DNASTAR
Lasergene® 11

Core Suite Tutorials

for Windows and Macintosh

DNASTAR, Inc. 2013

Contents

Before You Begin

We're Here to Help!

If you have any difficulties with or questions about installation, please contact a DNASTAR support representative:

E-mail: support@dnastar.com

Phone (Madison, WI, USA): 608-258-7420

In the USA and Canada, call toll free: 1-866-511-5090

In the UK, call free on: 0-808-234-1643

In Germany, call free on: 0-800-182-4747

About These Tutorials

These tutorials pertain to DNASTAR's Lasergene® Core Suite version 11.0, and were last updated on October 23, 2013.

Video Tutorials

In addition to these written tutorials, free Lasergene video tutorials are available through the Training section of our website.

Copyright Information

For copyright and trademark information, please see the Legal Information page of our website.

SeqBuilder

Overview

SeqBuilder enables you to edit nucleic acid and amino acid sequences and view the sequences and related items in multiple ways. Available views include the following:

- Sequence view

- Feature List view

- Comment view

- Linear map

- Circular map

- Primer Design view

- Primer List view

- Minimap

- Site Summary

You can split SeqBuilder windows into multiple panes and view a sequence several ways simultaneously. Within a view, you can show or hide many meaningful display elements, including features, ORFs, translations, restriction sites, and comments.

SeqBuilder's rich, high-quality graphical displays can be easily customized to suit your preferences. In addition, all graphical views can be copied and pasted into standard drawing programs, making the production of publication quality figures a simple task.

SeqBuilder enables you to perform automated virtual cloning using several different cloning methods. SeqBuilder supports virtual restriction enzyme cloning as well as TA Cloning, Directional TOPO® Cloning, Gateway® Cloning and Gateway MultiSite Cloning. An extensive vector catalog to use for cloning is installed with your software, and custom vectors can easily be created and added to it.

SeqBuilder can assist you with designing primers for regions of interest on your sequence. Once a primer pair has been selected and modified as desired, SeqBuilder enables you to cut and clone the PCR insert into a vector using one of the cloning methods mentioned above.

SeqBuilder uses dynamic links between the sequence and its annotations to automatically update feature coordinates as you edit a sequence. Additional functionality includes the ability to reverse complement, translate and back-translate sequences, identify ORFs and perform BLAST and Entrez text searches directly to NCBI.

SeqBuilder's editing capabilities are integrated into most other Lasergene applications. SeqBuilder edits are shared automatically with other Lasergene applications that analyze the same data, making SeqBuilder an integral member of the powerful Lasergene Suite.

10 Quick-Start Steps for Using SeqBuilder

Here are ten basic steps for using SeqBuilder:

1) Create a new SeqBuilder project (**File > New Project**) or open an existing sequence file or project (**File > Open**).

2) Choose from among the nine views available for DNA sequences or from the three views available for protein sequences.

3) Modify the views by resizing, dividing, or closing panes and curtains, opening multiple views simultaneously, turning elements within the views on or off, rearranging elements on the display, or changing colors and styles of features so that they are easier to differentiate.

4) Select the restriction sites to display using the options in the **Enzymes Displayed** folder located in the curtain (DNA Sequences only). Or, setup new selectors (**Enzymes > New Selector**) to define a group of restriction enzymes to display based on specific criteria.

5) Search the BLAST or Entrez databases for matches to your sequence or other interesting sequences using the options from the **Net Search** menu.

6) Use the graphic results to form a hypothesis about your sequence, and then annotate any interesting features (**Features > New Feature**).

7) Design primers for regions of interest on your sequence (**Priming > Create Primer Pairs**).

8) Use the options from the **Cloning** menu to perform automated virtual cloning with any of the following methods: restriction enzyme cloning, TA Cloning, Gateway Cloning, or Directional TOPO Cloning.

9) Open SeqBuilder files in another Lasergene application. Edits made in SeqBuilder, Protean, GeneQuest, PrimerSelect, or MegAlign will be shared universally between those programs.

10) Save this file (**File > Save**) and open another to begin again, or exit the program.

SeqBuilder Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example SeqBuilder project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Create a new feature

Objectives: To open a sequence file and then to create and modify annotated features in a sequence.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP®:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OS® X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

Note: You will be working with the same sequence for this tutorial and the next, *Display restriction enzyme sites*.

1) Launch SeqBuilder.

2) Select **File > Open**.

3) Select **pGEM-T Easy.seq** from the **Demo Data** folder. Click **Open**. The sequence appears as follows.
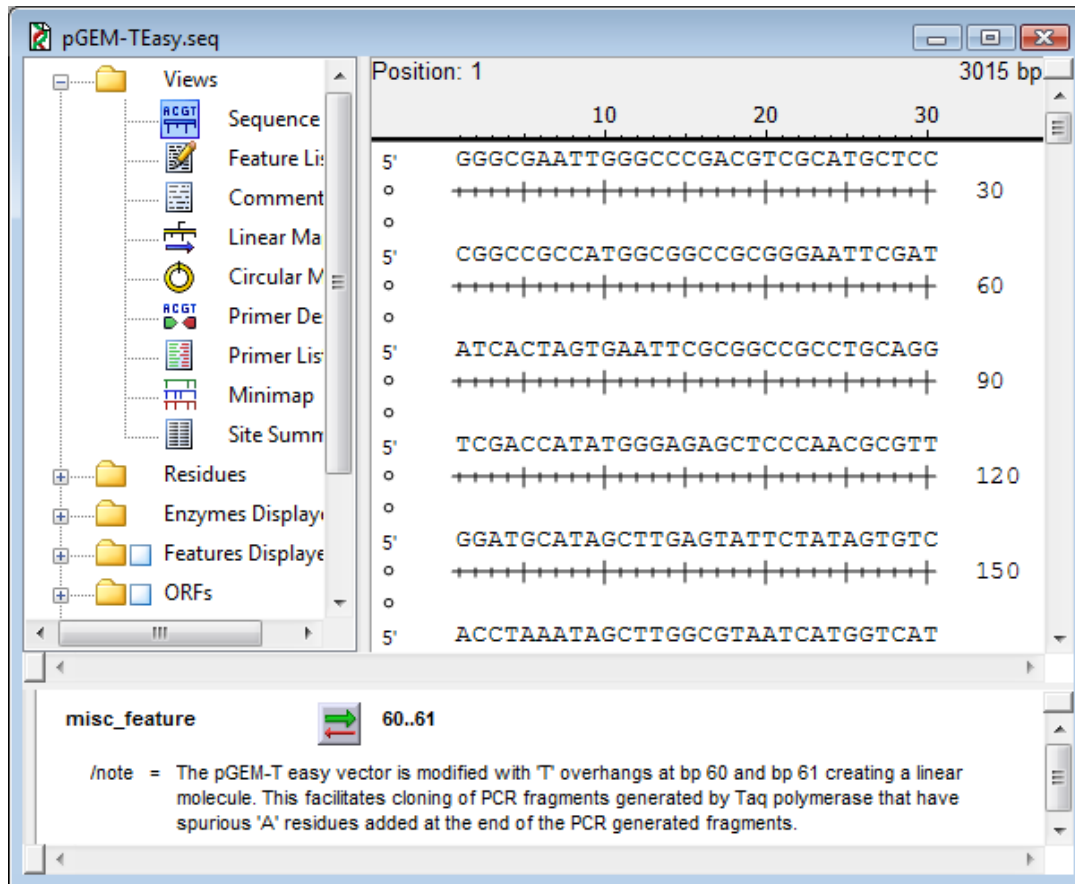
- The upper pane of the window contains the Sequence view. This view displays the base level DNA sequence. You can also display restriction sites, ORFs, and features in this view by using options from the curtain. For protein sequences, this view displays the ruler and protein sequence.

- The lower pane of the window contains the Feature List view. This view displays features that have previously been identified in the DNA or protein sequence. Features are annotations that provide additional information about a particular section of DNA or protein. Features may be user-defined, or they may be part of the input file from GenBank or from other Lasergene applications.

4) Select **Edit > Go To Position**. The Go To Position dialog appears.
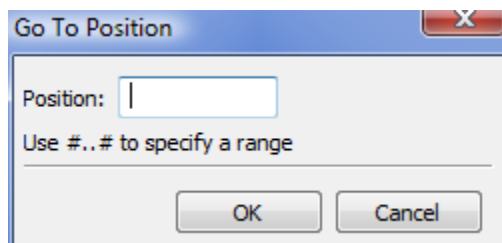
5) In the Position box, enter **10..128**. Click **OK**. The region from base pair 10 to base pair 128 is now highlighted in the Sequence view, the Linear Map, and Circular Map views.

