Lasergene

Sequence Analysis Software
for Macintosh and Windows

GETTING STARTED

An Introductory Tour of the LASERGENE System

Version 5.1 July 1999

DNASTAR, Inc.
1228 South Park Street
Madison, Wisconsin 53715
(608) 258-7420

Copyright © 1999 by DNASTAR, Inc.

All rights reserved. Reproduction, adaptation, or translation without prior written permission is prohibited, except as allowed under the copyright laws or with the permission of DNASTAR, Inc.

Fifth Edition, July 1999
Printed in Madison, Wisconsin, USA

Trademark Information
DNASTAR, Lasergene, Lasergene99, SeqEasy, SeqMan, SeqMan II, EditSeq, MegAlign, GeneMan, Protean, MapDraw, PrimerSelect, GeneQuest, GeneFont , and the Method Curtain are trademarks or registered trademarks of DNASTAR, Inc. Macintosh is a trademark of Apple Computers, Inc. Windows is a trademark of Microsoft Corp. 373 DNA Sequencer is a registered trademark of Applied Biosystems Inc. ALF is a registered trademark of Pharmacia Biotech AB.

Disclaimer & Liability
DNASTAR, Inc. makes no warranties, expressed or implied, including without limitation the implied warranties of merchantability and fitness for a particular purpose, regarding the software. DNASTAR does not warrant, guaranty, or make any representation regarding the use or the results of the use of the software in terms of correctness, accuracy, reliability, currentness, or otherwise. The entire risk as to the results and performance of the software is assumed by you. The exclusion of implied warranties is not permitted by some states. The above exclusion may not apply to you.

In no event will DNASTAR, Inc. and their directors, officers, employees, or agents (collectively DNASTAR) be liable to you for any consequential, incidental or indirect damages (including damages for loss of business profits, business interruption, loss of business information and the like) arising out of the use of, or the inability to use the software even if DNASTAR Inc. has been advised of the possibility of such damages. Because some states do not allow the exclusion or limitation of liability for consequential or incidental damages, the above limitations may not apply to you.

DNASTAR, Inc. reserves the right to revise this publication and to make changes to it from time to time without obligation of DNASTAR, Inc. to notify any person or organization of such revision or changes. The screen and other illustrations in this publication are meant to be representative of those that appear on your monitor or printer.

CONTENTS

UPDATING & INSTALLING
ON MACINTOSH

Updating via the Internet

Software Installation

Network Installation

Troubleshooting

UPDATING VIA THE INTERNET

If you have previously installed ***Lasergene99*** and have a current Updates and Support Contract, you may update your software through our website. Each module of the ***Lasergene99*** system is stored in a self-extracting archive. You select and download each application individually.

What You Need

- Your user name and member number. Both are printed on your **DNASTAR Installation Disk for Macintosh**.

Downloading Updated Applications

- Create a temporary folder on your hard drive for backing up your existing ***Lasergene99*** software.
- Locate the application file(s) that you would like to update, and move them from the DNASTAR folder to the temporary backup folder.
- Connect to the Internet and type the address http://www.dnastar.com into the navigation bar.
- Choose **Customers** from the menu options.
- Click **Lasergene Updates** to open the Enter Password dialog. Fill in your registered user name and password, which is the same as your member number. The download page opens.
- Scroll to the section entitled **Macintosh Software**.
- If desired, you may read about the release version of a particular module by clicking the **More Info** button. If you do so, be sure to click the **Back** button on your browser to return to the downloads page.
- Click the blue link located to the right of an application to open the Download Manager and download the application to the location of your choices.

- When you have finished downloading the desired applications, you may close your web browser.

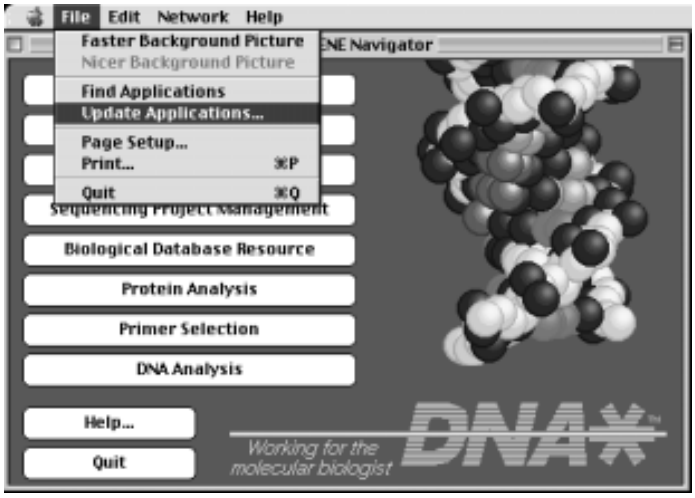
- For each application, double-click the self-extracting archive files (.sea) on the desktop to open the Select Destination Folder dialog.

- Double click folders until the folder DNASTAR appears to the right of the word **Folder**.

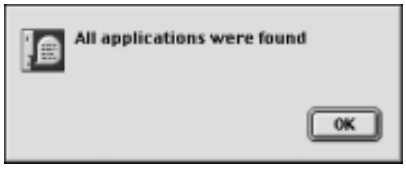
- Click **Extract**. If you did not remove the old version of the application from the DNASTAR folder, you will receive the message: **Application name already exists**. Click **Replace ALL Duplicates** to continue extracting the application.

- Use the Macintosh Finder to open the Lasergene Navigator, located in the DNASTAR folder.

- From the **FILE MENU**, choose **Update Applications**, as shown at right. You will be prompted to insert your installation disk if you have not already done so.



- After a brief wait, the message below should appear, informing you that all updated applications have been located.



SOFTWARE INSTALLATION

This section describes the procedure for installing ***Lasergene99*** for Macintosh from a CD. Note that the installation procedure requires you to quit all open applications and directs you to restart your computer upon completion.

What You Need

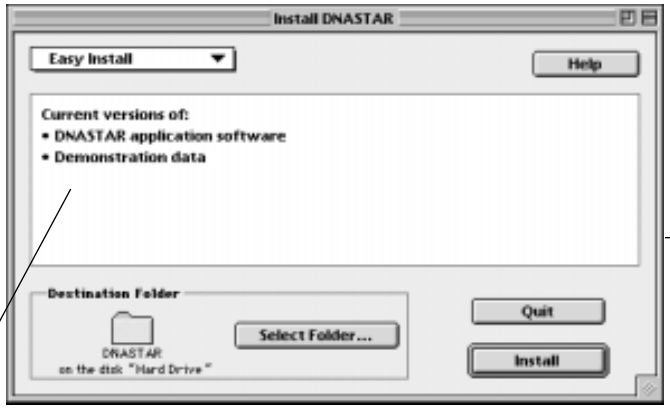
- A personal **Lasergene Installation Disk for Macintosh**.
- A **Lasergene Applications** CD.
- Sufficient hard disk space and memory. For a full ***Lasergene99*** system you will need at least 30 Mb hard disk space and a recommended minimum of 32 Mb RAM.

Lasergene99 Installation from a CD

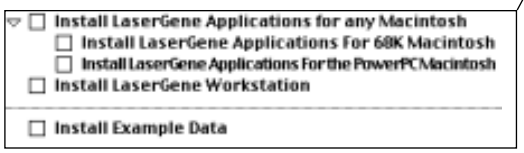
- Insert the **Lasergene Installation Disk for Macintosh** and the **Lasergene Applications** CD.
- Double-click on the installation disk icon.  This opens the window shown at right.
- Click **Continue** to open the Install DNASTAR window shown on the next page.

There are two different methods for installing ***Lasergene99*** software: **Easy Install** and **Custom Install**. **Easy Install** installs applications capable of running on both the 68K-family and Power Macintoshes. **Custom Install** allows you to install workstation files, application files or sample data in any combination you choose, or to install either the 68K-family only or Power Macintosh only versions. If disk space is not of great concern, we recommend choosing **Easy Install**. If you are using exclusively the 68K-family or Power Macintoshes, however, you can save about 5 Mb of disk space by choosing **Custom Install**.

- Using the pull-down menu at the top left of the window, choose between **Easy Install** and **Custom Install**. If you choose **Custom Install**, you'll need to follow the additional step described in the box, below, before proceeding:



If you choose **Custom Install**, the center portion of the window changes to offer the following options:



- Click the triangle to the left of **Install Lasergene Applications for any Macintosh** to reveal two additional options for custom installation, as shown. Check the **Install Lasergene Applications for any Macintosh** box to install the default version of the software, or the 68 K or PowerPC Macintosh box to install versions optimized for those computers. Check **Install Lasergene Workstation** to install the Navigator, DNASTAR fonts and other required system resources. You must select this option if you have not installed ***Lasergene99*** on this computer before. Check **Install Example Data** to install optional demonstration sequences for all modules.

- By default, ***Lasergene99*** is installed on your hard drive in a folder named DNASTAR. If desired, you may change the destination folder as follows:

  - Click **Select Folder** to open the Select Installation Folder dialog.
  - Double-click folders until you have located the desired destination folder, or click **New** to create a new folder.
  - Click **Select "Destination Folder"** to return to the Install DNASTAR window.

- Click **Install** and follow the on-screen instructions. When you have finished installing ***Lasergene99***, the dialog box below appears, informing you that the installation was successful.

- Click **Restart** to complete the installation.

If you have trouble with any part of this installation, refer to the *Troubleshooting* section on page 11, or contact DNASTAR.

NETWORK INSTALLATION

The ***Lasergene99*** network system runs on any Macintosh on a local area network (LAN) comprised of a file server connected to one or more client Macintoshes. Choose a fast machine with a large hard drive to be your file server. A DNASTAR hardware device called a dongle is used to monitor the number of workstations currently using ***Lasergene99*** applications.

What You Need

- A **Lasergene Installation Disk - Server Version**.
- A **Client Disk Set**, which includes the **Installation Disk - Client Version** and **Disk 1 - Client Version**.
- A **Lasergene Applications** CD.
- A DNASTAR network dongle.
- Sufficient disk space and memory. A **Server Installation** requires 30 MB of free hard disk space and 8Mb RAM. A **Client Installation** requires 1 MB of free hard disk space and a recommended 32 Mb RAM.

Definitions

Before installing your ***Lasergene99*** network system, you should be familiar with the following terms:

Applications - For the purpose of this discussion, **Applications** refers to EditSeq, GeneMan, GeneQuest, MapDraw, MegAlign, PrimerSelect, Protean, and SeqMan II. The **Navigator** is an important part of the network system, but will not be discussed as an application.

Application Server -The machine where **Applications** are stored. It is usually the same machine as the **Dongle Server**, but may be different. There can be multiple application servers for one network system. This is often the case when people install the network applications on their local hard drive. The **Application Server** does not have to be a Macintosh. The machine that serves the applications does not need the ability to run the software, only to store it.

Client Installation - (also known as **Workstation Installation**). Installs the minimum software needed to run the network version of ***Lasergene99***. It includes the following: **LicenseKeeper**, **Navigator**, and system resources.

Client - This can be any machine on the network that is not the **Dongle Server**. A **Client** is any machine that has gone through at least a **Client Installation**, so that it contains at least the minimal software to run the ***Lasergene99*** applications.

Dongle - A hardware device required by the ***Lasergene99*** network. Each network system requires a single **Dongle**. The dongle is a small rectangular device with a short cable on one end. There are two different types of dongles. One has a 9-pin male connector that connects to the modem port on the back of the computer, while the other has a 4-pin male connector that connects between the mouse and the keyboard.

Dongle Server - This is the machine that has the **Dongle** attached to it. It must be a Macintosh that has gone through at least a **Client Installation**. This machine should remain on at all times.

Navigator - A tool designed to make launching the ***Lasergene99*** applications easier. It contains a series of alias buttons that launch the desired application. **Navigator** will automatically mount a network drive, if necessary, and also contains diagnostic tools for the network.

LicenseKeeper - A proprietary system extension used to communicate with the **Dongle**. It tells the dongle which application is being launched so that the dongle can give permission to the machine attempting to launch the application.

Server Installation - This is a full installation of ***Lasergene99*** software, including all applications, **LicenseKeeper**, **Navigator**, and system resources.

Dongle Installation

The first operation for establishing a ***Lasergene99*** network is installing the dongle. Before you proceed, please determine whether your dongle has a 9-pin connector or a 4-pin connector.

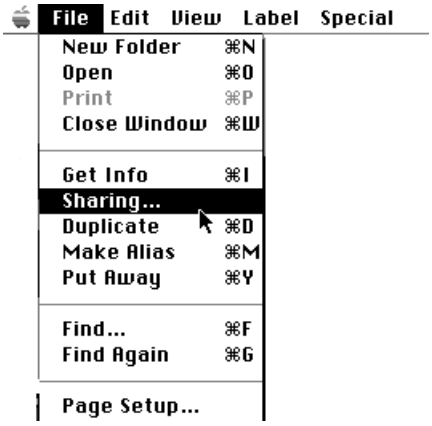
- Turn off the power to the Macintosh that you are going to establish as the dongle server.
- If your dongle has a 9-pin connector, attach it to the modem port on your computer. The modem port is an external 9-pin receptacle located near the icon of a phone handset on the back panel of your Macintosh.

- If your dongle has a 4-pin connector, disconnect the mouse from the keyboard. Connect the dongle to the keyboard, then connect your mouse to the dongle.

- Restart your Macintosh.

Once your dongle has been connected to the Macintosh, you must establish it as the dongle server by performing a software installation. If you plan to use this machine as a file server, follow the instructions for *Server Installation*, below. If you do not plan to use this machine as a file server, follow the instructions for *Client Installation*. After performing a successful software installation on the dongle server, the server icon at right should be displayed during startup.

Server Installation



This installation is the same as the installation procedure described in *Software Installation* on page 5. Since most networks contain many different types of machines (both 68K and Power Macintosh machines), we recommend that you choose **Easy Install**.

After completing the software installation, be sure to "share" the folder containing the ***Lasergene99*** applications so that other network users will have access. If you are not sure how to do this, contact your network administrator.

Client Installation

Client Installation must be performed on all machines that require access to the ***Lasergene99*** network but that are not acting as file servers.

- Insert the **Installation Disk - Client Version** and follow all instructions given under *Software Installation* on page 5.

- Continue by mounting the file server on the client machine using the **Chooser** from the Apple menu. Some networks might use other protocols for mounting volumes. If you are having problems with this step, contact your network administrator.

- Launch Navigator from the DNASTAR folder on the client machine.

- From the **FILE MENU**, select **Find Applications**. You should receive the message: **All applications were found by the Navigator**.

Lasergene99 applications can be now launched from the file server by clicking the appropriate button on the Navigator.

TROUBLESHOOTING

System Extension Conflicts

Some system extensions may conflict with the installation process. If you are unable to install successfully from the ***Lasergene99*** CD, it may help to turn off unnecessary extensions:

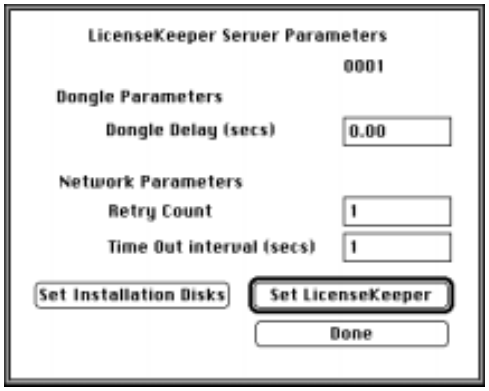
- Use the Extensions Manager, located in the Macintosh Control Panel, to disable any unnecessary extensions.
- Restart your machine and continue with the installations.

Network Troubleshooting

An unrecognized dongle server is the most common problem associated with the network system. You may discover this problem when you try to launch a DNASTAR application and instead receive the message: **DNASTAR License Server is not responding, try again later**. A dongle server may be unrecognized for several reasons:

- If the Macintosh containing the dongle is shut down or crashed, restart the dongle machine to enable the network.
- If the dongle is not connected to a Macintosh or is not seated properly in the modem port or mouse port, reinstall the dongle on a Macintosh following the *Dongle Installation* instructions on page 8. Remember always to shut down the Macintosh before removing or inserting a dongle.
- If your dongle is connected to IIfx, Quadra 900, or Quadra 950 Macintoshes, you need to perform the following steps in addition to the setup already described:
  - Choose **Control Panels** from the Apple menu, then double-click on **Serial Switch**. A window appears as shown at right.
  - Set the switch setting from **Faster** to **Compatible** and close the window.
  - Restart the Macintosh.

- If the dongle is connected to the modem port properly but is not identified by the computer, it could be a sign that your computer is not recognizing the modem port during the booting process. Allow LicenseKeeper more time to find the modem port as follows:

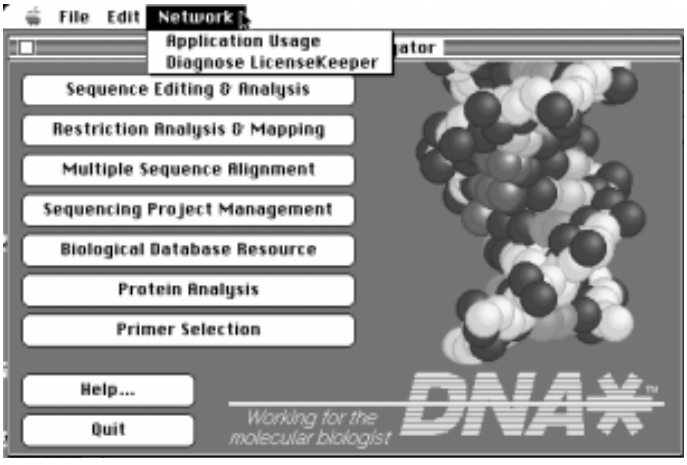


- Start **Navigator**.
- While holding down the **Option** key, go to the **NETWORK MENU** and choose **Set LicenseKeeper** to open the dialog box at right.
- Set the **Dongle Delay** parameter to **3 seconds**.
- Click **Set LicenseKeeper**, then click **Done**.
- Restart the Macintosh.

If the problem persists, contact DNASTAR.

Other Tools for Solving Network Problems

To open a window displaying the status of the licensing system on your computer and all DNASTAR dongle servers accessible over the network, go to the **NETWORK MENU** and choose **Diagnose LicenseKeeper**. The status information reports whether a dongle is attached to your computer and whether the LicenseKeeper system extension is present. This report also shows the site identification and version numbers for the LicenseKeeper. Regardless of the status of the licensing system on your computer, Navigator displays a report of all accessible dongle servers. The License Server Report is organized by network zone and shows the user name, site identification and version numbers of all dongle servers responding on your network. Usually, your LAN contains a single dongle server but there may be several on large networks.



If your computer is connected to a large LAN, you may experience a delay opening this report. Locating all LicenseKeeper servers requires that a network request is sent to each zone on the LAN. It takes at least one second to determine that no answer is forthcoming from unanswered requests.

LicenseKeeper Error Reports

If the licensing system has not been installed or is encountering difficulty communicating with the dongle server, Navigator diagnoses the problem and reports an error message. Below, error messages are shown in bold type, followed by their respective causes.

- **The LicenseKeeper system extension is not in the system folder or the extensions folder. Reinstall the Lasergene99 client software**. Navigator could not communicate with the system extension and did not find one in the system folder. For System 7 or later Macintoshes, this problem can also occur if you restart the computer while holding the **Shift** key down.

- **The mouse button or the Shift key was held down when the machine was started. Restart your machine.** Navigator could not communicate with the system extension because the system extension was disabled at startup.

- **A very old version of AppleTalk is installed on your machine. Install AppleTalk version 53 or later on your machine.** The system extension indicated that AppleTalk protocols required for its operation are not supported by the AppleTalk version on your computer. On small networks or closely-knit installations, this will not be a problem, since the dongle server will be in the same zone as all clients. If necessary, DNASTAR can supply the drivers and installation script for a more recent version of AppleTalk.

- **Could not register LicenseKeeper on the network. Contact DNASTAR.** The system extension could not gain access to the network due to too many other network processes on your computer. This is extremely rare, but may occur if your computer is acting as a network bridge.

- **There is insufficient system memory for the LicenseKeeper system extension. Remove unneeded system extensions.** This may occur if your operating system is using all of the memory available on your computer.

- **The system folder is protected. Unprotect the system folder and restart your machine**. Files in the system folder and used by the system extension could not be written. You may encounter this problem with some disk compression and password protections extensions.

- **Reinstall the Lasergene99 workstation software.** The files in the system folder which are used by the system extension were not present.

Application Usage

To open a window displaying the number of users connected, how long they have been connected and which applications they are using, go to the **NETWORK MENU** and choose **Application Usage**. The window also presents general information about the licensing system such as client/server status, site identification and version numbers of the LicenseKeeper system extension. Application usage information is updated by Navigator every ten seconds while this window is open.

Further LicenseKeeper Information

At startup, the LicenseKeeper queries the network for zone information and looks for a dongle server in each zone. Only dongle servers with an identification number matching the LicenseKeeper system extension are recognized. This allows more than one dongle server to operate within the same network. Should the network administrator decide to change the name of the zone containing the dongle server and the dongle server is restarted, each client outside the zone must restart to be able to find the dongle server.

When the system extension successfully loads as a Dongle Server, it displays this icon at startup.

When the system extension successfully loads as a client, it displays this icon at startup.

When the system extension has been installed but fails to load due to low memory or is disabled (by holding down the mouse button during startup) it displays this icon at startup.

UPDATING & INSTALLING
ON WINDOWS

Updating via the Internet

Software Installation

UPDATING VIA THE INTERNET

If you have previously installed ***Lasergene99*** and have a current Updates and Support Contract, you may update your software through our website. Each module of the ***Lasergene99*** system is stored in a self-extracting archive. You must select and download each application indivually.

What You Need

- Your user name and member number. Both are printed on your **DNASTAR Installation Disk for Windows**.

Downloading Updated Applications

- Create a temporary folder on your hard drive for backing up your existing ***Lasergene99*** software.
- Copy the executable (.exe) files you plan to update into your temporary backup directory. For example, if you wish to update the MapDraw application, copy the mapdraw.exe file from the DNASTAR\MapDraw folder into the temporary folder. Do not simply move the files, as successful installation depends upon the presence of the original files being in their expected locations.
- Connect to the Internet and type the address http://www.dnastar.com into the navigation bar.
- Choose **Customers** from the menu options.
- Click **Lasergene Updates** to open the Enter Password dialog. Fill in your registered user name and password, which is the same as your member number. The download page opens.
- Scroll to the section entitled **Windows 95/98/NT Software**.
- If desired, you may read about the release version of a particular module by clicking the **More Info** button. If you do so, be sure to click the **Back** button on your browser to return to the downloads page.
- To download an application, click the blue link located to its right to open the file download wizard.

- Click **Save this file to disk** to open the Save As dialog.
- Locate and open the folder where you store your DNASTAR applications, then the folder of the application that you are downloading.
- Click **OK**. A message should appear: **Download complete**.
- Click **OK** again.
- When you have finished downloading the desired applications, you may close your web browser.
- Open Windows Explorer and locate the downloaded applications. These applications may appear with the extension "exe" and/or a red icon shaped like a box. Double-click each application to open the extraction wizard.
- Click **Finish**. You shoud see the message: **"Application name" has been updated**.
- Click **OK**.

