilus influenzae*. 1. Purification and general properties // *J. Mol. Biol.* 1970. Vol. 51. P. 379–391.
- Stephens R.S., Kalman S., Lammel C., Fan J., Marathe R., Aravind L., Mitchell W., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis* // *Science*. 1998. Vol. 282. P. 754–759.



- Szybalski W.* Proposal for sequencing DNA using ligation of hexamers to generate sequential elongation primers (SPEL-6) // *Gene*. 1990. Vol. 90. P. 177–178.
- Szybalski W.* From the double-helix to novel approaches to the sequencing of large genomes // *Gene*. 1993. Vol. 135. P. 279–90.
- Tang C.M., Hood D.W., Moxon E.R.* *Haemophilus* influence: the impact of whole genome sequencing of microbiology // *Trends Genet*. 1997. Vol. 13. P. 399–404.
- Tateno Y., Fukami-Kobayashi K., Miyazaki S., Sugawara H., Gojobori T.* DNA Data Bank of Japan at work on genome sequence data // *Nucl. Acids Res*. 1998. Vol. 26. P. 16–20.
- Tatusov R.L., Koonin E.V., Lipman D.J.* A genomic perspective on protein families // *Science*. 1997. Vol. 278. P. 631–637.
- Tatusov R.L., Mushegian A.R., Bork P., Brown N.P., Hayes W.S., Borodovsky M., Rudd K.E., Koonin E.V.* Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Wesherichia coli* // *Curr. Biol*. 1996. Vol. 6. P. 279–291.
- Thomas K.* Yeasties and beasts: 7 years of genome sequencing // *FEBS Lett*. 1996. Vol. 396. P. 1–6.
- Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., Fleischmann R.D., Ketchum K.A., Klenk H.P., Gill S., Dougherty B.A., Nelson K., Quackerbush J., Zhou L., Kirkness E.F., Peterson S., Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A., McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K., Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M., Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M., Venter J.C.* The complete genome sequence of the gastric pathogen *Helicobacter pylori* // *Nature*. 1997. Vol. 388. P. 539–547; Errata // *Nature*. 1997. Vol. 389. P. 412.
- Venter J.C., Smith H.O., Hood L.* A new strategy for genome sequencing // *Nature*. 1996. Vol. 381. P. 364–366.
- Wang S.M., Rowley J.D.* A strategy for genome-wide gene analysis: Integrated procedure for gene identification // *Proc. Nat. Acad. Sci. USA*. 1998. Vol. 95. P. 11909–11914.
- Watson J.D., Crick F.H.C.* A structure for deoxyribose nucleic acid // *Nature*. 1953a. Vol. 171. P. 737–738.
- Watson J.D., Crick F.H.C.* Genetic implication of the structure of deoxyribonucleic acid // *Ibid*. 1953b. Vol. 171. P. 964–967.
- Wu R., Taylor E.* Nucleotide sequence analysis of DNA. 2. Complete nucleotide sequence of the cohesive ends of bacteriophage  $\lambda$  DNA // *J. Mol. Biol*. 1971. Vol. 57. P. 491–511.
- Xu Y., Mural R.J., Uberbacher E.E.* Correcting sequencing errors in DNA coding regions using a dynamic programming approach // *Comput. Appl. Biosci*. 1995. Vol. 11. P. 117–124.

### Дополнительная литература

- Беляева Н.Н.* О государственной научно-технической программе “Геном человека” // *Молекуляр. биология*. 1991. Т. 25. С. 835–851.
- Этико-правовые аспекты проекта “Геном человека”: (Международные документы и аналитические материалы) / Ред.-сост. *В.И. Иванов, Б.Г. Юдин*. 1998. М.: РНКБ РАН, 1998. 190 с.
- Cantor C.R.* Orchestrating the Human genome project // *Science*. 1990. Vol. 248. P. 49–51.
- Science*. 1993. Vol. 262, N 5130: (Genome issue).
- Science*. 1994. Vol. 265, N 5181: (Genome issue).
- Science*. 1995. Vol. 270, N 5235: (Genome issue).
- Science*. 1996. Vol. 274, N 5287: (Genome issue).
- Science*. 1997. Vol. 278, N 5338: (Genome issue).
- Science*. 1998. Vol. 282; N 5389: (Genome issue).
- Watson J.D.* The Human genome project: past, present and future // *Science*. 1990. Vol. 248. P. 44–49.

## СЕКВЕНИРОВАНИЕ ДНК ПОСРЕДСТВОМ СИНТЕЗА

Собственно и сам классический метод ферментативного секвенирования ДНК по Сэнгеру осуществляется посредством синтеза новой цепи ДНК по уже существующей одноцепочечной матрице. Однако его главное отличие от описываемых здесь состоит в том, что в методе Сэнгера синтез новой комплементарной цепи ДНК происходит в условиях специфической терминации, достигаемой включением дидезоксинуклеотидтрифосфатов, блокирующих ее дальнейший рост, причем дальнейшее возобновление синтеза не предусматривается. В то же время существуют подходы к ферментативному секвенированию ДНК, где построение новой ДНК происходит без участия терминаторов, а используются так называемые дельмиторы, или ограничители, представляющие собой модифицированные аналоги дНТФ, такие как тио- или борпроизводные [Labeit et al., 1986, 1987; Porter et al., 1997]. Они не прерывают дальнейший синтез, но при этом служат местами, по которым после завершения синтеза осуществляется образование гетерогенных по 3'-концу секвенируемых фрагментов ДНК, предназначенных к их разделению по размеру с помощью электрофореза. И вот здесь лежит еще одно отличие рассмотренных выше методов ферментативного секвенирования ДНК от метода секвенирования ДНК посредством синтеза, что дает весомые основания все же выделить для этого подхода отдельную главу. Более того, в англоязычной литературе уже появился термин *sequencing by synthesis* по аналогии с *sequencing by hybridization* (последний, впрочем, имеет уже и используемую аббревиатуру – SBH). Таким образом, данная глава будет посвящена методам секвенирования ДНК, продукты ферментативных реакций которых не требуют их анализа посредством электрофореза.

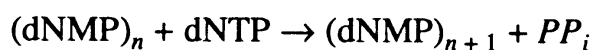
Несмотря на ряд ограничений и низкую (пока) производительность рассматриваемых ниже методов, они имеют весьма важное преимущество. При секвенировании ДНК классическими способами определение последовательности нуклеотидов осуществляется путем разделения смеси гетерогенных фрагментов ДНК по их размеру с помощью электрофореза, но какого бы высокого разрешения он не проводился, все же эффективное разделение фрагментов ДНК крупнее 2000 нуклеотидов представляется маловероятным. А для секвенирования ДНК посредством синтеза теоретически предела “чтения” нуклеотидной последовательности секвенируемой матрицы ДНК не существует (разве что при достижении сверхвысокого количества новосинтезированной ДНК в реакционном сосуде) и поэтому, раз начавшись с какого-либо места протяженного фрагмента ДНК, его секвенирование может продолжаться “бесконечно” долго. Впрочем, в пиросеквенировании количество присоединяемых нуклеотидов зависит от местоположения биотинилированной метки, которая может быть как на 5'-, так и на 3'-конце ма-

трицы. (Достаточно подробное описание способов мечения одного из концов секвенируемого фрагмента ДНК содержится в разделе 1.2) Так, если молекула биотина “пришита” к 5'-концу, то протяженность секвенируемой ДНК будет зависеть от расстояния, на котором расположено место отжига олигонуклеотидного праймера, а если к 3'-концу, то от общей длины одноцепочечной матрицы ДНК.

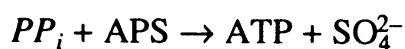
## 12.1. ПИРОСЕКВЕНИРОВАНИЕ ДНК

Метод пиросеквенирования ДНК основан на многошаговой детекции неорганического пирофосфата, образующегося из дНТФ (или ддНТФ) в процессе присоединения ДНК-полимеразой соответствующего нуклеотидмонофосфата к 3'-концу растущей цепи ДНК, иммобилизованной на твердом носителе [Ronaghi et al., 1998]. Чтобы понять принцип работы данной системы, в основе которой лежит разработанный ранее метод детекции полимеразной активности через анализ высвобождающегося пирофосфата в режиме реального времени [Nyren, Lundin, 1985; Nyren, 1987], видимо, вполне уместным будет приведение схемы последовательно протекающих ферментативных реакций:

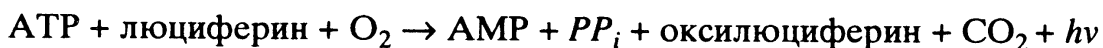
ДНК-полимераза



АТФ-сульфурилаза



Люцифераза



где,  $(dNMP)_n$  – ДНК, а  $n$  – число нуклеотидных остатков,  $PP_i$  – неорганический пирофосфат,  $APS$  – аденозин 5'-фосфосульфат,  $h\nu$  – квант света.

Как видно из приведенной серии последовательных реакций, образующийся в процессе присоединения очередного нуклеотида неорганический пирофосфат затем под действием АТФ-сульфурилазы преобразует аденозинфосфосульфат в молекулу АТФ, которая, в свою очередь, необходима для осуществления ферментативной реакции окисления люциферина люциферазой с выделением квантов света.

В основе секвенирования ДНК, названной авторами ELIDA (Enzymatic Luminometric Inorganic pyrophosphate Detection Assay), лежит та же схема ферментативных реакций, предполагающая последовательные присоединения очередных дНТФ и детекцию высвобождающихся пирофосфатов. Затем неорганический пирофосфат под действием АТФ-сульфурилазы превращается в АТФ и уже молекулы АТФ детектируются с помощью люциферазы жука светлячка, однако было обнаружено, что замена дАТФ его тиопроизводным значительно снижает нежелательный фон, поскольку ДНК-полимераза способна использовать дАТФ $\alpha$ S в качестве субстрата, а люцифераза – нет [Ronaghi et al., 1996].

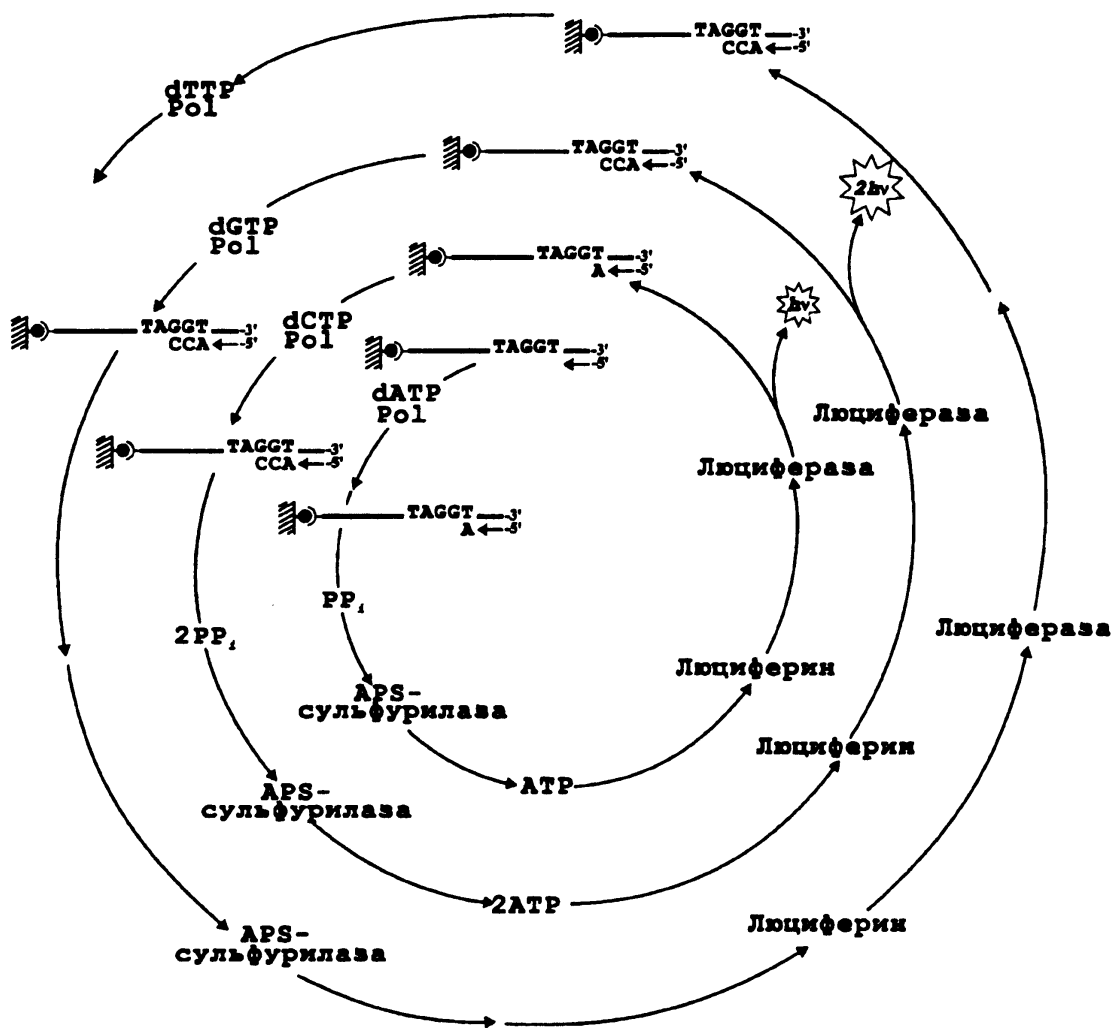


Рис. 12.1. Схема последовательных реакций в секвенировании ДНК ELISA-методом с помощью детекции пирофосфата

Как можно видеть из приведенной на рис. 12.1. схемы секвенирования ДНК посредством понуклеотидной синтеза комплементарной цепи, правильнее говорить не о циклическом характере данного процесса, а скорее как о развивающемся по спирали. К такому выводу можно прийти ввиду того, что каждый (почти каждый) новый цикл начинается с молекулы ДНК, отличающейся от таковой в предыдущем цикле на один или более нуклеотидов.

Ранее этими же авторами был предложен несколько отличающийся подход, названный ими минисеквенированием (рис. 12.2) [Nyren et al., 1993]. Его особенностью было то, что на первом этапе в реакционной смеси присутствовали только ддНТФ, каждый в своей пробирке. Второй этап заключается в удлинении комплементарной цепи ДНК в тех пробирках, где не произошло включения ддНМФ. За счет этого достигалось значительное увеличение сигнала, поскольку при этом происходило включение сразу нескольких или даже многих дНМФ. Тип нуклеотида, находящегося в матрице сразу после праймера, определялся по тому, в какой из пробирок не происходило образование на второй ста-

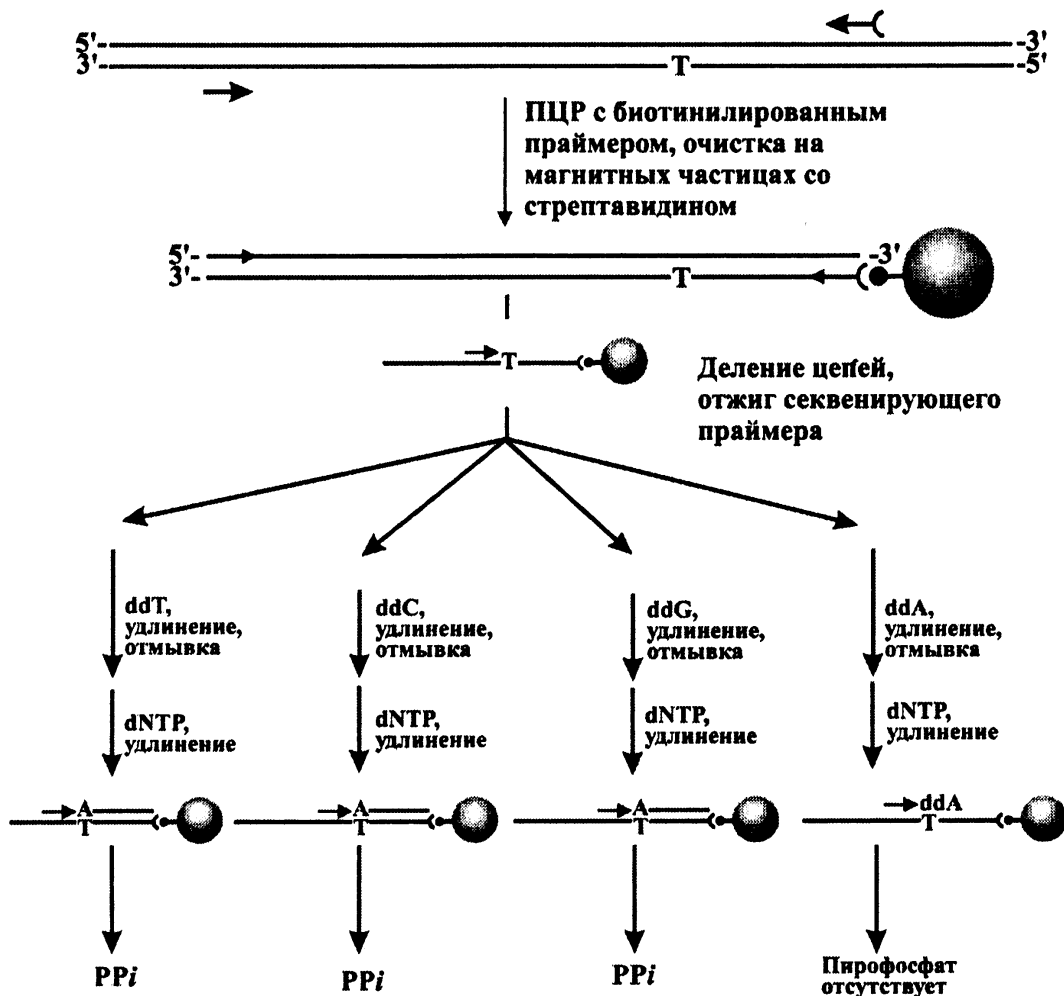


Рис. 12.2. Схема минисеквенирования ДНК посредством синтеза комплементарной цепи и детекции пирофосфата

дии пирофосфата. В качестве контроля этого процесса вводились на первой стадии две дополнительные пробирки со всеми четырьмя ддНТФ одновременно и всеми дНТФ соответственно. Определив таким образом первый нуклеотид, приходилось готовить новые пробирки с добавлением в них этим же дНТФ, удлинять праймер и уже затем проводить опять все описанные выше процедуры по выявлению каких-либо трех невключенных в синтезируемую цепь ДНК ддНТФ.