6) Select **Features > New Feature**. A feature appears in the Sequence view in the top pane as an orange box and in the Feature List view as a new line entry, as shown below.
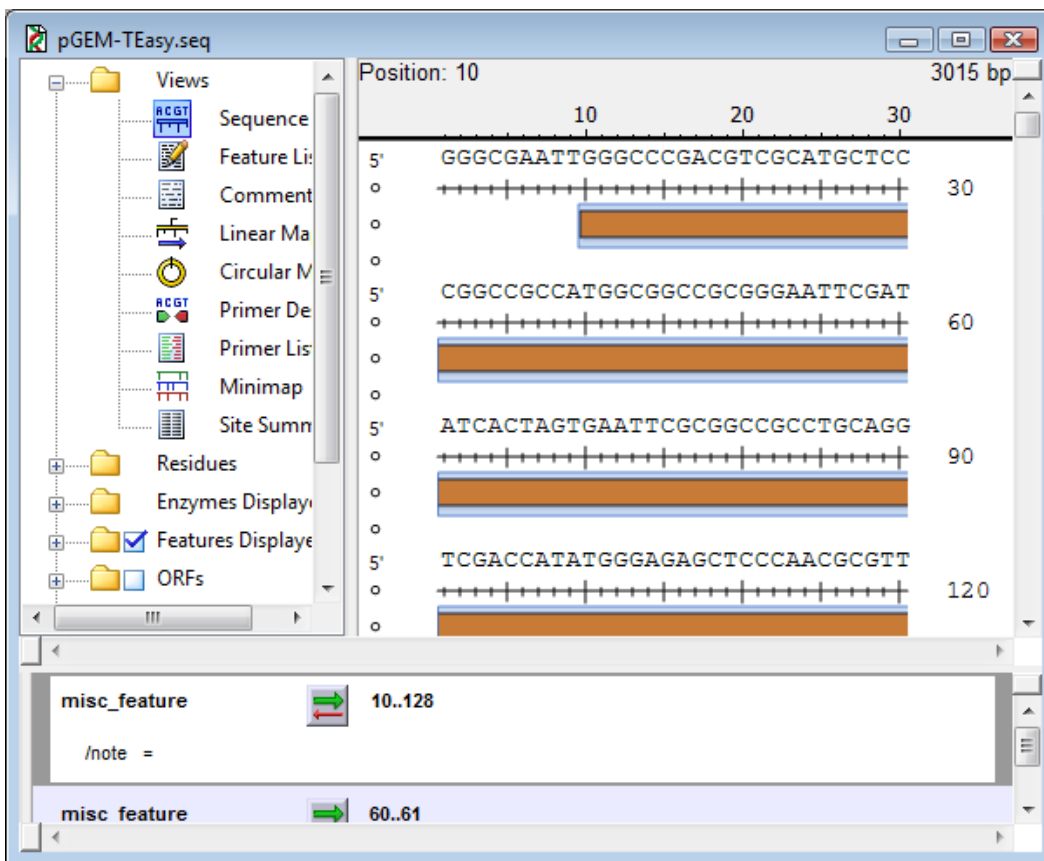
7) A blue box outlining the feature (the orange box) indicates that the feature has been selected. If you do not see a blue box, double-click the new feature to select it.

8) Select **Features > Edit Feature Style**. The Feature Style panel appears.

9) On the Feature Style panel: from the Shape list, select **Filled Arrow**; uncheck **Shadow**. The Feature Style panel should appear as follows:
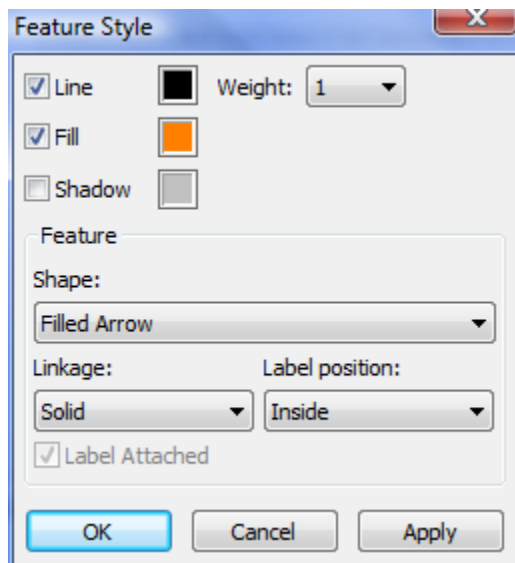
10) Click **Apply**, and then click **OK**.

11) In the bottom pane of the screen, click to the right of the equal sign next to the word **/note** on the new feature. Type **MCR** for multiple cloning region and press **Enter/Return**; this notation will now appear as the label for the feature in the Sequence view, as shown below.
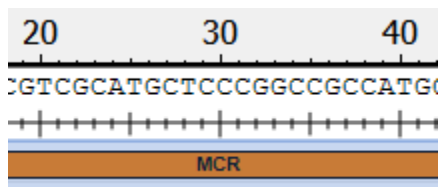
12) To make this the permanent label for the feature, click on **/note** and choose **Features > Use As Feature Label**. Alternatively, right-click on **/note** and choose the same command. Then press **Enter/Return**.

Proceed to the next tutorial.

Display restriction enzyme sites

Objective: To display restriction enzyme sites in a sequence.

Note: Before beginning this tutorial, you should complete the previous tutorial, *Create a new feature*.

1) Click in the upper pane (Sequence view) to make it the active pane.
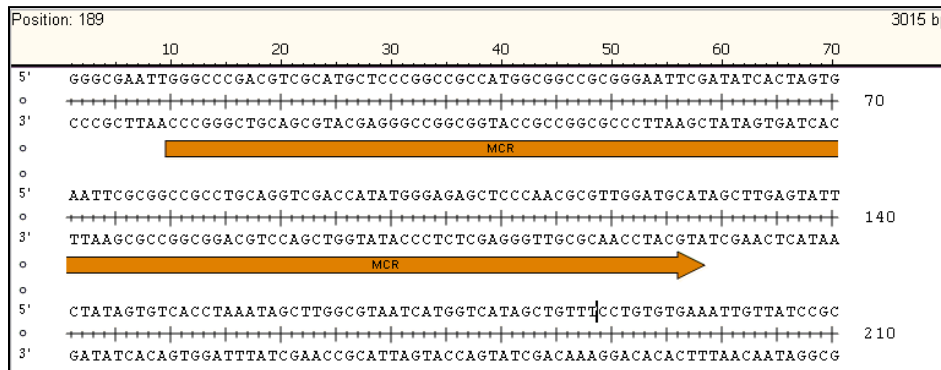
2) Locate the EcoRI enzyme by using the following technique.

   a) Next to the **Enzymes Displayed** folder in the upper pane's curtain, click ▶ (Mac) or ⊞ (Win) to expand the folder. You can now see the **Filter by Selectors** and **All Enzymes - Alphabetical** folders under the Enzymes Displayed folder.

   b) Click ▶ (Mac) or ⊞ (Win) next to the **All Enzymes-Alphabetical** folder to expand it.

   c) Navigate through the alphabetically named folders to find EcoRI.

3) Check the box next to the EcoRI enzyme to display it. Note that both of the EcoR1 sites fall within the feature defined in the previous tutorial, as shown below.
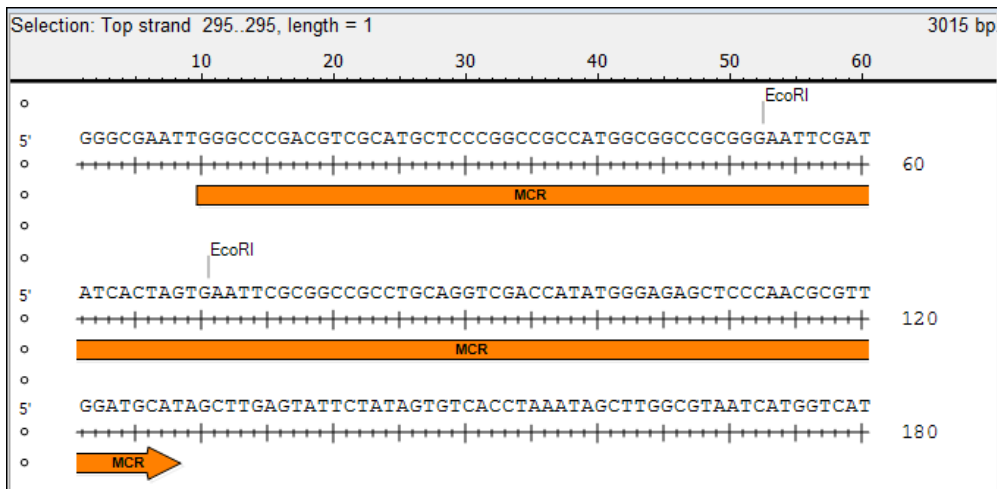
4) Click ▼ (Mac) or ▢ (Win) next to the **Enzymes Displayed** folder in the curtain to collapse the folder. Note that the selected enzyme EcoRI remains displayed in the Sequence view.

5) Go to **File > Close**. SeqBuilder will ask if you would like to save the file.

6) Click **No**.

Entering an Entrez sequence

Objective: To enter a histone DNA sequence from NCBI's Entrez database.

Note: In order to complete this tutorial, your computer must be connected to the Internet.

1) Launch SeqBuilder if it is not already open.

2) Select **File > Open Entrez Sequence**. The Open Entrez dialog appears.

3) Under Enter Sequence ID or Locus Name, enter **X05543.1**.

4) In the list box under that field, select **Nucleotide**.
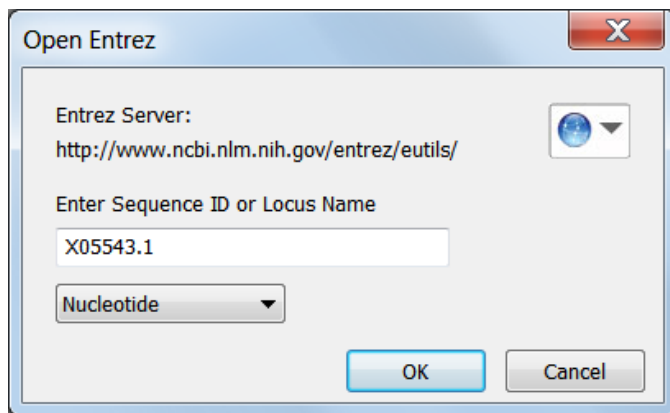
5) Click **OK**. A Save dialog appears.

6) Find a location to store the sequence and click **Save**. The sequence appears.

Proceed to the next tutorial.