SOFTWARE INSTALLATION

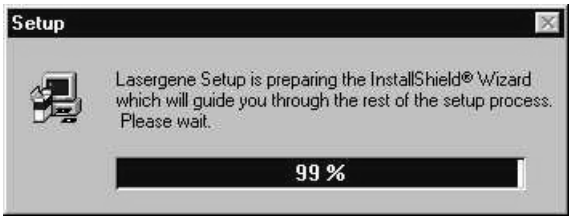
This section describes the procedure for installing ***Lasergene99*** for Windows 95/98/NT from a CD.

What You Need

- A personal **Lasergene Installation Disk for Windows**.
- A **Lasergene Applications** CD.
- Sufficient hard disk space and memory. For a full ***Lasergene99*** system you will need at least 30 Mb hard disk space and a recommended minimum of 32 Mb RAM.

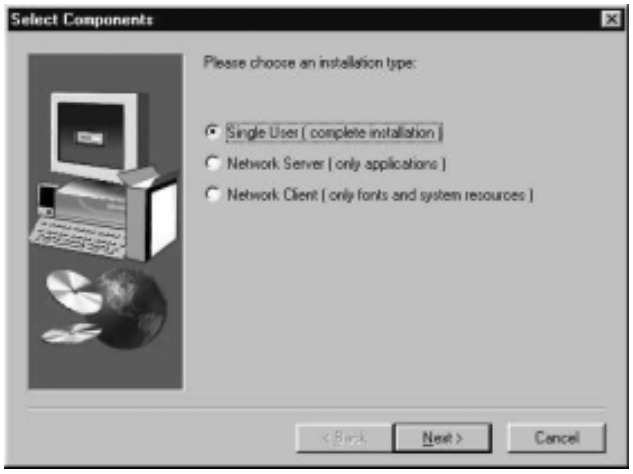
Installation from a CD

- Insert the **Lasergene Installation Disk for Windows** and the **Lasergene Applications** CD.
- From the **Start** menu at the bottom left of your screen, choose **Run**.
- Type **A:\setup.exe**. Click **OK** to begin the installation wizard. You will see the message below.



In a few moments, the welcome screen appears, as shown at right. The top portion of the screen displays the name and institution of the licensee. If this information is incorrect, please contact DNASTAR.

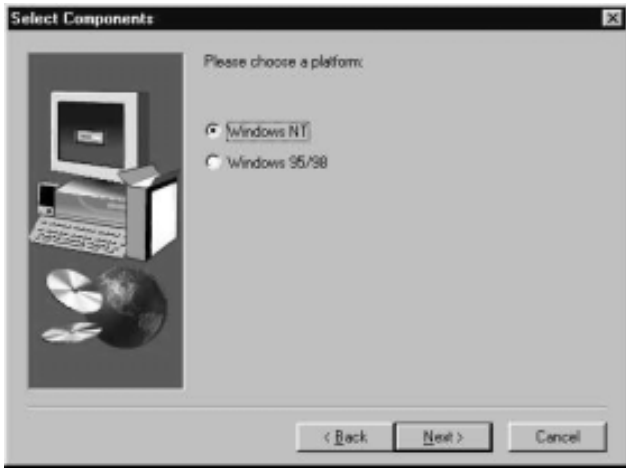
✦ Click **Next**.



Use the dialog at left to set up your ***Lasergene99*** system as a single user, network server, or network client system. **Single User** installs all required components and applications to the local computer. **Newtork Server** installs only the ***Lasergene99*** applications to the local machine, but does not install other required components. **Network Client** installs the required components, but does not install the ***Lasergene99*** applications.

✦ Click the desired option and click **Next**.

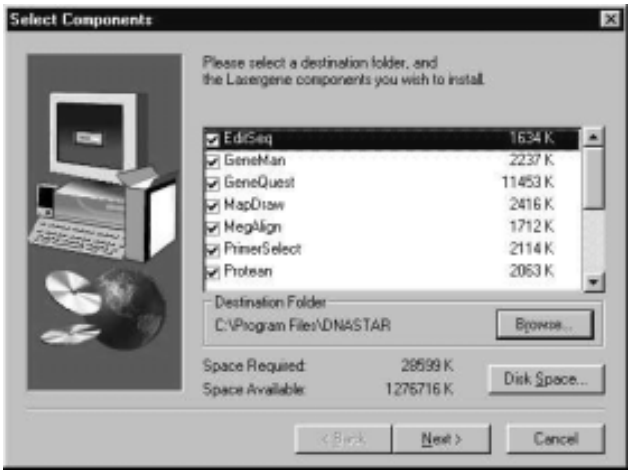
Use the dialog at right to select the platform you will be using to run the ***Lasergene99*** software. The default will vary based on the system software of the computer you are using.



✦ Select the desired platform and click **Next**.

The next screen, shown on the following page, presents you with a list of applications that may be installed. This list will vary based on the customized system purchased from DNASTAR. The default is to install all of the available applications.

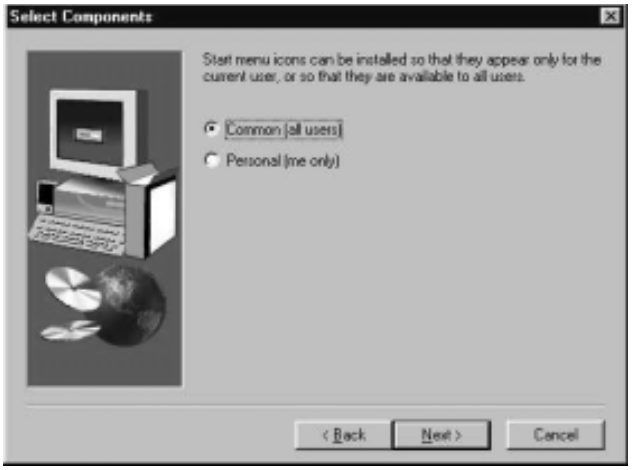
- To omit an application, click to the left of its name to remove the checkmark.



You may also use this window to select or create a destination folder. This is the folder on your hard disk in which the ***Lasergene99*** applications will be placed. The default is **C:\Program Files\DNASTAR**.

- To change the default, click **Browse** and use the menu system to select a new location.

- Click **Next**.

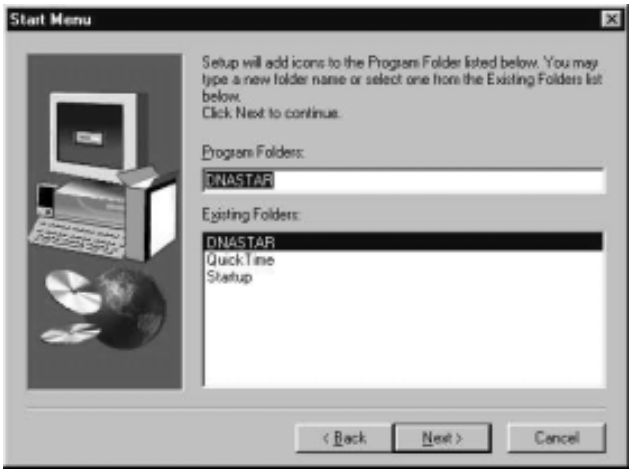


You will only see the dialog at left if you have administrator privileges on the computer that you are using. If you don't see this window, proceed to the next step. If you do see this window, you have the option of creating Lasergene99 icons in the Start menu so that they are available to everyone who logs on to this computer, or only to you.

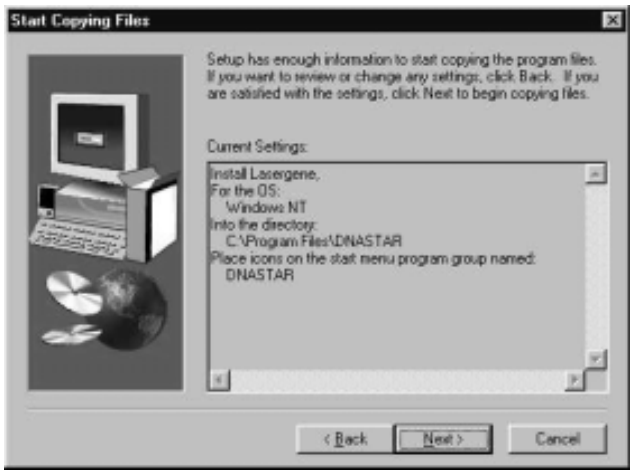
- Click **Common** to create icons for all users or **Personal** to create icons for yourself only.

- Click **Next**.

Use the window at right to specify where to create the program icons that will be used to launch the ***Lasergene99*** applications. You can select a folder by clicking on it or by typing the folder name into the **Program Folders** text box. A folder will be created automatically if one with the name you type in does not already exist. The default folder is called DNASTAR.



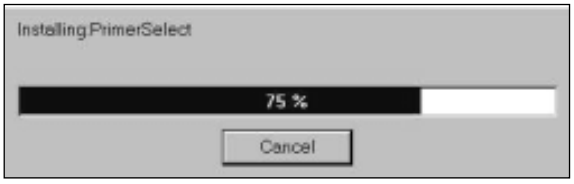
- Select a folder and click **Next**.



The final window, shown at left, allows you to review your choices. If you would like to change any of the information that you provided, click **Back** until you get to the screen you want to change.

- Click **Next** to continue with the installation.

You will see the progress bar shown at right. When the installation is finished, a message appears informing you that it was successful.



- Click **Finish**.

Getting Started with EDITSEQ

Welcome to this brief overview of ***Lasergene99's*** sequence editor for Windows and Macintosh platforms.

EditSeq is a tool for rapidly and accurately entering DNA or protein sequence data into your computer, and for modifying existing sequences. Each EditSeq document is divided into two editable sections, with sequence data in the upper section and annotations and comments in the lower section.

EditSeq can read sequence data from most formats--including FASTA, GenBank, ABI and GCG--as well as ASCII text. You may open or import local sequence files using menu commands or via drag and drop. Sequences may also be entered manually by using the keyboard, by copying and pasting data from another source, or by using the SEQ-EASY II digitizer to read bases directly from an autoradiogram. Sequences located via Entrez or BLAST searching may be downloaded and opened directly from the Internet or Intranet server.

Once a sequence has been opened, EditSeq can translate or reverse translate it using standard or custom genetic codes, search for open reading frames, and provide audible proofreading. EditSeq can also export sequences to GenBank, FASTA and GCG formats.

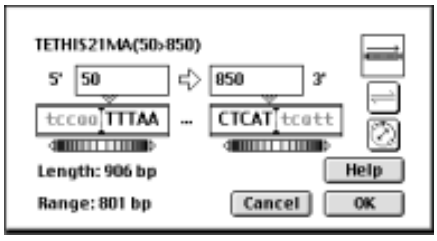
For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

Opening an Existing Sequence

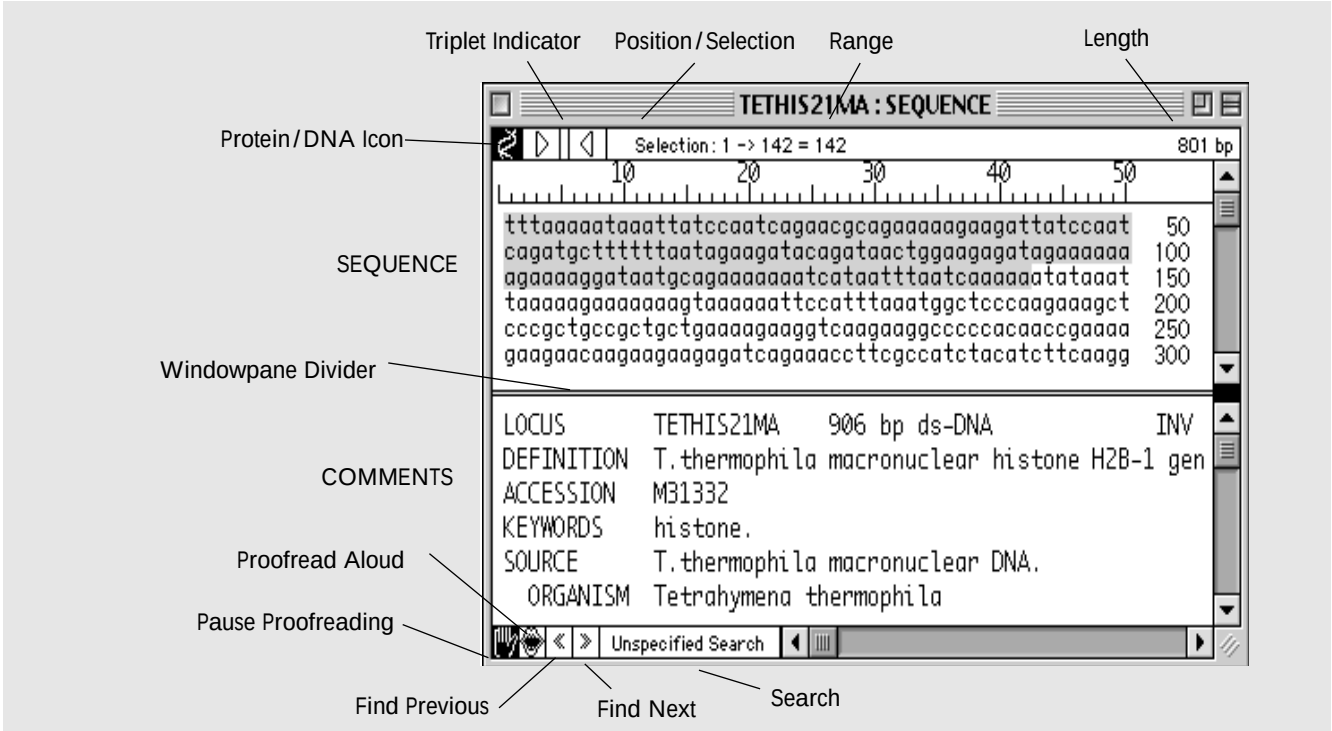
We'll begin by opening the sequence called "tethis21.seq" on Windows and "TETHIS21MA" on Macintosh. Imagine that we suspect the ends of the sequence to be contaminated with vector. We'll use the **Set Ends** command to trim off the 5' and 3' ends of the sequence at the same time we open it in EditSeq.

- From the **FILE MENU**, choose **Open**.
- Open the folder named "Demo Sequences" and single click on the sequence "TETHIS21" to highlight it.
- Click the **Set Ends** button, located at the bottom right corner of the Open dialog. The Set Ends dialog box opens as shown at right.



- Type **50** and **850** into the 5' and 3' end boxes and click **OK**.
- Click **Open** (Macintosh) or **OK** (Windows) to open the sequence. When the EditSeq window opens, notice the length in the upper right corner. By "setting ends," you now have an 801 bp fragment of the original sequence. The **Set Ends** option is available in all ***Lasergene 99*** applications.

Anatomy of an EditSeq Document

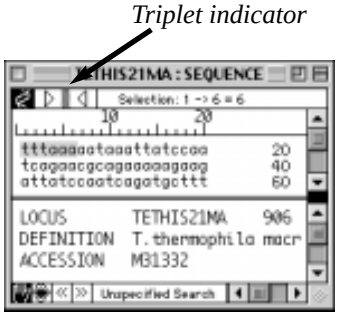


Locating Open Reading Frames

In this section of the tutorial, we will locate the largest ORF in the sequence and translate it.

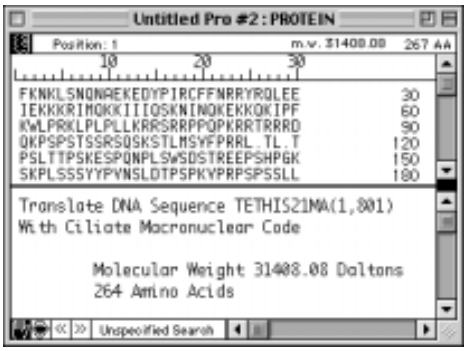
- Go to the **SEARCH MENU** and select **Find ORF** to open the small dialog box at right.
- Click the **Find Next** button to close the dialog and locate the first ORF.
- Click the alternate **Find Next** button (twin right arrowheads) at the bottom left of the EditSeq window until you have located the ORF at position 54-551. The coordinates for the highlighted ORF are displayed near the top of the EditSeq window.

Translating DNA Sequences



We'll be translating our ORF in this tutorial, but any "in-frame" portion of sequence may be translated using the procedure described below. If your selection is in-frame with codon triplets, the triplet indicator bar appears as a solid black line (see figure at left). If you are out of frame, either the left arrow or the right arrow will be solid black, indicating that a one-base pair move is necessary to get in frame.

- If the ORF described in the previous section is not still highlighted, highlight it now.
- From the **GOODIES MENU**, choose **Translate.** Your translated sequence appears in a new untitled protein window like that shown at right. It has been translated using the Standard Genetic Code. In the next sections, we'll experiment with using alternate genetic codes and editing an existing code.
- To continue with this tutorial, close the protein translation window.



Using Alternate Genetic Codes

Depending on the organic source of your sequence, you may wish to select a non-standard genetic code to use when translating or reverse-translating. In this tutorial, we will switch from the Standard Genetic code to the Ciliate Macronuclear code.

✥ From the **GOODIES MENU**, choose **Genetic Codes** to open the submenu shown at left.

✥ Click on "Ciliate Macronuclear." EditSeq may prompt you to locate the code file, which is in the "Genetic Codes" folder in the DNASTAR directory.

Until you select another code, all translation in EditSeq will now be done using the Ciliate Macronuclear code.

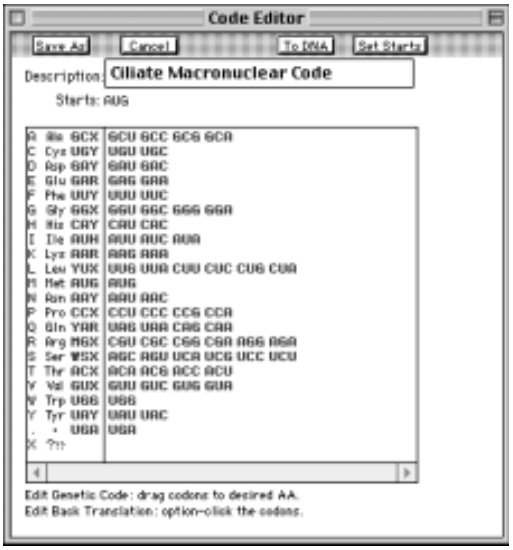
Modifying Genetic Codes

We'll continue this tutorial by modifying the Ciliate Macronuclear code chosen in the section above.

✥ From the **GOODIES MENU**, choose **Edit Selected Code**. This opens a window, like that at right, showing how the genetic code translates DNA and RNA sequences.

✥ To view the code as DNA, click the **To DNA** button.

✥ To edit the code, click on any codon to the right of the vertical black line and drag it from its existing position to a new amino acid assignment. Moved codons appear black, while codons in their original positions are red. Note that the codons to the left of the vertical black line show the overall ambiguity code for each amino acid. These codes change as you move codons on the right from one row to another.



- To use alternate start codons in your code, click the **Set Starts** button. This opens a second genetic code window designed to change start codons. Click on any amino acid (or codon position) to change it to a start codon. The codon position now has an arrow to its left and appears in green. To remove the start designation from a codon, click it.

- To continue with this tutorial, cancel any changes you made and exit from the Code Editor by pressing the **Cancel** button. (If you had wanted to save your changes, you would have clicked the **Save As** button prior to exiting.)

Opening the Reverse Complement or Reversing Sequence Order

Besides translating our sequence, we can also view the reverse complement of the TETHIS21 sequence or reverse its order. The latter procedure is useful for correcting autoradiogram data that have been entered in reverse.

- Highlight a portion of the sequence.

- From the **GOODIES MENU**, choose **Reverse Complement** or **Reverse Sequence**. Choosing **Reverse Complement** opens a new window containing the reverse complement (minus strand) of the highlighted portion of sequence. Choosing **Reverse Sequence** opens a window containing the sequence written in reverse order.

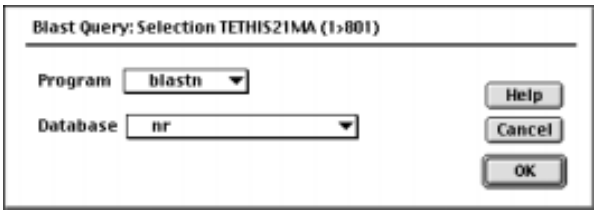
- To continue with this tutorial, close the reverse complement and/or reverse sequence window(s).

Performing a BLAST Search

Next, we'll search for matches to the TETHIS21 sequence on the BLAST server at the National Center for Biotechnology Information. Note that you must have Internet access to perform this BLAST search. If you are not connected to the Internet, skip this part of the tutorial and continue with the next section.

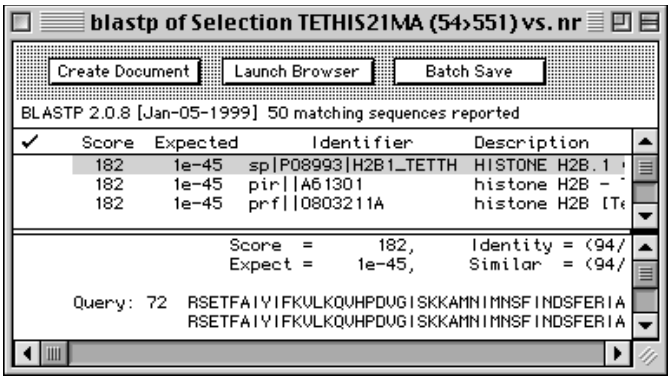
- Remove any highlighting by clicking once anywhere on the sequence. Removing highlighting tells EditSeq to perform a BLAST search on the entire sequence.

- From the **SEARCH MENU**, choose **BLAST Search**. A BLAST dialog box will appear as at right.



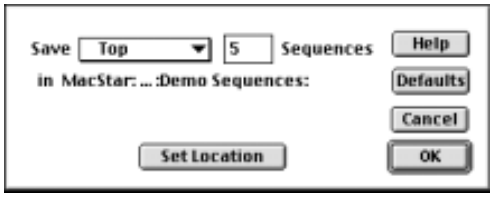
- Leave the defaults so that the program is *blastn*, and the database is *nr*. (For more information on different BLAST programs and available databases, click the **Help** button.)

- Click **OK** to begin the search. EditSeq displays search results in a two-paned BLAST Results window like that shown at left. The upper pane contains the names of possible matches in order of probability, while the lower pane contains the alignment of the query (upper sequence) to the highlighted database entry (lower sequence). Detailed information about how "score" and "expectation" are derived is available at NCBI's website, http://www.ncbi.nlm.nih.gov/BLAST. In general, a higher score and lower expectation connote a better match.



The three buttons at the top of the BLAST Results window are used to open or save the sequence matches or allow you to view more information about them.

We'll start by using the "Create Document" button to open the top five sequences as separate EditSeq windows:



- Click **Create Document**. A small dialog box appears as at right.

- A pull down menu at the top left shows **Top** as the default. Leave the default and type the number **5** into the text box provided to the right of the pull-down menu.