Даже простое перечисление всех этих этапов само за себя говорит о невероятной громоздкости данной процедуры. В связи с этим авторы говорят о применении этого метода исключительно для выявления предполагаемых мутаций и единичных различий между разными образцами ДНК с заранее известной последовательностью. Более простой подход для выявления замен отдельных нуклеотидов, основанный также на детекции в режиме реального времени образовавшегося пирофосфата, был предложен ими позже [Nyren et al., 1997]. Суть его заключается в использовании двух праймеров, отличающихся своими дистальными нуклеотидами на 3'-конце, что позволяет наблюдать или не наблюдать удлинение цепи ДНК из-за образования или не образования спаренных нуклеотидов на 3'-конце праймера с матрицами ДНК соот-

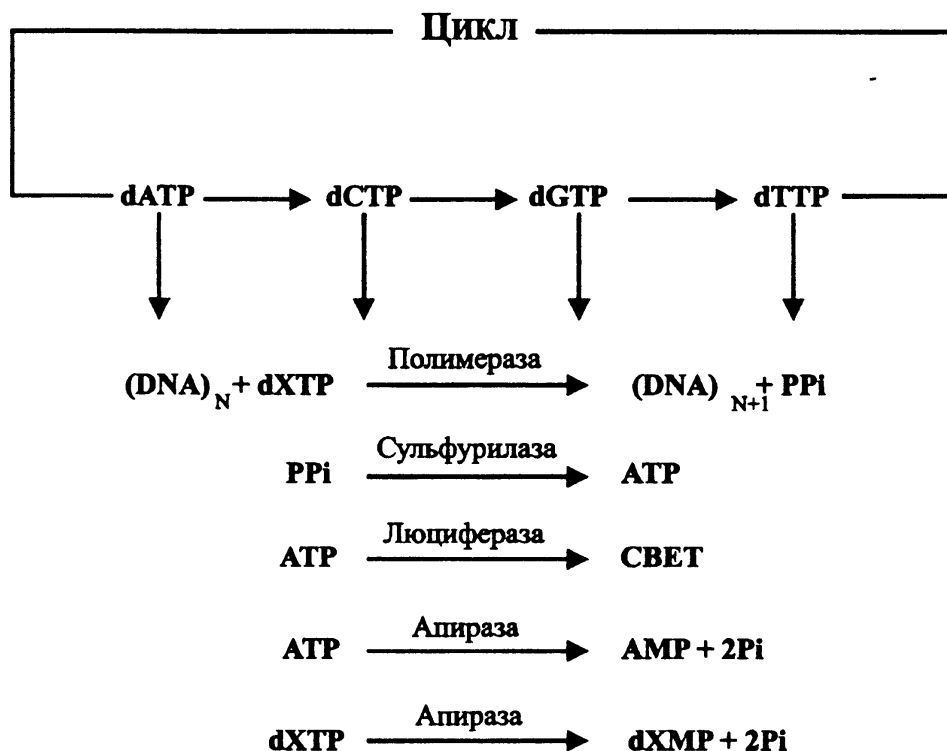


Рис. 12.3. Упрощенная схема этапов пиросеквенирования ДНК

ветственно. Главным условием правильного проведения этого процесса является использование ДНК-полимеразы, не обладающей 3' → 5'-экзонуклеазной редактирующей активностью, способной удалить неспаренные нуклеотиды, что привело бы к недостоверному результату.

В своих предыдущих работах эти авторы использовали в качестве ДНК-полимераз ферменты с отсутствием экзонуклеазных активностей – секвеназу, версия 2.0 и Кленовский фрагмент ДНК-полимеразы I, экзо<sup>-</sup>. Недавно они успешно применили эту технику и для РНК-зависимой ДНК-полимеразы – ревертазы вируса миелобластома птиц [Karamohamed et al., 1998].

Однако наибольший интерес вызывает работа этих авторов, где ими сделаны существенные улучшения процесса секвенирования ДНК ELIDA-методом, что дало им основания и назвать новый подход пиросеквенированием ДНК [Ronaghi et al., 1998]. Так, в цитируемой работе в описанную выше схему было введено важное дополнение в виде нового фермента – апиразы, деградирующей нуклеотидтрифосфаты. Таким образом, в пиросеквенировании ДНК теперь задействовано 4 фермента (рис. 12.3). Роль первых трех известна, а апираза же позволяет автоматизировать этот процесс, удаляя между циклами избыток как образующихся в процессе реакции пирофосфоролиза молекул АТФ, не задействованных люциферазой, так и не включившихся в растущую цепь ДНК дНТФ. За счет этого не требуется проведения трудоемкого и отнимающего значительное время этапа промывки с целью удаления упомянутых выше нежелательных ингредиентов.

Авторы сообщают о созданном ими автоматическом инструменте для пиросеквенирования ДНК, позволяющем проводить одновремен-

ный анализ большого числа образцов с минимальным ручным трудом. К недостаткам метода можно отнести его не очень высокую производительность, а также некоторые трудности секвенирования гомополимерных участков, где происходит одновременное (в течение одного цикла) включение более чем одного одинакового дНМФ. Причиной последнего является непропорциональное увеличение свечения при включении подряд двух и более одинаковых нуклеотидов, однако, как считают сами авторы, специальная компьютерная программа могла бы в значительной мере снять эту проблему.

Справедливости ради следует отметить, что, основываясь на ранних работах этой группы авторов [Nyren, Lundin, 1985; Nyren, 1987], другим экспериментатором была предложена и осуществлена несколько отличающаяся схема секвенирования ДНК путем детекции пирофосфата в режиме реального времени [Hupan, 1988]. Одним из главных отличий было то, что в предложенном им способе реакции протекали не в растворе, а в стеклянном капилляре, представляющем собой одновременно хроматографическую колонку и реакционный сосуд, где ДНК и ферменты были иммобилизованы на трезил-сефарозе 4В. Другим новшеством было добавление к ДНК-полимеразе, АТФ-сульфурилазе и люциферазе еще двух ферментов – глицерокиназы и гексокиназы, с помощью которых удалялось обычно всегда присутствующее в препаратах аденозинфосфосульфата загрязняющее количество АТФ путем превращения последнего в АДФ, что заметно снижало уровень “шумов”. Модельный эксперимент по секвенированию поли(дА · дТ) позволил выявить ряд нюансов при использовании данного метода. Так, была показана линейная зависимость увеличения свечения при одновременном (за один этап) включении нескольких одинаковых нуклеотидов и соответственно увеличенной концентрации высвобождающегося при этом пирофосфата, но при относительно низком количестве иммобилизованной люциферазы. Отрицательным моментом такой линейности явилась более низкая чувствительность самого метода.

## 12.2. СЕКВЕНИРОВАНИЕ ДНК ПУТЕМ ДЕТЕКЦИИ ОДИНОЧНЫХ ФЛУОРЕСЦЕНТНО МЕЧЕННЫХ дНМФ, ПОСЛЕДОВАТЕЛЬНО РЕАЛИЗУЕМЫХ С 3'-КОНЦА ФРАГМЕНТА ДНК ПОД ДЕЙСТВИЕМ ЭКЗОНУКЛЕАЗЫ

Еще один метод секвенирования ДНК, не требующий электрофоретического анализа продуктов реакции, основан на детекции флуоресцентно меченных дНМФ с помощью проточного цитофлуориметра [Jett et al., 1989; Davis et al., 1991; Harding, Keller, 1992]. Основная идея метода заключается в детекции меченных разными флуоресцентными группами дНМФ, последовательно отщепляемых с 3'-конца молекулы ДНК под действием экзонуклеазы. Как можно видеть из несколько упрощенной схемы этого метода, приведенной на рис. 12.4., первым этапом яв-

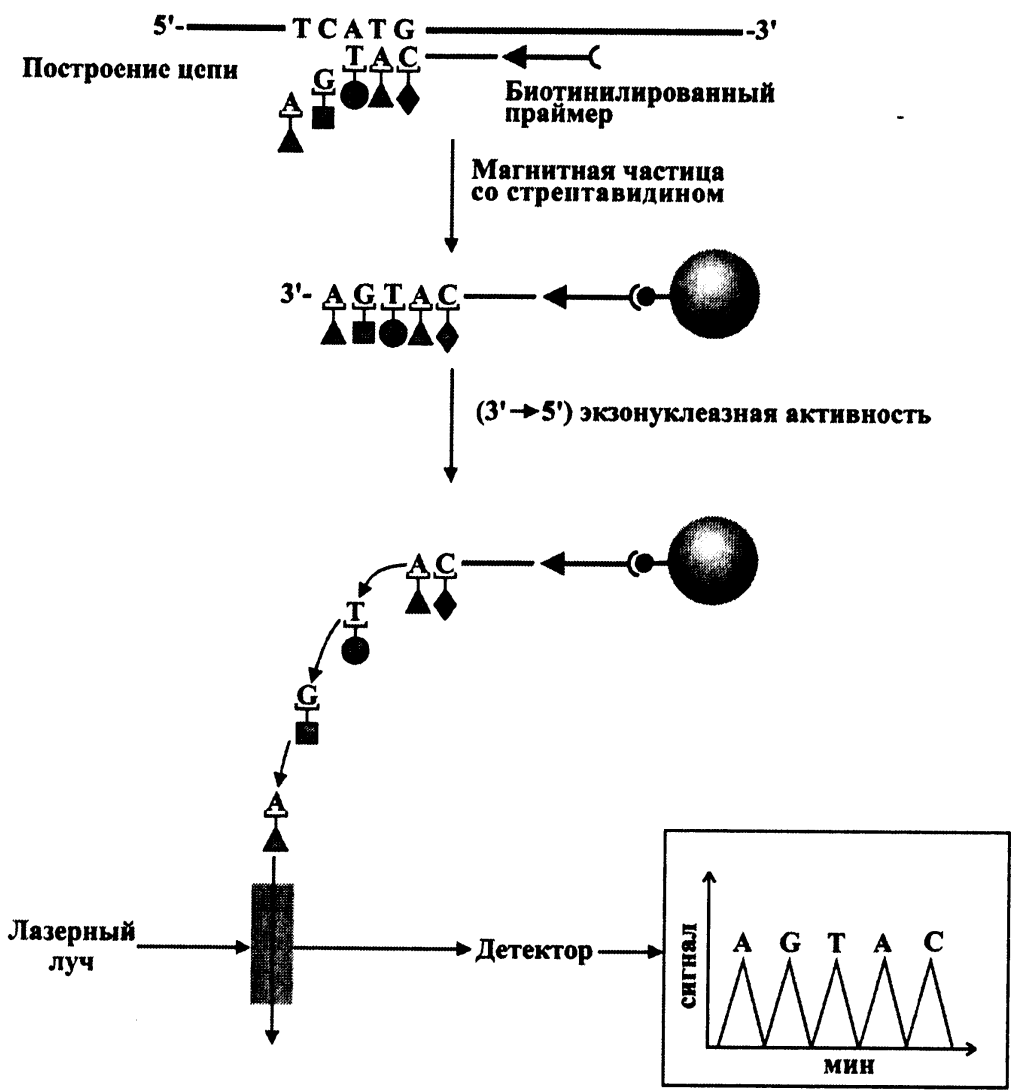


Рис. 12.4. Схема секвенирования ДНК путем детекции одиночных флуоресцентно меченных дНМФ, последовательно реализуемых с 3'-конца фрагмента ДНК под действием экзонуклеазы

ляется построение комплементарной цепи ДНК по существующей матрице с использованием в качестве затравки биотинилированного праймера. Главной особенностью этого процесса является включение в растущую цепь ДНК дНМФ, несущих различные флуоресцентные метки (для каждого типа оснований – своя метка). После денатурации ДНК и сорбции новосинтезированной цепи, несущей молекулу биотина на магнитные частицы со стрептавидином, и тщательного удаления ингредиентов первой реакции следует процесс последовательного отщепления по-разному меченных дНМФ с 3'-конца секвенируемого фрагмента ДНК под действием экзонуклеазы и их детекция с помощью проточного флуориметра. Чередование типов флуоресцентной метки, соответствующих разным нуклеотидам и улавливаемых данным прибором после возбуждения лазерным лучом, позволяет определить первичную структуру секвенируемого фрагмента ДНК.

Несмотря на привлекательность этого подхода к секвенированию ДНК, он имеет ряд недостатков. К главным из них можно отнести тре-



бование присутствия в реакционной смеси новосинтезированных цепей ДНК, характеризующихся одинаковым 3'-концом. Так, если во всех остальных методах ферментативного секвенирования ДНК неизменным условием является гомогенный 5'-конец, то здесь роль "первого" нуклеотида отводится находящемуся на 3'-конце. Причем получение молекул ДНК с гомогенным 5'-концом, принадлежащим в большинстве случаев олигонуклеотидным праймерам, даже несколько проще. Гомогенный же 3'-конец должен образоваться в результате ферментативного действия высокопроцессивной ДНК-полимеразы при построении новой комплементарной цепи ДНК, да еще и состоящей из одних флуоресцентно меченных нуклеотидов, что может заметно снижать эффективность работы фермента.

К достоинствам метода авторы относят его высокую скорость, поскольку экзонуклеаза отщепляет нуклеотиды со скоростью 100–1000 нуклеотидов/с. Другой важной чертой является возможность секвенирования весьма крупных молекул ДНК, не прибегая к необходимости получения субклонов отдельных фрагментов, однако, чем крупнее фрагмент, тем больше вероятность присутствия значительного количества более коротких участков ДНК с различными гетерогенными 3'-концами. Усовершенствование приборов детекции и повышение их чувствительности, обнаруженные новые ДНК-полимеразы и экзонуклеазы заставляют авторов с оптимизмом смотреть в будущее [Roslaniec et al., 1997].

### 12.3. СЕКВЕНИРОВАНИЕ ДНК ПУТЕМ ВОЗОБНОВЛЕНИЯ СИНТЕЗА КОМПЛЕМЕНТАРНОЙ ЦЕПИ ДНК ПОСЛЕ УДАЛЕНИЯ ТЕРМИНАТОРНОЙ ГРУППЫ У КОНЦЕВОГО НУКЛЕОТИДА

Другой, пока только теоретически существующий способ определения нуклеотидной последовательности ДНК без применения секвенирующего гель-электрофореза, названный авторами BASS (Base Addition Sequencing Scheme), основан на использовании особого рода терминаторов цепи ДНК, которые под действием УФ-света или какого-либо другого воздействия теряют блокирующую дальнейший синтез ДНК-группу и, таким образом, рост цепи ДНК возобновляется [Metzker et al., 1994]. В разделе 2.6 уже упоминался чувствительный к ультрафиолетовому свету терминатор построения комплементарной цепи ДНК – 3'-О-(2-нитробензил)-дАТФ. На рис. 12.5. показана принципиальная схема такого способа секвенирования ДНК.

В BASS-способе олигонуклеотидный праймер сорбирован на твердом носителе и после отжига на нем секвенируемого фрагмента ДНК под действием ДНК-полимеразы осуществляется понуклеотидный циклический синтез новой комплементарной цепи ДНК. К особенностям такого секвенирования ДНК путем синтеза новой цепи ДНК следует отнести отсутствие какой-либо радиоактивной или флуоресцентной метки у праймера, а также отсутствие в реакционной смеси стандарт-

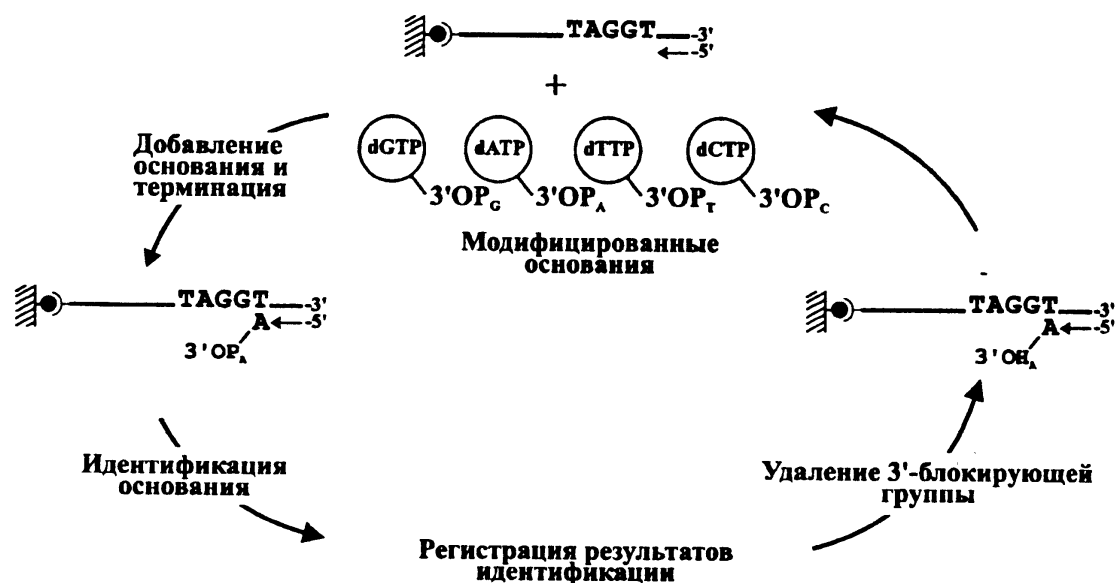


Рис. 12.5. Теоретическая схема определения нуклеотидной последовательности ДНК BASS-способом

ных субстратов для ДНК-полимераз, каковыми являются дНТФ. Таким образом, новосинтезируемая цепь ДНК формируется исключительно из модифицированных по 3'-концу дНТФ. Детекция очередного присоединенного нуклеотида и его определение происходят спектроскопически за счет того, что специфические модифицированные дНТФ должны нести (теоретически) уникальные блокирующие группы для каждого типа азотистых оснований, позволяющих осуществить их достоверное распознавание. После этапа депротекции под действием УФ-света дальнейший синтез цепи возобновляется и присоединяется очередной тот или иной комплементарный нуклеотид. Последовательное циклическое проведение всех этих стадий позволит (теоретически) определить нуклеотидную последовательность секвенируемой матрицы ДНК.