Displaying restriction sites based on frequency

Objective: To create and apply a frequency selector that displays only enzymes that cut the sequence twice or less.

Note: Before beginning this tutorial, you should complete the previous tutorial, *Entering an Entrez sequence*.

1) Select **Enzymes > New Selector…**. The Enzyme Selector Manager appears.

   Note: New Selector is highlighted in the Selector list and the Name is **New Selector**.

2) In the Name box, type **Two cuts max**.

3) From the Type drop-down menu, select **Frequency**.

4) Overwrite the number **1** in the Max box with the number **2**. Leave all other parameters at their default settings.
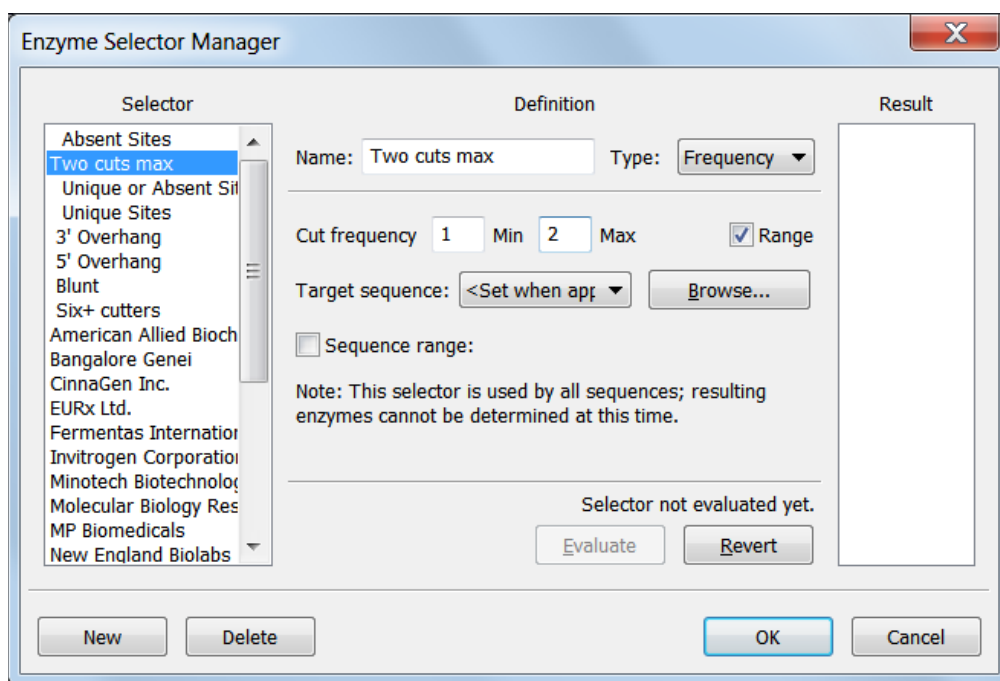
5) Click **OK** to create the new selector and return to the main view.

6) Locate the new selector in the upper pane's curtain by clicking ▶ (Mac) or ⊞ (Win) next to the **Enzymes Displayed** folder. You can now see the **Filter by Selectors** and **All Enzymes - Alphabetical** subfolders.

7) Click ▶ (Mac) or ⊞ (Win) next to the **Filter by Selectors** folder to expand it.

8) Put a checkmark in the **box** next to the **Two cuts max** selector to view the restriction enzymes that will cut this sequence two times or less.

9) Click ▼ (Mac) or ⊟ (Win) next to the **Enzymes Displayed** folder to collapse the hierarchy.

Proceed to the next tutorial.

Displaying manually chosen restriction sites

Many restriction sites remain. You can now set some more stringent criteria.

Note: Before beginning this tutorial, you should complete the previous tutorial, *Displaying restriction sites based on frequency*.

Objective: To use one of the remaining sites to excise the histone-coding region from the X05543.1 sequence.

1) In the upper pane's curtain, select **Linear Map**.

2) In the same curtain, click the box to the left of **Features Displayed**.

3) On the linear map, scroll up or down to locate the **CDS** feature named **Histone H2B-1**. This is a sequence feature that was annotated in the original GenBank entry. The information for the feature was imported with the sequence when you opened it.

4) Single-click the feature to highlight the feature's range in all views.
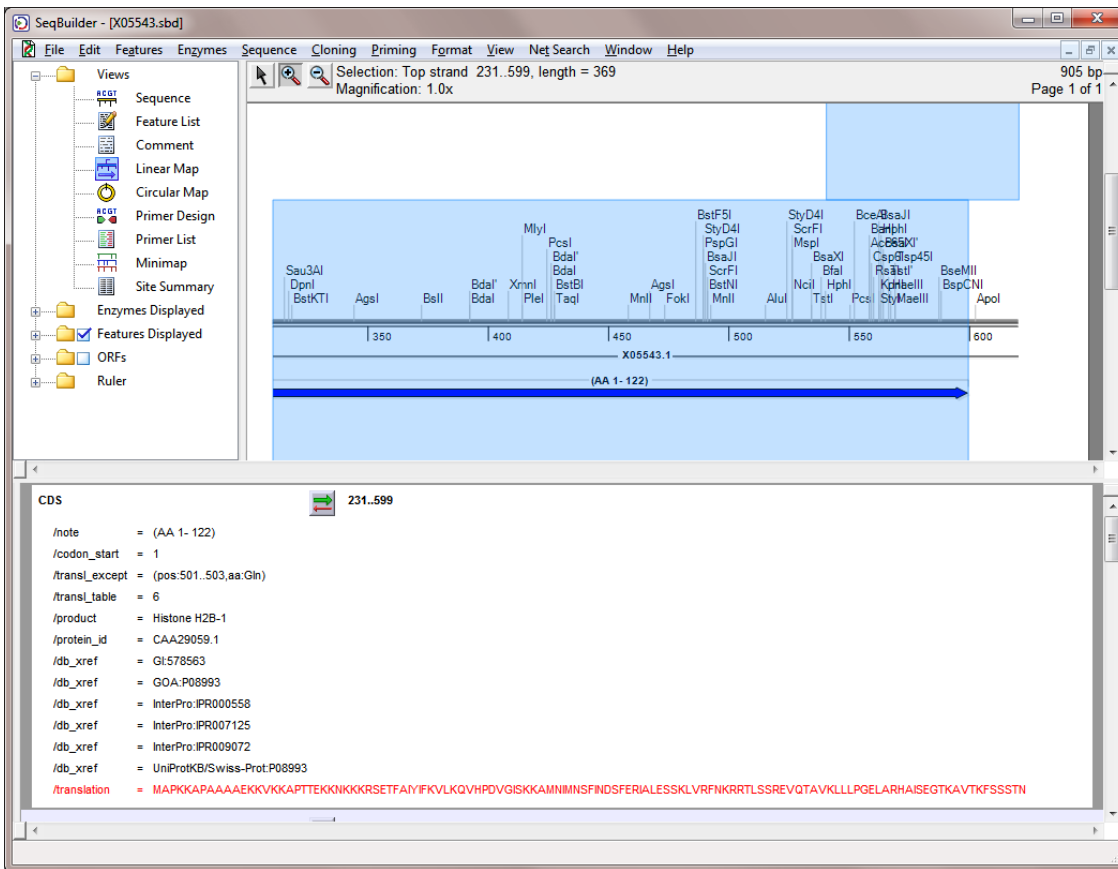
5) Next, select **Minimap** from the upper pane's curtain. The Minimap view displays a smaller version of the current DNA sequence. The Minimap is scaled to show the whole sequence in a single line. In this view, restrictions enzyme names appear as individual rows in the left-hand column. Their corresponding cut sites are shown in the same row on the right.
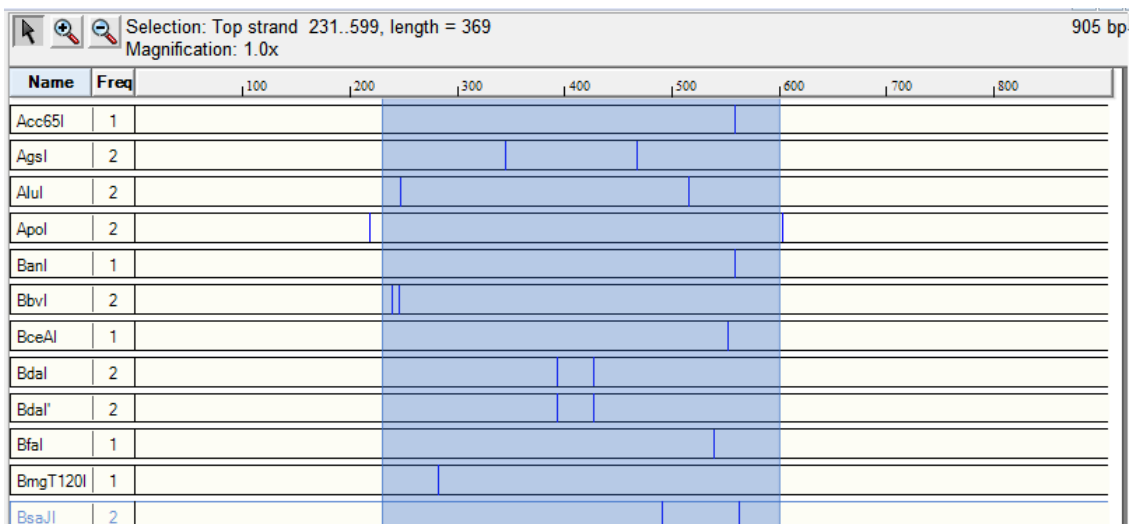
6) At the top of the Minimap view, click on the ruler.