- Click **OK.** EditSeq automatically checks for duplicate sequences. If EditSeq informs you that two or more sequences are identical, respond by clicking **OK**, then **OK** again. EditSeq will retrieve the unique sequences from the Internet database and open them as separate EditSeq windows.

- To continue with this tutorial, close the newly created EditSeq windows.

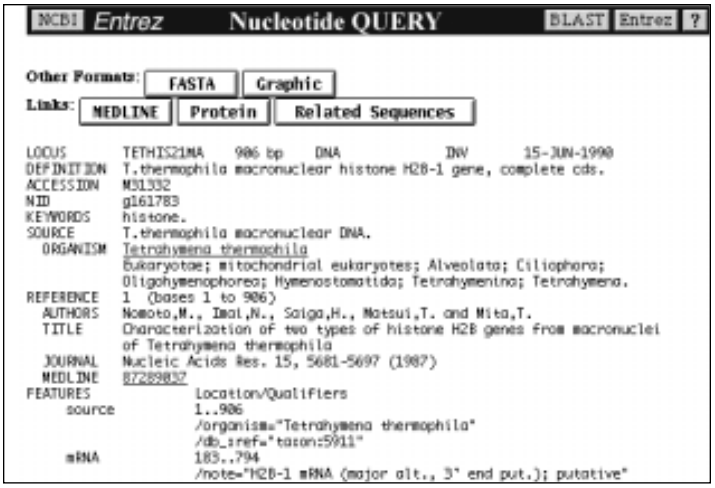
We'll continue by using the "Batch Save" button to save sequences 3-10 as EditSeq documents:

- Highlight the sequence third from the top in the BLAST Results window.

- Click **Batch Save**. A small gray dialog box appears.
- Select **Next** from the pull-down menu whose default is "Top." Type the number **8** into the text box provided to the right of the pull-down menu.
- Click the **Set Location** button to open the Browse for Folder dialog.
- Double-click on the names of folders until you have opened the location where you'd like to save the sequences.
- Click **OK** to return to the gray dialog box.
- Click **OK** to save the sequences in the location you just specified. Sequences are saved under their locus names plus the file extension ".seq". During the downloading process, EditSeq automatically checks for duplicate sequences. If EditSeq informs you that two or more sequences are identical, respond by clicking **OK**, then **OK** again. EditSeq responds by downloading and saving separate sequence files for each unique sequence. Unless you receive an error message, you can assume that your sequences were successfully downloaded.

Finally, we'll view detailed information about a particular sequence by opening its Entrez entry using the "Launch Browser" button:

- Highlight the desired sequence.
- Choose **Launch Browser**. Your web browser will open to display an NCBI Entrez database entry like that shown at right.
- Once you have finished viewing the Entrez entry, close it to continue the tutorial.



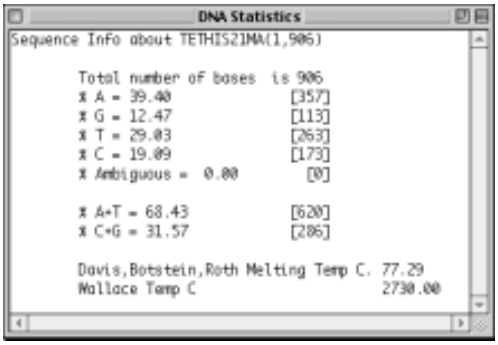
Viewing Sequence Information

Now we will view information about our open TETHIS21 sequence using an EditSeq menu command.

- Highlight a portion of the sequence. If you prefer to select the entire sequence, go to the **EDIT MENU** and choose **Select All.**

- From the **GOODIES MENU**, choose **DNA Statistics**. A window will open as shown at right.

- Once you have finished viewing the statistics window, close it to continue the tutorial.



Proofreading a Sequence

Before we learn how to save and export sequences, let's try out EditSeq's proofreading feature. This feature can help you identify misreadings in your sequencing gels.

- Highlight a portion of the sequence.

- Click the proofread aloud icon (open mouth) at the bottom of the sequence window, or go to the **DIGITIZER MENU** and choose **Proof-Read Sequence**. An electronic voice will begin reading off bases beginning at the start of the highlighted region. (Note: If you do not hear any sound, make sure that your computer speaker is turned on and set to maximum volume).

- To change the speed of voice read-back, go to the **DIGITIZER MENU** and choose **Faster** or **Slower**. Choosing **Speed**, instead, launches a control panel (shown at right) where you can control proofreading speed by dragging the knob up or down.

- To stop proofreading, click the pause proofreading icon (hand) or go to the **DIGITIZER MENU** and choose **Proof-Read Sequence**.

Saving and Exporting Sequences

First, we'll create a new EditSeq sequence to save for this tutorial.

- From the **FILE MENU**, choose **New**, then **New DNA** or **New Protein**.

- Type a string of bases using standard IUB nucleotide or protein codes. Your computer will beep a warning if you attempt to enter illegal characters.

Next, we'll save the sequence as an EditSeq document:

- Go to the **FILE MENU** and choose **Save**.
- Locate a folder in which to store the sequence.
- Give the sequence a name in the **Save Sequence As** box (Windows) or **Name** box (Macintosh).
- Click **Save** (Macintosh) or **OK** (Windows).

To save one or more sequences in GenBank or GCG formats:

- From the **File** menu, choose **Export**.
- Locate a folder in which to store the sequence(s).
- Select a format for the sequence(s).
- Give the sequence(s) a name in the **Save Sequence As** box (Windows) or **Name** box (Macintosh).
- Click **Save** (Macintosh) or **OK** (Windows).

To save one or more sequences in FASTA format:

- From the **FILE MENU**, choose **Export** (for one sequence) or **Export All As One** (for multiple sequences). If both DNA and protein files are open when using **Export All As One**, make sure that the active (top) window matches the sequence type you wish to save. EditSeq will save only the open files of that type to the FASTA file.
- Locate a folder in which to store the sequence(s).
- Select the FASTA format.

- Give the sequence(s) a name in the **Save Sequence As** box (Windows) or name box (Macintosh).

- Click **Save** (Macintosh) or **OK** (Windows).

Getting Started **with GENEQUEST**

Welcome to this brief overview of ***Lasergene99's*** genetic sequence analysis software for Windows and Macintosh platforms.

GeneQuest helps you discover and annotate genes and other features of biological interest in DNA sequences. Such features include open reading frames (ORFS), splice site junctions, transcription factor binding sites, sequence repeats, restriction sites, and many more. Features are located by applying "methods" to the sequence and viewing the graphical display of the method results. You may annotate any features you discover. As with all ***Lasergene99*** applications, GeneQuest also offers an integrated BLAST search function.

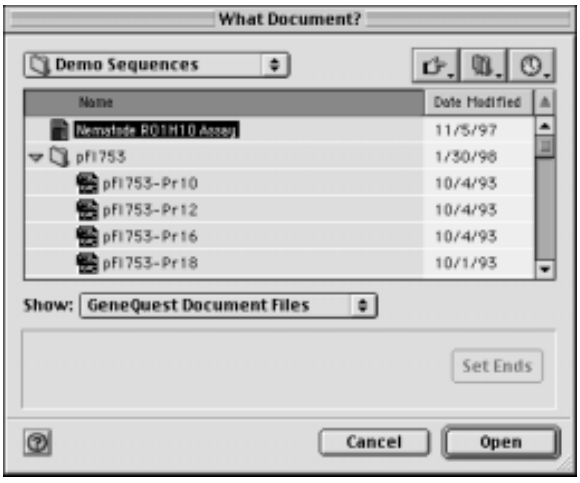
GeneQuest can open DNASTAR, ABI and GenBank files directly. Sequence files in other formats may be converted to DNASTAR format using EditSeq. If you know the accession number or locus name of a Genbank sequence, you may open the sequence directly. Otherwise, you may perform a text search of the entire Entrez database in order to locate and import sequences of interest.

For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

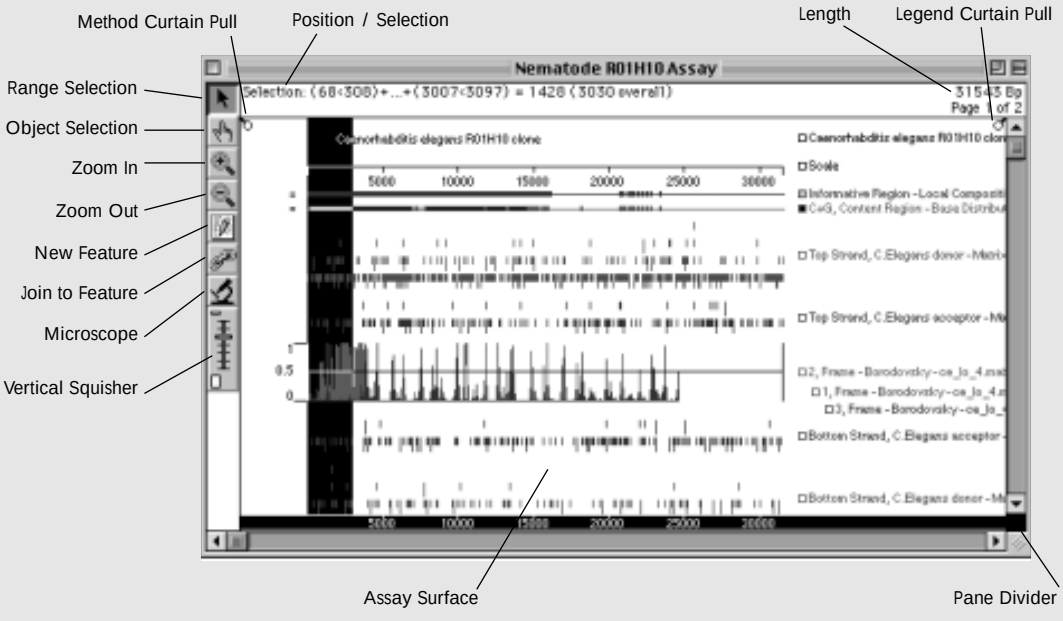
Opening an Assay Document

In this tutorial, we will work with an existing GeneQuest document (also known as a GeneQuest assay) called "Nematode R01H10."



- From the **FILE MENU**, choose **Open** to access a dialog similar to that shown at right.
- On Macintosh, select GeneQuest Document Files from the **Show** menu. On Windows, select GeneQuest Documents from the **Files of Type** menu.
- Using your file management system, open the folder named "Demo Sequences."
- Double-click on Nematode R01H10 to open an assay document similar to the one below.

Anatomy of a GeneQuest Assay Document



GeneQuest DNA Analysis Methods

Once you have opened a sequence in GeneQuest, the next step is to choose and apply methods. The graphic displays of method results are what help you determine features of interest on the sequence. Upon opening your sequence, you will observe that only a few of the methods listed below appear on the assay surface. In the next section, we will learn how to apply additional methods to our sequence.

- **Title** --Adds a title to the document.
- **Ruler** --Adds a ruler to the document.
- **Sequence** --Displays the sequence on the document.
- **Patterns - Matrix** --Uses files containing log odds or positional frequency matrix patterns.
- **Patterns - Signal** -- Uses a text-based pattern database of known DNA transcription factor sites.
- **Patterns - Type-In Pattern** -- Uses patterns that have been typed in using the keyboard.
- **Repeats - Inverted Repeats** -- Locates inverted repeats.
- **Repeats - Dyad Repeats** -- Locates dyad repeats and palindromes.
- **Repeats - Direct Repeats** -- Locates direct repeats.
- **Gene Finding - DNA Finder** -- Indentifies regions that match a user-specified DNA sequence file. Displays results separately for the top and bottom strands.
- **Gene Finding - Protein Finder** -- Identifies regions whose translation matches a user-specified protein sequence file. The display combines results for all six reading frames.
- **Enzymes - Restriction Map** -- Creates restriction maps using any of the enzymes from the DNASTAR enzymes catalog.
- **Coding Prediction - Borodovsky** --Gene discovery tool that uses Borodovsky's Markov method to identify potential coding regions and plot them graphically.
- **Coding Prediction - Starts Stops ORFs** -- Locates and graphically summarizes open reading frames longer than a user-specified minimum length. You can designate whether or not a start codon is required. Starts and stops are displayed independently for all frames.
- **Coding Prediction - Local Compositional Complexity** -- Uses the Shannon information theory to find informationally rich regions and display them graphically.
- **Base Contents - Base Distribution** -- Plots the local frequency of the four bases and A+T and G+C pairings. It also plots the local strand skew of the complements AT and GC.
- **Bent DNA - Bending Index** -- Predicts the bending of free DNA by showing localized angles of helix trajectory computed for a specified window size.

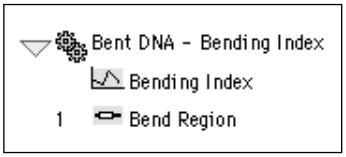
Working with Analysis Methods

To apply new GeneQuest methods, or to add extra copies of methods in use, you move them from the **More Methods** menu to the method curtain, and from there to the assay surface. In this tutorial, we will practice working with analysis methods by applying the **Bent DNA - Bending Index** method to our assay.

- Open the method curtain by going to the **ANALYSIS MENU** and choosing **Show Available Methods** or by dragging it open using the ring located to the right of the **Object Selector** palette tool. The method curtain contains all methods that are ready to be applied to your sequence.

- You will notice that the **Bent DNA - Bending Index** method does not appear in the method curtain. Go to the top of the method curtain and click **More Methods** to open a submenu containing the complete list of available methods. Click once on **Bent DNA - Bending Index** to move it into the main body of the method curtain.

- To determine if a method in the method curtain has already been applied, click the triangle next to its name to display a list like the one at right. If a number appears next to a display icon, as it does in the graphic at right, that display has already been applied to the assay. Since we have not yet applied our method, clicking the triangle next to **Bent DNA - Bending Index** method reveals an absence of numbers next to either Bending Index or Bend Region.



- Click in the white space to the left of the method to ungroup the two display types.

- Click on "Bend Region," and drag the display name to the right until your mouse is where you want the method to appear on the assay surface. Release the mouse button. Regions of your sequence that may bend are displayed as a string of boxes on the assay surface.

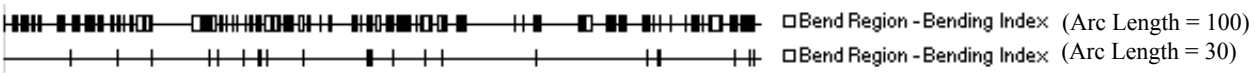
Changing Method Parameters

Next, we'll experiment with changing the parameters for the **Bent DNA - Bending Index** method and comparing the new results to our original results.

- Move a second **Bent DNA - Bending Index** method from the **More Methods** menu to the method curtain and apply its "Bend Region" plot to the assay surface. You should now have two identical region plots showing the bent regions of your DNA sequence.

- Double-click on the name of one of the two **Bent DNA - Bending Index** methods you moved into the method curtain or on the method display itself. This opens the parameters dialog for that method.

- Change the **Arc Length** parameter to 30 and click **OK**. On the assay surface, the corresponding region plot will change to reflect this new formula for calculating bent DNA regions, as shown below. (Note: If you had simply dragged a second copy of the original method onto the assay surface, parameter changes would have automatically been applied to *both* plots.)



Optimizing the Display

To organize or move displays:

- Click the **Object Selector** palette tool (hand icon), located on the left side of the GeneQuest window.

- Select a method display from the assay surface and use the mouse to drag it to the desired location.

To change method formatting:

- Click the **Object Selector** palette tool (hand icon), then select any method on the assay surface.

- From the **OPTIONS MENU**, choose **Line Color** to open a submenu of color choices. When you select a color from the submenu, both the method title and display will change to that color.

- You may also wish to experiement with the similar commands: **Line Weight**, **Fill Color** and **Fill Pattern**.

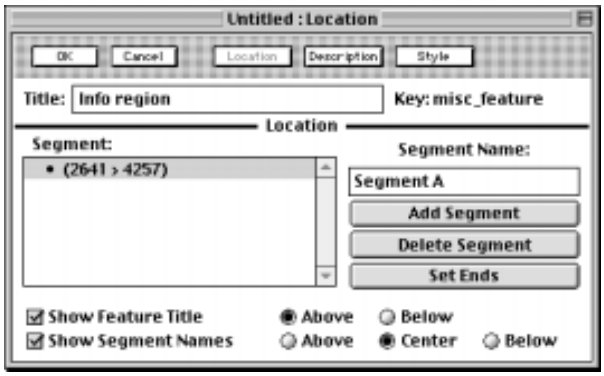
To remove a method:

- Click the **Object Selector** palette tool, select the method display and hit the **Backspace** or **Delete** key. Experiment with this procedure by removing the **Bent DNA - Bending Index** methods from the assay surface. Remove the same methods from the method curtain by selecting them in the curtain and following the same procedure. Remember that any method you remove can always be accessed again from the More Methods menu.

Annotating Features

When you import a sequence with a properly formatted Genbank features table, features from the document are automatically placed as annotation methods at the bottom of the method curtain. These features can be applied just like any other method, by dragging them onto the assay surface. GeneQuest also lets you create new annotations to label areas of interest in your DNA sequence:

- Click the **Range Selector** palette tool (arrow icon), located on the left side of the main GeneQuest window.
- Scroll to the top of the assay window until you see the plot for **Informative Region - Local Compositional Complexity**.
- Locate the bar on this plot which is located at position 2641-4257, and click on it to automatically highlight the range it covers.
- Select the **New Feature** palette tool (pencil icon) or go to the **SITES & FEATURES MENU** and choose **New Feature**. Both open the three-part Feature Editor dialog at right.



- When you enter the Feature Editor, you initially will be in the **Location** window. Type "Info region" into the **Title** box and "Segment A" into the **Segment Name** box, then click the radio buttons at the bottom of the dialog box to show where you wish the title and segment name to appear in relation to the feature.
- Click the **Description** button near the top of the window to switch to the next Feature Editor window. Type in a note about the feature, if you wish, and chose a description key from the menu at the bottom.
- Click **Style** to move to the final window. Chose the font type, size, and color, and the type of graphic you would like for your feature.
- Click **OK** to close the Feature Editor. A graphic of our "Info region" feature will appear at the bottom of the assay surface.

Next, we will link our newly composed feature to an existing feature in the sequence.

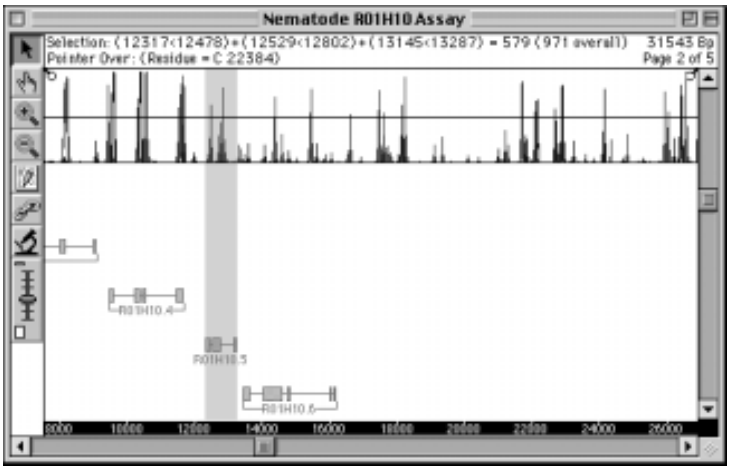
- Click the **Range Selector** palette tool (arrow icon), located at the left side of the GeneQuest window.
- Select the first feature by clicking on the "Info region" feature you just created.
- Click the **Joining Tool** palette tool (chain icon) on the left of the window, or go to the **SITES & FEATURES MENU** and choose **Join To Feature** to open the Join dialog.
- Click the feature named "R01H10.4 -- CDS."
- Click **Join**. Your "Info region" feature will continue to exist as a separate method, but a copy of it will appear as a segment of the R01H10.4 feature to which you linked it.

Performing a BLAST Search

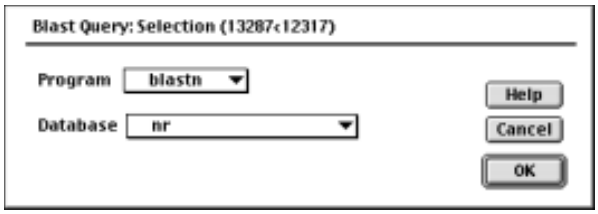
GeneQuest provides two different ways to query Internet or Intranet databases for matches to your sequence. The **BLAST Selection** command allows you to use any portion of sequence as a query. **BLAST ORF** requires you to first select an ORF, then automatically translates it and searches it against a protein database. In this tutorial, we'll search for matches to the Nematode R01H10 sequence on the BLAST server at the National Center for Biotechnology Information. Note that you must have Internet access to perform this BLAST search. If you are not connected to the Internet, skip this section of the tutorial.

- Click the **Range Selector** palette tool (arrow icon), located on the left side of the GeneQuest window.

Click anywhere on the feature called "R01H10.5" to select it, as shown at right.



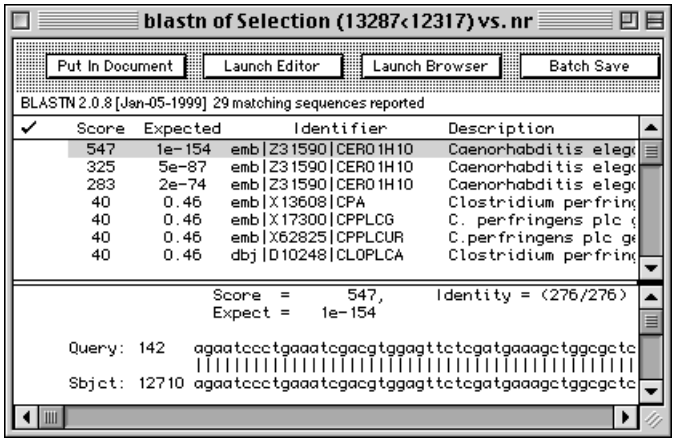
- Although the whole sequence spanning the feature is highlighted, the "Selection" line on the assay header shows that only the segments (exons) have been selected.



✥From the **ANALYSIS MENU**, choose **BLAST Selection**. A BLAST dialog box will appear as at left.

✥Leave the defaults so that the program is *blastn*, and the database is *nr*.

✥ Click **OK** to begin the search. GeneQuest displays search results in a two-paned BLAST Results window like that shown below. The upper pane contains the names of possible matches in order of probability, while the lower pane contains the alignment of the query to the highlighted database entry. For more information on BLAST programs and available databases, see http://www.ncbi.nlm.nih.gov/BLAST. In general, a higher score and lower expectation denote a better match.



The four buttons at the top of the BLAST Results window are used to open or save the sequence matches or to view more information about a particular sequence.

The "Put In Document" button lets you apply a sequence as a **Gene Finding - DNA Finder** method in the current GeneQuest document. (If we had used blastp or blastx, the sequence would be applied as a **Gene Finding - Protein Finder** method). This method is used to determine which DNA regions in your GeneQuest sequence correspond to those in the selected BLAST match.

✥ Highlight an entry in the BLAST Results window.

✥ Click **Put In Document**. A Save dialog window will open.

✥ Locate the location where you wish to store the sequence.