Аналогичная схема секвенирования ДНК путем временного прерывания и дальнейшего возобновления синтеза комплементарной цепи ДНК была предложена и другими авторами [Canard, Sarfati, 1994]. В данной работе была синтезирована серия 2'-дезоксирибонуклеотид трифосфатов (дАТФ, дГТФ, дТТФ и дЦТФ), у которых остаток гидроксила в 3'-положении дезоксирибозы был замещен производными антрацилина, причем с отличающимися спектрами эмиссии для каждого типа азотистых оснований, давая, таким образом, возможность их регистрации. Было показано, что эти соединения с достаточно высокой эффективностью могут служить субстратами для различных ДНК-полимераз при построении новой цепи ДНК. Так, после распознавания включенного модифицированного дНМФ и его регистрации проводилась обработка 0,1M NaOH с последующей нейтрализацией и синтез комплементарной цепи ДНК возобновлялся. Однако, к сожалению, в отличие от УФ чувствительного 3'-O-(2-нитробензил)-дАТФ, описанного выше, здесь для дальнейшего удлинения праймера требовалось больше усилий и дополнительные стадии в виде добавления различных химических ре-

агентов. Некоторой альтернативой химической депротекции может быть, как сообщают авторы, ферментативная обработка с помощью какой-либо специфичной эстеразы, также способной удалить блокирующую флуоресцентную группу.

#### 12.4. ПРОЧИЕ МЕТОДЫ СЕКВЕНИРОВАНИЯ ДНК БЕЗ ИСПОЛЬЗОВАНИЯ ЭЛЕКТРОФОРЕТИЧЕСКОГО РАЗДЕЛЕНИЯ

Известно, что сканирующий туннельный микроскоп позволяет получать изображения молекул ДНК с высоким разрешением [Veebe et al., 1989; Briscoll et al., 1990]. Так, в литературе описана экспериментальная схема секвенирования протяженных фрагментов ДНК с его помощью [Kelson, Nussinov, 1994]. В основе предложенного метода лежит синтез комплементарной цепи ДНК, где один из дНТФ в реакционной смеси представлен исключительно в меченном виде, причем высокоэнергетическим радионуклидом  $^{32}\text{P}$ . Таким образом, первая стадия подготовки препарата ДНК к секвенированию заключается в параллельном проведении 4 реакций с разным типом меченых нуклеотидных оснований в каждой. Далее, вновь синтезированные цепи ДНК сорбируются на соответствующую подложку и идет период ожидания, необходимый для полного распада короткоживущего радионуклида  $^{32}\text{P}$  и превращения его в стабильный изотоп  $^{32}\text{S}$ . Образовавшиеся атомы серы не могут уместиться в места, оставленные им атомами фосфора, что приводит к разрыву цепи ДНК и образованию набора коротких участков, ограниченных в каждом типе реакций соответствующими нуклеотидами. Авторы считают, что определение линейных размеров (что позволяет перейти к числу азотистых оснований в цепи) таких коротких участков ДНК с помощью сканирующего туннельного микроскопа и сравнительный компьютерный анализ теоретически могут позволить восстановить нуклеотидную последовательность анализируемой молекулы ДНК.

Весьма интересны 2 недавно предложенных метода секвенирования ДНК с использованием рестрикционных эндонуклеаз класса II, расщепляющих последовательности ДНК вне своего сайта узнавания [Jones, 1997]. Суть данных методов заключается в по-шаговом секвенировании единичных нуклеотидов с помощью лигирования секвенируемого фрагмента ДНК (предварительно расщепленного этим же ферментом) с набором специальных адапторов, несущих сайт узнавания используемой рестрикционной эндонуклеазы, а также три неспаренные вырожденные нуклеотида на 5'-конце. В результате инкубации секвенируемой матрицы на предшествующем этапе в растворе ддНТФ, несущих метку в виде  $^{33}\text{P}$  или в виде флуоресцентных красителей (различных для каждого типа оснований) и последующего распознавания включившегося определенного ддНМФ (или флуорометрией, или с помощью жидкостного сцинтилляционного счетчика) происходит сдвиг "чтения" дальнейших нуклеотидов. Циклический характер подобного

секвенирования позволяет, таким образом, осуществлять по-шаговое "чтение" очередных нуклеотидов и определять последовательность до 150 нуклеотидов длиной за достаточно короткий промежуток времени. Проведенные автором подсчеты показали теоретически очень высокую производительность данного метода при условии проведения всех процедур в микропланшетах с 384 лунками. Она может еще более возрасти при использовании биочипового формата в сопровождении с соответствующим детектором.

Целая серия статей посвящена способу ферментативного секвенирования ДНК, где электрофоретическое разделение продуктов терминирующих реакций заменено их детекцией с помощью масс-спектрометрического метода (matrix assisted laser desorption/ionization или иначе MALDI) [Tang et al., 1994, 1995; Fu et al., 1996; Koster et al., 1996; Roskey et al., 1996; Ху et al., 1997; Berkenkamp et al., 1998]. Однако следует признать, что несмотря на сложность самого приборного оснащения, секвенировать удается обычно не более 40–50 нуклеотидов. Таким образом, данный подход для секвенирования *de novo* не пригоден и может использоваться, например, для диагностических целей или генотипирования [Laken et al., 1998]. В то же время справедливости ради следует отметить, что имеется единичное сообщение, где авторам удалось с помощью данного метода детектировать 500 нуклеотидов [Tang et al., 1994].

### Цитируемая литература

- Beebe T.P., Jr., Wilson T.E., Ogletree D.F., Katz J.E., Balhorn R., Salmeron M.B., Siekhaus W.J. Direct observation of native DNA structure with the scanning tunneling microscope // *Science*. 1989. Vol. 243. P. 370–372.
- Berkenkamp S., Kirpekar F., Hillenkamp F. Infrared MALDI mass spectrometry of large nucleic acids // *Science*. 1998. Vol. 281. P. 260–262.
- Canard B., Sarfati R.S. DNA polymerase fluorescent substrates with reversible 3'-tags // *Gene*. 1994. Vol. 148. P. 1–6.
- Davis L.M., Fairfield F.R., Harger C.A., Jett J.H., Keller R.A., Hahn J.H., Krakowski L.A., Marrone B.L., Martin J.C., Nutter H.L., Ratliff R.L., Shera E.B., Simpson D.J., Soper S.A. Rapid DNA sequencing based upon single molecule detection // *Genet. Anal. Techn. Appl.* 1991. Vol. 8. P. 1–7.
- Driscoll R.J., Youngquist M.G., Baldeschwieler J.D. Atomic-scale imaging of DNA using scanning tunnelling microscopy // *Nature*. 1990. Vol. 346. P. 294–296.
- Fu D.J., Koster H., Smith C.L., Cantor C.R. Efficient preparation of short DNA sequence ladders potentially suitable for MALDI-TOF DNA sequencing // *Genet. Anal.* 1996. Vol. 12. P. 137–142.
- Harding J.D., Keller R.A. Single-molecule detection as an approach to rapid DNA sequencing // *Trends Biotechnol.* 1992. Vol. 10. P. 55–57.
- Hyman E.D. A new method of sequencing DNA // *Anal. Biochem.* 1988. Vol. 174. P. 423–436.
- Jett J.H., Keller R.A., Martin J.C., Marrone B.L., Moyzis R.K., Ratliff R.L., Seitzinger N.K., Shera E.B., Stewart C.C. High-speed DNA sequencing: An approach based upon fluorescence detection of single molecules // *J. Biomol. Struct. Dyn.* 1989. Vol. 7. P. 301–309.
- Jones D.H. An iterative and regenerative method for DNA sequencing // *Biotechniques*. 1997. Vol. 22. P. 938–946.
- Karamohamed S., Ronaghi M., Nyren P. Bioluminometric method for real-time detection of reverse transcriptase activity // *Ibid.* 1998. Vol. 24. P. 302–306.
- Kelson I., Nussinov S. A scheme for sequencing large DNA molecules by identifying local nuclear-induced effects // *Proc. Nat. Acad. Sci. USA*. 1994. Vol. 91. P. 6963–6966.

- Koster H., Tang K., Fu D.J., Braun A., van den Boom D., Smith C.L., Cotter R.J., Cantor C.R.* A strategy for rapid and efficient DNA sequencing by mass spectrometry // *Nat. Biotechnol.* 1996. Vol. 14. P. 1123–1128.
- Laken S.J., Jackson P.E., Kinzler K.W., Vogelstein B., Strickland P.T., Croopman J.D., Friesen M.D.* Genotyping by mass spectrometric analysis of short DNA fragments // *Ibid.* 1998. Vol. 16. P. 1352–1356.
- Metzker M.L., Raghavachari R., Richards S., Jacutin S.E., Civitello A., Burgess K., Gibbs R.A.* Termination of DNA synthesis by novel 3'-modified-deoxyribonucleoside 5'-triphosphates // *Nucl. Acids Res.* 1994. Vol. 22. P. 4259–4267.
- Nyren P.* Enzymatic method for continuous monitoring of DNA polymerase activity // *Anal. Biochem.* 1987. Vol. 167. P. 235–238.
- Nyren P., Karamohamed S., Ronaghi M.* Detection of single-base changes using a biolumino-metric primer extension assay // *Ibid.* 1997. Vol. 244. P. 367–373.
- Nyren P., Lundin A.* Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis // *Ibid.* 1985. Vol. 151. P. 504–509.
- Nyren P., Pettersson B., Uhlen M.* Solid phase DNA minisequencing by an enzymatic lumi-nometric inorganic pyrophosphate detection assay // *Ibid.* 1993. Vol. 208. P. 171–175.
- Porter K.W., Briley J.D., Shaw B.R.* Direct PCR sequencing with boronated nucleotides // *Nucl. Acids Res.* 1997. Vol. 25. P. 1611–1617.
- Ronaghi M., Karamohamed S., Pettersson B., Uhlen M., Nyren P.* Real-time DNA sequen-cing using detection of pyrophosphate release // *Anal. Biochem.* 1996. Vol. 242. P. 84–89.
- Ronaghi M., Uhlen M., Nyren P.* A sequencing method based on real-time pyrophosphate // *Science.* 1998. Vol. 281. P. 363, 365.
- Roskey M.T., Juhasz P., Smirnov I.P., Takach E.J., Martin S.A., Haff L.A.* DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry // *Proc. Nat. Acad. Sci. USA.* 1996. Vol. 93. P. 4724–4729.
- Roslaniec M.C., Bell-Prince C.S., Crissman H.A., Fawcett J.J., Goodwin P.M., Habbersett R., Jett J.H., Keller R.A., Martin J.C., Marrone B.L., Nolan J.P., Park M.S., Sailer B.L., Sklar L.A., Steinkamp J.A., Cram L.S.* New flow cytometric technologies for the 21st century // *Human. Cell.* 1997. Vol. 10. P. 3–10.
- Tang K., Fu D.J., Kotter S., Cantor C.R., Koster H.* Matrix-assisted laser desorption/ioniza-tion mass spectrometry of immobilized duplex DNA probes // *Nucl. Acids Res.* 1995. Vol. 23. P. 3126–3131.
- Tang K., Taranenko N.I., Allman S.L., Chang L.Y., Chen C.H.* Detection of 500-nucleotide DNA by laser desorption mass spectrometry // *Rapid Commun. Mass Spectrom.* 1994. Vol. 8. P. 727–730.
- Xu L., Bian N., Wang Z., Abdel Baky S., Pillai S., Magiera D., Murugaiah V., Wang P., O'Keeffe T., Abushmaa I.J., Kutney I., Church G., Carson S., Smith D., Park M., Wronka J., Laukien F.* Electrophore mass tag dideoxy DNA sequencing // *Anal. Chem.* 1997. Vol. 69. P. 3595–3602.

## СЕКВЕНИРОВАНИЕ ДНК ПОСРЕДСТВОМ ГИБРИДИЗАЦИИ

Продолжающийся поиск альтернативных возможностей определения нуклеотидной последовательности ДНК привел к разработке принципиально нового метода секвенирования ДНК посредством гибридизации. В его основу было положено представление, что вся последовательность ДНК состоит из коротких подпоследовательностей. Знание порядка нуклеотидов в них, а также расположение этих подпоследовательностей друг относительно друга позволяют реконструировать весь исследуемый фрагмент ДНК. Принцип определения нуклеотидной последовательности секвенируемого фрагмента ДНК этим методом основан на образовании Уотсон–Криковских пар в результате проведения этапа гибридизации фрагмента ДНК с набором олигонуклеотидных зондов. Анализ характера гибридизации конкретных олигонуклеотидов позволяет провести реконструкцию секвенируемого фрагмента ДНК.

В конце 1980-х годов, практически одновременно, четыре исследовательские группы (две в Англии, одна в России и еще одна в Югославии) предложили различные варианты этого непривычного подхода к секвенированию ДНК [Лысов и др., 1988; Bains, Smith, 1988; Dumanac et al., 1989; Southern, 1992]. Главное их отличие между собой заключалось в том, что одни авторы для выявления присутствия в ДНК конкретных олигонуклеотидных блоков использовали в качестве меченого зонда сам секвенируемый фрагмент, находящийся в растворе, тогда как другие, напротив, сорбировали исследуемый фрагмент на мембрану и проводили гибридизацию с находящимися в растворе мечеными олигонуклеотидами. Последний вариант известен еще как формат 1, тогда как упомянутый здесь первым носит обозначение “формат 2”.

В настоящее время такой способ секвенирования ДНК, хотя и не стал массовым, уже широко известен и в англоязычной литературе существует специальная аббревиатура SBH или SbH (Sequencing by Hybridization). В отечественной литературе на протяжении ряда лет использовалось сокращение СГОМ (Секвенирование Гибридизацией с Олигонуклеотидной Матрицей, или Микрочипом), позже переименованное в ГНОМ (Гибридизация На Олигонуклеотидной Матрице).

Надо отметить относительную простоту процедур, требующихся для определения последовательности нуклеотидов ДНК данным методом, особенно по сравнению с секвенированием ДНК, основанным на полимеризации (метод Сэнгера) и деградации (метод Максама–Гилберта). Однако, к сожалению, существует ряд ограничений, из-за которых метод секвенирования ДНК гибридизацией так и не стал основным. Пожалуй, главными препятствиями этому служат серьезные проблемы при секвенировании повторяющихся элементов генома и “чтение” относительного малого числа нуклеотидов в ходе одного эксперимента. (Теоретически существующие возможности повышения разрешающей

способности данного метода сдерживает нынешний технологический уровень.) В то же время при исследовании полиморфизма отдельных генов с уже известной последовательностью, выявлении мутаций в определенных генах секвенирование гибридизацией имеет заметное преимущество в скорости и простоте по сравнению с остальными методами, что позволяет использовать его в диагностических целях [Drobyshev et al., 1997; Case-Green et al., 1998]. Учитывая, что секвенирование ДНК-гибридизацией основано на другом принципе, чем методы, ставшие уже классическими, то оно может быть применено для подтверждения нуклеотидных последовательностей, хранящихся в международных банках данных или выявления в них возможных ошибок.

Словом, потенциал секвенирования ДНК гибридизацией достаточно велик, что и объясняет повышенный интерес ряда фирм и ведущиеся споры вокруг патента на этот метод, о правах на который в 1988 г. заявил английский ученый Э. Саузерн [Vains, 1997; Butler, 1997].

### 13.1. ГИБРИДИЗАЦИЯ МЕЧЕНЫХ ОЛИГОНУКЛЕОТИДОВ С ФИКСИРОВАННОЙ СЕКВЕНИРУЕМОЙ ДНК

Метод молекулярной ДНК/ДНК-гибридизации какого-либо зонда, представляющего собой определенный фрагмент одноцепочечной ДНК, с фиксированной на мембранном фильтре другой одноцепочечной ДНК, известен достаточно давно и применяется в молекулярной биологии при решении большого числа задач чрезвычайно широко. Однако цели в виде определения последовательности нуклеотидов сорбированной ДНК при этом ранее никогда не ставилось. Подобная попытка была совершена югославскими авторами, детально описавшими в своей статье теорию самого метода [Drmanac et al., 1989], причем ранее в 1987 г. ими был получен югославский патент на данный метод. Так, они показали, что характер гибридизации сорбированного на нейлоновом фильтре исследуемого фрагмента ДНК с короткими олигонуклеотидными зондами (длиной от 11 до 20 звеньев) позволяет реконструировать последовательность нуклеотидов в этом секвенируемом фрагменте. В дальнейшем ими была продемонстрирована возможность секвенирования с помощью данного метода фрагмента ДНК протяженностью 100 пн [Strezoska et al., 1991]. Предложенная автоматизация всего процесса гибридизации (секвенирования) теоретически могла позволить определять до 100 мпн в год [Drmanac et al., 1992]. Но при этом предполагался гигантский объем экспериментальной работы по подготовке самих образцов ДНК, их гибридизации с олигонуклеотидными зондами и последующим анализом полученных результатов, требующим каждодневного обсчета около 5 млн гибридизационных сигналов. В своей другой статье эти авторы предположили, что секвенирование ДНК посредством гибридизации может ускорить прохождение фазы картирования генома человека при выполнении этого глобального проекта [Drmanac et al., 1993].