This will sort the enzymes in order of those with cuts closest to the selection, in this case, the CDS feature.

Note: The restriction sites are sorted so that those cutting closest to and outside the feature appear near the top, those cutting further away from the outside appear further down, and those cutting inside the feature's range appear at the bottom.

7) Leaving the feature highlighted, scroll to the top of the window. Note that the ApoI restriction site cuts just to the left and right of the histone gene, as shown below, and thus may be used for excising the gene.

8) Select **Enzymes > New Selector**. The Enzyme Selector Manager appears.

9) In the Name field, enter **ApoI selector**.

10) In the Type box, select **Pick**.

11) In the Remaining Enzymes list, scroll down to and select **ApoI**.

12) Click . ApoI now appears in the Selected Enzymes list, as seen below.
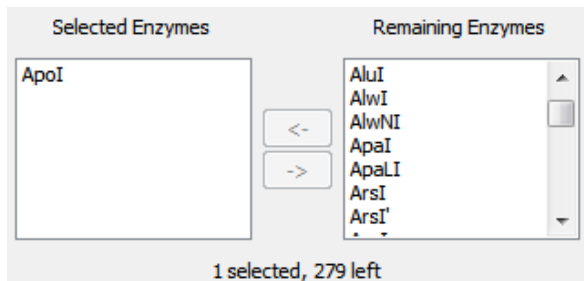
13) Click **OK**.

14) Apply the new selector using the following technique.

- Next to the **Enzymes Displayed** folder in the upper pane's curtain, click ▶ (Mac) or ⊞ (Win) to expand the folder. You can now see the **Filter by Selectors** and **All Enzymes - Alphabetical** folders under the Enzymes Displayed folder.

- Click ▶ (Mac) or ⊞ (Win) next to the **Filter by Selectors** folder to expand it.

- Uncheck the box next to **Two cuts max** to unselect it.

- Check the box next to the **ApoI selector** to display it. The Minimap view now shows only the ApoI restriction sites.
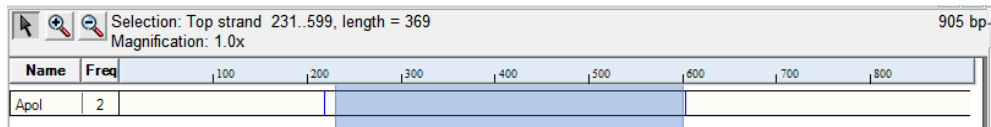
15) Select **File > Close**. SeqBuilder will ask if you would like to save the file.

16) Click **No**.

Note: The Enzyme Selectors that you created will be saved, and can be accessed later either from the curtain or by going to **Enzymes > Selector Manager**.

Working with SeqBuilder cloning projects

In addition to restriction enzyme cloning, SeqBuilder supports these three following commonly used methods:

- Gateway Cloning
- Directional TOPO Cloning
- TA Cloning

Note: Restriction cloning and TA cloning methods are discussed further in the section *Designing Primers in SeqBuilder*.

Beginning a New Cloning Project

To start a new cloning project, select **File > New Cloning Project**:
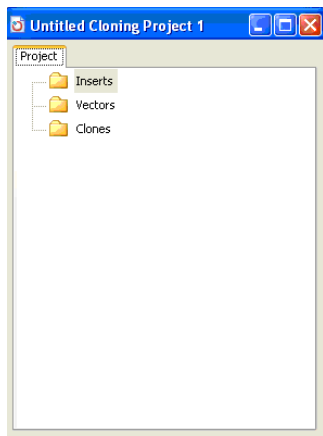
This New Project window is initially named Untitled Cloning Project, but you can rename it when you select **File > Save Project As**. All cloning projects contain three folders: Inserts, Vectors, and Clones.

When you create a project, you can either select inserts before you start cloning or select them during the process. With vectors, however, it is best to move any custom vectors into the Vector Catalog prior to including them in a project. The reason is that all but the restriction cloning projects use the vectors listed in the Vector Catalog.

Here is an example of what the folder looks like after cloning a PCR fragment:
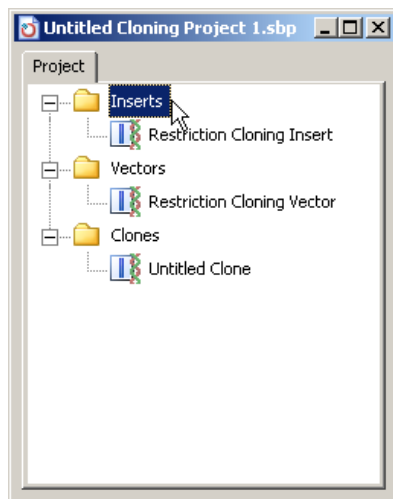
Note: As stated earlier, if you have custom vectors that you plan to use, you can either create them during the course of the project or you must import them into the Cloning Vector Catalog before you can use them in a project. To do this, see the information in the next two tutorials.

Adding custom vectors to a project

Objective: To create a custom TA or Directional TOPO vector for use with your cloning project.

1) Launch SeqBuilder.

2) Go to **File > New Cloning Project**. The Untitled Cloning Project window will be displayed.

3) Click once on the **Inserts** folder to highlight it.

4) Select **File > Import Sequences(s) to Project** and choose the file *N.thermophilus_his.sbd* from the Lasergene 11 Data\Demo Data folder.

5) Double-click on the file to open it in the Sequence View.

6) Single-click on the *PAS/PAC sensor signal transduction* feature to highlight it.

7) Do one of the following:

- To create a T-Vector: Hold down the **Alt** (Windows) or **Option** (Mac) key. Simultaneously, click and hold on the **Cloning** menu, then select **Create TA Cloning Insert**. This adds the linearized T-vector to the front-most project.

  Note: Failing to hold down the **Alt/Option** (Win/Mac) key will instead create a Taq amplified insert.

  Or

- To create a Directional TOPO vector: Hold down the **Alt** (Windows) or **Option** (Mac) key. Simultaneously, click and hold on the **Cloning** menu, then select **Create Directional TOPO Insert**. This adds the linearized TOPO vector to the project.

  Note: Failing to hold down the **Alt/Option** (Win/Mac) key will instead create a directional TOPO insert.

Holding down the **Alt/Option** (Win/Mac) key while selecting one of the menu options listed above changes the function of the menu item and acts as a shortcut for automatically creating linearized vectors with suitable overhangs.

A linearized T-vector contains 3' T overhangs at both ends. A linearized TOPO Vector contains a 5' GTGG overhang and a blunt 3' end. Custom vectors without the appropriate overhangs will not be recognized in the Directional TOPO or TA Cloning Advisors.

Note: If you create a linearized custom vector manually (without using the shortcut), you will likely also need to manually adjust the overhangs (**Cloning > Change Overhangs**) to meet the requirements listed above.

Adding vectors to the Cloning Vector catalog

The vectors listed in the Gateway Cloning, Directional TOPO Cloning, and TA Cloning advisors come from both your project and the *CloningVectors.sbp* catalog. To make your custom vector files available to the cloning advisors in all projects, you will need to add them to this special catalog. Otherwise, you will only be able to access your vector in the project(s) that contain it.

Objective: To add a vector to the Cloning Vector catalog.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11

1) Launch SeqBuilder

2) Select **File > Open** to display the Open File dialog.

3) Highlight and double-click the **CloningVectors.sbp** from the location listed above. *Cloning Vectors.sbp* is the vectors catalog and is a cloning project.

4) Do *one* of the following:

   - If the vector file is a sequence file, click on one of the folders, select **File > Import Sequences(s) to Project** to navigate to the file you want, and then double-click the sequence to open.

   - If the vector file is a sequence file, open the file, select **File > Add 'x'.seq** to CloningVectors.sbp. A copy of the sequence file is added to the project file.

   - If the vector file is in another project file, open that project by selecting **File > Open**, change the Files of Type to a SeqBuilder project (.sbp) and select your project. Copy and paste the file from the project to the Cloning Vector catalog.

Creating a Gateway clone

Gateway Cloning is a widely used molecular technique that allows one to transfer DNA inserts between different cloning vectors while maintaining the reading frame and orientation of the insert. This technology eliminates the need for restriction enzyme digestion, ligation and colony screening for recombinants and is based on the site-specific recombination system of phage I.

Objective: To create a virtual Gateway clone.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch SeqBuilder

2) Select **File > New Cloning Project**. The Untitled Cloning Project window appears.

3) Add a sequence to the **Inserts** folder by first highlighting the folder.

4) Select **File > Import Sequence(s) to Project** to display the Open File window. Look for the **Tn5wPCR.sbd** sequence in the **Demo Data** folder, and double-click the file to add the file to the **Inserts** folder.

5) Open and select the insert sequence of interest. For this tutorial, click on the neomycin/kanamycin resistance feature labeled **aminoglycoside-3'-O-phosphotransferase** to select that range of sequence.

6) Select **Cloning > Create Gateway Insert/Entry clone**. The Create Gateway Entry Clone dialog appears.

7) Choose your donor vector from pull down list (use **PDONR™201**), then click **Create** to add an Entry clone to your cloning project.

Note: Features in the Circular Map view can be moved by double-clicking a feature and then dragging it either up or down.

8) Go to **Window > Untitled Cloning Project 1** to return to the project window.