✥ Type a name for the sequence in the text box provided.

✥ Click **Save**. The sequence you chose will automatically be applied as a method at the bottom of the assay document. Unless you are zoomed in very close to the sequence, the method will likely appear as a horizontal black line with several short vertical bars. The vertical bars show where the BLAST match aligns with your sequence.

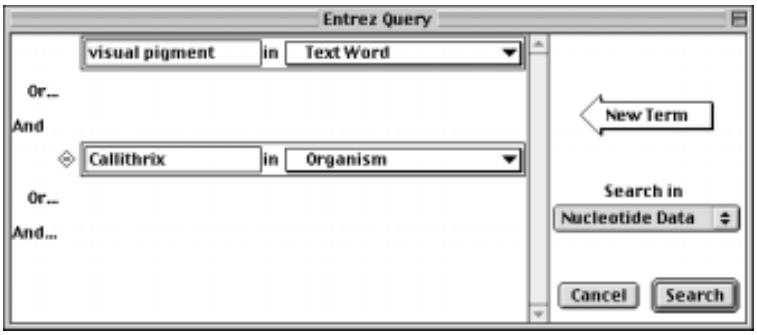
- Zoom in on a bar by clicking the **Zoom In** palette tool (magnifying glass with + sign), then dragging a small box around one of the vertical bars to expand that region. Repeat this procedure until you can see the sequence clearly.

The **Launch Browser** and **Batch Save** buttons work just as they were described in *Getting Started with EditSeq* on page x. The **Launch Editor** button opens EditSeq, then works identically to EditSeq's **Create Document** button.

Searching an Entrez Database

GeneQuest allows you to perform a text search on an Internet or Intranet Entez database. Once you have located matches, you can import them into GeneQuest (or EditSeq) or save them as sequence documents. In this tutorial, we will search for sequences related to visual pigment in marmosets, then learn how to open one of the sequences in GeneQuest or EditSeq.

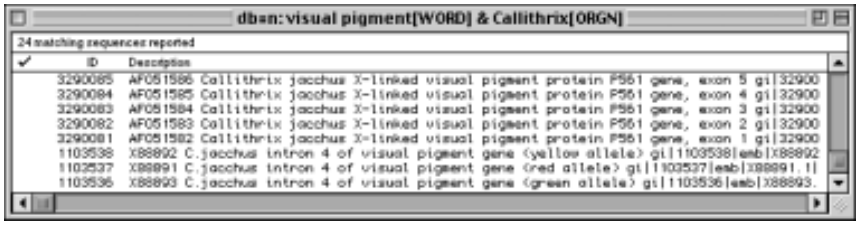
- From the **ENTREZ SEARCH MENU**, choose **New Text Search** to open the Entrez Query window shown at right.



- In the text box whose default is "New Term," type the search query "visual pigment."

- Using the pull-down menu whose default is "All Fields," select "Text Word."

- Add a second term to the search by using the mouse to drag the New Term arrow, located on the right of the window, next to the word **And** on the left. Type "Callithrix" into the "New Term" box and select "Organism" from the "New Fields" menu.



- Click **Search**. The results for your search appear in an Entrez Results window like the one at left.

The Internet site for the National Center for Biotechnology Information (NCBI) features information about each sequence in the database .

- Double-click on any sequence in the Entrez Results window to view details about it using your web browser.

If you wish to open one of the Entrez sequences in GeneQuest or EditSeq:

- From the Entrez Results window, click on the entry that you would like to open.
- From the **ENTREZ SEARCH MENU**, choose **Open with GeneQuest** or **Open with EditSeq**.
- Find a location to save the document using the **Save In** menu (Windows) or directory window (Macintosh). GeneQuest provides the name of the Entrez entry as the default file name, but you may instead elect to enter a new name in the **File Name** box (Windows) or name box (Macintosh).
- Click **OK** (Windows) or **Save** (Macintosh).

If you choose **Open with GeneQuest**, GeneQuest will now open and its default methods will automatically be applied to the sequence. If you choose **Open with EditSeq**, the Entrez sequence will be displayed in the main window, where it may then be viewed or modified.

Additional GeneQuest Features

To view a portion of the sequence as folded RNA:

- Highlight the region of the sequence that you want to fold, go to the **ANALYSIS MENU**, choose **Fold as RNA**.

To simulate multiple digestion experiments, and to examine fragment sizes in predicted agarose gel migration patterns:

- Open the method curtain using the curtain pull.
- Click on **More Methods** and choose **Enzymes - Restriction Map** to add the method to the method curtain.
- Click the blue triangle to the left of the method name to open a list of available enzymes. Note that the method curtain displays only those enzymes that cut the DNA sequence at least once. Upon opening the list, all enzymes are grouped together and highlighted.

- Ungroup the enzyme list by clicking once in the white margin to its left.
- Drag one or more of the enzyme names to the assay surface to display restriction sites for that enzyme.
- From the **SITES & FEATURES MENU**, choose **Agarose Gel Simulation**. A new window opens showing the predicted electrophoretic separation of the restriction fragments.

Saving an Assay Document

- From the **FILE MENU,** choose **Save**.
- Locate a folder in which to store the document.
- Give the sequence a name in the **File Name** box (Windows) or name box (Macintosh).
- Click **Save** to save the assay document. The sequence and all the methods you have applied and displayed will be saved.

Getting Started **with MAPDRAW**

Welcome to this brief overview of ***Lasergene99's*** restriction site analysis software for Windows and Macintosh platforms.

MapDraw produces six types of restriction maps for experimental design, analysis, and presentation of experimental results. The restriction maps range from simple linear minimaps to annotated circular maps and full-featured linear illustrations with up to six reading frames and their translations displayed. MapDraw also displays annoted features for any sequences imported from GenBank.

You may choose the restriction sites you wish to locate and display based on the frequency of sites in your DNA template, the complexity of the site, availability and compatibility with other restriction sites. You may also choose any combination of sites manually. Restriction site filters may be combined using Boolean operators.

MapDraw's tools enable you to plan restriction and cloning experiments and generate summary or fully detailed views of the results. As with all ***Lasergene99*** applications, MapDraw also offers an integrated BLAST search function.

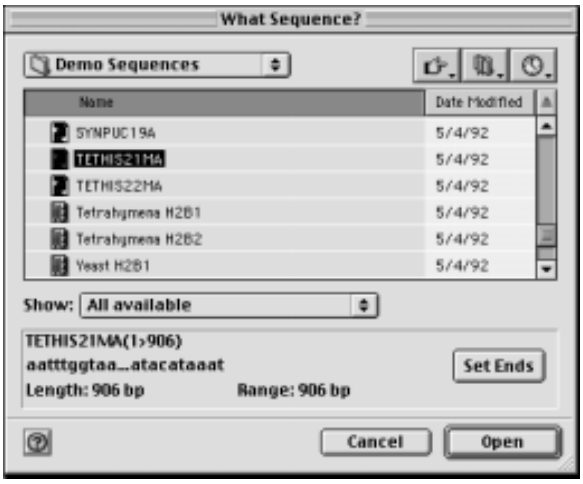
For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

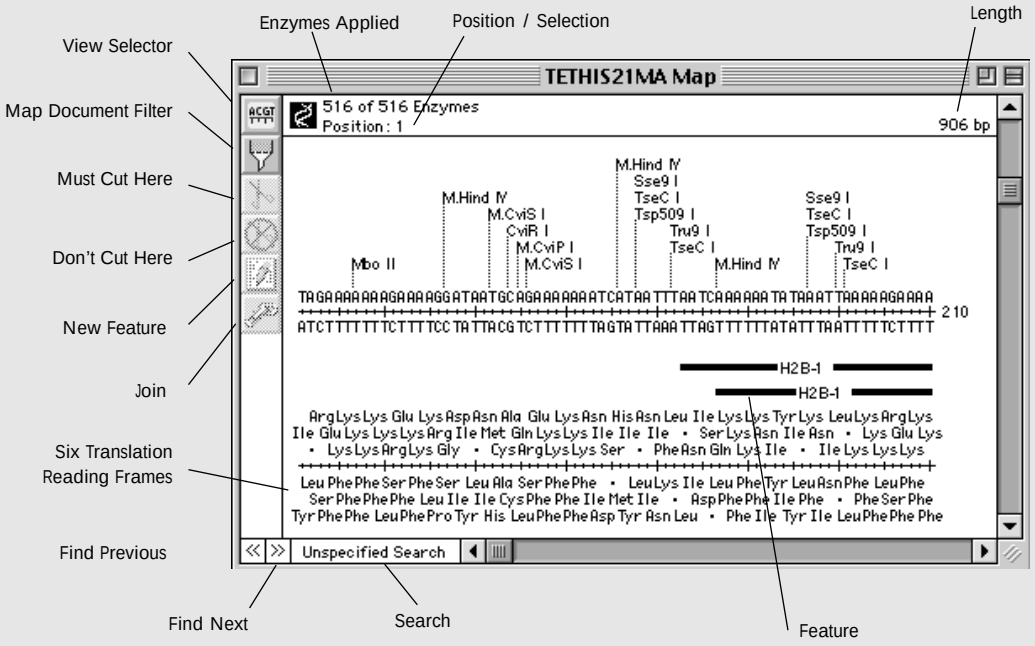
Creating a New Restriction Map

We'll begin by opening a histone DNA sequence called "tethis21.seq" on Windows and "TETHIS21MA" on Macintosh. Our first goal in this tutorial is to find restriction sites that can be used to excise the portion of DNA comprising the sequence's histone gene.

- From the **FILE MENU**, choose **New** to access a dialog similar to that shown at right.
- Using your file management system, open the folder named "Demo Sequences."
- Double-click on the TETHIS21 sequence to open a Site and Sequence View similar to the one below.



Anatomy of a MapDraw Document



Types of Filters

A filter is a subset of the available restriction sites that you define. Until you create a filter, every restriction site is displayed for your sequence. You may combine filters with **And**/**Or** Boolean operators. MapDraw's built-in filters are as follows:

An **Overhang** filter groups restriction sites according to a set of overhang criteria you define. These criteria include ranges for 3' and 5' overhangs, compatibility with any other restriction sites, and use of degenerate overhangs. Use this filter for finding compatible sites in a cloning project.

A **Frequency** filter subsets restriction sites based on the frequency with which they appear in a specified range in the sequence. We'll be using this filter type in the next section of this tutorial.

A **Class & Complexity** filter groups restriction sites by class. These classes include: class I (random), class II (precise) or both. This filter also allows you to subset sites based on site complexity, cost, and availability.

A **Manual Pick** filter allows you to choose restriction sites by name. We'll learn how to design a manual pick filter in a later section.

The **Must Cut Here** and **Don't Cut Here** palette tools also act as filters and will be covered later in the tutorial.

Applying a Frequency Filter

Let's start by using a frequency filter to eliminate from consideration any restriction site that cuts our sequence more than twice.

- From the **ENZYME MENU**, choose **New Filter**, then **Frequency** to open a parameters dialog.
- Input the numbers 1 and 2 in the **Min** and **Max** boxes, but leave all other parameters at their default settings. Our Min/Max settings will automatically eliminate any site present in our sequence more than twice.
- In the **Filter Name** box, type "Two-cuts-max."

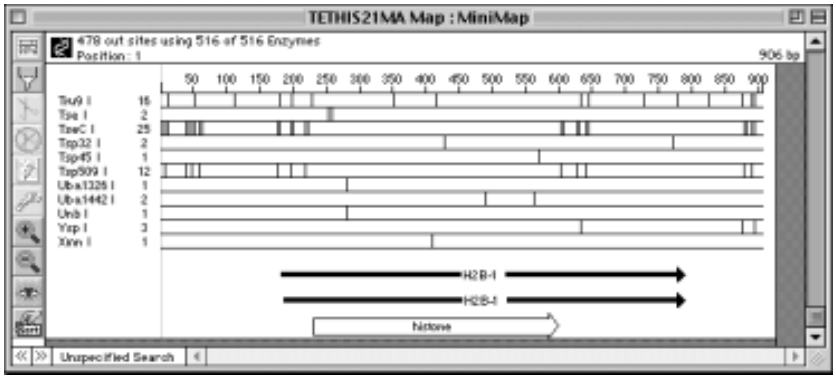
- Click **Apply** to apply the filter to the sequence, then **OK** to exit the parameters dialog. You will notice that the number of sites displayed on the map is now greatly reduced.

Applying a Manual Pick Filter

We've reduced the number of sites by about two-thirds, but over a hundred enzymes still remain. We can afford to set some more stringent demands. Let's see if one of our remaining enzymes could be used to neatly excise the histone coding region from our TETHIS21 sequence.

- From the **MAP MENU**, choose **Linear Minimap**. This opens a view showing cut sites for each restriction enzyme.

- Scroll to the bottom of the window until you see the word "histone," written inside a large, right-pointing arrow (see picture at right). "Histone" is a sequence feature that was annotated in the original Genbank entry. We imported the information for the feature along with the sequence when we opened it.



- Click once on the feature to highlight it.

- Click the **Sort** palette tool on the left of the screen, then choose **Sort by Cuts Close to Selection**. The restriction sites instantly sort themselves so that those cutting closest to and outside the feature appear higher up in the window.

- Scroll to the top of the window. You will notice that there are two restriction sites near the top, Acs I and Apo I, just to the left and right of the histone gene. Either restriction site would be ideal for our purposes.

We will now create a **Manual Pick** filter that includes only the Apo I restriction site.

- From the **ENZYME MENU**, choose **New Filter**, then **Manual Pick**. This opens the Restriction Site Filter Editor shown on the next page.

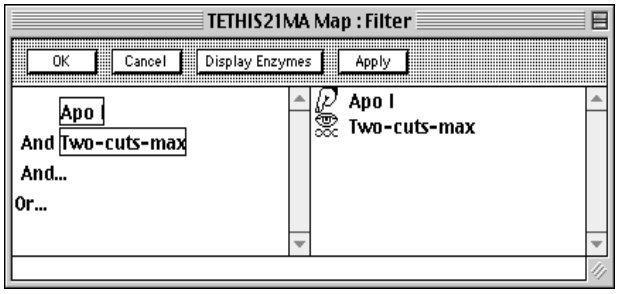
✥Drag Apo I from the right pane into the left pane.

✥Type a name for the filter. In this tutorial, we'll simply call the filter "Apo I."

✥Click **Apply**, then **OK**, to apply and save the filter you have made and close the window. Notice that the linear minimap now consists only of the Apo I restriction site and the histone features.

Using the Filter List

So far in this tutorial, we have applied two filters: a Frequency filter called "Two-cuts-max" and a Manual Pick filter called "Apo I." MapDraw automatically linked the two filters with the Boolean operator "AND," so that a prospective site would have to meet both conditions: Apo I AND cut in one or two places. Since our manual pick filter contained a single-site subset of the existing frequency filter, it automatically "overrode" that filter. The Filter List, shown below, allows us to apply or remove filters from the sequence or to substitute an OR operator for the automatic AND operator.



- From the Map Document, click the **Filter** palette tool (funnel icon at top left of the window) to open a Filter List like that shown at right. All filters currently applied to the map document appear in the left pane.

- To remove the Apo I filter from active use, highlight it, then drag it from the left to the right pane and click **Apply**. The number of sites displayed for the sequence instantly increases since now only the Two-cuts-max filter is applied.

- Drag the Apo I filter back into the left pane and place it next to the **And** operator. Click **Apply** to reapply the Apo I filter and reduce the number of restriction sites on the document. (If we had instead placed the filter next to the word **Or**, the number of sites displayed would not change, since Apo I is also a subset of the Two-cuts-max filter).

- To continue with this tutorial, drag both the enzymes to the right pane so that neither filter is applied, then click **Apply** and **OK**.

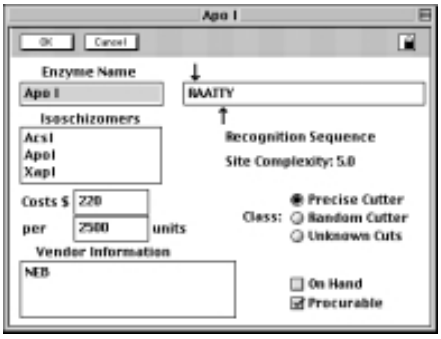
Using the Must Cut Here / Don't Cut Here Palette Tools

An alternative method for filtering restriction sites is to use the **Must Cut Here** and **Don't Cut Here** palette tools. We'll use this procedure to again find which sites best excise the histone coding region from the rest of the sequence.

- From the **MAP MENU**, choose **Site & Sequence**.
- Scroll down until you see the word "histone" written inside a large, right-pointing arrow. Click once on the feature to highlight it.
- Exclude any restriction site that cuts the histone gene by clicking on the **Don't Cut Here** palette tool, the icon that looks like a pair of scissors with a red circle and slash. (Clicking on the **Must Cut Here** tool, the scissors icon, would have excluded all sites except those which cut inside the histone coding region.)

Restriction sites are filtered automatically so that only those not cutting the feature remain.

Displaying Enzyme Information



The Enzyme Editor allows you to view detailed information about an enzyme, including its name, recognition sequence, isoschisomers, class, cost, vendor, and availability. Using the editor, you may also edit, add or delete information from the Enzyme Library. To open an Enzyme Editor window like that at left, double-click on the name of an enzyme from within any view. The arrows pointing toward the recognition sequence show the cut sites for the enzyme.

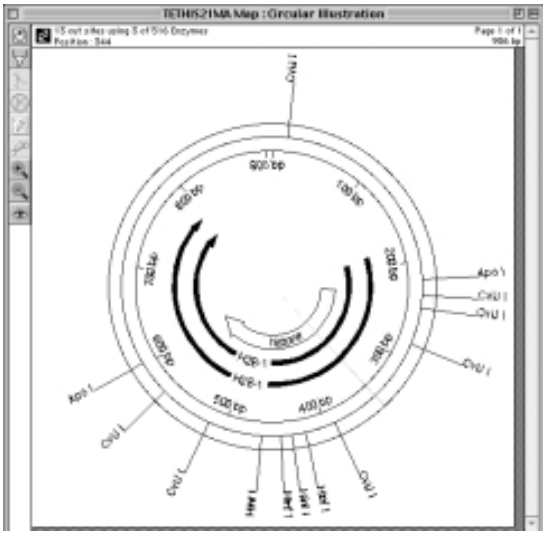
Documenting a New Feature

Besides displaying features you have imported along with your sequence, you may add your own annotations to a sequence by following the directions in "*Getting Started with GeneQuest*" on page 38. Annotations can be made in either application using the same procedure.

Viewing a Circular Illustration

You may view a circular illustration for a circular or linear sequence, but MapDraw will warn you first if the sequence is a linear one. You will usually want to filter the restriction sites for circular illustrations. In our case, the **Don't Cut Here** filter we applied a few sections ago should be sufficient.

- Verify that only the **Don't Cut Here** filter is currently applied by clicking the **Filter** palette tool at the left side of the window. The **Don't Cut Here** filter should be in the left pane of the Filter List and all other filters should be in the right pane. If this is not the case, use your mouse to drag filters to the appropriate pane.



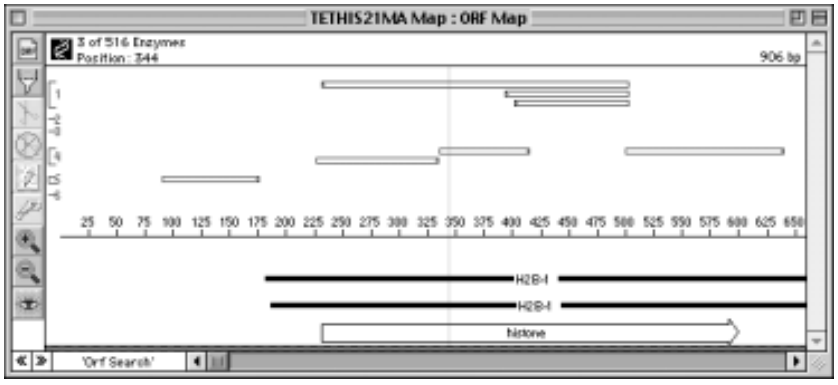
- From the **MAP MENU**, choose **Circular Illustration** to view a picture like that at left. (If there were too many restriction sites at this point, you could apply additional filters).

- Reduce the drawing size by going to the **OPTIONS MENU** and choosing **Drawing Size**. The corner where the two red lines meet shows the minimum size that you may make your drawing. Drag the lower right corner of the rectangle containing the map upwards and toward the left until the corner meets the intersection of the red lines.

Click **OK**.

Viewing an ORF Map

- From the **MAP MENU**, choose **ORF Map** to display open reading frames for all six frames as shown at right. The default stop and start codons are determined using the chosen genetic code. To learn how to choose alternate genetic codes, see *Getting Started with EditSeq* on page 26.



- Set preferences by going to the **OPTIONS MENU** and choosing **ORF Criteria.** You can adjust the minimum ORF length, define an upstream promoter and distance from an upstream promoter, if desired, then click **OK**.

- Display the translated sequence by zooming in on an ORF. To do this, click the **Zoom In** palette tool (magnifying glass icon with +), then click the display. Repeat these two steps until you can see the translation.

Selecting Display Options

MapDraw's **OPTIONS MENU** provides a variety of commands that allow you to do things such as:

- Select 1 or 3 letter amino acid codes (**3 Letter Amino Acid Codes**)
- Display different combinations of reading frames (**Reading Frames** or **Line Layout**)
- Select and edit the genetic code used for translation (**Genetic Codes** and **Edit Selected Code**)
- Select a vertical or horizontal enzyme display (**Enzyme Display**)
- Linearize a circular sequence (**Linearize Sequence**)
- Display uncertain sites (**Include Uncertain Sites**)

Saving and Exiting

- From the **EDIT MENU,** choose **Save Map**.

- Locate a folder in which to store the sequence.

- Give the sequence a name in the **Save Sequence As** box (Windows) or name box (Macintosh).

- Click **Save** (Macintosh) or **OK** (Windows).

- From the **FILE MENU**, choose **Exit**. You will be prompted to save any changes to the enzymes and filters file you used to make your map. Click **Save** to save any changes, new filters, or new enzymes, or **Discard** to discard them.

Getting Started **with MEGALIGN**

Welcome to this brief overview of ***Lasergene99's*** sequence alignment software for Windows and Macintosh platforms.

MegAlign is used to make pairwise and multiple alignments of DNA and protein sequences, providing a choice of six alignment methods. Multiple alignments may be viewed and adjusted in MegAlign's worktable. Phylogenetic trees based on your alignments are generated quickly, and data on sequence distances and residue substitutions are readily tabulated. Customized presentations of multiple alignments may be generated in the alignment view, where similarities and differences may be distinctly highlighted, and where colored histograms of sequence similarity and divergence may be displayed. As with all ***Lasergene99*** applications, MegAlign also offers an integrated BLAST search function.