Главным недостатком этого варианта секвенирования ДНК посредством гибридизации, известного также как формат 1, является необхо-

димось очень большого числа повторных операций по гибридизации нуклеиновых кислот и прочих необходимых при этом процедур. Следует отметить, что в упоминавшихся работах этих авторов в качестве метки использовался радионуклид  $^{32}\text{P}$ . В недавней работе другими исследователями было применено уже мультиплексное секвенирование ДНК посредством гибридизации, ставшее возможным за счет использования в качестве метки стабильных изотопов олова с различными массовыми числами, позволяющими произвести их отдельную регистрацию при помощи ионизационной масс-спектрометрии [Arlinghaus et al., 1997a]. Однако такой мультиплексный подход, несмотря на свое удобство, все же лишь незначительно решает проблему многочисленных ДНК/ДНК-гибридизаций.

### 13.2. ГИБРИДИЗАЦИЯ МЕЧЕНОЙ ДНК С ФИКСИРОВАННЫМИ ОЛИГОНУКЛЕОТИДАМИ—“ДНК-чип” ТЕХНОЛОГИЯ

Другой альтернативный подход к секвенированию ДНК посредством гибридизации, так называемый формат 2, заключается в инкубации сорбированных на подходящем носителе в строго определенных местах коротких олигонуклеотидов с находящейся в растворе секвенируемой меченой ДНК [Лысов и др., 1988; Bains, Smith, 1988; Khrapko et al., 1989; Southern, 1992]. Определение последовательности нуклеотидов этим методом состоит из нескольких экспериментальных стадий и анализа полученных результатов. Так, после завершения гибридизации и последующей отмывки происходит выявление среди множества остальных только тех олигонуклеотидов, которые сформировали совершенные дуплексы с одноцепочечной ДНК, находящейся в растворе. Поскольку последовательность самих олигонуклеотидов (которые могут быть представлены как подпоследовательность конкретного участка ДНК) известна, то анализ олигонуклеотидов, вступающих в гибридизацию, позволяет определить среди них таковые с перекрывающимися последовательностями и рассчитать их расположение друг относительно друга, что и приводит в итоге к реконструкции последовательности нуклеотидов секвенируемого фрагмента ДНК. Примечательной особенностью этого метода можно считать возможность “чтения” последовательности ДНК в обоих направлениях и с любого произвольного места, последовательно находя перекрывающиеся олигонуклеотидные блоки. К другой важной черте секвенирования ДНК-гибридизацией можно отнести то, что даже в одном эксперименте практически каждый нуклеотид в секвенируемом фрагменте ДНК “читается” многократно. Так, например, при использовании октануклеотидной матрицы для большинства нуклеотидов повторное “чтение” будет осуществляться еще 7 раз за счет перекрытия последовательностей данных олигонуклеотидов с таковой у выбранного в качестве первого. К недостатку, присущему, главным образом, формату 2, можно отнести влияние сильной вторичной структуры отдельных участков секвенируемого фрагмента ДНК,



не позволяющей образовывать с ней дуплексы олигонуклеотидам, фиксированным в матрице.

Таким образом, основными составляющими процесса секвенирования ДНК посредством гибридизации являются сама секвенируемая ДНК, несущая подходящую метку, и полный набор олигонуклеотидов определенной длины, формирующих специальную олигонуклеотидную матрицу. За годы, прошедшие со времени разработки данного метода, все эти элементы претерпели существенные изменения. Так, в ранних экспериментах в качестве метки использовались радиоактивно меченные молекулы, что требовало этапа радиоавтографии или применения специальных приборов, как, например, PhosphorImager (Molecular Dynamics, США) или ему подобных. Впоследствии радиоактивная метка была заменена флуорофорами, что позволило в еще большей степени автоматизировать весь процесс сбора данных, использовав при этом флуоресцентный микроскоп, и создать таким образом прототип автоматического секвенатора ДНК, включавший также CCD-камеру и компьютер [Мирзабеков, 1992; Mirzabekov, 1994; Yershov et al., 1996]. Следует отметить специально написанную компьютерную программу DNA-SPEKTRUM, позволяющую проводить реконструкцию последовательности нуклеотидов секвенированного фрагмента ДНК и обладающую целым рядом других сервисных функций для данного метода секвенирования ДНК [Belyi, Pevzner, 1997].

Центральным элементом метода секвенирования ДНК-гибридизацией служит олигонуклеотидная матрица, или микрочип, на котором в строго определенных местах фиксированы конкретные олигонуклеотиды с известной последовательностью. Если для использования данного метода в диагностических целях можно ограничиться лишь несколькими конкретными олигонуклеотидами, то для секвенирования *de novo* необходимо иметь полный набор олигонуклеотидов, состоящий из  $4^n$  вариантов. При этом считается, что длина фрагмента ДНК, который можно “просеквенировать” в одном эксперименте, приблизительно равна квадратному корню из числа всех возможных олигонуклеотидов в данном наборе [Southern et al., 1992]. В работе Певзнера и соавт. [Певзнер и др., 1991] приведено приблизительное соотношение длины применяемых олигонуклеотидов в матрице и протяженности фрагмента ДНК, нуклеотидная последовательность которого может быть однозначно установлена. Так, например, около 2,5 тыс. нуклеотидов теоретически можно “прочитать” в одном эксперименте с помощью полного набора олигонуклеотидов, состоящих из 12 звеньев. Однако сама матрица при этом будет состоять из огромного числа  $16\,777\,216$  вариантов. Набор из  $262\,144$  9-мерных олигонуклеотидов, хотя теоретически и позволяет однозначно определять нуклеотидную последовательность фрагмента ДНК протяженностью около 500 нуклеотидов, в настоящее время слишком велик для изготовления подобной матрицы. Достаточно оптимальной можно считать матрицу с длиной олигонуклеотидов, равной 8 звеньям, что составляет всего  $65\,536$  всех возможных вариантов таких молекул, изготовление которой также весьма не простой процесс. Приведенная на рис. 13.1 несколько упрощенная схема иллю-

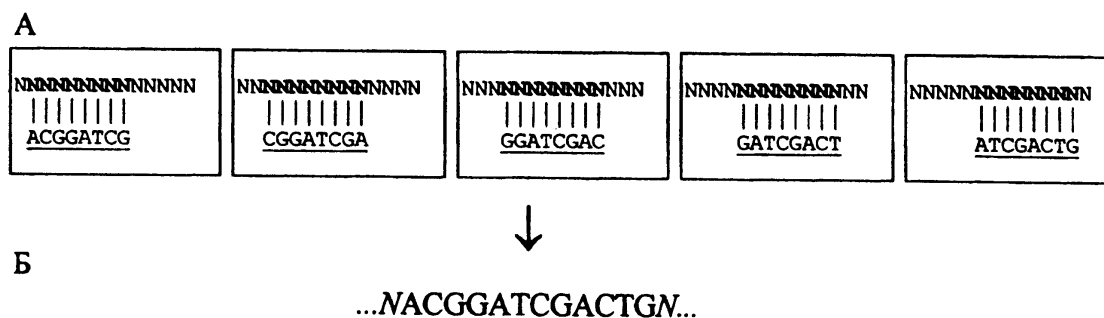


Рис. 13.1. Реконструирование нуклеотидной последовательности с помощью метода секвенирования ДНК посредством гибридизации

А – отдельные ячейки олигонуклеотидной матрицы, содержащие различные октануклеотиды (подчеркивание символизирует их фиксацию на подложке), с произошедшей там гибридизацией с исследуемым фрагментом ДНК (вертикальные черточки изображают связи между комплементарными олигонуклеотидами, выделенными жирным шрифтом); Б – реконструированная часть нуклеотидной последовательности секвенируемого фрагмента ДНК

стрирует процесс восстановления последовательности нуклеотидов исследуемого фрагмента ДНК именно с помощью октануклеотидной матрицы.

Как будет видно из дальнейшего изложения, в процессе продолжающейся разработки метода секвенирования ДНК посредством гибридизации предлагались различные варианты модификации олигонуклеотидов как иммобилизованных в матрице, так и образования на их основе за счет стэкинг-взаимодействий некоего гибридизационного комплекса с участием более коротких олигонуклеотидов, как несущих, так и не несущих метку.

Не осталась без изменения и твердая фаза, на которой фиксировались синтезированные олигонуклеотиды. Так, первоначально некоторый недостаток секвенирования ДНК-гибридизацией с помощью формата 2 по сравнению с описанным в предыдущем разделе форматом 1 заключался в плохой сорбции коротких олигонуклеотидов ДНК на мембранных фильтрах. Чтобы преодолеть это препятствие, разными авторами были применены отличающиеся способы их фиксации. Так, в некоторых случаях в качестве носителя выступало модифицированное стекло [Maskos, Southern, 1992] или полипропилен с аминогруппами [Matson et al., 1995; Weiler, Hoheisel, 1996]. Описывается применение процесса фотолитографии для создания матриц с 256 различными октануклеотидами [Fodor et al., 1993; Pease et al., 1994]. Весьма сложным расположением ячеек характеризуется олигонуклеотидная матрица, созданная Саузерном и соавт. [Southern et al., 1992; Southern et al., 1994; Southern, 1996]. Целая серия работ этой группы авторов была посвящена анализу особенностей взаимодействия олигонуклеотидов, фиксированных на данной матрице, при их гибридизации с мишенью [Southern et al., 1994; Williams et al., 1994]. Специальное исследование было направлено на оценку стабильности образующихся дуплексов мишени с олигонуклеотидами, содержащими дезоксиинозин как универсальное основание [Case-Green, Southern, 1994]. В недавней работе этих авторов изуча-

лись стерические факторы, воздействующие на эффективность гибридизации олигонуклеотидов, фиксированных при помощи специального спейсера на полипропиленовой матрице, содержащей функциональные аминогруппы [Shcherpinov et al., 1997]. Было обнаружено, что оптимальная длина спейсера в 40 атомов повышала эффективность гибридизации в 150 раз.

Отечественные ученые под руководством А.Д. Мирзабекова оставили свой выбор на использовании в качестве носителя тонкого слоя 8%-ного полиакриламидного геля, приполимеризованного к стеклу, обработанному  $\gamma$ -метакрилоксипропилтриметоксисиланом [Храпко и др., 1991]. К преимуществам данного подхода можно отнести хорошую прозрачность и прочность этого носителя, а также его значительную емкость и, как следствие, увеличенное количество молекул сорбируемых олигонуклеотидов и более высокий уровень гибридизационного сигнала. В то же время возникла проблема проникновения меченого секвенируемого фрагмента ДНК внутрь такого матрикса, что, впрочем, было решено этими авторами путем некоторой фрагментации исследуемой ДНК. Сорбцию окисленных периодатом натрия олигонуклеотидов, содержащих 3-метилуридин в качестве линкера, проводили на матрице, полиакриламидный матрикс которой был предварительно подготовлен обработкой гидразингидратом [Храпко и др., 1991]. В дальнейшем, кроме полиакриламидного геля, исследовалась возможность применения для этих целей и сополимерных полидиметил-акриламидных гелей, содержащих в качестве реакционноспособных альдегидную или аминную группы [Timofeev et al., 1996]. Так, активная альдегидная группа показала более высокую эффективность иммобилизации олигонуклеотидов. Полидиметил-акриламидный гель в отличие от обычного полиакриламидного позволил провести его дополнительную обработку в органических растворителях.

Недавно Мирзабековым и соавт. [Василисков и др., 1998] разработан новый способ изготовления микрочипов с улучшенными свойствами, заключающийся в сополимеризации 4%-ного полиакриламида с синтезированными олигонуклеотидами, несущими на 5'-конце аллильную группу, добавляемую на последнем этапе обычного синтеза в виде аллильного производного фосфорамидита. За счет этого было достигнуто более равномерное распределение олигонуклеотидов по всему объему гелевой ячейки и процесс полимеризации (изготовления ячеек) и иммобилизация ("пришивка" олигонуклеотидов) проходили в одну стадию в отличие от формирования ячеек с помощью робота, описанного в другой работе этих же авторов [Yerskov et al., 1996].

Сам размер олигонуклеотидной матрицы, или микрочипа, за эти годы изменился также весьма существенно. Так, первые конструкции изготавливались вручную, механически удаляя часть полимеризованного геля с целью получения квадратов со сторонами в 200 мкм, разделенных полосками чистого стекла шириной 300 мкм [Храпко и др., 1991]. Толщина гелевого слоя составляла 30 мкм. Совершенствование подобных чипов для процесса секвенирования ДНК посредством гибридизации привело к созданию ячеек толщиной 20 мкм и размером 40×40 мкм,

разделенных промежутками в 80 мкм [Yershov et al., 1996]. Для изготовления таких миниатюрных чипов, имеющих общий размер 1×1 см, и способных нести в ячейках гелевого матрикса, полимеризованного на стекле, от 20 до 30 тыс. вариантов олигонуклеотидов, был создан специальный робот. Авторы сообщают, что подобный микрочип выдерживал до 15 и в отдельных случаях до 50 гибридизаций, а перед использованием мог храниться в высушенном виде при 4 °С в течение года. Для иммобилизации же этим способом полного набора всех возможных октануклеотидов, состоящего из 65536 вариантов, потребовалась бы стеклянная пластинка с размерами около 3×3 см.

Относительно недавно этой же группой авторов была предложена простая процедура, позволяющая ручным способом изготовить аналогичный микрочип с еще меньшими размерами ячеек (25×25 мкм), разделенных полосками гидрофобизированной поверхности покровного стекла к микроскопу, на котором и размещались данные ячейки [Guschin et al., 1997]. Дальнейшая миниатюризация подобных ДНК-чипов привела к созданию с помощью специально сконструированного микроскопа фрагмента матрицы с гелевыми ячейками размером всего 10×10×5 мкм [Василисков и др., 1998].

Молекулярная гибридизация иммобилизованных олигонуклеотидов с меченой ДНК проводится при температуре около 0°С и после серии отмывок при температуре гибридизации с дальнейшим дискретным повышением на 5°С, обеспечивающих удаление несвязавшейся мишени и диссоциации несовершенных дуплексов осуществляется детекция флуоресцирующих (или в ранних экспериментах дающих сигнал при радиоавтографии) ячеек, свидетельствующая об образовании содержащимися в них олигонуклеотидами совершенных дуплексов с секвенируемой ДНК. Одним из ключевых моментов метода секвенирования ДНК с помощью гидридизации с олигонуклеотидными матрицами является четкое разграничение тех ячеек с совершенными дуплексами от таких, несущих отдельные неспаренные нуклеотиды. Учитывая важность постгибридизационной отмывки, ей была посвящена отдельная статья, детально рассматривающая различные теоретические аспекты этого процесса [Лившиц и др., 1992; Livshits et al., 1994]. Не меньший интерес представлял и расчет кинетики гибридизации секвенируемой ДНК с олигонуклеотидами, иммобилизованными в слое полиакриламидного геля. Специальное исследование данного аспекта метода указало пути оптимизации этого процесса, позволяющие осуществить лучшую дискриминацию совершенных дуплексов от несовершенных [Лившиц, Мирзабеков, 1996]. Использование в модельном эксперименте вместо секвенируемого фрагмента ДНК флуоресцентно меченных олигонуклеотидов и их гибридизации с иммобилизованными в микрочипе показало некоторую зависимость прочности образующихся дуплексов от наличия выступающих на 3'- и/или 5'-концах нескольких нуклеотидов, которые могли быть или вырожденными или универсальными в виде 5-нитроиндола [Fotin et al., 1998].

Одной из серьезных проблем, присущих секвенированию ДНК посредством гибридизации, является разная температура диссоциации для

АТ- и GC-богатых олигонуклеотидов. Для ее преодоления предлагались различные варианты, включая увеличение длины АТ-богатых октануклеотидов еще на один нуклеотид. Однако более простое решение заключается в добавлении в гибридизационный буфер бетаина или тетраметиламмонийхлорида, известных тем, что они сглаживают различия в температурах плавления АТ- и GC-пар [Wood et al., 1985; Rees et al., 1993; Bains, 1994]. Хотя следует отметить, что в литературе содержатся сведения о том, что тетраметиламмонийхлорид не универсален и с некоторыми последовательностями ДНК его эффект малозаметен [Riccelli, Benight, 1993]. Так, было показано, что добавление тетраметиламмонийхлорида позволило использовать один гибридизационный буфер только для 90% всех олигонуклеотидов с разными АТ- и GC-составами при условии обеспечения температурного различия в 2°C для плавления совершенных и несовершенных дуплексов [Bains, 1994]. Однако в отсутствие тетраметиламмонийхлорида при стандартных условиях гибридизации подобное условие выполняется лишь для 60% олигонуклеотидов.

Обнаруженный эффект увеличения температуры диссоциации дуплексов на 5°C при трехкратном увеличении их концентрации в матриксе, представленном полиакриламидным гелем, был использован Мирзабековым и соавт. для выравнивания температур плавления октануклеотидов с разными GC-составом, что необходимо для лучшей идентификации совершенных и несовершенных дуплексов [Храпко и др., 1991; Khrapko et al., 1991]. Так, одним из “тонких” мест данного метода секвенирования ДНК является как раз детекция только совершенных дуплексов, не содержащих неспаренных нуклеотидов, поскольку в противном случае будет внесена ошибка, делающая дальнейшую реконструкцию ДНК или невозможной или неправильной. Считается, что наибольшие трудности подстерегают экспериментатора при выявлении несовершенных дуплексов в случае наличия в них терминальных неспаренных нуклеотидов и, в частности, пар G•T и G•A [Mirzabekov, 1994].