9) Go to the **Clones** folder, select **Untitled Entry Clone** and go to **Edit > Rename**. Rename the sequence **GA clone1**.
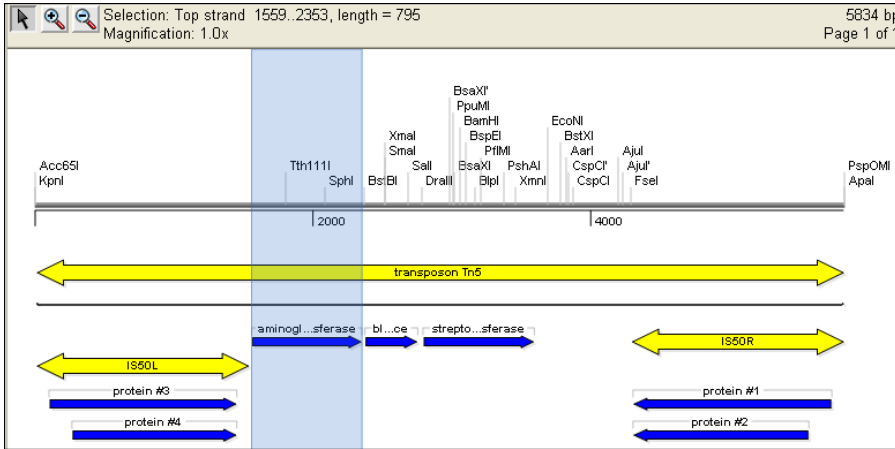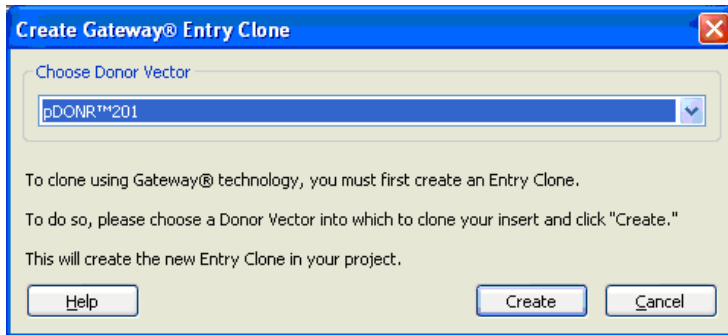
10) Select **Cloning > Gateway Cloning Advisor**. The advisor dialog appears.
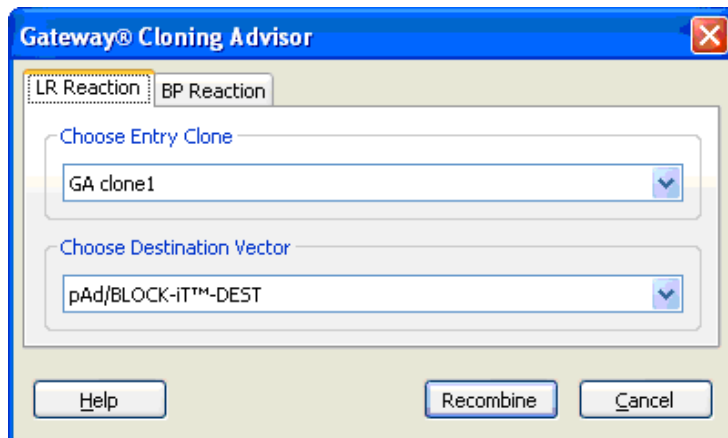
11) Select a suitable entry clone, (use **GA clone1**), and destination vector (use **pAd/BLOCK-iT™-DEST**) for recombination, then click **Recombine**.
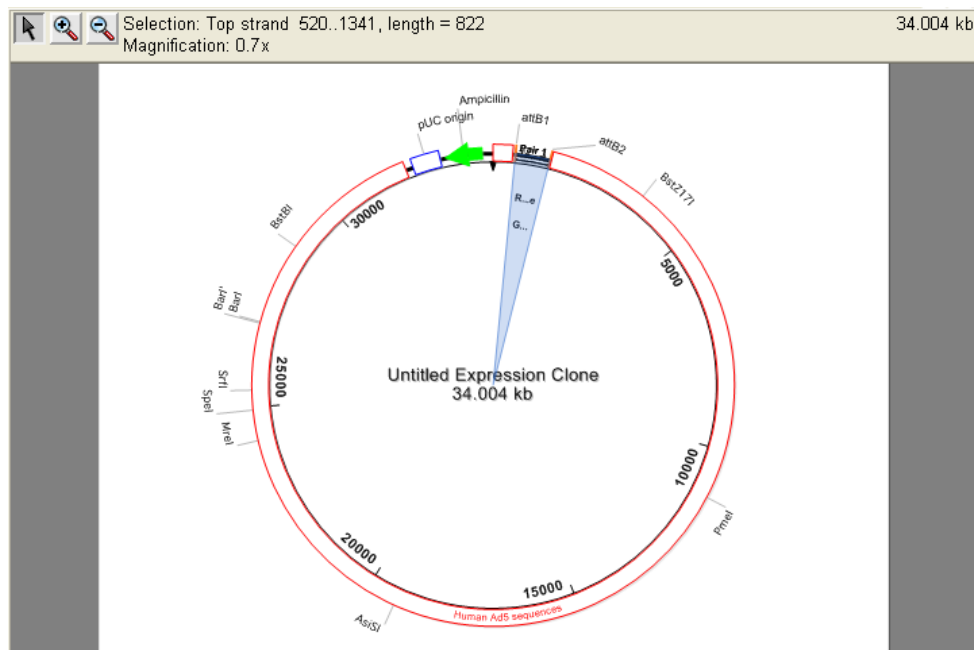
12) Save the project and the information in its subfolders by selecting **File > Save Project As**.

13) Enter a new name in the File Name field and choose a file location, then click **Save**.

14) Go to **File > Close**.

Creating a Gateway MultiSite clone

Multisite Gateway technology enhances the classic Gateway cloning protocol by allowing you to clone up to four regions at one time. In Multisite Gateway cloning, three additional types of recombination sites have been engineered to recombine uniquely between the two ends of a source sequence and a host vector.

Objective: To create a clone that has a Pol promoter, two multiple cloning sites, and a reporter gene.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\ Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch SeqBuilder

2) Select **File > Open**. Navigate to the **Demo Data** folder, select the *MultiSiteExample.sbp* file, and click **Open**. The MultiSite Example cloning project is displayed.

3) Expand the **Inserts** folder so that the project appears as follows:
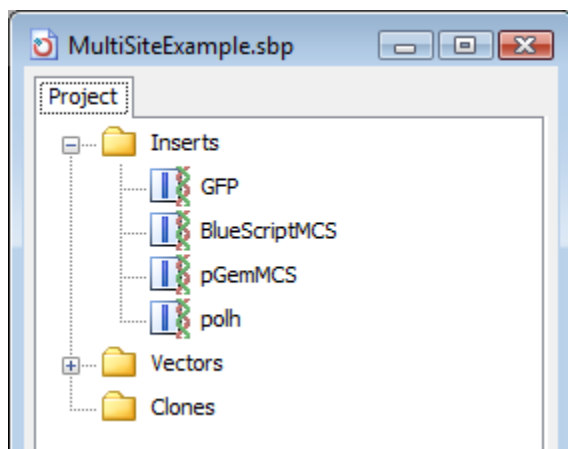
4) To set up the entry clone for Green Fluorescent Protein, the reporter gene, open the **GFP** sequence by double-clicking on it. The GFP sequence is displayed in the Sequence View.

5) Select **Edit > Select All** to select the entire GFP sequence.

6) Then, select **Cloning > Create Gateway Insert/Entry Clone**. The Create Gateway Entry Clone dialog appears.
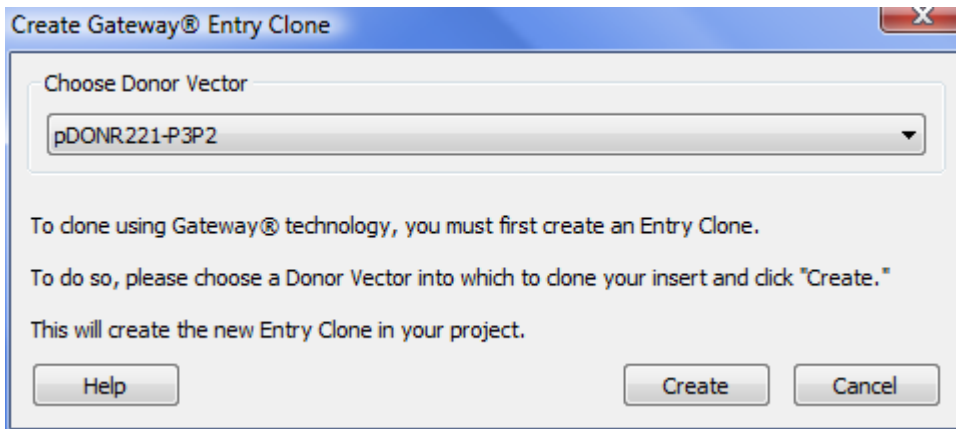
7) In order to place the GFP reporter gene into the last of the four insert positions, we need to select the donor vector with the appropriately annotated P3-P2 *att* sites. Select **pDONR221-P3P2** from the dropdown list, and then click **Create**. An entry clone will be created and displayed containing the GFP insert, and automatically added to your project in the **Clones** folder.