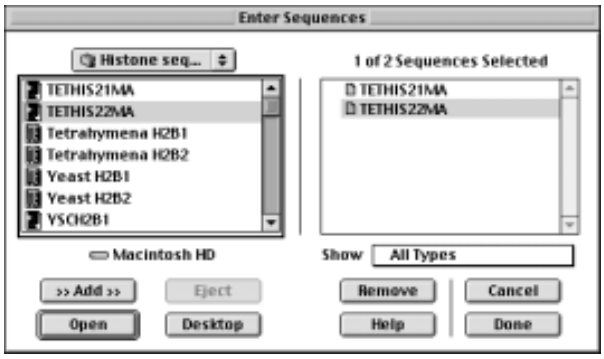
For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

Creating an Alignment Project

MegAlign offers two basic groups of alignment methods: pairwise and multiple. Pairwise alignments may be performed on any two selected sequences, while multiple alignments align every sequence in the Worktable. In the first part of our tutorial, we'll be learning to use two different types of pairwise alignment methods. We'll begin by creating a MegAlign project and entering the two histone sequences we wish to align.

- From the **FILE MENU**, select **Enter Sequences** to open the Enter Sequences dialog box.



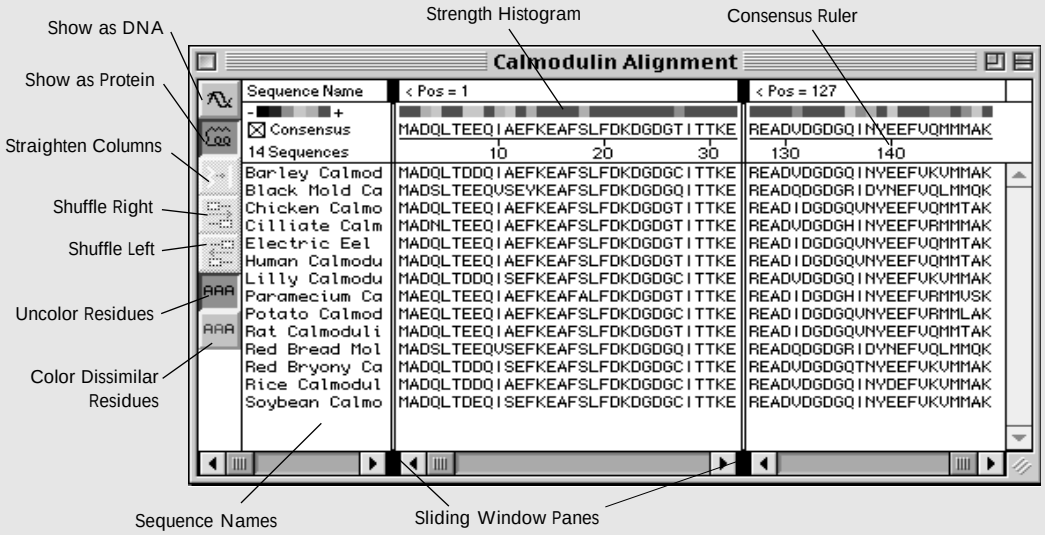
- Locate and open the folder called "Demo MegAlign," located in your DNASTAR directory.

- Double click on the folder called "Histone Sequences." The top two sequences on the left are the ones we will add to the MegAlign Worktable.

- Click TETHIS21, then the **Add** button. Click TETHIS22, then the **Add** button again. The two sequences will appear now in the right pane, as well as the left.

- Click **Done** to move the sequences into the Worktable.

Anatomy of a MegAlign Worktable



- From the **OPTIONS MENU**, use the **Size** command to increase the font size until characters in the Worktable are legible.

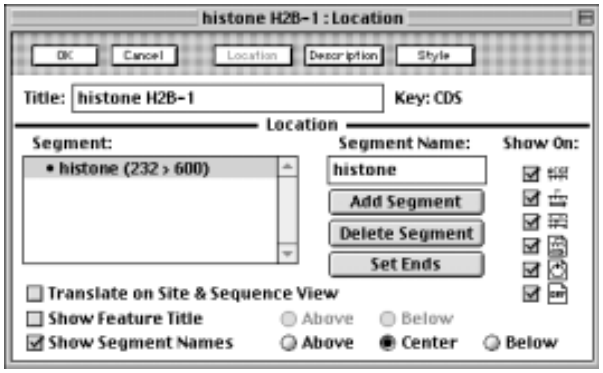
Setting Sequence Limits

Next, we'll practice subranging the two histone sequences that we entered in the previous section. The ability to limit the endpoints of sequences is not important in our particular case, but becomes very important when aligning 1) sequences that may not match well over their entire lengths, or 2) a mixture of protein and DNA sequences. In the latter situation, each DNA sequence must be in-frame for correct translation to occur.

Limits for sequences can be set manually, by typing in the endpoints, or automatically, by limiting the sequence to the range contained within a specified annotated feature. Here, we will limit sequences using the histone H2B-1--CDS feature.

- Click the name of the first sequence in the list (TETHIS21) to highlight it.

- From the **OPTIONS MENU**, choose **Set Sequence Limits,** then **From Feature Table**. This opens the Set Sequence Ends window at right. The **histone H2B-1--CDS** feature in the upper pane is already highlighted. The lower pane shows the feature name and range of the highlighted sequence.



- Click **Change the Rest** to limit the second histone sequence using the same feature. You will automatically be returned to the MegAlign Worktable, where subranges of the original sequences appear.

Performing a Pairwise Alignment

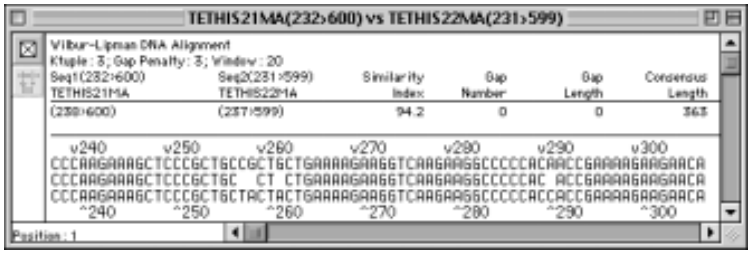
MegAlign offers a variety of pairwise alignment methods. Since our sequences are DNA, the methods available to us are Wilbur-Lipman, Martinez-Needleman-Wunsch and Dotplot. In this tutorial, we will be using the Wilbur-Lipman method.

- Hold down the **Shift** key while you click on each of the two sequences to highlight them. (Even if there are only two sequences on the worktable, you must highlight them before doing a pairwise alignment).

- From the **ALIGN MENU**, choose **One Pair,** then **By Wilbur-Lipman Method.**

- Click **OK** to use the default parameters for the alignment. MegAlign calculates the alignment, then displays results in a separate Alignment window like that below.

The window header displays the similarity index (percent of all residues that are matching), gap number, total gap length and consensus length. We can change the color scheme to more easily distinguish matches from mismatches.



- Click on the **Alignment Color** palette tool (box icon with "x"), to open the Alignment Color dialog shown below.

- Click the dark blue box, then click the black box to the right of **Match Color** to change it to blue, as well. All residues that match the consensus immediately turn dark blue on the Alignment View. Repeat, choosing red for **Mismatch Color** and green for **Consensus Color**.

- Check and uncheck each of the four boxes just under the color selection boxes, and notice how the Alignment View changes.

- Click **Vertical Bars** to show agreement with the consensus as green vertical lines instead of letters.

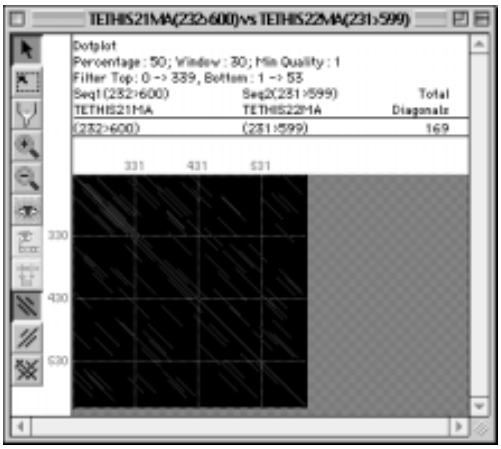
- Once you have finished experimenting with the Alignment Color dialog, close it and the Alignment View to continue with the tutorial.

Using the Dot Plot Method

The Dot Plot method compares sequences by overlapping them, counting mismatches, and shifting the sequences one residue relative to one another.

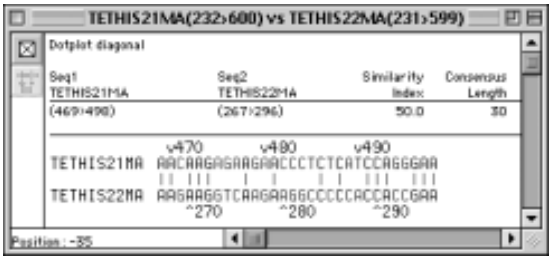
- Hold down the **Shift** key while you click on each of the two sequences to highlight them.

- From the **ALIGN MENU**, choose **One Pair**, then **Dot Plot** to open the resulting plot in a separate window like that at right.



- Click **OK** to use the default parameters for the alignment. MegAlign calculates the alignment, then displays results in a separate window.

Each match having the specified similarity within a particular group of residues (both designated in the parameters dialog) is displayed in blue on the Dot Plot. The red diagonal line marks where both sequences are aligned, starting at their left ends.



- Double click on any blue line to open a window like that at left, showing the underlying alignment of the selected diagonal.

- Once you have finished viewing the underlying alignment window, close it and the Dot Plot.

Performing a Multiple Alignment

To demonstrate the multiple alignment feature in MegAlign, we'll close the current two-sequence project and open a project containing fourteen related calmodulin sequences. This larger data set will give us the opportunity to create phylogenetic trees and to explore additional MegAlign features.

We'll begin by closing the current MegAlign project:

- From the **FILE MENU**, choose **Close**. When prompted to save the document, click **Discard**.

Next, we'll create a new project containing the fourteen calmodulin sequences:

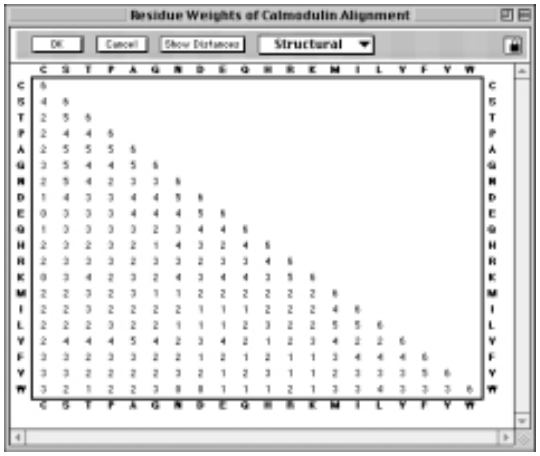
- From the **FILE MENU**, choose **Open** to access the Which MegAlign Document? dialog..

- Locate and open the folder called "Demo MegAlign," located in your DNASTAR directory.

- Double-click on the file named "Calmodulin Alignment" to open it in the MegAlign worktable.

- From the **OPTIONS MENU**, use the **Size** command to increase the font size until characters in the Worktable are legible.

We now need to choose one of MegAlign's two multiple alignment methods: Clustal or Jotun Hein. Jotun Hein is recommended if sequences are known to share homology, and Clustal is preferred if there is no a priori knowledge of relatedness. The sequences we are using are all known calmodulins, so we'll use the Jotun Hein method.

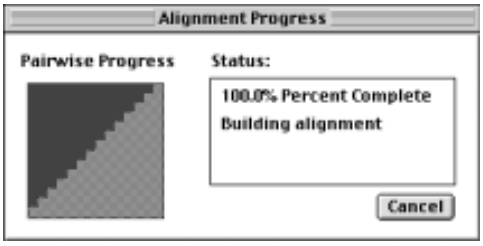


Before we perform our alignment, we should choose a weight table. MegAlign's residue weight tables are used in scoring multiple alignments so that mismatched residues that are chemically similar score higher than chemically different residues. Since our sequences are protein and we will be using the Jotun Hein method, the "Structural" table is the best choice.

- From the **ALIGN MENU**, choose **Set Residue Weight Table** to open the table shown at left.

- Choose **Structural** from the pull-down menu, located at the top, just to the left of the padlock icon.

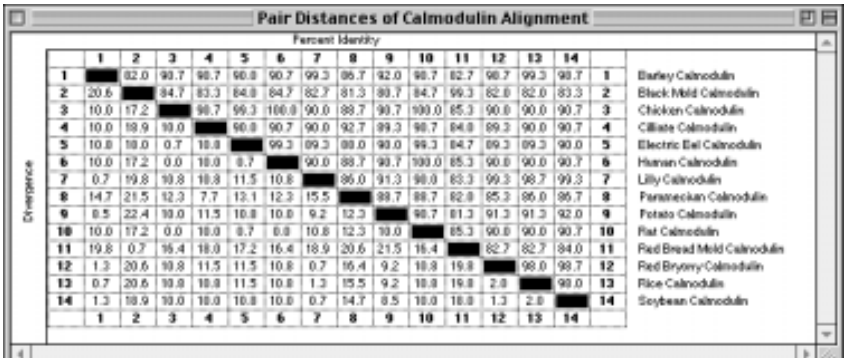
- Click **OK**.



We are now ready to align our calmodulin sequences.

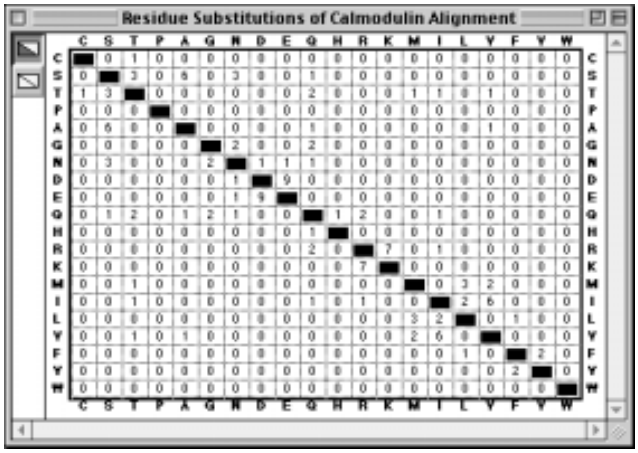
- From the **ALIGN MENU**, choose **By Jotun Hein Method**.

The Alignment Progress window shows what percentage of the project has been aligned. When the alignment is complete, as shown above at right, the Worktable reappears with the aligned sequences.



- View the divergence and similarity of the sequences by going to the **VIEW MENU** and choosing **Sequence Distances**. A window like that at left opens to show the sequence distances in tabular form.

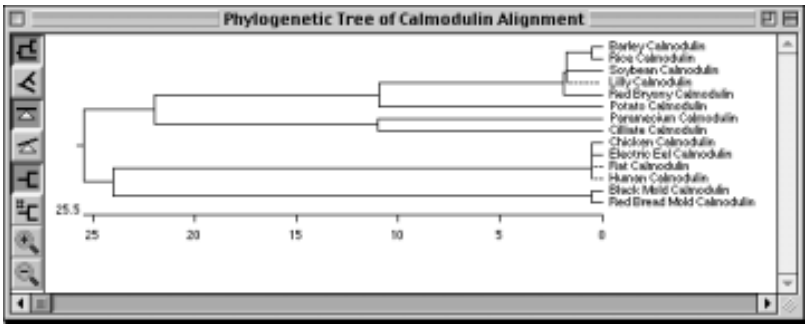
- View the number of ancestral changes that occur in the residues by going to the **VIEW MENU** and choosing **Residue Substitutions**. A second window like that at right opens showing the number of residue changes.



- To continue with this tutorial, close the Residue Substitutions and Sequence Distances windows.

Viewing the Phylogenetic Tree

- From the **VIEW MENU**, choose **Phylogenetic Tree** to open a window like that below. The default is for the **Balanced Branches** palette tool (third from top) to be active and for the results to be drawn as a phenogram. In a phenogram, distance lengths are approximate.



- To view the results as a cladogram, click the **Unbalanced Branches** palette tool (fourth from top). In a cladogram, the branch length is an estimate of divergence from an ancestral node.

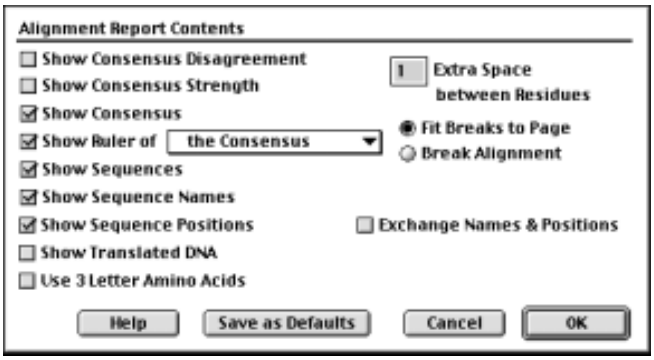
- To continue with this tutorial, close the Phylogenetic Tree window.

Viewing the Alignment Report

Next, we will view detailed results of the multiple alignment by viewing, then modifying, the Alignment Report.

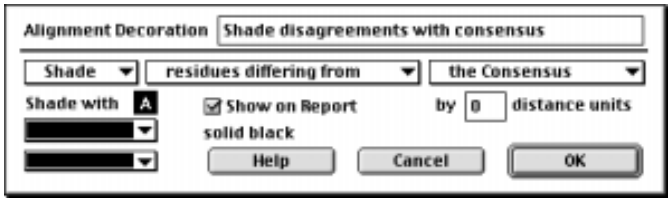
- From the **VIEW MENU**, choose **Alignment Report** to open a separate window containing the results of the alignment.

✦ MegAlign allows you to change the look of the Alignment Report. Go to the **OPTIONS MENU** and choose **Alignment Report Contents** to open the dialog box at right Check boxes 3-7 and 9, as shown, and leave all other settings at their defaults. Click **OK** to update the report.



Creating Decorations and Consensi

We can also optimize the Alignment Report by adding "decorations" and "consensi." Decorations are graphics that box, shade, or hide residues that agree or disagree with the sequence you choose. In this tutorial, we will create a decoration that shades those amino acids that disagree with the consensus sequence.



✦ From the **OPTIONS MENU**, choose **New Decoration** to open the dialog shown at left.

✦ In the title box, type a name such as "Shade disagreements with consensus."

✦ The next row contains three pull-down menus. Change the first pull-down menu to **Shade** and the middle menu to **residues differing from**. Leave the third menu, the one on the right, at the default setting, **the Consensus**. Choosing **Shade** from the first menu caused two additional pull-down menus to open below it.

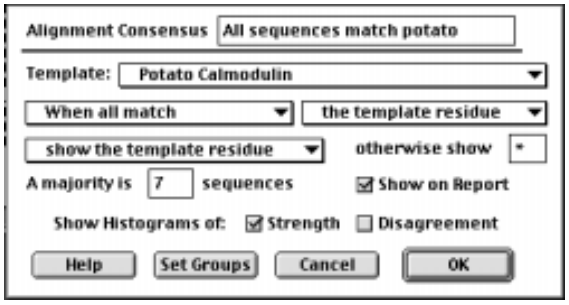
✦ Choose a color from the upper menu and a shading scheme from the lower menu.

✦ Leave the distance units box at "0."

✦ Click **OK**. The Alignment Report now shades any constituent sequence residues that do not match those in the consensus sequence.

Consensi are additional displays of the consensus sequence that can be used to highlight ambiguous residues and other features of interest graphically. Agreement and disagreement histograms for the consensus may also be displayed in the Alignment Report. We will be adding a consensus that shows asterisks whenever there are disagreements about a particular amino acid. Also, we will display a histogram showing the strength of the consensus.

- From the **OPTIONS MENU**, choose **New Consensus** to open the dialog at right.



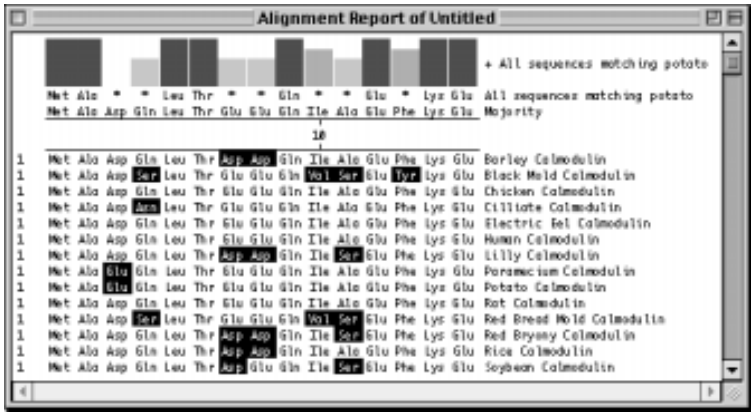
- In the title box, type a name such as "All sequences matching potato."

- The next rows contain four pull-down menus. Change the first menu to **Potato Calmodulin**. Leave the other menus at their defaults: **When all match**, **the template residue** and **show the template residue**.

- In the box to the right of the words "otherwise show," delete the period and type an asterisk (*).

- Check the box next to the word "Strength."

- Click **OK** to add the new consensus to the already open Alignment Report.



The final report, shown at left, now features the "shading" decoration we added earlier, plus a new consensus called "All sequences matching potato." Ambiguous residues for the new consensus are shown by asterisks, while three-letter amino acid codes appear wherever all 13 other sequences match the potato calmodulin sequence. A color histogram shows the strength of the new consensus graphically.

Saving a MegAlign Project

- From the **FILE MENU,** choose **Save**.
- Locate a folder in which to store the sequence.
- Give the sequence a name in the **File Name** box (Windows) or name box (Macintosh).
- Click **Save** (Macintosh) or **OK** (Windows). Changes to the alignment and the Alignment Report are both saved.

Getting Started **with PRIMERSELECT**

Welcome to this brief overview of ***Lasergene99's*** primer selection and design software for Windows and Macintosh platforms.

PrimerSelect helps you design primers and probes for PCR, sequencing, and hybridization experiments. The project begins when you enter your DNA, RNA, or back-translated protein template sequence. PrimerSelect processes the template for melting temperatures, free energy, and terminal free energy in pentamer windows. These primary calculations are all controlled by user-specified initial conditions, including primer concentration, salinity and ΔG calculation temperature.

After template processing, PrimerSelect locates and scores primers by user-defined characteristics, then offers advice on the best upper and lower primers for the template sequence. Primer "workbenches" allow you to modify your primer and see how your edits affect translated reading frames, secondary structure, false priming sites and restriction sites. Once you have selected and optimized your primer, PrimerSelect lets you create documents recording the theoretical activity of the templates, primers, and amplifiers. As with all ***Lasergene99*** applications, PrimerSelect also offers an integrated BLAST search function.