Выявление с помощью ДНК-лигазы и ДНК-полимеразы подобных неспаренных нуклеотидов (пары G•T и G•A) в образовавшихся дуплексах при секвенировании ДНК посредством гибридизации, контролируемое путем подсчета радиоактивности, было предпринято в модельном эксперименте другими авторами [Broude et al., 1994]. Под действием ДНК-лигазы проводилось лигирование встык отоженной мишени на выступающем конце двуцепочечного олигонуклеотида, а ДНК-полимераза вызывала удлинение образовавшегося в результате отжига углубленного 3'-конца. Этот подход был основан на известном факте худшего узнавания подобных конструкций данными ферментами, однако полученные ими данные вряд ли могут быть прямо проецированы на более обычный вариант секвенирования гибридизацией, поскольку здесь в качестве твердой фазы служили магнитные частицы, покрытые стрептавидином, а один из олигонуклеотидов содержал молекулу биотина. Более того, одноцепочечный участок двух комплементарных друг другу олигонуклеотидов, доступный для гибридизации с мишенью, был представлен всего 5 нуклеотидами. (Такое внимание к этой работе вызвано тем, что она будет упомянута еще в конце данной главы.)

Как уже отмечалось выше, предел “чтения” с помощью теоретически приемлемой ныне октануклеотидной матрицы невелик. На использование же матриц с увеличенной длиной иммобилизованных олигонуклеотидов за счет значащего числа звеньев в них в настоящее время невозможно даже рассчитывать. Причем за счет увеличения длины секвенирующих олигонуклеотидов в матрице на одно звено предел “чтения” повысится всего в 2 раза, тогда как число ячеек в самой матрице возрастет уже в 4 раза. Некое решение проблемы уменьшения числа вариантов олигонуклеотидов в матрице при одновременном увеличении их длины заключалось в добавлении во время синтеза в их центральную часть вырожденных нуклеотидов. Подобные “окта”-нуклеотиды, названные авторами “разнесенными”, имели вид ACGTNNACGT, где N означает смесь всех четырех нуклеотидов [Певзнер и др., 1991; Pevzner et al., 1991]. Таким образом, иммобилизованный в геле подобный “окта”-нуклеотид содержал в каждой ячейке все 16 возможных вариантов декамеров. Авторы считают, что за счет этого удастся в 5–15 раз снизить число вариантов используемых олигонуклеотидов без уменьшения разрешающей способности метода. Описано и применение октануклеотидов, ставших 12-мерами за счет добавления сразу четырех вырожденных нуклеотидов [Лысов и др., 1995; Southern, 1996]. Это привело к присутствию в каждой ячейке уже 256 вариантов олигонуклеотида, что в свою очередь рождает новые проблемы. В другой работе в результате компьютерной симуляции при анализе известной последовательности ДНК человека было рассчитано, что использование 65536 11-меров с 3 вырожденными нуклеотидами позволит определять уже до 800 нуклеотидов в одном эксперименте [Bains, 1993]. Принимая во внимание открытую рамку считывания, при секвенировании кДНК с помощью матрицы с подобными вырожденными 11-мерами можно в идеальных случаях повысить предел “чтения” до 2400 нуклеотидов [Bains, 1993 a].

Иной путь увеличения длины олигонуклеотидов заключался в добавлении во время гибридизации секвенируемого фрагмента ДНК с иммобилизованными октануклеотидами еще и коротких пентануклеотидов, как и сам секвенируемый фрагмент ДНК, находящихся в растворе [Лысов и др., 1993; Khrapko et al., 1991; Lysov et al., 1994]. За счет образующихся стэкинг-взаимодействий это позволило как бы увеличить длину октануклеотида, находящегося в ячейке матрицы или микрочипа, что, в свою очередь, привело к повышению достоверности и эффективности секвенирования ДНК. Данный способ теоретически позволяет с помощью 1024 ( $4^5$ ) пентамеров и 65536 ( $4^8$ ) октамеров в какой-то степени имитировать огромную библиотеку 13-меров, которая бы состояла из 67108864 ( $4^{13}$ ) вариантов. Так, повторные гибридизации встык с различными наборами пентануклеотидов увеличивают длину секвенируемого с помощью октануклеотидной матрицы фрагмента ДНК с 200 до нескольких тысяч нуклеотидов. Использование одного или двух пентамеров, несущих различные флуорофоры, и гибридизирующихся встык с основным октануклеотидом и первым пентамером соответственно позволило увеличить эффективную длину секвенирующего олигонуклеотида до 13 и даже до 18 звеньев [Pavlov et al., 1996]. Включение в состав

данных пентамеров или всех четырех нуклеотидов в виде вырожденно-го нуклеотида или 5-нитроиндола, как универсального основания, привело к сокращению числа дополнительных раундов гибридизации. Объединение этих двух подходов с использованием олигонуклеотидов с вырожденными основаниями и гибридизации встык с находящимися в растворе пентамерами показывает возможность сокращения числа дополнительных гибридизаций примерно вдвое с сохранением той же эффективности реконструкции секвенируемого фрагмента ДНК [Лысов и др., 1995].

Выше уже отмечалось, что значительные трудности подстерегают экспериментатора при секвенировании посредством гибридизации фрагментов ДНК, содержащих повторяющиеся последовательности. Так, при реконструкции нуклеотидной последовательности подобных фрагментов ДНК возникают неоднозначности в выборе того или иного октануклеотидного варианта и образования, таким образом, мест ветвления [Лысов и др., 1993]. Показано, что их частичное преодоление может быть достигнуто путем формирования серии из нескольких коротких разных (меченых и немеченных) олигонуклеотидов, первые из которых гибридизируются встык с основным олигонуклеотидом, а вторые уже с ними [Лысов и др., 1994б]. Дополнительная информация в виде предварительно измеренной длины секвенируемого фрагмента, содержащего повторы, или с помощью ПЦР, или рестриктазным анализом также позволяет повысить эффективность секвенирования подобной ДНК методом гибридизации [Лысов и др., 1994а; Lysov et al., 1996]. Полностью процедура реконструкции последовательности ДНК в подобном случае может быть представлена следующей схемой, приведенной в работе Лысова и соавт. [Лысов и др., 1993] и здесь несколько сокращенной. Так, после гибридизации секвенируемого фрагмента ДНК на олигонуклеотидной матрице и анализа образовавшихся совершенных дуплексов рассчитывается полный СГОМ-спектр, представляющий собой список всех олигонуклеотидов, встречающихся в секвенируемом фрагменте ДНК, с учетом количества вхождений каждого такого олигонуклеотида в анализируемую последовательность. Следующим этапом является анализ этого СГОМ-спектра на наличие повторов, приводящих к образованию точек ветвления. Для того чтобы прийти к однозначной реконструкции, проводится составление списков так называемых предысторий и продолжений в выявленных точках ветвления. Затем следуют подбор вариантов следующей гибридизации “встык” с определенными пентамерами, ее проведение и дальнейший анализ, приводящий или к однозначно “прочитанной” последовательности секвенируемого фрагмента ДНК, или к повторению этапов гибридизации “встык”. В литературе содержатся сведения также и о других подходах к решению проблемы неоднозначного “чтения” нуклеотидной последовательности ДНК, содержащей повторы при секвенировании гибридизацией [Кузнецова и др., 1994; Dubiley et al., 1997].

Интересные результаты были получены при компьютерном анализе нуклеотидной последовательности протяженного фрагмента ДНК человека размером 522 тпн [Bains, 1994]. Было показано, что 75% 12-

мерных и 89% 22-мерных олигонуклеотидов являются уникальными для анализируемого фрагмента, однако произвольной случайной ДНК аналогичного размера были бы присущи несколько иные соотношения. Так, уже 89% 12-мерных олигонуклеотидов были бы в этом случае уникальными, а для 22-мерных эта цифра составила бы все 100%. Подобное несоответствие можно объяснить как имеющимися повторами в геномной ДНК человека, так и просто “случайно” встречающимися олигонуклеотидными блоками, не входящими в состав специфических повторяющихся элементов. Однако олигонуклеотидная матрица, состоящая из более чем 17 триллионов вариантов всех 22-меров, причем также не обеспечивающая фронтальное секвенирование геномов, просто нереальна.

К сожалению, общая производительность метода секвенирования ДНК посредством гибридизации пока не очень невысока [Chetverin, Kramer, 1993]. Считается, что с помощью полного набора октануклеотидов (65536 вариантов) можно секвенировать *de novo* фрагменты ДНК протяженностью лишь около 200 пн [Hannenhalli et al., 1996]. Увеличение длины олигонуклеотидов, фиксированных на матрице, позволяет увеличить длину чтения до 400 нуклеотидов и более, но число вариантов самих олигонуклеотидов возрастает при этом катастрофически. Смешанный вариант секвенирования ДНК посредством гибридизации и обычным ферментативным методом, теоретически позволяющий за 5–7 экспериментов определить последовательность фрагмента ДНК размером 10 тпн, был предложен Бэйном [Bains, 1993b]. Этот же автор ранее отмечал, что метод секвенирования ДНК-гибридизацией значительно проще ферментативного и требует меньшего количества обученного персонала [Bains, 1991].

В литературе описан интересный способ детекции сигналов гибридизации на ДНК-биочипе [Arlinghaus et al., 1997b]. Главной отличительной особенностью данного чипа является то, что вместо стандартных олигонуклеотидов на нем фиксированы химерные молекулы олигонуклеотидов, так называемые PNA (Peptide Nucleic Acids), где скелет такой молекулы образуется благодаря наличию N-(2-аминоэтил)глициновых остатков и фосфодиэфирные связи между азотистыми основаниями, таким образом, заменены амидными. Детекция сигналов гибридизации, проводящаяся с помощью ионизационной масс-спектрометрии, рассчитана на детекцию атомов фосфора, присутствующих только в тех участках биочипа, где произошла гибридизация, поскольку все остальные ячейки его не содержат. Здесь, видимо, следует отметить, что подобные пептидные нуклеиновые кислоты (ПНК) были синтезированы уже достаточно давно и была показана возможность формирования двойной спирали с комплементарными им обычными олигонуклеотидами [Nielsen et al., 1991; Egholm et al., 1993; Wittung et al., 1994]. Так, Саузерн в своей обзорной статье [Southern, 1996] упоминает о подобных соединениях и о потенциальной возможности их использования в качестве олигонуклеотидных матриц, однако отмечает, что подобные сведения в литературе на тот момент отсутствовали. В настоящее время сфера применения этих химерных соединений в молекулярной биологии и биотехнологии



весьма широка и продолжает расширяться, рассмотрению чего посвящен целый ряд обзорных статей [Buchardt et al., 1993; Corey, 1997; Uhlmann, 1998; Wang, 1998]. Так, ПНК нашли применение в блот-гибридизации по Саузерну в качестве гибридизационных зондов [Perry-O'Keefe et al., 1996]. Оказалось, что гибридизация ДНК с ПНК происходит даже быстрее, чем с обычными зондами, и образующиеся при этом связи заметно прочнее [Iyer et al., 1995]. В то же время наличие единичных неспариваемых нуклеотидов оказывает заметное влияние на этот процесс, что привело к их использованию в качестве зондов для выявления единичных мутаций [Wang et al., 1997]. Так, для подобных целей были разработаны специальные матрицы с иммобилизованными ПНК [Wang et al., 1997; Weiler et al., 1997]. Для лучшего понимания происходящих взаимодействий данных химер, содержащих пептидный остов с настоящими нуклеиновыми кислотами, при формировании двуцепочечных и трехцепочечных структур был проведен теоретический анализ, позволивший предположить существование двух стадий этого процесса [Lomakin, Frank-Kamenetskii, 1998].

Другое теоретически возможное усложнение метода секвенирования ДНК-гибридизацией, имеющее целью повысить его производительность, видится в использовании отдельных элементов из метода секвенирования ДНК путем синтеза. Так, увеличение числа нуклеотидов, например, в ячейке октануклеотидной матрицы с 8 до 9, возможно путем присоединения с помощью ДНК-полимеразы соответствующего ддНМФ, комплементарного очередному нуклеотиду секвенируемого фрагмента ДНК. Непременным условием является применение ддНТФ, меченных различными флуорофорами, отличающимися своими спектрами эмиссии друг от друга, а также от основной метки, принадлежащей самому секвенируемому фрагменту ДНК, что позволит (теоретически) осуществить и этап гибридации и удлинение данного олигонуклеотида, выступающего здесь в качестве праймера в одну стадию с последующей детекцией флуоресцентных сигналов, принадлежащих и ддНМФ и фрагменту ДНК, с помощью соответствующих фильтров. Что касается ДНК-полимеразы, то учитывая низкую температуру гибридации, необходимо использовать эффективно работающий при подобных температурах фермент, выделенный из какого-либо психрофильного микроорганизма. Кроме этого, видимо, необходимо предварительное удаление 3'-5'-экзонуклеазной активности этой ДНК-полимеразы. При этом можно предполагать даже улучшенную дискриминацию совершенных и несовершенных дуплексов, причем наиболее трудных в выявлении концевых неспаренных нуклеотидов за счет показанной ранее избирательности ДНК полимеразы к подобным затравочным комплексам [Broude et al., 1994]. Неким отрицательным моментом данного способа можно считать некоторое увеличение общего фона за счет присутствия большого количества флуоресцентно меченных ддНТФ, хотя этот недостаток может быть и преодолен. Другим препятствием к использованию этого подхода можно считать наличие самого матрикса в виде полиакриламидного геля, который может не дать возможность ДНК-полимеразе проявить свое ферментативное действие.

Неким решением проблемы в этом случае может стать снижение концентрации полиакриламида и уменьшение его "сшивки".

Следует отметить, что ранее подобное ферментативное удлинение олигонуклеотида в ячейке полиакриламидного геля было невозможно из-за применявшегося способа иммобилизации посредством расположенного на 3'-конце 3-метилуридина [Храпко и др., 1991]. Разработанный вариант сополимеризации полиакриламида с олигонуклеотидами, несущими на 5'-конце аллильную группу [Василисков и др., 1998], оставляет свободным их 3'-конец и позволяет, таким образом, каждому конкретному олигонуклеотиду служить в качестве затравки для полимеразной реакции.

Завершая рассмотрение метода секвенирования ДНК посредством гибридизации, следует отметить, что развитие формата 2 активно продолжается в отличие от формата 1, который менее пригоден для автоматизации и значительно сложнее в исполнении. Наблюдаемый в последние годы перенос технологий из микроэлектронной промышленности в биотехнологию в виде создания биочипов уже оказывает революционизирующее воздействие в этой сфере [Chee et al., 1996; Sosnovski et al., 1997; Edman et al., 1997; Case-Green et al., 1998; Livache et al., 1998].

### *Цитируемая литература*

*Василисков В.А., Тимофеев Э.Н., Суржиков С.А., Дробышев А.Л., Шик В.В., Мирзабеков А.Д.* Метод получения микрочипов с помощью сополимеризации с акриламидом // Молекуляр. биология. 1998. Т. 32. С. 923–925.

*Кузнецов С.А., Каневский И.Э., Флорентьев В.А., Мирзабеков А.Д., Шабарова З.А.* Секвенирование ДНК гибридизацией с олигонуклеотидами, иммобилизованными в геле. Химическое лигирование как способ расширения возможностей метода // Там же. 1994. Т. 28. С. 290–299.

*Лившиц М.А., Иванов И.Б., Мирзабеков А.Д., Флорентьев В.Л.* Секвенирование ДНК гибридизацией с олигонуклеотидной матрицей (СГОМ). Теория отмывки ДНК после гибридизации // Там же. 1992. Т. 26. С. 1298–1313.

*Лившиц М.А., Мирзабеков А.Д.* Расчет кинетики гибридизации ДНК с олигонуклеотидами, фиксированными в слое геля // Там же. 1996. Т. 30. С. 1158–1166.

*Лысов Ю.П., Миронов А.А., Гнучев Ф.Н., Черный А.А., Мирзабеков А.Д.* Измерение расстояний между участками ДНК для увеличения эффективности секвенирования с помощью гибридизации на олигонуклеотидной матрице // Там же. 1994а. Т. 28. С. 652–657.

*Лысов Ю.П., Флорентьев В.Л., Хорлин А.А., Храпко К.Р., Шик В.В., Мирзабеков А.Д.* Определение нуклеотидной последовательности ДНК гибридизацией с олигонуклеотидами: Новый метод // Докл. АН СССР. 1988. Т. 303. С. 1508–1511.

*Лысов Ю.П., Черный А.А., Балаев А.А., Битти К.Л., Флорентьев В.Л., Мирзабеков А.Д.* Реконструкция секвенируемой последовательности по результатам гибридизации встык с олигонуклеотидной матрицей // Молекуляр. биология. 1993. Т. 27. С. 1126–1138.

*Лысов Ю.П., Черный А.А., Балаев А.А., Гнучев Ф.Н., Битти К.Л., Мирзабеков А.Д.* Эффективность проведения секвенирования с помощью гибридизации встык на олигонуклеотидных матрицах с различной длиной иммобилизованных нуклеотидов // Там же. 1994в. Т. 28. С. 832–839.

*Лысов Ю.П., Черный А.А., Балаев А.А., Гнучев Ф.Н., Битти К.Л., Мирзабеков А.Д.* Использование гибридизации встык при секвенировании с помощью модифицированных олигонуклеотидных матриц // Там же. 1995. Т. 29. С. 104–113.

*Мирзабеков А.Д.* Новые стратегии секвенирования ДНК и определения расположения белков на геномной ДНК // Биоорганич. химия. 1992. Т. 18. С. 1361–1374.