8) Close the newly created Untitled Entry Clone, and then select it from the Project Window. Rename it to **GFP L3-L2** by selecting **Edit > Rename**.
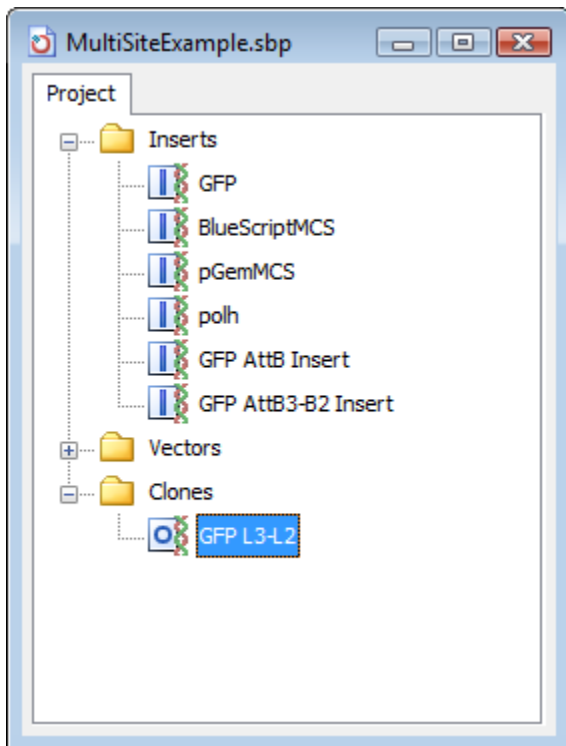
9) Repeat steps 4-8 for the other 3 insert sequences.

- For **BlueScriptMCS**, select **pDONR221-P5P4** for the Donor Vector and rename the newly created clone **BlueScriptMCS L5-L4**.

- For **pGemMCS**, select **pDONR221-P4rP3r** for the Donor Vector and rename the newly created clone **pGemMCS L4-L3**.

- For **polh**, select **pDONR221-P1P5r** for the Donor Vector and rename the newly created clone **polh L1-L5**.

10) Once all the entry clones are constructed, we are ready to perform the final recombination. Expand the Vectors folder from the Project Window, and select the destination vector **pYES-DEST52**.
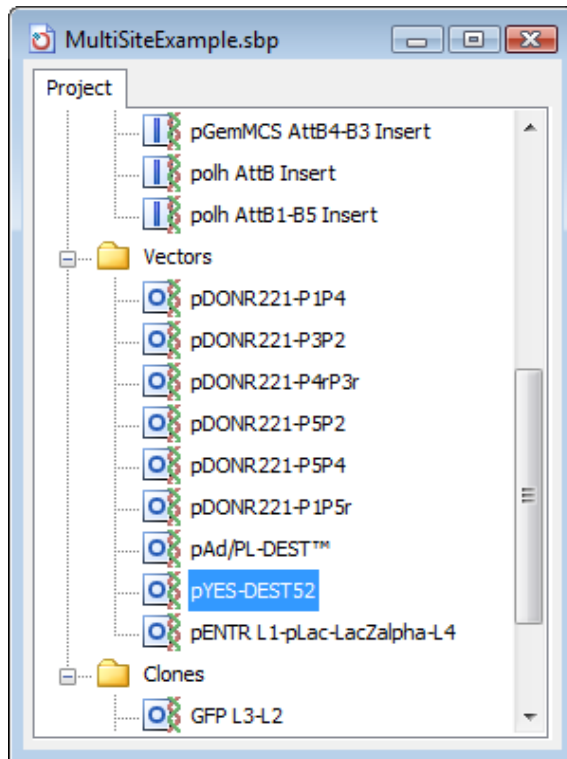
11) Select **Cloning > Gateway Multisite Cloning Advisor**. The Gateway MultiSite Cloning Advisor is displayed and lists the destination vector we had selected from the Project Window:
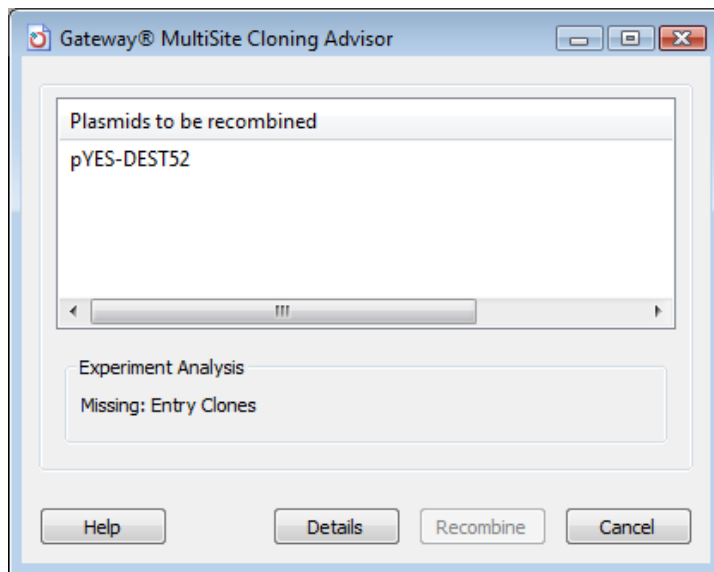
12) Notice in the Experiment Analysis section of the dialog, the Advisor lets us know that we need to specify entry clones before recombination can occur. To do this, go back to the Project Window, and click on the **Clones** folder to select it. Then, drag and drop the **Clones** folder onto the **Gateway MultiSite Cloning Advisor** window. All of the newly created entry clones from the **Clones** folder will be added to the Advisor window:
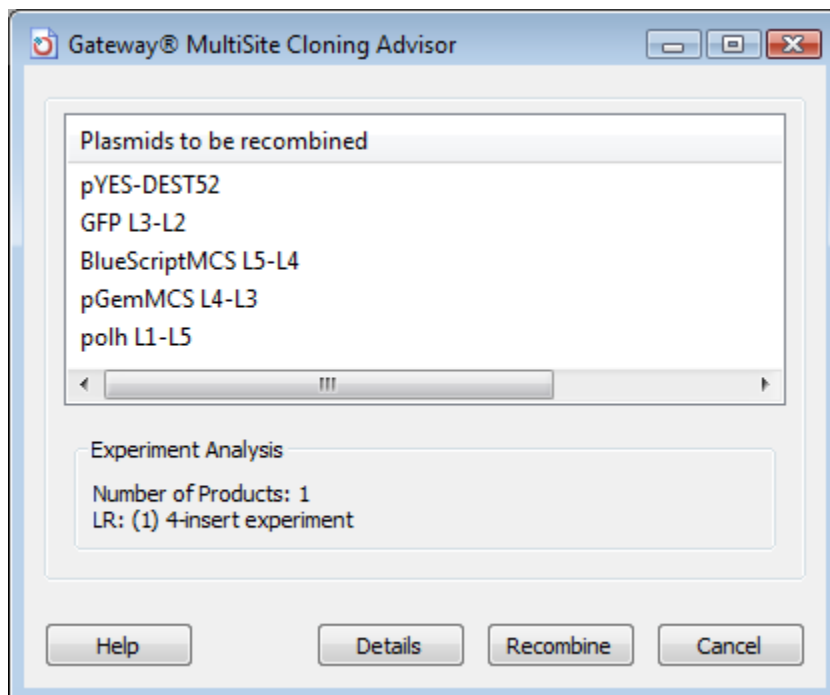
13) Notice that in the Experiment Analysis section of the window, a message appears to let us know that recombination of the specified plasmids will result in *one* clone with *four* inserts. Since this is the result we expect, click the **Recombine** button to proceed with recombination. The expression clone containing all four inserts is created and displayed in the Circular Map view:
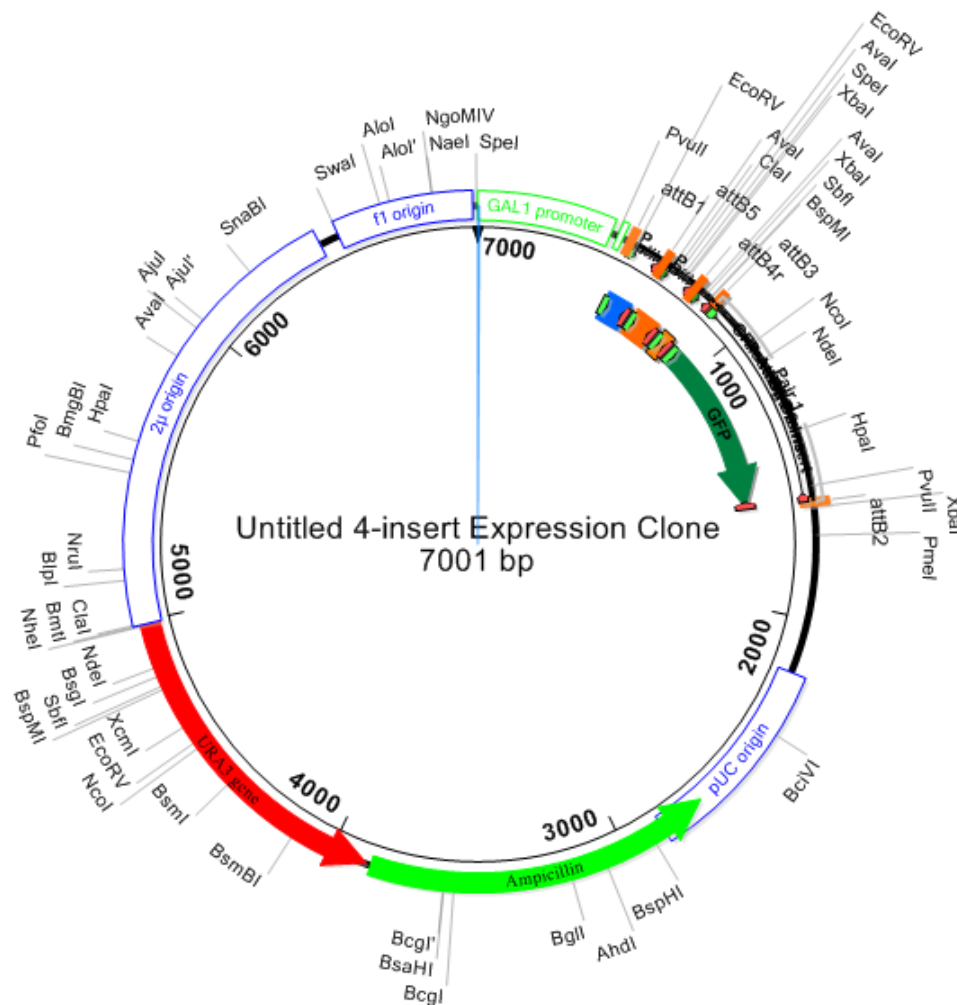
Creating a Directional TOPO clone

Directional TOPO Cloning utilizes DNA topoisomerase I, which allows researchers the ability to directionally clone DNA targets. The enzyme functions as both a restriction enzyme and ligase. The covalent attachment of DNA topoisomerase I to the vector allows researchers the ability to directionally clone targets with a high degree of specificity.