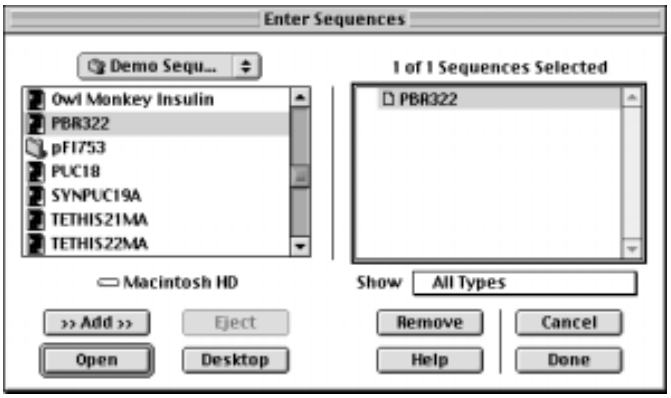
For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

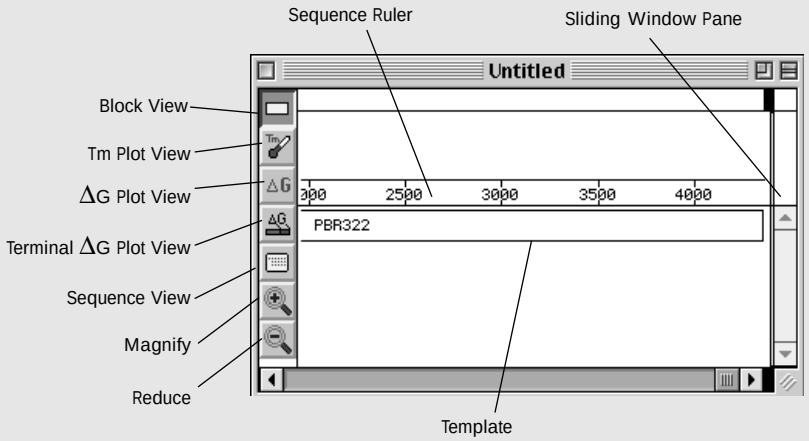
Creating a PrimerSelect Project

Our template DNA sequence will be the cloning vector pBR322. In this tutorial, we'll be selecting primers to amplify the H-strand Y effector site that runs from position 2013-2169. We'll begin by opening our template sequence.

- From the **FILE MENU**, choose **New** to create a new PrimerSelect document.
- From the **FILE MENU**, choose **Enter Sequence**. An Open dialog like the one below will appear.
- Locate and open the folder called "Demo Sequences", located in the DNASTAR directory.
- Select the file named pbr322.seq (Windows) or PBR322 (Macintosh) and click **Add** to move the sequence into the **Selected Sequences** window.
- Click **Done** to open the selected sequence in PrimerSelect's Document Window.

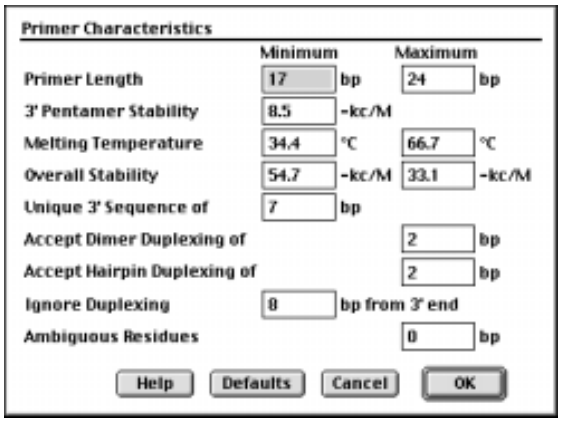


Anatomy of the PrimerSelect Document Window



Defining Primer Characteristics and Locations

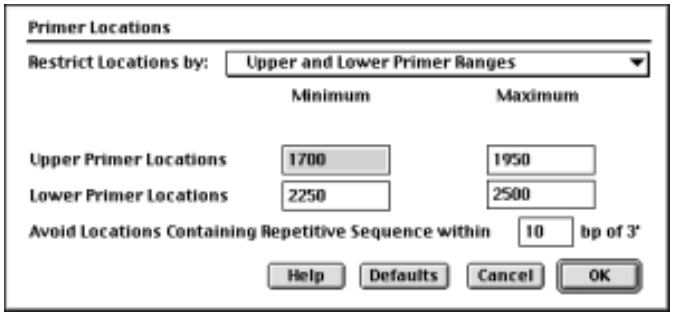
Next, we need to specify what types of primers we are looking for.



✦From the **CONDITIONS MENU**, choose **Primer Characteristics** to open the dialog at left. This dialog can be used to set variables such as primer length, maximum 3' pentamer stability, and acceptable duplexing. Note that the default primer length has been set to 17-24 bp, and that dimer and hairpin duplexing of 1 or 2 bp will be accepted.

✦Leave the default settings and click **Cancel** to exit the dialog.

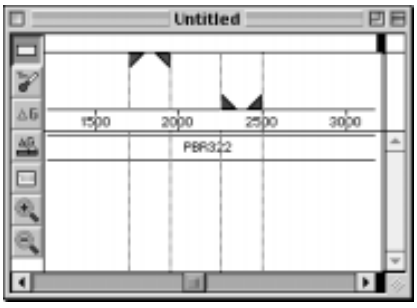
✦ From the **CONDITIONS MENU**, choose **Primer Locations** to open the dialog shown at right. This dialog allows us to restrict primer locations based on a given product length, a range in the template, or a combination of both.



✦ Go to the **Restrict Locations By** pull-down menu and choose **Upper and Lower Primer Ranges**. This is the best choice when designing PCR primers for a coding region in a completely characterized sequence.

✦ Enter the numbers 1700 and 1950 into the **Minimum** and **Maximum** text boxes to the right of **Upper Primer Locations**. Type 2250 and 2500 in the boxes to the right of **Lower Primer Locations**.

✦ Click **OK** to return to the Document Window, which will now appear as below. Paired green and red triangles mark the allowable ranges for the upper and lower primers, respectively

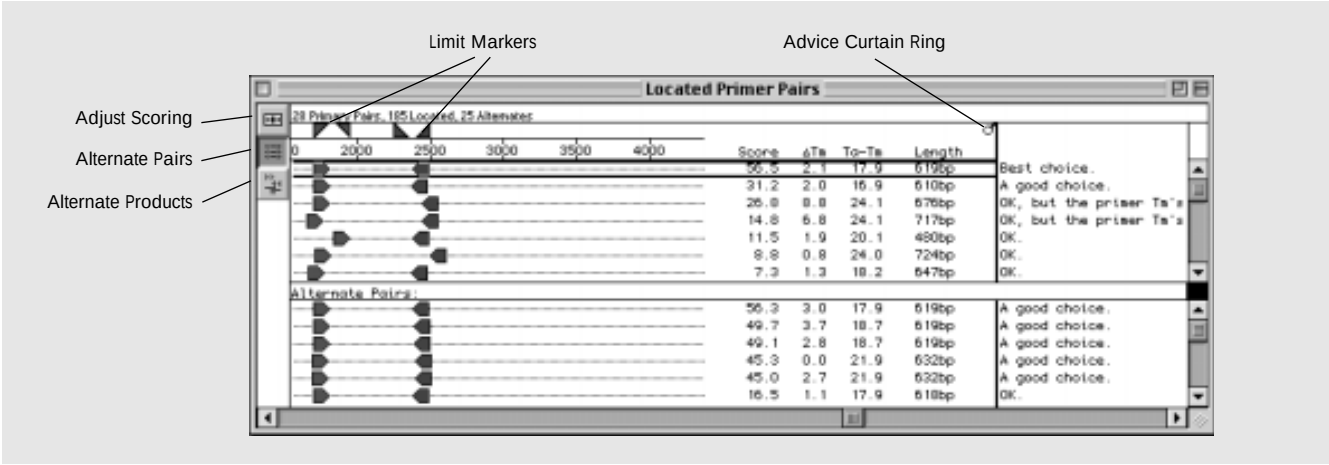


Searching for Primer Pairs

Now that we have specified what types of primers are acceptable, it's time to let PrimerSelect begin searching for our ideal primer pairs.

- From the **LOCATE MENU**, choose **PCR Primer Pairs**. The Located Primer Pairs window opens to show a ranked list of primer pair candidates, with the best choices shown at the top of the window.

Anatomy of a PrimerSelect Located Primer Pairs Window



The top pane of the Located Primer Pairs window shows a list of 28 "primary" primer pairs. Because the **Alternate Pairs** palette tool (middle tool at left of window) is active by default, the bottom pane currently shows a list of alternative pairs for the pair currently selected in the top pane.

- Open the advice curtain on the right side of the window by dragging the small circular curtain pull to the left. PrimerSelect's advice about each of the primer pairs is now visible in the right column of the window, as shown in the above graphic.

Note that the currently highlighted primer pair at the top of the upper pane is listed as the "Best choice." PrimerSelect chose this primer based on criteria established in the **Initial Conditions, Primer Characteristics** and **Primer Locations** dialogs. Looking at the lower pane of the window, you can see that PrimerSelect has listed 25 alternatives for the "Best choice" primer pair. These alternative pairs amplify the same region as the primary pair, but their sequences differ in ways that make them slightly less desireable.

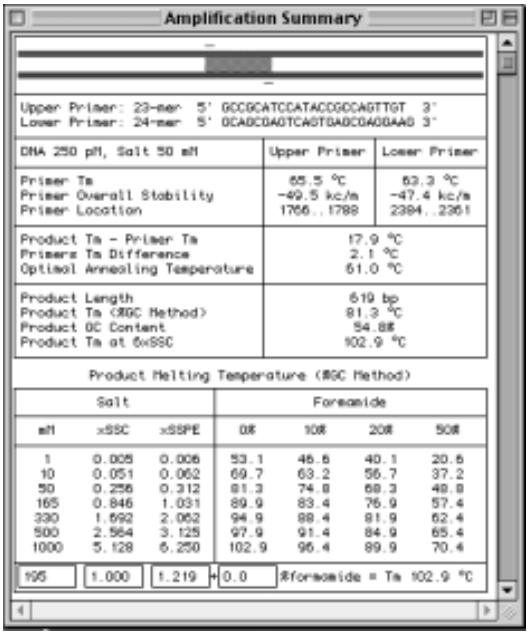
- Select the **Alternate Products** palette tool (bottom tool at left of window) to change the lower pane into a display showing intended and false primer sites and potential products for the highlighted primer pair.

The presence of a single thick black bar, like that at left, indicates that a large amount of amplification product is expected from our "Best choice" primers. The absence of multiple 5' or 3' end primer site arrows also shows that this pair has no false priming sites.

✦ Click on each successive primer pair in the top pane and observe how the product bar in the lower pane becomes thinner, denoting a decrease in the amount of product expected.

Note that eight of the primer pairs have a false lower primer site, indicated by a pink or light green arrow in the lower pane. In the primer pairs 13th and 22nd from the top, this is predicted to lead to false product, which is shown by the presence of a second product bar, as shown at right.

Viewing Additional Primer Information

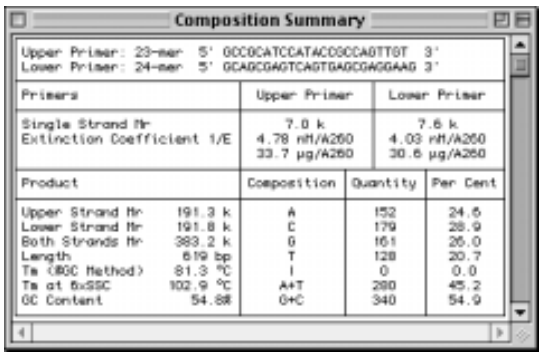


PrimerSelect compiles extensive tabular data about each primer pair and can predict the secondary structures of primer self-dimers, pair-dimers and hairpins.

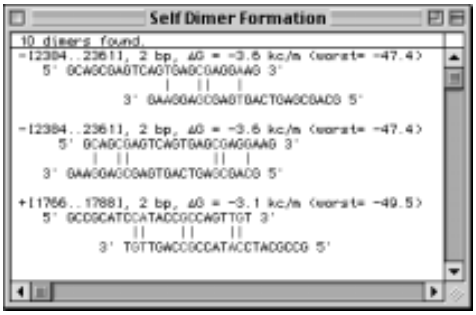
✦Double click on the "Best choice" primer at the top of the window to select it.

✦From the **REPORT MENU**, choose **Amplification Summary**. This opens a new window, shown at left, with a table showing how well the selected pair is expected to perform as a PCR amplifier.

✦ Choose **Composition Summary**. This opens a new window, shown at right, with a table describing the composition of the amplified region and the primer pairs.



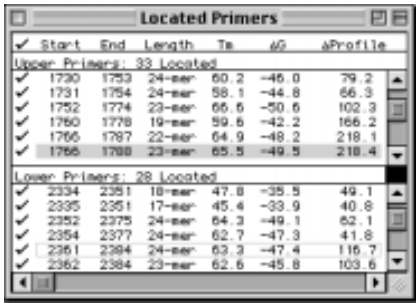
- From the **REPORT MENU**, choose **Primer Self Dimers**, **Primer Pair Dimers**, and/or **Primer Hairpins** to open windows showing the secondary structures predicted for each type. The Self-Dimer Formation window is shown at right.



- Close each of the newly created windows until only the Located Primer Pairs window remains open.

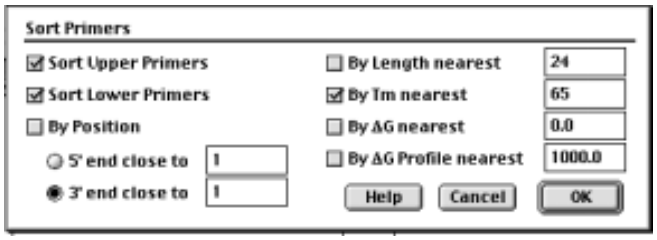
Sorting Primers by Characteristic

PrimerSelect considered all of our desired primer characteristics before automatically sorting primers in the Located Primer Pairs window from best to worst. The Located Primers & Probes window, on the other hand, allows us to sort the primers according to individual characteristics like position, length, melting point, or free energy. Let's experiment with this feature by sorting our primers according to melting point.



- From the **REPORT MENU**, choose **Located Primers & Probes** to open the window at left. The window displays a list of primers that matched our criteria. By default, primers are listed by location.

- From the **LOCATE MENU**, choose **Sort Primers** to open the dialog below.



- Leave the **Sort Upper Primers** and **Sort Lower Primers** boxes checked, and uncheck the **By Position** box.

- Check the **By Tm Nearest** box and enter the number 65 next to it.

- Click **OK**. Primers in both panes are now sorted by melting point, with those having melting points closest to 65^{o}C appearing at the top of each pane.

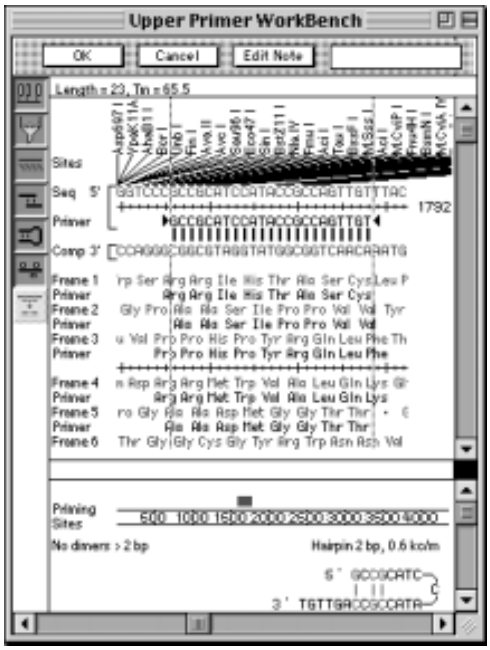
- From the **FILE MENU**, choose **Close** to return to the Located Primer Pairs window.

Changing Primer Length

PrimerSelect's Workbenches allow you to modify primers by introducing restriction sites, creating mutations, changing codon usage, examining primer secondary structure, and locating false priming sites. In this tutorial, we will experiment by using the Upper Primer Workbench to extend our primer length.

- Double-click on the "Best choice" primer pair at the top of the window to make it the active primer.

- From the **EDIT MENU**, choose **Work on Upper Primer** to open the Upper Primer Workbench shown below.



- Make sure that the **Dimer**, **Hairpin** and **Priming Sites** palette tools (on the left of the window, 4-6 from the top) are active.

Let's look at the Workbench in more detail. About a third of the way down the window, you will see the primer sequence written in black. The triangular "handles" at either end allow you to shorten or lengthen the sequence. The header in the upper left corner of the window tells us that the primer is 23 nucleotides long and has a melting point of 65.5^{o}C, while a ruler near the bottom of the window shows our single possible annealing site as a green bar. Below the ruler are messages informing us that there are no dimers larger than 2 bp, and that the primer is able to form only one, energetically unfavorable 2 bp hairpin. To resize the panes, drag the black box between the upper and lower scroll bars.

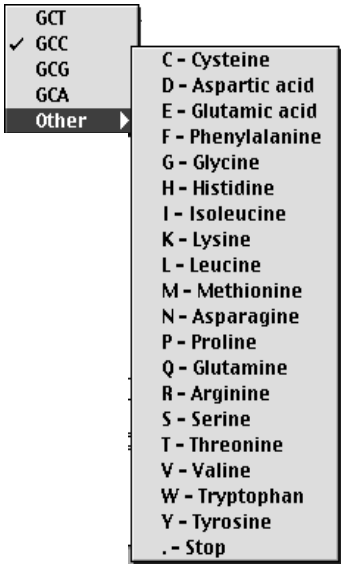
- Click the black triangle on the left side of the primer sequence and drag it to the left until the length has been extended to 33 nucleotides.

The current length is displayed in the header at the top left of the window. The header also shows that the melting point for the primer has increased to 78.6^{o}C, which would make it difficult to use in PCR. Looking at the lower part of the window, notice the new, energetically favorable dimers and hairpins, which PrimerSelect has labeled with a warning (BAD!). This increased likelihood of longer primers having stable dimers and hairpins is another reason that shorter primers are usually preferred.

- Drag the triangle to the right again until the primer has been returned to its original length of 23 nucleotides. The window appears as it did when we first opened it.

Introducing a Primer Mutation

The PrimerSelect Workbench makes it easy to test the consequences of introducing primer mutations. Our "Best choice" upper primer sequence, currently open in the Workbench, codes for a single Alanine residue in frame 1. In this tutorial, we'll convert this Alanine into Threonine by modifying codons in the translated reading frames. Once the conversion is complete, we will observe the effects of our mutation.



✥ Click on the Alanine (**Ala**) residue in frame 1, written in green just below the primer. A box opens, showing all four possible Alanine codons and the word "**Other.**" The check mark shows us that GCC is the triplet encoding this particular Alanine residue.

✥ Select **Other** to open a submenu with a list of amino acids, and click on "T - Threonine."

The **Ala** code in the original line of green text does not change, but the duplicate sequence that appears below it in black now shows a **Thr** where the **Ala** used to be. The **Thr** is written in red to show that it is a mutation from the original primer sequence. Looking at the bottom left corner of the screen, you will see that the mutated primer forms at least one stable, and therefore undesireable, dimer.

There's a chance that we can eliminate this unacceptable dimer by making a silent mutation to our new **Thr** residue. We'll try changing the triplet that codes for Threonine to ACC. This triplet differs from the one that coded for Alanine (GCC) by a single nucleotide.

✥ Click on the red **Thr** residue and select ACC from the menu, instead of the current ACT.

Looking at the lower part of the window, you'll notice that what is now a single nucleotide change from G to A has eliminated the problem dimer. On the other hand, the melting point for the upper primer has now decreased from 65.5 to 53.7^{o}C. The new melting point might be workable with the original lower primer, but it's probably better to seek a more suitable lower primer to pair with our mutated upper one.

✥ Name the mutant primer by typing a name into the white text box at the top right of the workbench.

✥ Click the **OK** button at the top left of the workbench to save the mutant primer and return to the Located Primer Pairs window.

- Update the window by going to the **LOCATE MENU** and choosing **PCR Primer Pairs**. Additional primer pairs are added to the window, with light green arrows representing mutant upper primers.

- Open the advice curtain again by dragging the circular curtain pull to the left. The first mutated primer pair that is not associated with a warning is the one eleventh from the top. This primer pair is rated as "OK" due to the small amount of product it produces, but it would be our best choice if circumstances forced us to use the mutated primer.

Creating a New Primer

When you open PrimerSelect for the first time, the Primer Catalog is empty. If you wish, you may use the catalog to store any primers created or modified in PrimerSelect. The **Only Cataloged Primers** command, found in the **LOCATE MENU**, then gives you the option of using only those primers that you have saved in the catalog. In this section of the tutorial, we'll learn how to create our own primers and enter them in a catalog.

- From the **LOG MENU**, choose **Primer Catalog** to open the primer catalog for our current project. Notice that our mutated upper primer is already present in the catalog. To the left of the primer is a checkmark identifying it as active, and a chevron (>>) (Windows) or bullet (Macintosh) showing that it has passed initial secondary structure tests.

- Press the **Return** key to create a field for entering a new primer.

- Type the words "Test Primer" into the **Name** field.

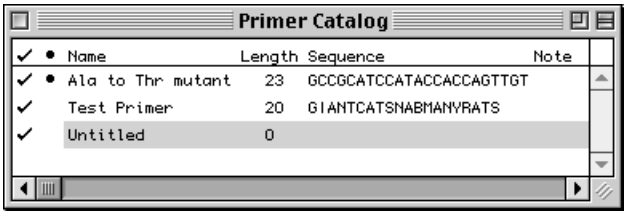
- Press the **Tab** key.

- Type "GIANTCATSNABMANYWARYRATS" (Giant cats nab many wary rats) into the **Sequence** field.

- Press the **Tab** key.

- Type the words "Contains 46% ambiguous bases" into the **Comments** field.

- Press the **Return** key to finish. The Primer Catalog should now appear as at right.



If our primer had passed initial secondary structure tests, a chevron or bullet would now appear just to the left of its name. Since no chevron or bullet is present, our test primer has failed. Let's examine why:

- From the **REPORT MENU**, choose **Primer Self Dimers**. Notice that twelve stable dimers have been found.

- From the **REPORT MENU**, choose **Primer Hairpins**. Our primer also has five stable hairpins.

As we might have predicted, the high percentage of ambiguous bases in our sequence have made it untenable as a primer candidate.

- Close each of the newly created windows until only the Located Primer Pairs window remains open.

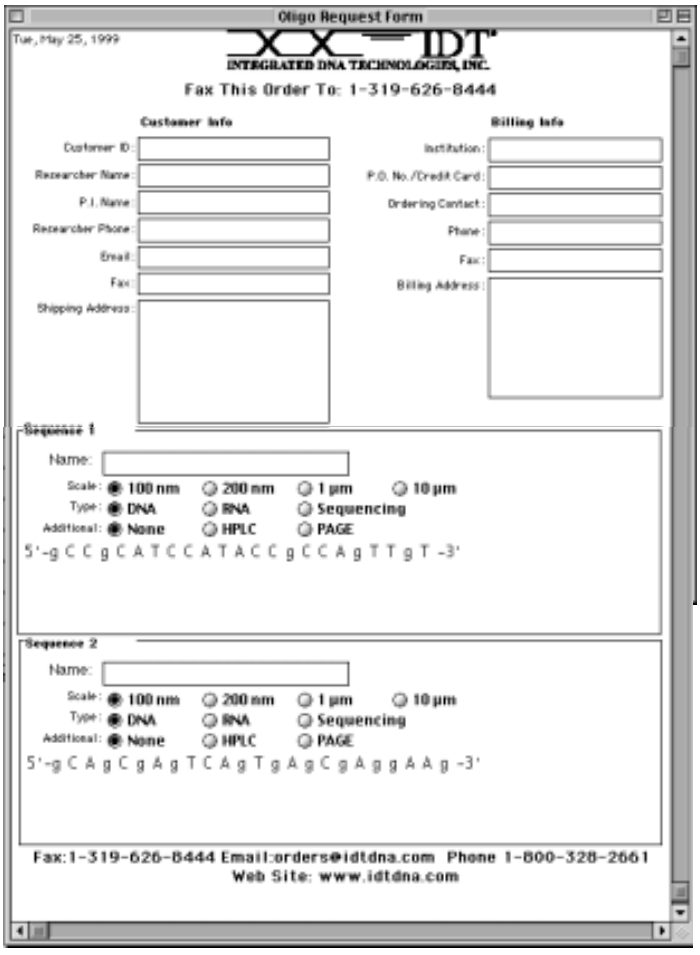
Using Oligonucleotide Order Forms

PrimerSelect provides two types of forms that you can print out and use to order oligonucleotides: Generic and IDT (Integrated Data Technologies, Inc.).