- Певзнер П.А., Лысов Ю.П., Храпко К.Р., Белявский А.В., Флорентьев В.Л., Мирзабеков А.Д.* Оптимальные чипы для мегабазного секвенирования ДНК // Молекуляр. биология. 1991. Т. 25. С. 552–561.
- Храпко К.Р., Хорлин А.А., Иванов И.Б., Чернов Б.К., Лысов Ю.П., Василенко С.К., Флорентьев В.Л., Мирзабеков А.Д.* Гибридизация ДНК с олигонуклеотидами, иммобилизованными в геле: Удобный метод регистрации одиночных замен оснований // Там же. 1991. Т. 25. С. 718–730.
- Arlinghaus H.F., Kwoka M.N., Guo X.-Q., Jacobson K.B.* Multiplexed DNA sequencing and diagnostics by hybridization with enriched stable isotope labels // *Anal. Chem.* 1997a. Vol. 69. P. 1510–1517.
- Arlinghaus H.F., Kwoka M.N., Jacobson K.B.* Analysis of biosensor chips for identification of nucleic acids // *Ibid.* 1997b. Vol. 69. P. 3747–3753.
- Bains W.* Hybridization methods for DNA sequencing // *Genomics.* 1991. Vol. 11. P. 294–301.
- Bains W.* Characterization and sequencing cDNA using oligonucleotide hybridization // *DNA Sequenc.* 1993a. Vol. 4. P. 143–150.
- Bains W.* Mixed hybridization and conventional strategies for DNA sequencing // *Genet. Anal. Techn. Appl.* 1993b. Vol. 10. P. 84–94.
- Bains W.* Selection of oligonucleotide probes and experimental conditions for multiplex hybridization experiments // *Ibid.* 1994. Vol. 11. P. 49–62.
- Bains W.* Litigation escalates for patents on a chip // *Nat. Biotechnol.* 1997. Vol. 15. P. 406.
- Bains W., Smith G.C.* A novel method for nucleic sequence determination // *J. Theor. Biol.* 1988. Vol. 135. P. 303–307.
- Belyi I., Pevzner P.A.* Software for DNA sequencing by hybridization // *Comput. Appl. Biosci.* 1997. Vol. 13. P. 205–210.
- Broude N.E., Sano T., Smith C., Cantor C.R.* Enhanced DNA sequencing by hybridization // *Proc. Nat. Acad. Sci. USA.* 1994. Vol. 91. P. 3072–3076.
- Buchardt O., Egholm M., Berg R.H., Nielsen P.E.* Peptide nucleic acids and their potential applications in biotechnology // *Trends Biotechnol.* 1993. Vol. 11. P. 384–386.
- Butler D.* Truce likely in battle over 'DNA-chip' patent rights // *Nature.* 1997. Vol. 387. P. 221.
- Case-Green S.C., Mir K.U., Pritchard C.E., Southern E.M.* Analysis genetic information with DNA arrays // *Curr. Opin. Chem. Biol.* 1998. Vol. 2. P. 404–410.
- Case-Green S.C., Southern E.M.* Studies on the base pairing properties of deoxyinosine by solid phase hybridisation to oligonucleotides // *Nucl. Acids Res.* 1994. Vol. 22. P. 131–136.
- Chee M., Yang R., Hubbell E., Berno A., Huang X.C., Stern D., Winkler J., Lockhart D.J., Morris M.S., Fodor S.P.* Accessing genetic information with high-density DNA arrays // *Science.* 1996. Vol. 274. P. 610–614.
- Chetverin A.B., Kramer F.R.* Sequencing of pools of nucleic acids on oligonucleotide arrays // *Biosystems.* 1993. Vol. 30. P. 215–231.
- Corey D.R.* Peptide nucleic acids: Expanding the scope of nucleic acid recognition // *Trends Biotechnol.* 1997. Vol. 15. P. 224–229.
- Drmanac R., Drmanac S., Labat I., Crkvenjakov R., Vicentic A., Gemmell A.* Sequencing by hybridization: Toward an automated sequencing of one million M13 clones arrayed on membranes // *Electrophoresis.* 1992. Vol. 13. P. 566–573.
- Drmanac R., Drmanac S., Strezoska Z., Paunesku T., Labat I., Zeremski M., Snoddy J., Funkhouser W.K., Koop B., Hood L., Crkvenjakov R.* DNA sequence determination by hybridization: A strategy for efficient large scale sequencing // *Science.* 1993. P. 1649–1652; Correction // *Science.* 1994. Vol. 263. P. 596.
- Drmanac R., Labat I., Brukner I., Crkvenjakov R.* Sequencing of megabase plus DNA by hybridization: Theory of the method // *Genomics.* 1989. Vol. 4. P. 114–128.
- Drobyshev A., Mologina N., Shik V., Pobedimskaya D., Yershov G., Mirzabekov A.* Sequence analysis by hybridization with oligonucleotide microchip: Identification of beta-thalassemia mutations // *Gene.* 1997. Vol. 188. P. 45–52.
- Dubiley S., Kirillov E., Lysov Y., Mirzabekov A.* Fractionation, phosphorylation and ligation on oligonucleotide microchips to enhance sequencing by hybridization // *Nucl. Acids Res.* 1997. Vol. 25. P. 2259–2265.

Edman C.F., Raymond D.E., Wu D.J., Tu E., Sosnowski R.G., Bulter W.F., Nerenberg M., Heller M.J. Electric field directed nucleic acid hybridization on microchips // *Ibid.* 1997. Vol. 25. P. 4907–4914.

Egholm M., Buchardt O., Christensen L., Behrens C., Freier S.M., Driver D.A., Berg R.H., Kim S.K., Norden B., Nielsen P.E. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules // *Nature.* 1993. Vol. 365. P. 566–568.

Fodor S.P.A., Rava R.P., Huang X.C., Pease A.C., Holmes C.P., Adams C.L. Multiplexed biochemical assays with biological chips // *Ibid.* 1993. Vol. 364. P. 555–556.

Fotin A.V., Drobyshev A.L., Proudnikov D.Y., Perov A.N., Mirzabekov A.D. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips // *Nucl. Acids Res.* 1998. Vol. 26. P. 1515–1521.

Guschin D., Yershov G., Zaslavsky A., Gemmell A., Shick V., Proudnikov D., Arenkov P., Mirzabekov A. Manual manufacturing of oligonucleotide, DNA, and protein microchips // *Anal. Biochem.* 1997. Vol. 250. P. 203–211.

Hannenhalli S., Feldman W., Lewis H.F., Skiena S.S., Pevzner P.A. Positional sequencing by hybridization // *Comput. Appl. Biosci.* 1996. Vol. 12. P. 19–24.

Iyer M., Norton J.C., Corey D.R. Accelerated hybridization of oligonucleotides to duplex DNA // *J. Biol. Chem.* 1995. Vol. 270. P. 14712–14717.

Khrapko K.R., Lysov Yu.P., Khorlin A.A., Shick V.V., Florentiev V.L., Mirzabekov A.D. An oligonucleotide hybridization approach to DNA sequencing // *FEBS Lett.* 1989. Vol. 256. P. 118–122.

Khrapko K.R., Lysov Yu.P., Khorlin A.A., Ivanov I.B., Yershov G.M., Vasilenko S.K., Florentiev V.L., Mirzabekov A.D. A method for DNA sequencing by hybridization with oligonucleotide matrix // *DNA Sequenc.* 1991. Vol. 1. P. 375–388.

Livache T., Fouque B., Roget A., Marchand J., Bidan G., Teoule R., Mathis G. Polypyrrole DNA chip on a silicon device: Example of hepatitis C virus genotyping // *Anal. Biochem.* 1998. Vol. 255. P. 188–194.

Livshits M.A., Florentiev V.L., Mirzabekov A.D. Dissociation of duplexes formed by hybridization of DNA with gel-immobilized oligonucleotides // *J. Biomol. Struct. Dyn.* 1994. Vol. 11. P. 783–795.

Lomakin A., Frank-Kamenetskii M.D. A theoretical analysis of specificity of nucleic acid interactions with oligonucleotides and peptide nucleic acids (PNAs) // *J. Mol. Biol.* 1998. Vol. 276. P. 57–70.

Lysov Yu.P., Chernyi A.A., Balaeff A.A., Beattie K.L., Mirzabekov A.D. Calculation of continuous stacking hybridization efficiency // *J. Biomol. Struct. Dyn.* 1994. Vol. 11. P. 797–812.

Lysov Yu.P., Gnuchev F.N., Mironov A.A., Chernyi A.A., Beattie K.L., Mirzabekov A.D. Efficiency of sequencing by hybridization on oligonucleotide matrix supplemented by measurement of the distance between DNA segments // *DNA Sequenc.* 1996. Vol. 6. P. 65–73.

Maskos U., Southern E.M. Oligonucleotide hybridisations on glass support: A novel linker for oligonucleotide synthesis and hybridisation properties of oligonucleotides synthesis *in situ* // *Nucl. Acids. Res.* 1992. Vol. 20. P. 1679–1684.

Matson R.S., Rampal J., Pentoney S.L., Jr., Anderson P.D., Coassin P. Biopolymer synthesis on polypropylene supports // *Anal. Biochem.* 1995. Vol. 224. P. 110–116.

Nielsen P.E., Egholm M., Berg R.H., Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide // *Science.* 1991. Vol. 254. P. 1497–1500.

Parinov S., Barsky V., Yershov G., Kirillov E., Timofeev E., Belgovskiy A., Mirzabekov A.D. DNA sequencing by hybridization to microchip octa- and decanucleotides extended by stacked pentanucleotides // *Nucl. Acids Res.* 1996. Vol. 24. P. 2998–3004.

Pease A.C., Solas D., Sullivan E.J., Cronin M.T., Holmes C.P., Fodor S.P.A. Light-generated oligonucleotide arrays for rapid DNA sequence analysis // *Proc. Nat. Acad. Sci. USA.* 1994. Vol. 91. P. 5022–5026.

Perry-O'Keefe H., Yao X.W., Coull J.M., Fuchs M., Egholm M. Peptide nucleic acid pre-gel hybridization: An alternative to Southern hybridization // *Ibid.* 1996. Vol. 93. P. 14670–14675.

Pevzner P.A., Lysov Yu.P., Khrapko K.R., Belyavsky A.V., Florentiev V.L., Mirzabekov A.D. Improved chips for sequencing by hybridization // *J. Biomol. Struct. Dyn.* 1991. Vol. 9. P. 399–410.

Rees W.A., Yager T.D., Korte J., Hippel P.H. von. Betaine can eliminate the base pair composition dependence of DNA melting // *Biochemistry.* 1993. Vol. 32. P. 137–144.

- Riccelli P.V., Benight A.S.* Tetramethylammonium does not universally neutralize sequence dependent DNA stability // *Nucl. Acids. Res.* 1993. Vol. 21. P. 3785–3788.
- Shchepinov M.S., Case-Green S.C., Southern E.M.* Steric factors influencing hybridization of nucleic acids to oligonucleotide arrays // *Ibid.* 1997. Vol. 25. P. 1155–1161.
- Sosnowski R.G., Tu E., Butler W.F., O'Connell J.P., Heller M.J.* Rapid determination of single-base mismatch mutations in DNA hybrids by direct electric field control // *Proc. Nat. Acad. Sci. USA.* 1997. Vol. 94. P. 1119–1123.
- Southern E.M.* DNA chips: Analysis sequence by hybridization to oligonucleotides on a large scale // *Trends Genet.* 1996. Vol. 12. P. 110–115.
- Southern E.M., Case-Green S.C., Elder J.K., Johnson M., Mir K.U., Wang L., Williams J.C.* Arrays of complementary oligonucleotides for analysing the hybridisation behavior of nucleic acids // *Nucl. Acids Res.* 1994. Vol. 22. P. 1368–1373.
- Southern E.M., Maskos U., Elder J.K.* Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models // *Genomics.* 1992. Vol. 13. P. 1008–1017.
- Strezoska Z., Paunesku T., Radosavljevic D., Labat I., Drmanac R., Crkvenjakov R.* DNA sequencing by hybridization: 100 bases read by a non-gel-based method // *Proc. Nat. Acad. Sci. USA.* 1991. V. 88. P. 10089–10093.
- Timofeev E., Kochetkova S.V., Mirzabekov A.D., Florentiev V.L.* Regioselective immobilization of short oligonucleotides to acrylic copolymer gels // *Nucl. Acids Res.* 1996. Vol. 24. P. 3142–3148.
- Uhlmann E.* Peptide nucleic acids (PNA) and PNA–DNA chimeras: From high binding affinity towards biological function // *Biol. Chem.* 1998. Vol. 379. P. 1045–1052.
- Wang J.* DNA biosensors based on peptide nucleic acid (PNA) recognition layers: A review // *Biosens. Bioelectron.* 1998. Vol. 13. P. 757–762.
- Wang J., Nielsen P.E., Jiang M., Cai X., Fernandes J.R., Grant D.H., Ozsoz M., Beglieter A., Mowat M.* Mismatch-sensitive hybridization detection by peptide nucleic acids immobilized on a quartz crystal microbalance // *Anal. Chem.* 1997. Vol. 69. P. 5200–5202.
- Weiler J., Gausepohl H., Hauser N., Jensen O.N., Hoheisel J.D.* Hybridization based DNA screening on peptide nucleic acid (PNA) oligomer arrays // *Nucl. Acids. Res.* 1997. Vol. 25. P. 2792–2799.
- Weiler J., Hoheisel J.D.* Combining the preparation of oligonucleotide arrays and synthesis of high-quality primers // *Anal. Biochem.* 1996. Vol. 243. P. 218–227.
- Williams J.C., Case-Green S.C., Mir K.U., Southern E.M.* Studies of oligonucleotide interactions by hybridization to arrays: The influence of dangling ends on duplex yield // *Nucl. Acids. Res.* 1994. Vol. 22. P. 1365–1367.
- Wittung P., Nielsen P.E., Buchardt O., Egholm M., Norden B.* DNA-like double helix formed by peptide nucleic acid // *Nature.* 1994. Vol. 368. P. 561–563.
- Wood W.I., Gitsher L.A., Lasky L.A., Lawn R.M.* Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries // *Proc. Nat. Acad. Sci. USA.* 1985. Vol. 82. P. 1585–1588.
- Yershov G., Barsky V., Belgovsky A., Kirillov E., Kreindlin E., Ivanov I., Parinov S., Guschin D., Drobishev A., Dubiley S., Mirzabekov A.D.* DNA analysis and diagnostics on oligonucleotide microchips // *Ibid.* 1996. Vol. 93. P. 4913–4918.

### *Дополнительная литература*

- Bains W., Smith G.C.* A novel method for nucleic sequence determination // *J. Theor. Biol.* 1988. Vol. 135. P. 303–307.
- Drmanac R., Labat I., Brukner I., Crkvenjakov R.* Sequencing of megabase plus DNA by hybridization: Theory of the method // *Genomics.* 1989. Vol. 4. P. 114–128.
- Mirzabekov A.D.* DNA sequencing by hybridization – a megasequencing method and a diagnostic tool // *Trends Biotechnol.* 1994. Vol. 12. P. 27–32.
- Southern E.M.* DNA chips: Analysis sequence by hybridization to oligonucleotides on a large scale // *Trends Genet.* 1996. Vol. 12. P. 110–115.

## ЭВОЛЮЦИОННОЕ РАЗВИТИЕ МЕТОДОВ СЕКВЕНИРОВАНИЯ ДНК

Те высоты, которых достигла в последнее время молекулярная биология, стали во многом возможны благодаря разработанным во второй половине 1970-х годов быстрым методам секвенирования ДНК. Слишком важной и нужной оказалась информация о последовательности нуклеотидов в ДНК. Даже простое перечисление того, к чему уже привело ее знание у нескольких полных геномов, огромного числа генов и прочих фрагментов ДНК, заняло бы достаточно много места. И еще большего можно ждать от секвенирования ДНК в будущем, поскольку «самая главная молекула» еще раскрыла не все свои «тайны».

При этом нельзя не видеть, что нынешнее секвенирование ДНК уже только напоминает то, что было в самом начале. Сегодня, исходя из стоящих перед конкретными исследователями задач, можно сказать, что есть *разное* секвенирование ДНК. Так, как и в догеномную эру, многие лаборатории продолжают проводить обычное секвенирование отдельных генов и прочих фрагментов ДНК и получаемые в процессе этих экспериментов сведения не потеряли своей актуальности, несмотря на начавшееся несколько лет назад секвенирование полных геномов некоторых организмов, которое превратилось в настоящее производство, где многие операции полностью автоматизированы. Последнее является достижением уже не только молекулярной биологии, а и смежных областей науки и техники.

Словом, секвенирование ДНК, включая многочисленные его составляющие, за четверть века своего эволюционирования претерпело целый ряд многочисленных преобразований и улучшений. Как и в обычных эволюционных процессах возникали и тупиковые ветви, не оказавшие заметного воздействия на общее развитие того, что называется секвенированием ДНК. Однако отражение только нынешнего состояния дел (или применяемых подходов) в секвенировании ДНК просто невозможно, хотя бы по той причине, что оно очень разнообразно. Так, например, в обычном секвенировании ДНК на нынешнем этапе кто-то предпочитает использовать одноцепочечные матрицы ДНК, кто-то – двуцепочечные, другие же определяют последовательность ДНК с помощью только циклического секвенирования. Сходная ситуация наблюдается и с типом используемых меток, где выбор простирается от классического радионуклида  $^{32}\text{P}$  до самых последних флуоресцентных красителей. Для них, правда, требуется уже специальная дорогостоящая техника, тогда как разделение фрагментов ДНК, меченных радионуклидами, возможно в приборе весьма простой конструкции, где центральным элементом служат два стекла, которые должны быть все-таки высокого качества. Что касается автоматического секвенирования, то и тут есть выбор между привычным стандартным электрофорезом в пластинах геля и более скоростным капиллярным гель-электрофорезом.

Оказалось, что даже такой бурно развивающийся раздел молекулярной биологии, как секвенирование ДНК, все же можно (хотя где-то и весьма кратко) вместить в одну книгу и четверть века это еще тот период времени, который вполне поддается охвату, попытка чего и была сделана в данной книге. Причем все же следует отметить (не желая при этом показаться нескромными), что подобного описания методов секвенирования ДНК нам нигде прежде не встречалось. Однако, на наш взгляд, оно было бы незавершенным, если не попытаться отразить хотя бы основные направления их эволюционного развития, чему и посвящена эта глава.