Objective: To create a TOPO clone.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch SeqBuilder

2) Select **File > New Cloning Project**. The Untitled Cloning Project window appears.

3) Add a sequence to the **Inserts** folder by first highlighting the folder.

4) Select **File > Import Sequence(s) to Project** to display the Open File window. Look for the **Tn5wPCR.sbd** sequence in the **Demo Data** folder, and double-click the file to add the file to the Inserts folder.

5) Open the insert sequence, and select the sequence range of interest. For this tutorial, click on the neomycin/kanamycin resistance feature labeled **aminoglycoside-3'-O-phosphotransferase** to select that range of sequence.
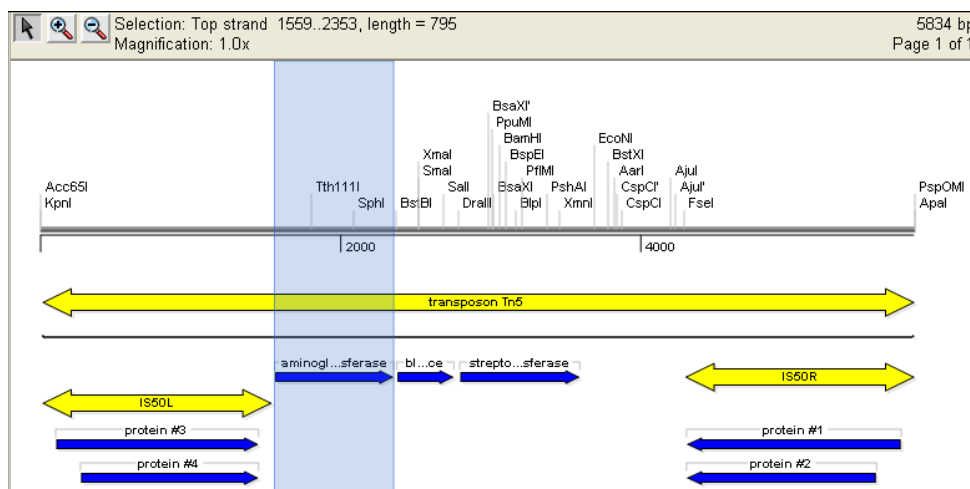
6) Select **Cloning > Create Directional TOPO Insert**. The new insert is placed in the Inserts folder, and the Directional TOPO Cloning Advisor dialog is displayed.
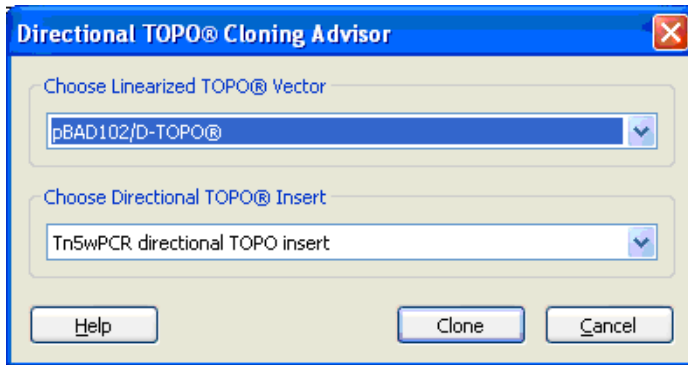
7) Verify the selected insert, and choose the appropriate cloning vector (use **pBAD102/D-TOPO** for this tutorial). Click **Clone**. The clone is displayed in a new window and it is stored in the Clone folder of the project window.

Note: The TOPO vector sequence should be linearized at the appropriate base pair prior to adding to the cloning project. See *Adding custom cloning vectors to a project*.

8) Click **Circular Map** in the View pane then click the **Features Displayed** checkbox within the curtain.
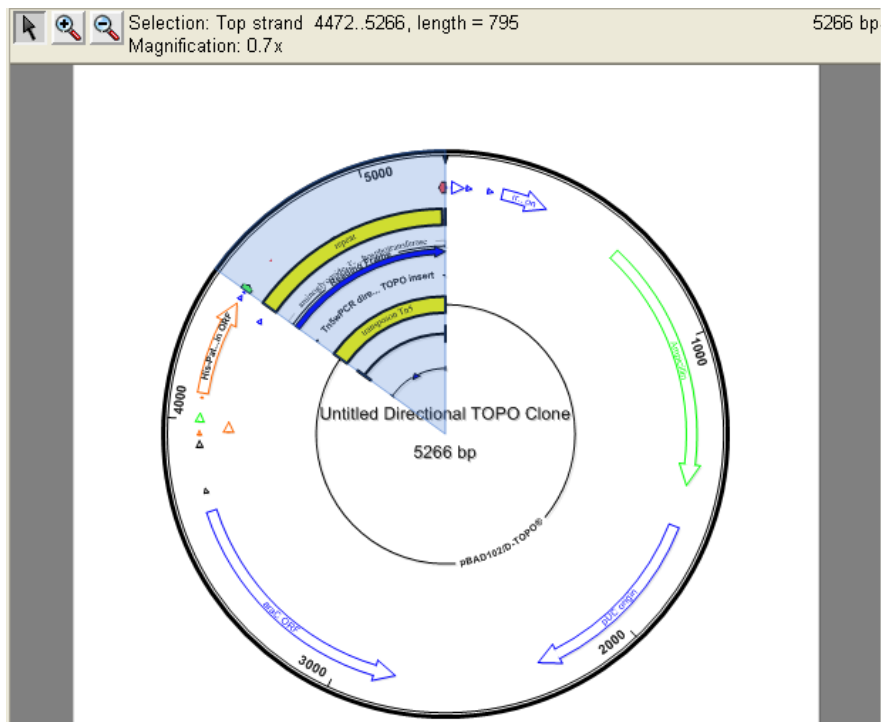
Note: Features in the Circular Map view can be moved by double-clicking a feature and then dragging it either up or down.

9) Save the project and the information in its subfolders by selecting **File > Save Project As**.

10) Enter a new name in the File Name field and choose a file location, then click **Save**.

Designing primers in SeqBuilder

SeqBuilder enables you to design primers for regions of interest on your sequence. If desired, once a primer pair is selected and modified, SeqBuilder allows you to cut and clone the PCR insert (with corresponding primer features) into a vector.

The data for the tutorials in this section can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 :Demo Data

Creating and modifying primer pairs for a region of interest

Objective: To create a primer pair for a region of interest on your sequence.

1) Launch SeqBuilder.

2) Select **File > Open** and choose **Tn5wPCR.sbd** from the **Demo Data** folder. The sequence opens in the Linear Map view.

3) Choose the neomycin/kanamycin resistance feature labeled **aminoglycoside-3'-O-phosphotransferase** as the region of interest by single clicking on that feature. The sequence range 1559-2353 will be selected.
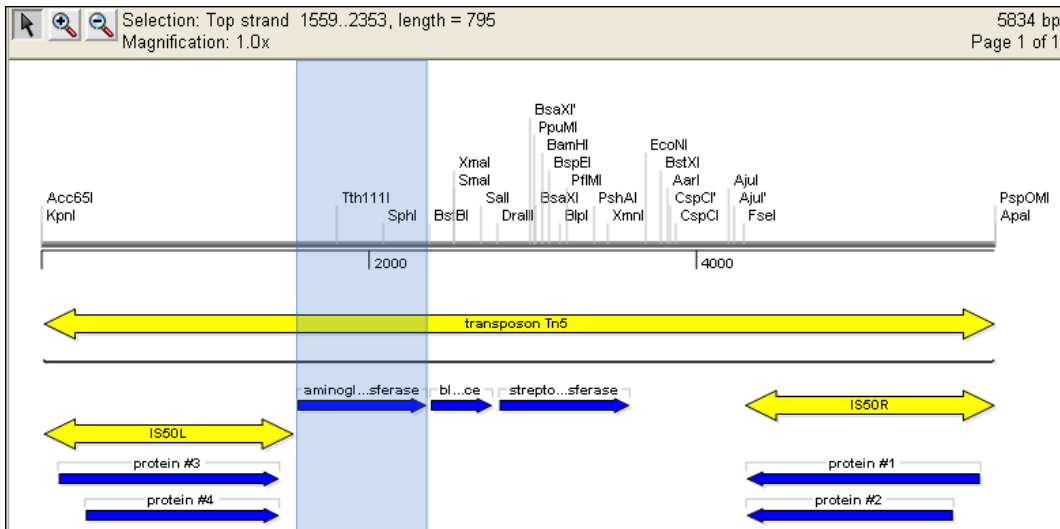
4) Select **Priming > Create Primer Pairs**. The Create Primer Pairs dialog appears.

5) Click the triangles next to **Conditions** and **Primer Characteristics** to expand those sections.