- Double-click on the "Best choice" primer pair at the top of the Located Primer Pairs window to select it.

- From the **LOG MENU**, choose **Oligo Request Form**, then **IDT** to open the IDT order form window shown at right. Note that our lower and upper primer sequences have been inserted into the form automatically. If we wanted to order the primers, we would simply fill out and print the form, then fax or e-mail it to the address at the bottom of the page.

- From the **FILE MENU**, choose **Close** to return to the Located Primer Pairs window.



Saving a PrimerSelect Project

- From the **FILE MENU,** choose **Save**.
- Locate a folder in which to store the sequence.
- Give the sequence a name in the **Save Sequence As** box (Windows) or name box (Macintosh).
- Click **Save** (Macintosh) or **OK** (Windows). Any changes to primers or to the Primer Catalog will be saved along with the project.

Getting Started with PROTEAN

Welcome to this brief overview of ***Lasergene99's*** protein sequence software for Windows and Macintosh platforms.

Protean helps you analyze and predict protein structures by applying multiple analytical methods to a sequence and displaying results grapically on a common horizontal scale. Analytical methods are organized into scientific concepts. Several methods may exist for a concept such as hydropathy, whereas others, such as flexibility, have only one available method. You can display method results on the Protean document in any order you like. In addition, Protean interprets and displays features from publicly available protein databases, and also allows you to annotate new features. As with all ***Lasergene99*** applications, Protean also offers an integrated BLAST search function.

For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

Contents

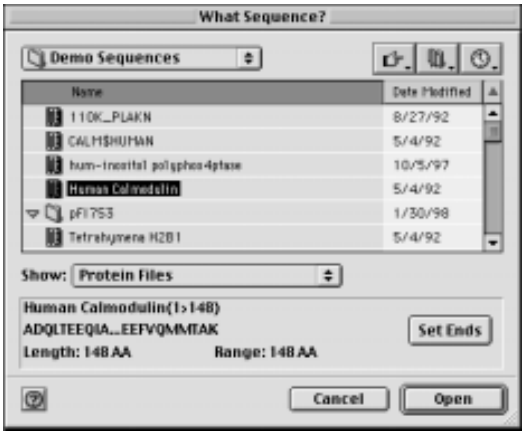
Creating a Protein Assay Document

In this tutorial, we will create a new Protean assay document for a human calmodulin sequence.

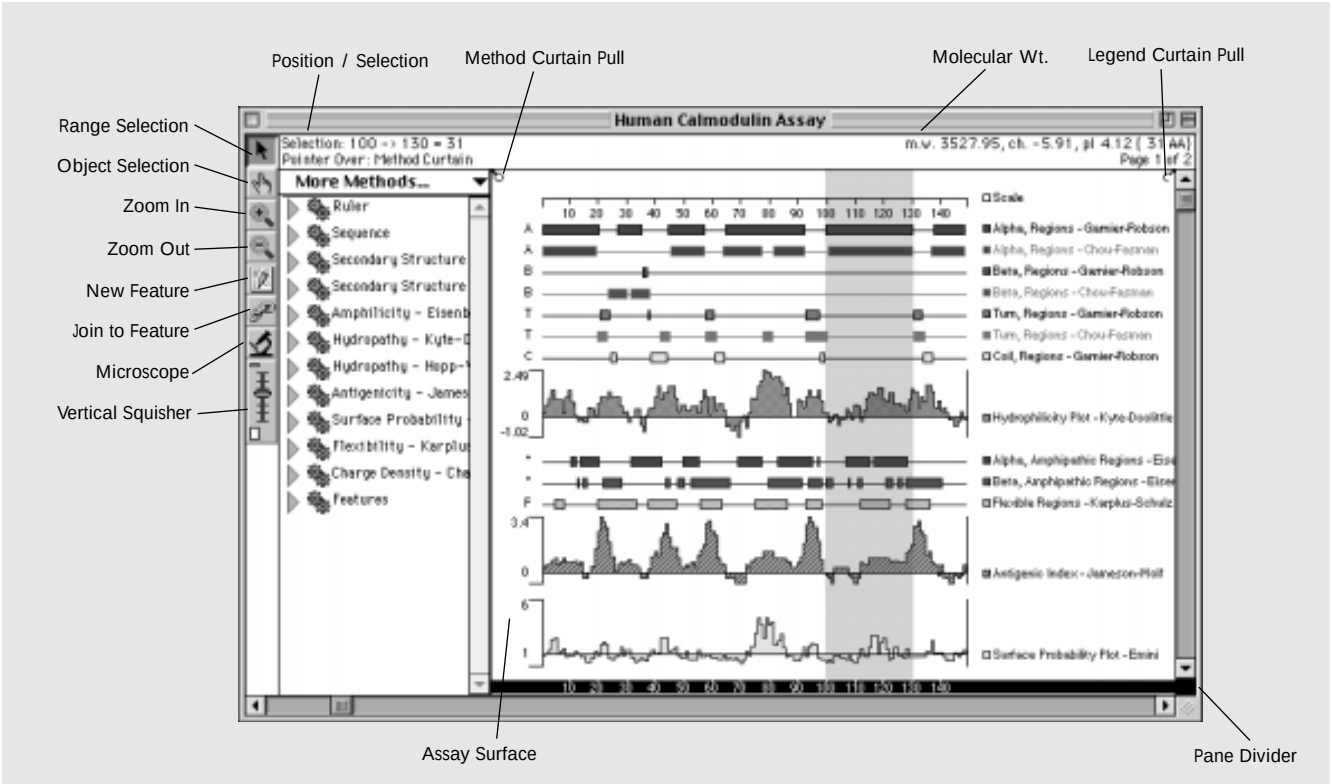
- From the **FILE MENU**, choose **New** to access a dialog similar to that shown at right.

- Using your file management system, open the folder named "Demo Sequences."

- Double-click on "Human Calmodulin" (or "Human Calmodulin.pro") to open the sequence in Protean. An assay document opens similar to that below:



Anatomy of a Protean Assay Document



Protean's Protein Analysis Methods

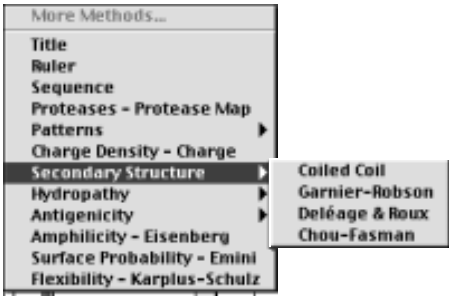
Once you have opened a sequence in Protean, the next step is to choose and apply methods. The graphic displays of method results are what help you determine features of interest on the sequence. Upon opening your sequence, you will observe that only a few of the methods listed below appear on the assay surface. The following methods are available:

- **Title** --Adds a title to the document.
- **Ruler** --Adds a ruler to the document.
- **Sequence** --Displays the sequence on the document.
- **Proteases - Protease Map** --Identifies proteolytic sites and displays them as a minimap.
- **Patterns - Prosite Database** -- Searches the Prosite database to find matches to your sequence.
- **Patterns - Ariadne File** -- Locates matches between a user-specified pattern descriptor and the sequence.
- **Charge Density - Charge** -- Predicts regions of positive and negative charge by summing charge over a specific range of residues.
- **Secondary Structure - Coiled Coil** -- Predicts transmembrane alpha helices.
- **Secondary Structure - Garnier-Robson** -- Examines the propensity of a given residue to exist in a certain structure.
- **Secondary Structure - Deleage-Roux** -- Uses an independent prediction of the protein's structural class to bias the prediction of its secondary structure.
- **Secondary Structure - Chou-Fasman**-- Predicts secondary structure of proteins from the crystallographic structures of their amino acid sequences.
- **Hydropathy - Goldman-Engleman-Steitz** -- Predicts non-polar alpha helices which may span a cell membrane.
- **Hydropathy - Kyte-Doolittle** --Predicts regional hydropathy of proteins from their amino acid sequences.
- **Hydropathy - Hopp-Woods** -- Finds protein antigenic determinants by searching protein sequences for the area of greatest local hydrophilicity.
- **Antigenicity - Sette MHC Motifs** -- Predicts peptide antigenic sites which interact with mouse MHC II haplotype d proteins.
- **Antigenicity - AMPHI** -- Predicts immunodominant helper T-lymphocyte antigenic sites from primary sequence data.
- **Antigenicity - Rothbard-Taylor** -- Locates potential T-lymphocyte antigenic determinants which contain a common sequence motif.
- **Antigenicity - Jameson-Wolf** -- Predicts potential antigenic determinants by combining existing methods for protein structural predictions.
- **Amphiphilicity - Eisenberg** -- Predicts the Eisenberg Moment.
- **Surface Probability - Emini** -- Predicts the probability that a given region lies on the surface of a protein.
- **Flexibility - Karplus-Schulz** -- Predicts backbone chain flexibility..

Applying Analysis Methods

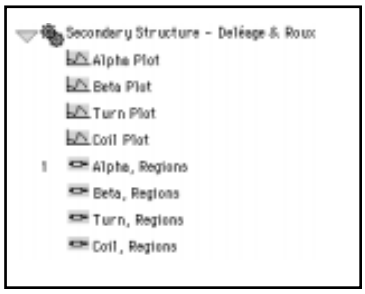
To apply new Protean methods, or to add extra copies of methods in use, you move them from the **More Methods** menu to the method curtain, and from there to the assay surface. In this tutorial, we will practice working with analysis methods by applying the **Secondary Structure - Deleage & Roux** method to our sequence.

- Open the method curtain by going to the **ANALYSIS MENU** and choosing **Show Available Methods** or by dragging it open using the ring located to the right of the **Object Selector** palette tool. The method curtain contains all methods that are ready to be applied to your sequence.



- You will notice that the **Secondary Structure - Deleage & Roux** method does not appear in the method curtain. Go to the top of the method curtain and click **More Methods** to open a submenu, shown at left, containing the complete list of available methods. Click **Secondary Structure** to open the secondary structure methods submenu, then click on **Deleage & Roux** to move the method into the main body of the method curtain. Now that the method is in the curtain, it may be applied to the sequence.

- To find out if a method in the method curtain has already been applied, click the triangle next to its name to display a list like the one at right. If a number appears next to a display icon, as it does in the graphic at right, that display has already been applied to the assay. Since we have not yet applied our method, clicking the triangle next to **Secondary Structure - Deleage & Roux** method reveals an absence of numbers next to any of the eight possible displays.



- Click in the white space to the left of the method to ungroup the eight sub-methods.

- Click on "Alpha, Regions," drag it to the assay surface and drop it wherever you want the graphical results to be displayed. Predicted alpha regions of your protein will be displayed as white boxes.

Changing Method Parameters

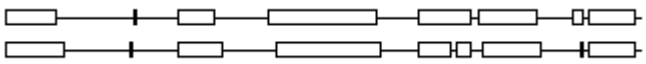
Next, we'll experiment with changing the parameters for the **Secondary Structure - Deleage & Roux** method and comparing the new results to our original results.

- Move a second **Secondary Structure - Deleage & Roux** method from the **More Methods** menu to the method curtain and move its "Alpha, Regions" plot to the assay surface. Drop the duplicate method just under the existing Deleage & Roux plot. You should now have two identical region plots showing the alpha regions of your protein.

- Double-click on the name of one of the two **Secondary Structure - Deleage & Roux** methods you moved into the method curtain. This opens the parameters dialog for that method.

- We'll choose Structural Class **A + B** for this experiment, so select that radio button and click **OK**.

On the assay surface, the corresponding region plot will instantly change to reflect this new formula for calculating alpha regions. You may now visually compare the recalculated results to the original results, as shown below. (Note: If you had simply dragged a second copy of the original method onto the assay surface, parameter changes would automatically have been applied to *both* plots.)



Optimizing the Display

We can optimize the look of our Protean assay document by using the following optional procedures.

To organize or move displays:

- Click the **Object Selector** palette tool, then using the mouse to drag displays where desired. Try superimposing displays by dragging the **Beta, Regions - Chou-Fasman** results on top of the **Alpha, Regions - Chou-Fasman** results.

To change method formatting:

- Click on the **Object Selector** palette tool (hand icon) then select any method on the assay surface. Here, you can experiment with either of the **Secondary Structure - Deleage & Roux** displays that you added earlier.

- From the **OPTIONS MENU**, choose **Line Color** to open a submenu of color choices. When you select a color from the submenu, both the method title and display will change to that color.

You may also wish to experiment with the similar commands: **Line Weight**, **Fill Color** and **Fill Pattern**.

To remove a method:

- Click on the **Object Selector** palette tool (hand icon) then select any plot on the assay surface. We'll experiment by selecting **Hydrophilicity Plot - Kyte-Doolittle**.

- Click the **Backspace** or **Delete** key to remove the plot from the assay surface.

- Remove the entire method from the method curtain by selecting **Hydropathy - Kyte-Doolittle** in the curtain and clicking **Backspace** or **Delete**. Remember that any method you remove from the curtain can always be accessed again from the **More Methods** menu.

Simulating Protease Digests and SDS PAGE Gels

Protean makes it easy to identify proteolytic sites and display them as minimaps.

- Move the **Protease - Protease Map** method from the **More Methods** menu to the method curtain.

- Click the blue triangle next to the method name to view the list of proteases.

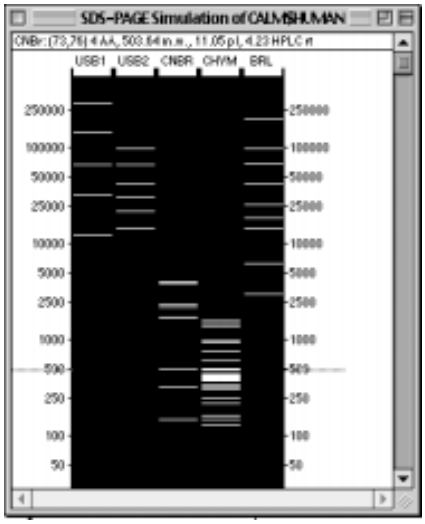
- Click in the white margin to the left of the protease list to remove highlighting from the group.

- Highlight "Chymotrypsin" and drag it onto the assay surface. Repeat for "CNBr." The results for these methods display recognition sites as vertical bars.

Now that at lease one protease method has been applied, we can simulate a polyacrylamide gel separation.

- Click on the **Range Selector** palette tool (arrow icon) to deselect the two protease methods that we just applied to the assay surface. Since we have not selected any protease methods, both of the applied methods will automatically appear in the gel simulation.

- From the **SITES & FEATURES MENU**, choose **SDS PAGE Gel Simulation**. A graphic of the simulated separation will appear in a new window, as shown at right.



Above each column in the gel is the name of the corresponding protease or molecular weight standard. When the cursor is over the gel, you can keep track of molecular weights by the horizontal cross-hairs.

- Move the cursor over any fragment on the gel. The header at the top of the window shows the range, size, molecular weight, isoelectric point and HPLC retention time of that fragment.

- To continue with the tutorial, close the SDS PAGE Gel Simulation window.

Annotating Features

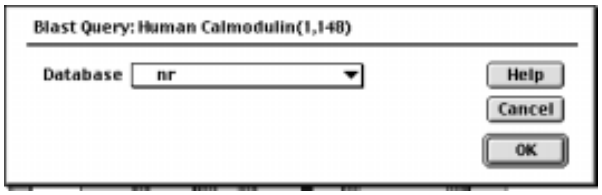
When you import an EditSeq or GenBank sequence into Protean, features from the document are automatically placed as annotation methods at the bottom of the method curtain. These features can be applied just like any other method, by dragging them onto the assay surface. Protean also lets you create new annotations to label areas of interest in your protein sequence. The procedure that you learned in *Getting Started with GeneQuest*, on page 38, works identically in Protean.

Performing a BLAST Search

Protean allows you to query sequence databases on the Internet or an Intranet for matches to your sequence. In this tutorial, we'll search for matches to our Human Calmodulin sequence on the BLAST server at the National Center for Biotechnology Information. Note that you must have Internet access to perform this BLAST search. If you are not connected to the Internet, skip this part of the tutorial and continue with the next section.

- Click the **Range Selector** palette tool (arrow icon).

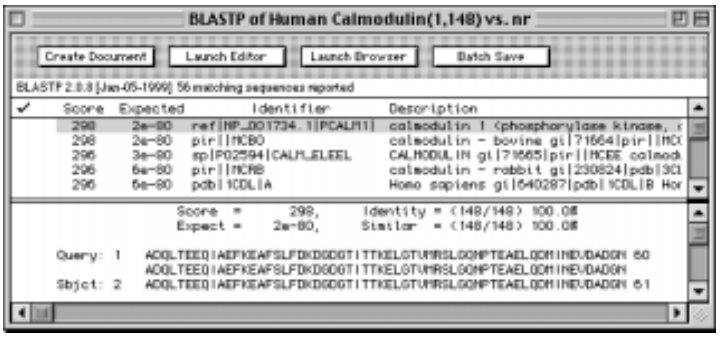
- Select the portion of the sequence that you wish to use as the search query. Since our sequence is a short one, we will select the entire sequence by going to the **EDIT MENU** and choosing **Select All**.



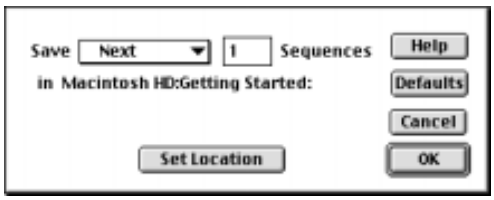
✦ From the **EDIT MENU**, choose **BLAST Search**. A BLAST dialog box will appear as at left.

✦ Leave the defaults so that the database is *nr*.

✦ Click **OK** to begin the search. Protean displays search results in a two-paned BLAST Results window like that shown below. The upper pane contains the names of matches in order of probability, while the lower pane contains the alignment of the query (upper sequence) to the highlighted database entry (lower sequence). Detailed information about how "score" and "expectation" are derived is available at NCBI's website, http://www.ncbi.nlm.nih.gov/BLAST. In general, a higher score and lower expectation denote a better match. The four buttons at the top of the window are used to open or save the sequence matches or to view more information about a particular sequence.



The "Create Document" button opens a new Protean assay document for one or more sequences. We will now select a single BLAST sequence and open it in Protean.



✦ Highlight an entry in the BLAST Results window.

✦ Click **Create Document**. The Save dialog window shown at left will open.

✦ Choose **Next** from the Save pull-down menu. Type the number "1" into the text box to the left of the word "Sequences."

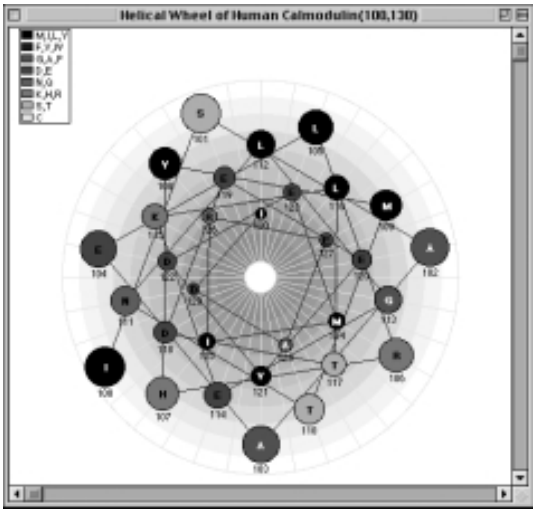
✦ Click **OK**. A new Protean assay document will open containing the sequence you selected.

The **Launch Browser** and **Batch Save** buttons work just as they were described on pages 28-29 of *Getting Started with EditSeq*. The **Launch Editor** button opens EditSeq, then works identically to EditSeq's **Create Document** button.

✦ To continue with the tutorial, close the newly opened document by going to the **FILE MENU** and choosing **Close**, then **Discard**. Now close the BLAST Results window.

Creating Models of Secondary Structures

Protean can display simple secondary structure models for elements such as helical wheels, helical nets, and beta sheets. It can also display the sequence of the protein as a linear space-fill model or as a chemical formula model. In this tutorial, we'll draw a helical wheel for one of the alpha regions of our sequence. The **Secondary Structure - Garnier-Robson** method has already been applied to our sequence, and its "Alpha, Regions" display conveniently shows alpha regions as bars on a region plot. Once we have selected one of these alpha regions, we can easily view its secondary structure as a helical wheel.



✥Click the **Range Selector** palette tool (arrow icon).

✥Click on one of the bars (predicted helical region) on the **Alpha, Regions - Garnier-Robson** plot at the top of the Protean document.

✥From the **ANALYSIS MENU**, choose **Model Structures**, then **Helical Wheel**. A window like the one at left will open showing the secondary structure for that portion of the sequence.

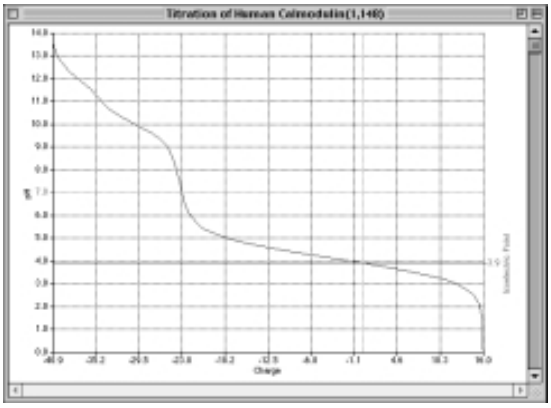
✥To continue with this tutorial, close the Helical Wheel window.

Displaying Titration Curves

Next, we'll view the titration curve for the entire protein sequence.

- ✥ Deselect the sequence by clicking once on the **Object Selector** palette tool (hand icon). If no part of the sequence is highlighted, Protean provides a titration curve for the entire protein.

- ✥ From the **ANALYSIS MENU**, choose **Titration Curve**.



A new window like that at right will open showing a graph of the titration curve. Protean adds blue "crosshairs" at the isoelectric point to show pH and charge.

Saving an Assay Document

- From the **FILE MENU,** choose **Save**.
- Locate a folder in which to store the sequence.
- Give the sequence a name in the **Save Sequence As** box (Windows) or name box (Macintosh).
- Click **Save** to save the assay document.