Еще одной особенностью данной главы является полное отсутствие ссылок на литературные источники. Пожалуй, главной причиной этого служит то, что в ходе всего предшествующего изложения различные новшества процесса секвенирования ДНК с приведением литературных ссылок уже упоминались. Подрисуночная расшифровка пронумерованных ветвей эволюционных древ лишь кратко характеризует какую-либо веточку. Особо следует оговорить то, что “толщина” деревьев и их веток достаточно произвольна и мало соответствует “годовым” кольцам того или иного подхода, способа или некоторого улучшения метода секвенирования ДНК. “Высота” деревьев также весьма условна, поскольку в подавляющем большинстве случаев невозможно точно определить время забвения какого-либо подхода или способа. Также следует заметить, что “абсолютно правильного” расположения ветвей нам достичь не удалось, поскольку для этого необходимо уже трехмерное пространство, а плоская схема есть плоская схема.

Чтобы не перегружать рисунки и не устраивать “непроходимую чащу” и одновременно не делать большое их количество, мы решили стандартное секвенирование ДНК уместить всего в 3 рисунках, условно разделив весь процесс на две стадии. К первой оказалось отнесено все то, что предшествует электрофоретическому разделению продуктов секвенирующих реакций. Вторая стадия состоит из самого электрофореза и некоторых последующих процедур. Таким образом, “нарисованными” оказались только все вариации двух классических методов секвенирования ДНК по Максаму–Гилберту и Сэнгеру.

Что касается заключительного этапа секвенирования ДНК в виде компьютерной обработки результатов, то его изображение в виде эволюционных древ привело бы к формированию таких раскидистых крон, что в них бы было чрезвычайно трудно разобраться. Заняться же их искусственной “обрезкой” не поднялась рука, в связи с чем эта схема оказалась не нарисованной. Также не оказались изображенными и все прочие методы секвенирования ДНК, отчасти по причине того, что многие из них представляют собой только молодую поросль, еще не сильно обросли ветками (модификациями), да и “высота” их будет весьма незначительна, хотя необходимо отметить, что потенциал некоторых из них весьма велик.

В качестве, видимо, некоторых необходимых пояснений представленных ниже рисунков следует отметить, что “главные стволы” представляют собой только сам принцип секвенирования ДНК тем или

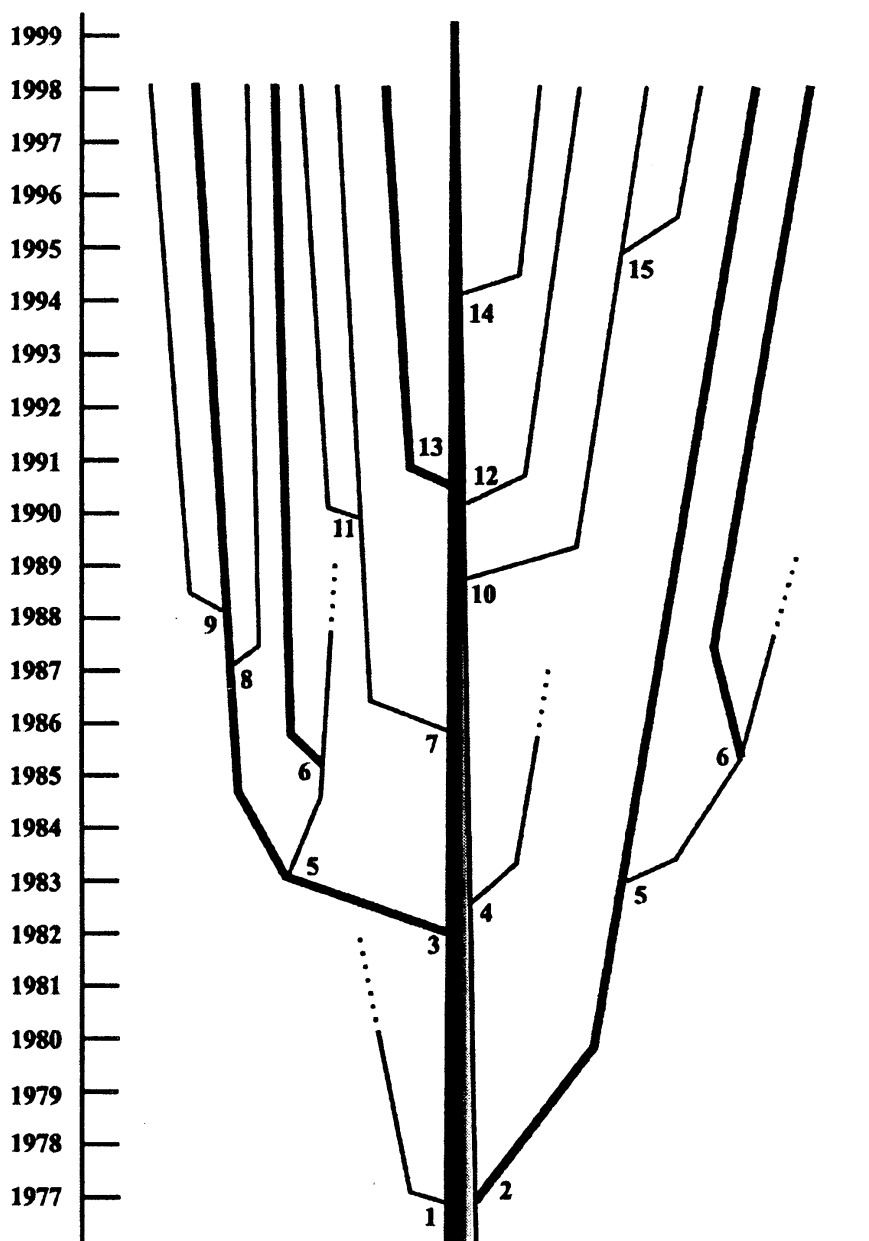


Рис. 14.1. Эволюционное “древо” развития метода секвенирования ДНК химической деградацией

1 – приготовление фрагмента ДНК для секвенирования путем деления цепей; 2 – приготовление фрагмента ДНК для секвенирования путем асимметричного расщепления рестрикционной эндонуклеазой; 3 – приготовление фрагмента ДНК для секвенирования путем его клонирования в специальном векторе с полилинкером; 4 – автоматическое проведение секвенирующих реакций с помощью химического робота; 5 – твердофазное секвенирование ДНК; 6 – твердофазное секвенирование ДНК (Hybond M & G); 7 – проведение химической модификации оснований и расщепления цепи ДНК в одну реакцию ( $A > C = G > T$ ); 8 – приготовление фрагмента ДНК для секвенирования путем его клонирования в специальном векторе с полилинкером, содержащем сайт рестрикционной эндонуклеазы *Tth111I*; 9 – мультиплексное секвенирование; 10 – приготовление фрагмента ДНК для секвенирования путем его амплификации с помощью ПЦР; 11 – проведение химической модификации оснований и расщепления цепи ДНК в одну реакцию ( $G > A > C > T$ ); 12 – секвенирование ДНК, меченной флуоресцентной меткой; 13 – автоматическое секвенирование флуоресцентно меченных фрагментов ДНК; 14 – автоматизация проведения реакций химической модификации и пиперидинового гидролиза с помощью лабораторного биоробота Biometek 1000; 15 – твердофазное секвенирование фрагментов ДНК, сорбированных на магнитных частицах



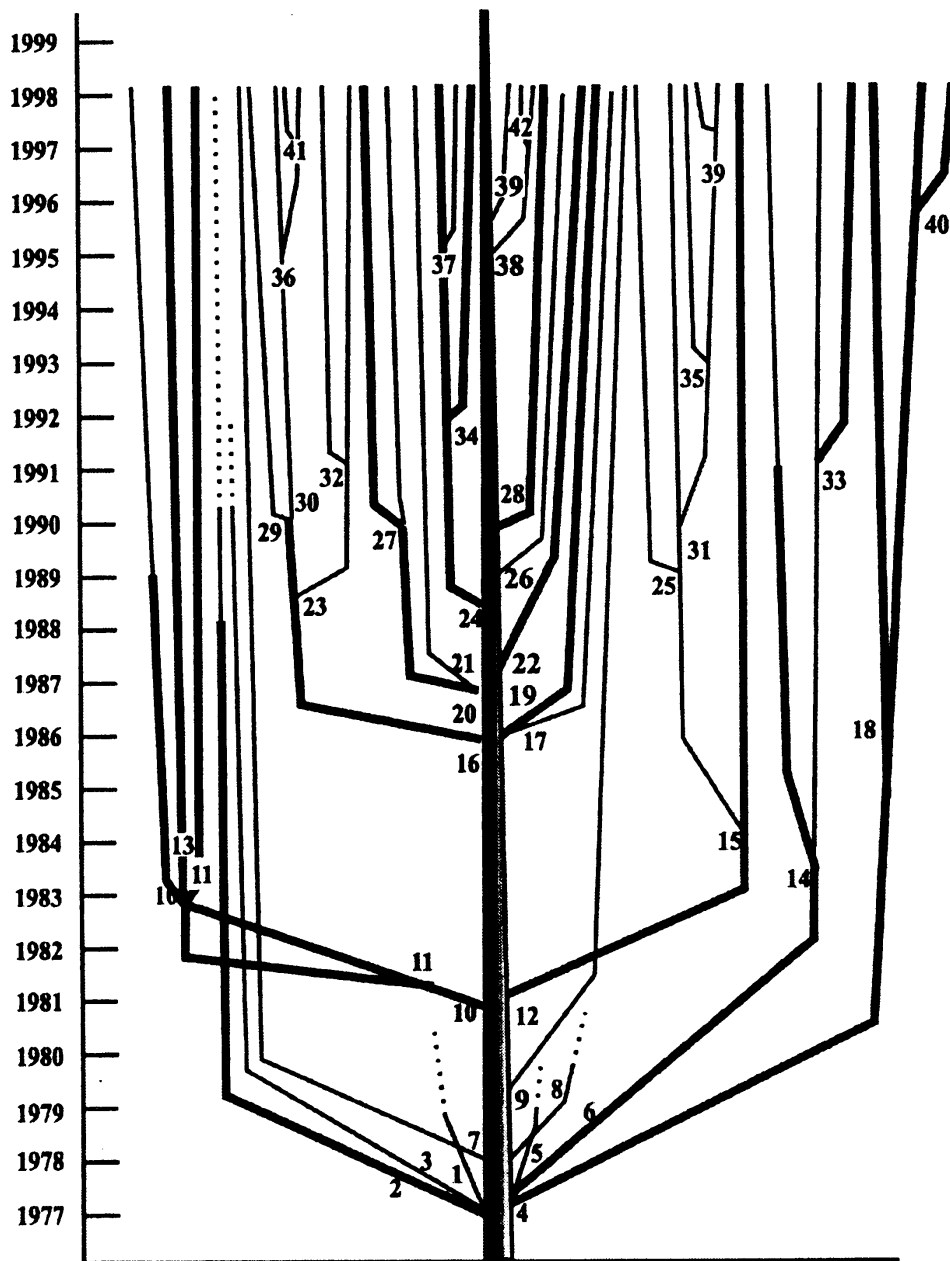


Рис. 14.2. Эволюционное “древо” развития метода секвенирования ДНК ферментативным построением комплементарной цепи в условиях специфической терминации

1 – природные одноцепочечные ДНК фагов; 2 – Кленовский фрагмент ДНК-полимеразы I; 3 – различные способы удаления одной из цепей ДНК; 4 – внутреннее мечение; 5 – короткие олигонуклеотиды; 6 – радионуклид  $^{32}\text{P}$ ; 7 – ревертаза; 8 – биологические праймеры; 9 – дИТФ; 10 – плазмидные вектора; 11 – фаг M13; 12 – универсальный праймер; 13 – фагмидные вектора; 14 – радионуклид  $^{35}\text{S}$ ; 15 – 5'-концевое мечение праймера; 16 – наработка матриц с помощью ПЦР; 17 – 7-деза-дГТФ; 18 – квазиконцевое мечение; 19 – флуоресцентное мечение праймера; 20 – секвенаса, версия 1.0; 21 – Bst ДНК-полимераза; 22 – флуоресцентно меченные ддНТФ; 23 – преимущественная наработка одной цепи ДНК с помощью асимметричной ПЦР; 24 – Taq ДНК-полимераза; 25 – библиотека коротких олигонуклеотидов; 26 – флуоресцентно меченные дНТФ; 27 – секвенаса, версия 2.0; 28 – хемилюминесцентное мечение; 29 – твердофазная ПЦР; 30 – циклическое секвенирование; 31 – модульные праймеры; 32 – температурная асимметричная ПЦР; 33 – радионуклид  $^{33}\text{P}$ ; 34 –  $\Delta\text{Taq}$  ДНК-полимераза; 35 – модифицированные модульные праймеры; 36 – полуэкспоненциальное секвенирование; 37 – термосеквенаса; 38 – окрашивание серебром; 39 – ддНТФ, меченные радионуклидом  $^{33}\text{P}$ ; 40 – одновременная амплификация и секвенирование двумя разными ДНК-полимеразами; 41 – дифференциальное удлинение праймера

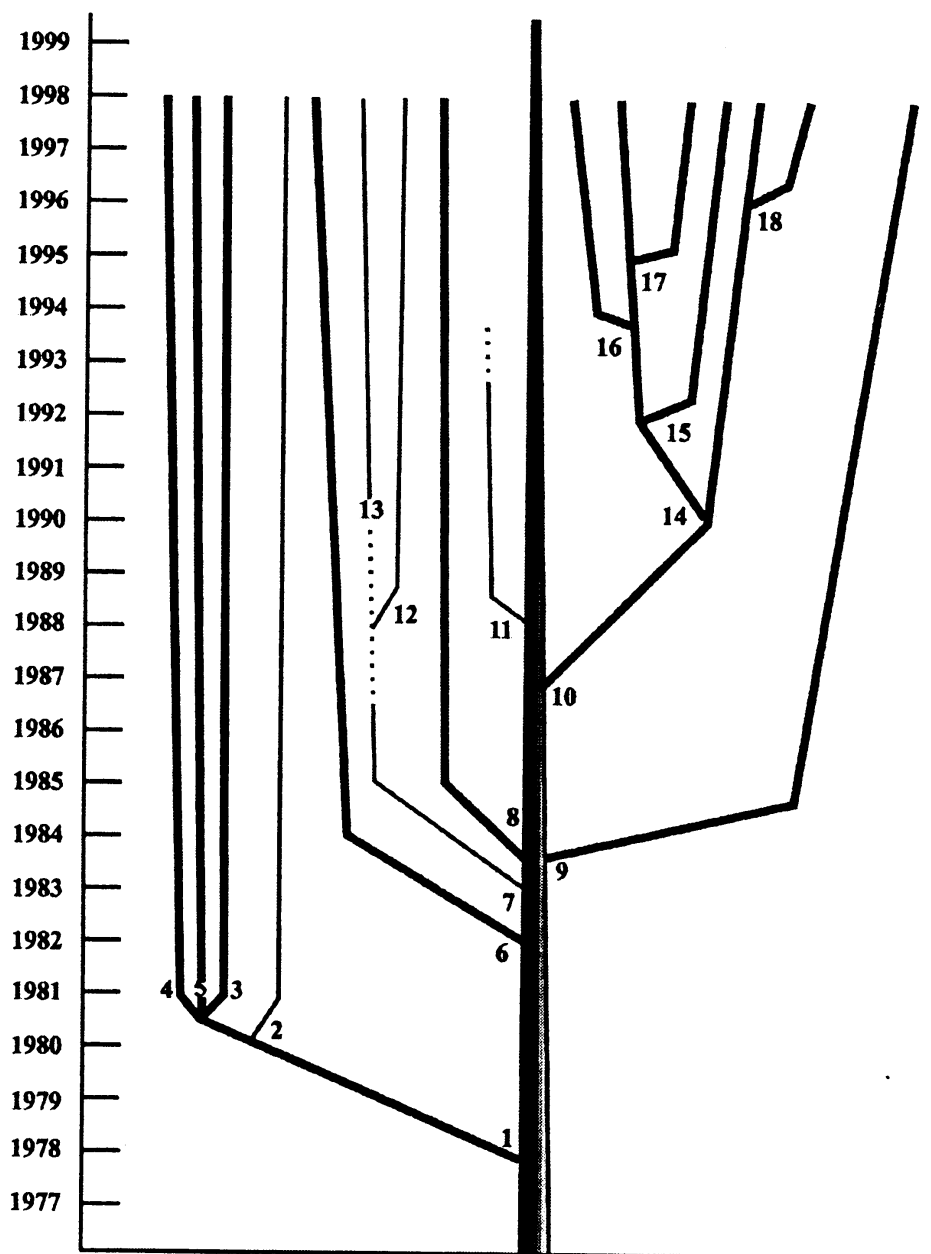


Рис. 14.3. Эволюционное “древо” развития метода электрофоретического разделения продуктов секвенирующих реакций, получающихся при секвенировании ДНК химической деградацией и ферментативным построением комплементарной цепи в условиях специфической терминации

1 – тонкий (0,4 мм) полиакриламидный гель; 2 – высушивание геля перед радиоавтографией; 3 – ультратонкие (0,2 мм и менее) полиакриламидные гели; 4 – “приклеивание” геля к стеклу во время полимеризации; 5 – термостатирование геля; 6 – гребешок “зубы акулы”; 7 – формирование в геле градиента буфера; 8 – клиновидный гель; 9 – электрофорез, сопряженный с блоттированием; 10 – разделение флуоресцентно меченных фрагментов ДНК с помощью автоматического секвенера; 11 – пульсирующий гель-электрофорез; 12 – формирование электролитического градиента между нижней и верхней камерами прибора; 13 – составные буферные системы; 14 – капиллярный гель-электрофорез; 15 – капиллярный гель-электрофорез, протекающий параллельно в нескольких капиллярах; 16 – применение “несшитого” линейного полиакриламидного геля; 17 – микрокапиллярный электрофорез; 18 – прибор для автоматического электрофореза с двумя типами лазеров

иным методом, а также электрофоретическое разделение продуктов секвенирующих реакций. Так, в основе метода секвенирования ДНК путем химической деградации по Максаму–Гилберту лежит ограниченное расщепление меченного по одному концу фрагмента ДНК под действием реагентов, специфичных к определенному типу оснований нуклеотидов. Это и есть сам ствол одного из деревьев (рис. 14.1). Другой принадлежит методу секвенирования ДНК, разработанному Сэнгером и соавт. и основанному на принципе ферментативного построения меченой комплементарной цепи ДНК по существующей одноцепочечной матрице с помощью ДНК-полимеразы при происходящем в разных местах цепи ДНК ингибировании ее дальнейшего роста за счет включения дидезоксинуклеотидов (рис. 14.2).

Третий ствол олицетворяет собой разделение по размеру меченых продуктов, образующихся как в ходе химической деградации в методе Максама–Гилберта, так и в результате специфической терминации цепи ДНК в методе Сэнгера, и осуществляется с помощью высоковольтного электрофореза в полиакриламидном геле высокого разрешения, позволяющего разделять фрагменты ДНК, отличающиеся между собой по длине всего на один нуклеотид, в достаточно широком диапазоне (рис. 14.3).

\* \* \*

Завершая и данную главу и всю книгу, вероятно, необходимо еще раз вернуться к высказыванию Эрвина Чаргаффа, упоминавшемуся во введении, о том, что чтение последовательности ДНК может стать задачей XXI века. Сейчас, наверное, уже можно считать, что он в действительности не ошибся в том плане, что в следующем столетии, по крайней мере в первой его половине, широкомасштабное секвенирование ДНК полных геномов большого числа организмов (наряду с обычным секвенированием отдельных генов и прочих фрагментов ДНК) будет, несомненно, продолжаться и станет одной из главных задач молекулярной биологии. Хотя строить прогнозы в науке вообще и в молекулярной биологии тем более, вещь неблагоприятная (кто в этом сомневается, может еще раз прочитать мысль, высказанную Чаргаффом, или в оригинале, или в нашем несколько вольном переводе (см. **Введение**)), однако можно все-таки предположить, что к середине следующего столетия накопленные знания в виде ДНКовых текстов полных геномов большого числа видов бактерий, растений, животных, а также и бесчисленных генов прочих организмов уже позволят вплотную подойти к пониманию взаимодействия всех генов или значительной их части даже сложноустроенного организма. Таким образом, главное внимание молекулярной биологии вероятно будет переключено именно на данное направление познания основ жизни. Что касается самого секвенирования ДНК, то оно будет, несомненно, востребовано и в дальнейшем, только вот какими методами будут добываться эти сведения, сейчас, наверное, даже невозможно представить.

## ПРИЛОЖЕНИЕ

### *Основные характеристики используемых в секвенировании ДНК радионуклидов*

Радионуклид	Период полураспада	Максимальная энергия $\beta$ -частиц (МэВ)	Максимальный пролет $\beta$ -частиц в воздухе (см)	Максимальный пролет $\beta$ -частиц в воде (мм)
$^{32}\text{P}$	14,3 дня	1,709	790	0,8
$^{33}\text{P}$	25,4 дня	0,249	49	0,6
$^{35}\text{S}$	87,4 дня	0,167	26	0,32

### *Пересчетные коэффициенты распада радионуклида $^{32}\text{P}$*

Дни	0	0,5	1	1,5	2	2,5	3	3,5
0	1,000	0,976	0,953	0,930	0,908	0,886	0,865	0,844
4	0,824	0,804	0,785	0,766	0,748	0,730	0,712	0,695
8	0,679	0,662	0,646	0,631	0,616	0,601	0,587	0,573
12	0,559	0,546	0,533	0,520	0,507	0,495	0,483	0,472
16	0,460	0,449	0,439	0,428	0,418	0,408	0,398	0,389
20	0,379	0,370	0,361	0,353	0,344	0,336	0,328	0,320
24	0,312	0,305	0,298	0,291	0,284	0,277	0,270	0,264
28	0,257	0,251	0,245	0,239	0,234	0,228	0,223	0,217
32	0,212	0,207	0,202	0,197	0,192	0,188	0,183	0,179
36	0,175	0,170	0,166	0,162	0,159	0,155	0,151	0,147
40	0,144	0,140	0,137	0,134	0,131	0,127	0,124	0,121
44	0,119	0,116	0,113	0,110	0,108	0,105	0,102	0,100

### *Пересчетные коэффициенты распада радионуклида $^{33}\text{P}$*

Дни	0	1	2	3	4
0	1,000	0,973	0,947	0,921	0,897
10	0,761	0,741	0,721	0,701	0,683
20	0,579	0,564	0,549	0,534	0,520
30	0,441	0,429	0,418	0,406	0,395
40	0,336	0,327	0,318	0,309	0,301
50	0,256	0,249	0,242	0,236	0,229
60	0,195	0,189	0,184	0,179	0,174
70	0,148	0,144	0,140	0,136	0,133
80	0,113	0,110	0,107	0,104	0,101

*Пересчетные коэффициенты распада радионуклида <sup>33</sup>P*

Дни	5	6	7	8	9
0	0,872	0,849	0,826	0,804	0,782
10	0,664	0,646	0,629	0,612	0,595
20	0,506	0,492	0,479	0,466	0,453
30	0,385	0,374	0,364	0,355	0,345
40	0,293	0,285	0,277	0,270	0,263
50	0,223	0,217	0,211	0,205	0,200
60	0,170	0,165	0,161	0,156	0,152
70	0,129	0,126	0,122	0,119	0,116
80	0,098	0,096	0,093	0,091	0,088

*Пересчетные коэффициенты распада радионуклида <sup>35</sup>S*

Дни	0	1	2	3	4	5	6
0	1,000	0,992	0,984	0,976	0,969	0,961	0,954
7	0,946	0,939	0,931	0,924	0,916	0,909	0,902
14	0,895	0,888	0,881	0,874	0,867	0,860	0,853
21	0,847	0,840	0,833	0,827	0,820	0,814	0,807
28	0,801	0,795	0,788	0,782	0,776	0,770	0,764
35	0,758	0,752	0,746	0,740	0,734	0,728	0,722
42	0,717	0,711	0,705	0,700	0,694	0,689	0,683
49	0,678	0,673	0,667	0,662	0,657	0,652	0,646
56	0,641	0,636	0,631	0,626	0,621	0,616	0,612
63	0,607	0,602	0,597	0,592	0,588	0,583	0,579
70	0,574	0,569	0,565	0,560	0,556	0,552	0,547
77	0,543	0,539	0,534	0,530	0,526	0,522	0,518
84	0,514	0,510	0,506	0,502	0,498	0,494	0,490

*Соотношение единиц измерения радиоактивности*

1 Бк (распад в сек) =  $2,7 \times 10^{-11}$  Ки

1 Ки =  $3,7 \times 10^{10}$  Бк = 37 ГБк

1 мКи = 37 МБк

1 мкКи = 37 КБк

*Однбуквенные коды, обозначающие одновременно  
два и более нуклеотидов*

R = А или G	Y = С или Т	M = А или С	K = G или Т
S = С или G	W = А или Т	H = А или С или Т	B = С или G или Т
V = А или С или G	D = А или G или Т	N = А или С или G или Т	

*Однобуквенные и трехбуквенные сокращенные обозначения  
аминокислот  
и их молекулярная масса (в Дальтонах)*

A	Ala	Аланин	Alanine	89
R	Arg	Аргинин	Arginine	174
N	Asn	Аспарагин	Asparagine	132
D	Asp	Аспарагиновая кислота	Aspartic acid	133
B	Asx	Аспарагин или Аспарагиновая кислота	Asparagine or Aspartic acid	
V	Val	Валин	Valine	117
H	His	Гистидин	Histidine	155
G	Gly	Глицин	Glycine	75
Q	Gln	Глутамин	Glutamine	146
E	Glu	Глутаминовая кислота	Glutamic acid	147
Z	Glx	Глутамин или Глутаминовая кислота	Glutamine or Glutamic acid	
I	Ile	Изолейцин	Isoleucine	131
L	Leu	Лейцин	Leucine	131
K	Lys	Лизин	Lysine	146
M	Met	Метионин	Methionine	149
P	Pro	Пролин	Proline	115
S	Ser	Серин	Serine	105
Y	Tyr	Тирозин	Tyrosine	181
T	Thr	Треонин	Threonine	119
W	Trp	Триптофан	Tryptophan	204
F	Phe	Фенилаланин	Phenylalanine	165
C	Cys	Цистеин	Cysteine	121
X	Xxx	Любая аминокислота*		

\* Подобное обозначение применяется в тех случаях когда достоверная последовательность нуклеотидов в каком-либо месте не известна и представлена в виде одного или нескольких N, например NTG, NNG или NNN.

$10^{18}$	Э	Экза	E	exa
$10^{15}$	П	Пета	P	peta
$10^{12}$	Т	Тера	T	tera
$10^9$	Г	Гига	G	giga
$10^6$	М	Мега	M	mega
$10^3$	к	кило	k	kilo
$10^{-3}$	м	милли	m	milli
$10^{-6}$	мк	микро	μ	micro
$10^{-9}$	н	нано	n	nano
$10^{-12}$	п	пико	p	pico
$10^{-15}$	ф	фемто	f	femto
$10^{-18}$	а	атто	a	atto
$10^{-21}$	ц	цепто	z	zepto
$10^{-24}$	й	йокто	y	yocto

1 тпн ДНК кодирует 333 аминокислотный полипептид молекулярной массой около  $3,7 \times 10^4$  кДа

Белок с молекулярной массой 10 кДа кодируется фрагментом ДНК протяженностью 270 пн

Белок с молекулярной массой 30 кДа кодируется фрагментом ДНК протяженностью 810 пн

Белок с молекулярной массой 100 кДа кодируется фрагментом ДНК протяженностью 2,7 тпн

$1A_{260}$  двуцепочечной ДНК = 50 мкг/мл

$1A_{260}$  одноцепочечной ДНК = 33 мкг/мл

1 мкг фрагмента ДНК размером 1000 пн = 1,52 пмоля или 3,03 пмоля концов

1 пмоль фрагмента ДНК размером 1000 пн = 0,66 мкг

Для расчета количества олигонуклеотида в нмолях необходимо 1000 разделить на молекулярную массу данного олигонуклеотида и затем умножить на известное количество олигонуклеотида в мкг.

Для расчета приблизительной молекулярной массы олигонуклеотида с известной последовательностью нуклеотидов применяется следующая формула:

$$MW = [(\#A's \times 312,2) + (\#C's \times 288,2) + (\#G's \times 328,2) + (\#T's \times 303,2) + (\#I's \times 335,2)] - 61,$$

где #A's, #C's, #G's, #T's и #I's обозначают количества остатков аденина, цитозина, гуанина, тимидина и инозина в данном олигонуклеотиде соответственно.

Для олигонуклеотидов с неизвестной последовательностью молекулярная масса одного азотистого основания может быть принята за 310 Да.

## ТАБЛИЦА ГЕНЕТИЧЕСКОГО КОДА (ДЛЯ ЯДЕРНОГО ГЕНОМА)

### Второе положение кодона

		U	C	A	G	
Первое положение кодона	U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
		UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
		UUA Leu	UCA Ser	UAA Ter	UGA Ter	A
		UUG Leu	UCG Ser	UAG Ter	UGG Trp	G
	C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
		CUC Leu	CCC Pro	CAC His	CGC Arg	C
		CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
		CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
	A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
		AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
		AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
		AUG Met	ACG Thr	AAG Lys	AGG Arg	G
	G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
		GUC Val	GCC Ala	GAC Asp	GGC Gly	C
		GUA Val	GCA Ala	GAA Glu	GGA Gly	A
		GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Третье положение кодона

*Аминокислотная последовательность условного пептида  
и кодирующие ее триплеты, включая терминирующий кодон*

Met	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Ter
ATG	GCN	CGN	AAU	GAY	TGU	CAR	GAR	GGN	CAY	ATH	CTN	AAR	TTY	CCN	TCN	ACN	UGG	TAY	GTN	TAR
	AGR							TTR							AGU					TGA

или

Met	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Ter
ATG	GCA	CGA	AAC	GAC	TGC	CAA	GAA	GGG	CAC	ATA	CTA	AAA	CCA	CCA	TCA	ACA	UGG	TAC	GTA	TAA
	GCC	CGC	AAT	GAT	TGT	CAG	GAG	GGC	CAT	ATC	CTC	AAG	TTT	CCC	TCC	ACC		TAT	GTC	TAG
	GCG	CGG					GGG	GGG	ATT	CTG	CTG			CCG	TCG	ACG			GTG	TGA
	GCT	CGT					GGT	GGT		CTT	CTT			CCT	TCT	ACT			GTT	
		AGA								TTA	TTA				AGC					
		AGG								TTG	TTG				AGT					



*Характер миграции красителей при электрофорезе  
в денатурирующем полиакриламидном геле*

<b>% Геля</b>	<b>Бромфеноловый синий</b>	<b>Ксиленцианол FF</b>
5,0	35	130
6,0	26	106
8,0	19	75
10,0	12	55
20,0	8	28

## ОТЕЧЕСТВЕННЫЕ ПОСТАВЩИКИ МАТЕРИАЛОВ И ОБОРУДОВАНИЯ ДЛЯ СЕКВЕНИРОВАНИЯ ДНК

ТОО “БИОССЕТ”, Новосибирск, (тел. 3832 39-62-25, 3832 32-53-34, факс 3832 39-62-25, 3832 35-68-11, email – sredin@modul.bioch.nsk.su) производит “Синтезатор ДНК ASM-102U”, являющийся автоматическим однореакторным прибором, приспособленным для синтеза нормальных и модифицированных олигонуклеотидов как фосфитамидным, так и Н-фосфонатным методами синтеза с масштабами синтеза: 0,05/0,10/0,25/0,50/1,0 мкмоль и средним выходом на цикл более 98,0%. По желанию заказчика синтезатор ДНК поставляется вместе с “Установкой для очистки олигонуклеотидов OPS-201”, предназначенной для очистки и детритилирования синтетических олигонуклеотидов.

ТОО Бион, Москва, (тел./факс 095 135-42-06, email–bion@glas.apc.org) в кратчайшие сроки поставляет высококачественные ферменты и реактивы американской фирмы Promega, а также пластиковую посуду швейцарской фирмы Greff.

ТОО “Хеликон”, Москва, (тел. 095 939-31-07, факс 095 939-31-81, email – vlad@energ.genebee.msu.su) поставляет широкий спектр лабораторного оборудования для электрофореза ДНК, включая секвенирующий электрофорез, пластиковую посуду ведущих фирм.

ТОО “Диа-М”, Москва (тел. 095 190-66-12, 095 190-39-48, факс 095 190-34-71, email – diam@dia-m.msk.su) поставляет широкий спектр лабораторного оборудования отечественного и зарубежного производства, биохимические реактивы.

ГНИИ Генетика, Москва, (тел./факс 095 315-00-65) поставляет высококачественные ферменты литовской фирмы MBI Fermentas по конкурентноспособным ценам.

“Сибэнзим”, Новосибирск, (тел. 3832 35-33-50, факс 3832 35-34-59, 3832 32-88-31, email – siben@isi.itfsnsk.su) производит высококачественные ферменты для молекулярной биологии (рестрикционные эндонуклеазы, ДНК-лигазы, ДНК-полимеразы и пр.).

НПФ “Литех”, Москва (тел. 095 246-96-46, факс 095 246-45-01, email – vak@lytech.msk.ru) осуществляет заказной синтез олигонуклеотидов проверенного качества.

НПФ “ЭВИОЧ-РОС”, Москва, (тел. 977-67-52/ 976-23-33, факс 977-09-47 email – vvk@agrobio.msk.su, sintol@iab.ac.ru, akuzubov@iab.ac.ru) осуществляет заказной синтез олигонуклеотидов проверенного качества и поставляет широкий спектр ферментов нуклеинового обмена (различные ДНК- и РНК-зависимые ДНК-полимеразы, ДНК-лигазы и др.).

НП АО “Силекс М”, Москва, (тел. 095 475-46-14, факс 095 187-77-03, email – sileks@glasnet.ru, <http://www.glasnet.ru/~sileks/>) осуществляет заказной синтез олигонуклеотидов проверенного качества и поставляет широкий спектр ферментов нуклеинового обмена (ДНК- и РНК-зависимые ДНК-полимеразы, ДНК-лигазы и др.), а также различные го-

товые наборы для выделения ДНК, мечения ДНК и др., включая секвенирование ДНК.

Филиал института биоорганической химии им. М.М. Шемякина и Ю.А. Овчинникова РАН, Пущино, (тел./факс 8-27-73-17-80 (из Москвы), 0967 73-17-80 из других городов, email – nucleus@fibkh.serpukhov.su или feo@fibkh.serpukhov.su) поставляет ряд рестрикционных эндонуклеаз, а также различные наборы для молекулярной биологии, включая набор для секвенирования ДНК.

ГосНИИ “Химфотопроект”, Казань, (тел. 8432 45-05-62, 8432 42-30-88) поставляет специализированную рентгеновскую пленку РА-1 для радиоавтографии секвенирующих гелей в рулонах.

Данный список, конечно же, не претендует на исчерпывающую полноту и авторы просят тех производителей, которые не оказались в данном списке, заранее извинить их и выражают надежду, что в случае повторных переизданий данной книги список будет значительно расширен.