Getting Started **with SEQMAN II**

Welcome to this brief overview of ***Lasergene99's*** sequence assembly software for Windows and Macintosh platforms. SeqMan II not only assembles up to 64,000 sequences into contigs, but also includes tools for trimming poor quality and contaminating data from your sequences prior to assembly, and provides sophisticated techniques for editing and exporting the consensus sequences. SeqMan II uses DNASTAR's unique Trace Quality Evaluation scheme to automatically generate the most accurate consensus sequence possible by evaluating the quality of trace data generated by automated sequencers.

A new project begins by entering individual sequence files comprised of trace files generated by automated sequencing instruments, text-format sequence files, or files in DNASTAR sequence format. You may also import sequences from an Internet Entrez or BLAST server. Preassembly options include the ability to trim poor quality data and remove specific vector or contaminant sequences. You can even specify that repetitive sequences be added last during assembly.

Once contigs have been assembled, the Strategy View displays contig coverage and quality graphically and shows where additional sequences may be needed. If desired, you can add more sequences or data from other SeqMan II projects to the current project and reassemble. The Alignment View allows you to edit and trim constituent and consensus sequences, restore previously trimmed data and adjust alignments. You may also view individual and summary quality scores. SeqMan II provides both automated and manual tools to help you determine whether multiple contigs may be joined. As with all ***Lasergene99*** applications, SeqMan II also offers an integrated BLAST search function.

For assistance in using this software, contact DNASTAR by phone at (608) 258-7420, by FAX at (608) 258-7439, by e-mail: at support@dnastar.com, or via DNASTAR's website at http://www.dnastar.com.

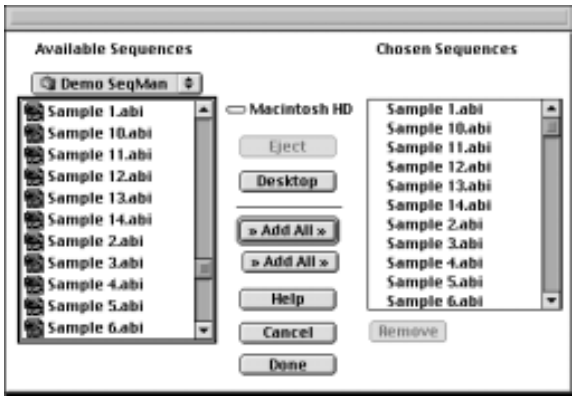
Contents

Entering Fragment Sequences

We'll begin our tutorial by opening the fourteen sample sequences located in the "Demo Seqman" folder in the DNASTAR directory. Each of the fragment sequences is a PE Applied Biosystems, Inc. automated sequencer trace file.

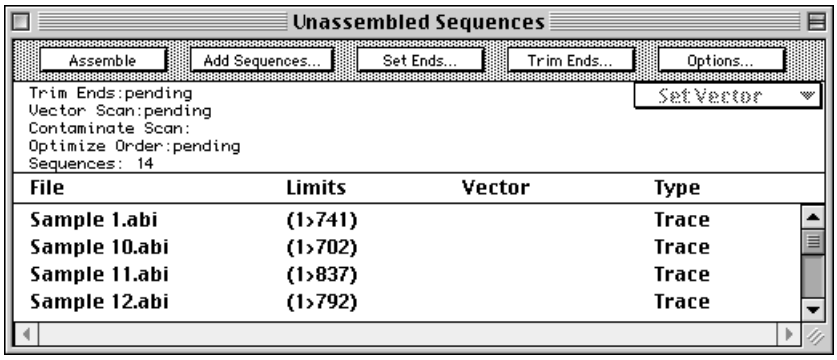
- From the **FILE MENU**, choose **New** to open the Unassembled Sequences window and an empty, untitled Project Summary window. The Project Summary window will remain empty until after you assemble your sequences.

- Open the Enter Sequences dialog, shown at right, by pressing the **Add Sequences** button at the top of the Unassembled Sequences window.



- Locate and open the DNASTAR folder, then double-click on the folder called "Demo Seqman." A list of sequences with the names Sample 1...Sample 14 will appear on the pane at the left of the dialog. These are the sequences we will be assembling.

- Transfer the fragment sequences to the right pane by clicking the **Add All** button in the lower right corner of the dialog. The fourteen sequence names should now appear in the right pane.



- Click the **Done** button. The Unassembled Sequences window, shown at left, now contains a list of the fourteen sequences. Expand the window by dragging its lower right corner until you can view all the sample names simultaneously.

Selecting Pre-Assembly Options

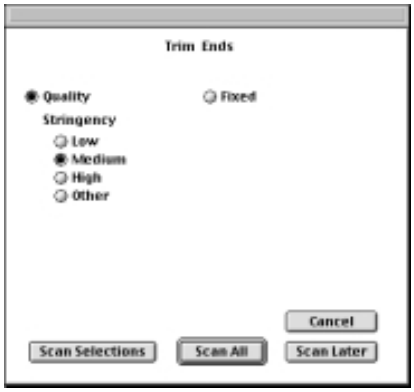
Before assembly, you may wish to consider a few pre-assembly options. You may remove vector and contaminant sequence from your fragments, optimize fragment assembly order, set fragment endpoints and flag sequence repeats.

In this tutorial, we will remove Janus vector sequence from the samples automatically and trim sequence ends based on SeqMan II's trace quality evaluation.

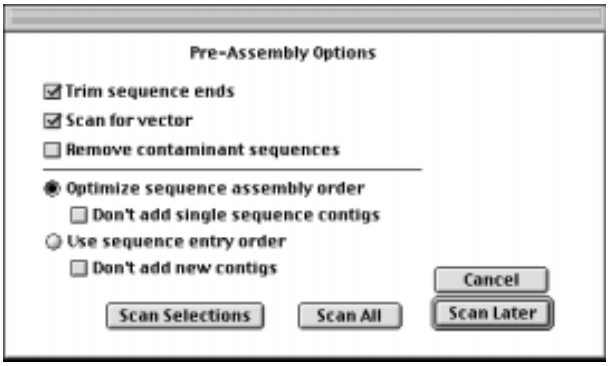
- Highlight all fourteen of the sequences by going to the **EDIT MENU** and choosing **Select All**.

- Using the **Set Vector** pull-down menu at the top right of the Unassembled Sequences window, choose "Janus" as the vector SeqMan II should search for.

- Click the **Trim Ends** button at the top of the Unassembled Sequences window to open the dialog shown at right. You will notice that by default, SeqMan II will perform quality trimming with medium stringency. Leave the settings as they are and exit the dialog by pressing the **Cancel** button.



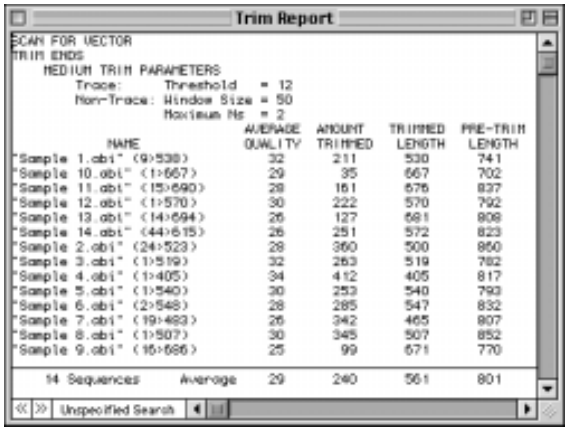
- Click the **Options** button located at the top right of the Unassembled Sequences window. The resulting dialog box, shown below, shows a list of all pre-assembly options. The default settings, which we will keep, tell SeqMan II to scan for the vector we selected, perform quality trimming on sequence ends and use the optimal sequence assembly order.



- Click **Scan All.** A Report window will be generated, but can be moved aside for now. After scanning, you will notice that the two center columns of the Unassembled Sequences window have changed. The vector column now shows, via a checkmark next to the vector name, that Janus vector was detected in all fourteen sequences. The limits column displays the new sequence endpoints, reflecting both quality trimming and vector removal.

Reviewing Trimmed Data

We may now view a report showing details of the quality trimming and vector removal performed during the scanning phase.



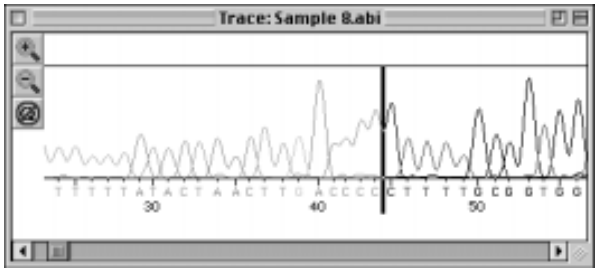
✥From the **PROJECT MENU**, choose **Trim Report**. The report, shown at left, displays current end-trimming parameters, followed by the name of each sequence, its average quality, amount trimmed, and both trimmed and pre-trimmed lengths.

✥Enlarge the Trim Report window by dragging its lower right corner until you can view the entire report.

✥After you have finished reviewing the report, close it.

To see the results of SeqMan II's quality trimming, let's view the trace data for one of the trimmed sequences.

- ✥ From the Unassembled Sequences window, double-click on Sample 8 to open its chromatogram in a separate window, as shown at right.



Data trimmed from the 5' end appears dimmed, denoting that it will not be included during sequence assembly. Notice the good quality of most of the trimmed portion of the sequence. It was trimmed here because the sequence originates from the Janus vector.

A vertical black bar appears at the intersection between trimmend and untrimmed sequence. You may drag this bar left, to unmask data that you feel should be included in the assembly, or right, to mask data that should not.

Next, let's look at the 5' end of the sequence:

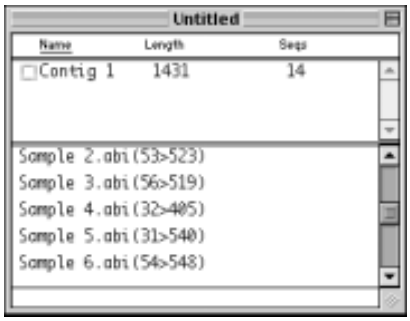
- ✥ Click anywhere on the consensus sequence.
- ✥ From the **EDIT MENU**, choose **Go To Position**.
- ✥ Type "500" into the **Position** text box.
- ✥ Activate the **Position Counts Consensus Gaps** button, and click **OK**.

Notice that the peak quality falls in this region, and that beyond base 507, SeqMan II determined that average peak quality falls below the medium stringency threshold.

✦ After you have finished reviewing the trace data, close that window to continue with the tutorial.

Assembling Sequences

We are now ready to launch the sequence assembly process.



✦ Click the **Assemble** button at the top left of the Unassembled Sequences window.

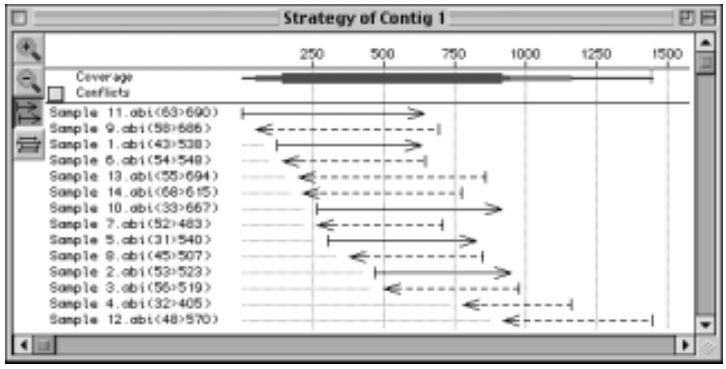
The **Unassembled Sequences** window immediately disappears and the **Report** window provides a running commentary on how contigs are being constructed. After assembly is complete, a **Project** window like that at left displays contigs in the upper pane and all constituent sequences in the lower pane. In our case, SeqMan II has assembled the 14 sequences into a single contig.

Checking Coverage and Construction

Now that we have assembled our fragments into a contig, we can examine the number of conflicts and degree of coverage for the contig using the Strategy View. This window graphically displays the position and orientation of every constituent sequence in the contig.

✦ Click on the contig in the Project window to highlight it. The Project window is the one which has the word "Untitled" at the top. Contigs are listed in the top pane of the window.

✦ From the **CONTIG MENU**, choose **Strategy View** to open a window like that at right.



- Enlarge the Strategy View by dragging its lower right corner. Notice the band just below the ruler, which summarizes sequencing coverage. The color and thickness of the band represent the amount of coverage. You will notice a thick green band between approximately 200-900. This denotes a region sequenced on both strands and above the minimum coverage threshold. (The coverage threshold may be changed by going to the **PROJECT MENU** and choosing **Parameters**, then **Strategy Viewing**). The medium blue line to its right indicates a region sequenced on one strand only. The thin red line appearing at each end of the contig means the region was sequenced only once.

- Check the **Conflicts** box at the top left of the window to view a histogram showing agreement between fragment sequences. Our histogram appears mostly black, indicating little or no conflict in most areas. The blue bands indicate regions where conflicts occur in 10-20% of the bases.

- Click once on the **Zoom In** palette tool (magnifying glass icon with +) at the left side of the window.

- Double-click on the blue area around base 135 to launch the Alignment View for that region. Conflicts are highlighted as red bases in Sample 9.

- Close the Alignment window.

Removing Conflicts and Gaps

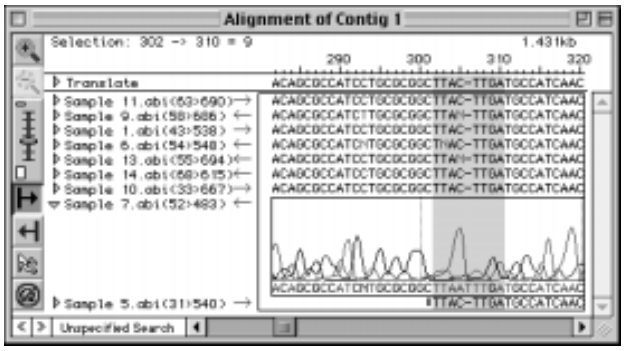
Our sample sequences combined neatly into a single contig because SeqMan II removed poor quality and vector data. In some cases, though, it may still be necessary to resolve conflicts and/or gaps in the data before a satisfactory consensus sequence is obtained. We will practice editing our sequences in the Alignment View.

- From the **CONTIG MENU**, choose **Alignment View**. The Alignment View displays the consensus sequence at the top of the window, with constituent sequences listed below.

- From the **EDIT MENU**, choose **Find Conflict**, then click **Find Next**. The cursor will jump to position 102. Click the **Find Next** button (**>>**) at the bottom left corner of the window twice to advance to the conflict located at position 306. Sample 7 reports a "T" at this location, while all other samples register it as a gap. We can resolve this conflict by viewing the trace data for Sample 7.

- Highlight Sample 7 from about position 302 to 310.

- Click on the gray triangle that appears just to the left of its sequence name to reveal trace data for Sample 7. The Alignment View now appears as at right.



- Click the **Zoom In** palette tool (magnifying glass icon with +) 2-3 times to view the chromatogram in greater detail.

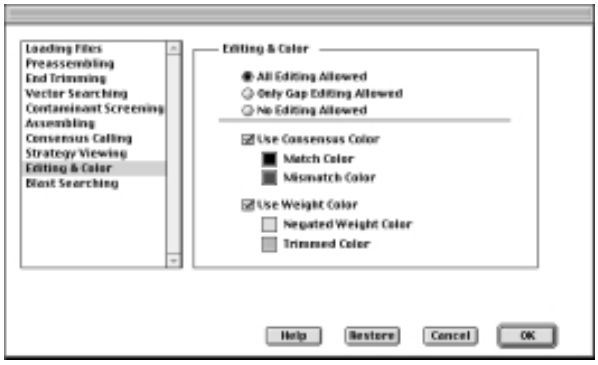
The trace data show that the purported "T" in Sample 7 is probably an artifact, as is the "A" to its right. You can see from the consensus sequence that SeqMan II's trace quality evaluation algorithm has automatically ignored both of these dubious bases. It is not necessary, therefore, to correct the sequence data by manually editing out the bases. For learning purposes, however, we will experiment by replacing the "T" with a gap, then removing the column of gaps from the consensus.

- Highlight the letter "T" at position 306 and replace it by typing a gap (hyphen).

- Click anywhere on the consensus, then go to the **EDIT MENU** and choose **Remove Consensus Gaps**. All gaps at position 306 disappear. Alternatively, you could highlight the gap in the consensus sequence and press the **Backspace** key (Windows) or **Delete** key (Macintosh).

Restoring Sequence Ends Manually

Prior to assembling our contig, SeqMan II automatically trimmed ends for the constituent sequences based on trace data quality and presence of Janus vector. Although sufficient data remained to assemble the sequences into a single contig, there are cases when restoring some of the trimmed data may allow SeqMan II to join multiple contigs into a single contig. In this tutorial, we will learn how to restore trimmed ends manually from the already open Alignment View.



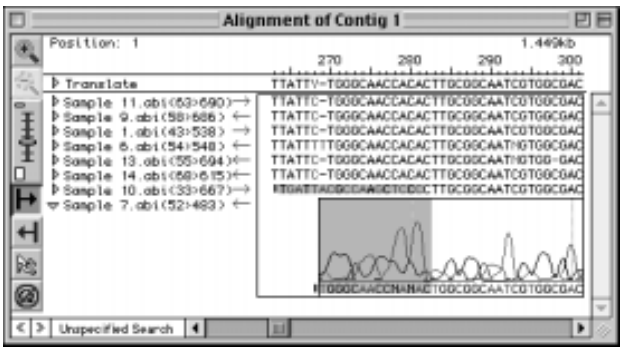
We'll want to differentiate between trimmed and untrimmed data, so we'll start by giving trimmed data a distinctive colored background.

- From the **PROJECT MENU**, choose **Parameters**, then **Editing & Color** to open the dialog at left.

- Make sure that the **Use Weight Color** box is checked, then click **OK**. Now when you manually restore trimmed data, it will appear with a bright yellow background.

Trimming is done in the Alignment View using trim bars, which appear as black vertical bars on the left ends and small black triangles on the right ends of sequences. For this tutorial, you may choose any of the fourteen sequences.

- To restore trimmed sequence ends, drag the bar away from the sequence. Notice that previously trimmed sequence is easily recognized by its bright yellow background, visible as shading in the graphic below. When you restore trimmed vector sequence, you will usually observe many conflicts (indicated by red letters) between the restored data and the consensus sequence, even though restored peaks are of high quality. You should not keep such data in your project. The data that may be useful in joining contigs is that which was trimmed due to inferior peak quality.



- Reset sequence ends to their original values by dragging the bar the opposite direction until the trimmed (yellow) portion of the sequence disappears from view.

Saving Projects and Exporting Sequences

To save a SeqMan II sequencing project:

- From the **FILE MENU,** choose **Save**.

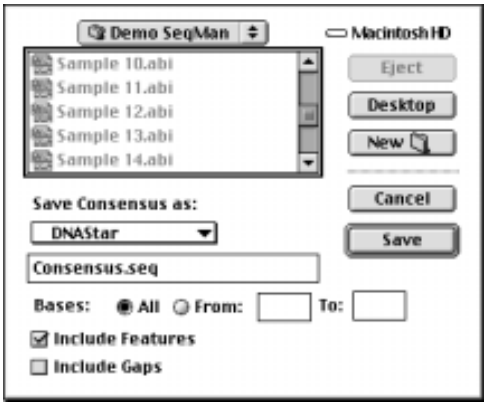
- Locate a folder in which to store the sequence.

- Give the sequence a name in the **File Name** box (Windows) or name box (Macintosh).

- Click **Save** to save the project, including any edits to contigs or fragments, and any vector, contaminant or repetitive sequences used.

To save a consensus sequence in a separate file:

- Highlight one or more contigs in the Project Summary window, the Alignment View or the Strategy View.
- From the **CONTIG MENU**, choose **Save Consensus**, then **Single File** to open the dialog shown at right.



- Locate a folder in which to store the sequence.
- Give the sequence a name in the **File Name** box (Windows) or "Untitled" box (Macintosh).
- Use the **Save As Type** (Windows) or **Save Consensus As** (Macintosh) pull-down menu to chose a format for the sequence. You may choose between DNASTAR, GenBank Flat File or FastA formats.
- Check **All** and **Add Features**, but leave **Include Gaps** unchecked. Checking **All** saves the entire sequence rather than a specified range. **Add Features** copies contig construction information into the comments pane of the sequence document. Leaving **Include Gaps** unchecked causes gaps in the consensus sequence to be discarded.
- Click **Save** to save the sequence.

To export edited constituent sequences:

- Highlight the sequence(s) you wish to export in the bottom pane of the Project window.
- From the **CONTIG MENU**, choose **Export Sequences**. This command is similar to **Save Consensus** and allows you to export edited sequences as sequence documents, Genbank Flatfiles or FastA files.

Further Exploration in SeqMan II

If you would like to experiment further with SeqMan II, we have provided two additional projects.

Project 1

This project will take you through the assembly of 1330 fragment sequences provided on your ***Lasergene99*** CD. Sequences may be found in the folders "Forward104k" and "Reverse104k," both located in the "Sample Data" folder. If you have enough space, the project will assemble faster if you begin by copying the sequence data to your hard drive. Follow the directions below to combine data into a single contig:

- Close any open SeqMan II projects, and create a new project by going to the **FILE MENU** and choosing **New**.
- You should leave all SeqMan II preassembly options and parameters set to their default values. If you edited the parameters earlier, check their settings by going to the **PROJECT MENU** and clicking **Parameters**. Set the top two values for **Assembling** to 12 and 80, and make sure that the **Trace Evidence** button in **Consensus Calling** is checked and its value set to 50. Click **OK**.
- Open the Windows Explorer or Macintosh Finder and locate the Forward 104k folder. Drag the Explorer or Finder window away from the Unassembled Sequences window so that you can view both at the same time.
- Use the mouse to drag the Forward 104k folder into the lower pane of the Unassembled Sequences window and drop it there. After a brief pause, each sequence file in the folder will appear individually in the Unassembled Sequences window. Please wait until all files have been listed before proceeding.
- Mark the already highlighted sequences to be scanned for Janus vector using the **Set Vector** pull-down menu.
- Locate the Reverse 104k folder in the Explorer or Finder window and drag it into the Unassembled Sequences window.
- Mark the newly added, already highlighted sequences to be scanned for InvJanus vector.
- Click the **Assemble** button to assemble the 1330 sequences into a single contig.

Project 2

This project illustrates the procedure for reassembling multiple contigs into a single contig:

- Create a new SeqMan II project and enter the original fourteen sequences as in the section "Entering Fragment Sequences."
- Click the **Trim Ends** button, select **Fixed**, and enter 300 and 500 into the **5' End** and **3'End** boxes. Click **Scan Later**.
- Click the **Assemble** button. Two contigs will be produced.
- Highlight both contigs in the Project View by holding down the **Shift** key while you select them.
- From the **CONTIG MENU**, choose **Extend Contig Ends**, then click **Extend**.
- Highlight both contigs again in the Project View.
- From the **CONTIG MENU**, choose **Reassemble Contig** to combine the two contigs into a single contig.