Note: The **Locations** section of the Create Primer Pairs dialog allows you to specify parameters that define where SeqBuilder will search for primers on your sequence. The **Conditions** section allows you to specify the initial salt concentration, which affects the calculation of predicted melting temperature. The **Primer Characteristics** section enables you to limit your search for primers based on primer length, target melting temperature, and primer interactions at the 3' end. You also have the option in this dialog to avoid known repetitive sequences that are likely to cause mispriming.

6) Adjust the **Target Tm** to **55.0°C**. Leave the rest of the values as they are, and then click **OK**.

SeqBuilder will choose the best primer pair with the characteristics specified that lies exactly within the selected region and display it in the Primer Design View. By default, the view focuses on the top strand primer first.

Note: While the default settings are useful in most cases, we have deliberately chosen an example in which the default **Primer length** values will not yield optimal results.

7) If necessary, resize the Primer Design View pane by clicking and dragging the dividing bar between panes so that you can see all 3 sections of the view: the Residue Pane, the Mispriming Pane, and the Alternate Pairs Pane:

Note: If desired, you can choose to display the Primer Design View in both the top and bottom panes, and then scroll to the Mispriming and Alternates Pairs sections in the bottom pane so that you can easily see all three sections at once. As you edit the primer in the Residue Pane, the other two panes are automatically updated to reflect the changes.

8) Notice in the Mispriming Pane, the **Most stable dimer** and **Most stable pair dimer** are both labeled "BAD!" This indicates that the final pentamer value of the dimer exceeds the **3' pentamer stability** threshold defined in the Primer Characteristics section of the Create Primer Pairs dialog. In general, you should avoid using primers with dimers or hairpins labeled "BAD!" unless you have experimental evidence that they will function as desired, or if you have no other option.

9) Switch to the Primer List view by selecting it from the curtain shown on the left side of the SeqBuilder screen. The single primer pair located by your search is displayed.

10) Click the triangle next to **Set 1** to expand the primer pair and view each individual primer within the pair:

	Set Name	Pair Name	T/B	Primer Sequence	Length	Tm	dG	Current
▼	Set 1	Pair 1		<ATGATTGAACAAGATGGATTGC>..<CCTTCTTGA	795	Tm=83.5, TaOPT=58.7, dTm=33.0		✔
			T	ATGATTGAACAAGATGGATTGC	22	50.8	-38.6	✔
			B	TCAGAAGAACTCGTCAAGAAGG	22	50.5	-38.2	✔

11) Notice that the melting temperatures for the individual primers are both close to 50° C, which is lower than our Target Tm of 55°C.

12) Switch back to the Primer Design view to view the top strand primer:
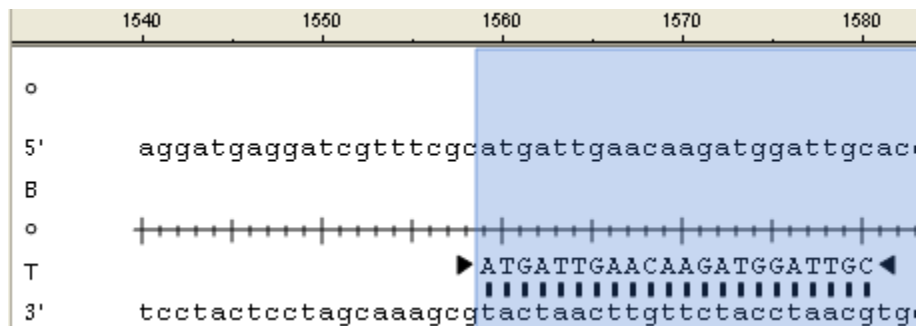
13) Select the codon change mode by clicking on the button in the Primer Design Toolbar. Notice that your cursor changes and that each codon is highlighted as you hover over the sequence.

14) Introduce a codon change by clicking on the **TTG** triplet in your primer, located at 1577-1579 of your sequence. From the list of codons given, select **CTA**. Notice that changes to the primer sequence and translation are shown in magenta, and that the Mispriming Pane has now updated to show that there is no longer a stable dimer conformation.

15) Display the bottom strand primer by clicking the button from the Primer Design Toolbar. The Primer Design View scrolls to show the bottom strand primer in the center of the view:
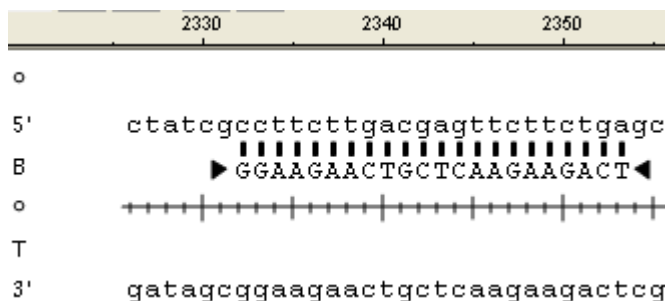
16) Increase the length of the bottom strand primer from **22 bp** to **23 bp** by clicking and dragging the triangle shown on the left (3') end of the primer sequence. The length of the primer is displayed and updated automatically in the Primer Design Toolbar shown at the top of the view:

17) Notice that once the primer length is increased to 23 bp, the melting temperature for the bottom strand primer increased to 54.4°C, which is much closer to our Target Tm. Also notice that the **Most stable pair dimer** has been updated in the Mispriming Pane and no longer has the indication: "BAD!".

18) Switch back to the top strand primer by clicking on the button. Increase the length of the top strand primer from **22 bp** to **31 bp** by clicking and dragging the triangle shown on the right (3') end of the primer sequence. Notice that the melting temperature of the top strand primer has now increased to 55.3°C.

19) Now that the primers have been modified to our satisfaction, change back to the Primer List view by selecting it from the curtain on the left.

20) Expand the pair again by clicking on the triangle next to **Set 1**. Notice that the edits to our primer pair are reflected here as well. Nucleotide changes to the primer sequence are shown in lowercase.

Set Name	Pair Name	T/B	Primer Sequence	Length	Tm	dG	Current
▼ Set 1	Pair 1		<ATGATTGAACAAGATGGAcTaCACGCAGGTT>..<	795	Tm=83.5, TaOPT=59.9, dTm=29.1		✔
		T	ATGATTGAACAAGATGGAcTaCACGCAGGTT	31	55.3	-50.1	✔
		B	TCAGAAGAACTCGTCAAGAAGGC	23	54.4	-41.4	✔

21) Name the modified pair by clicking on **Pair 1** in the Pair Name column and typing **Modified pair**.

Proceed to the next tutorial.

Creating a PCR insert for TA cloning

Now that the primer pair has been selected and modified as desired, you are ready to extract the PCR product for cloning.

Objective: To extract the PCR insert and create a virtual TA clone.

1) Switch to the Primer Design view by selecting **Primer Design** ![ACGT icon] from the curtain.

2) Choose **Priming > Create Insert by PCR** to extract the sequence of the PCR product and simulate its amplification with Taq polymerase including the addition of 3' A overhangs to each end. SeqBuilder will automatically create a new cloning project and display the new Taq amplified insert within the **Inserts** folder:
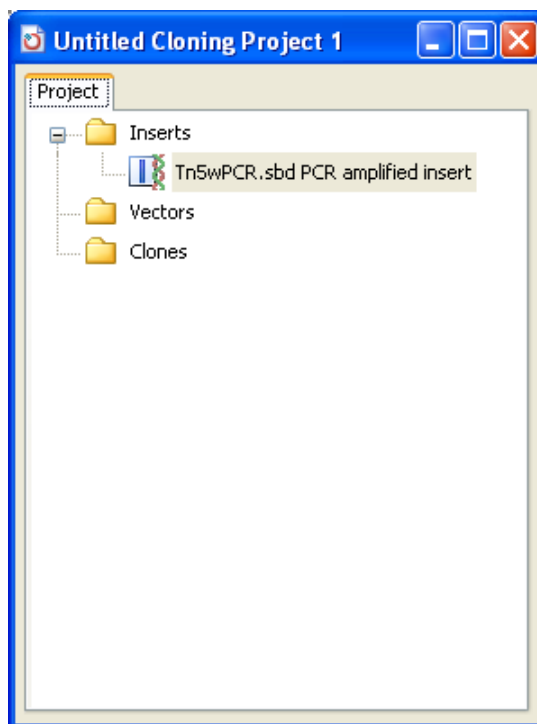
3) Select **Cloning > TA Cloning Advisor** to display the TA Cloning Advisor dialog:
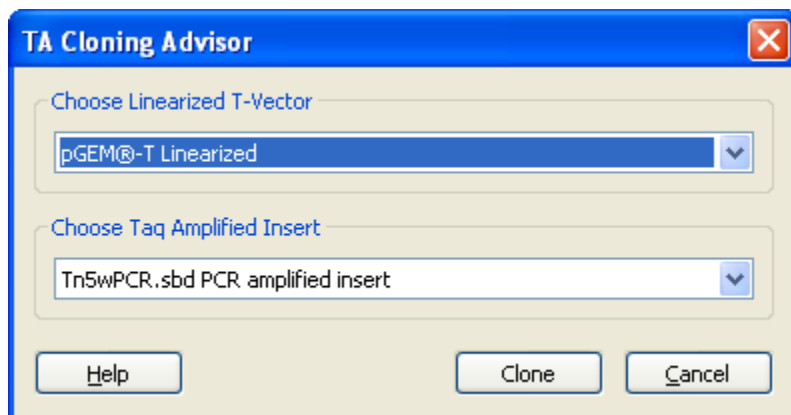
4) Notice that your PCR amplified insert is already selected in the Insert field. Select **pGEM-T Linearized** from the Linearized T-Vector dropdown list, and then click **Clone**. The clones, forward and reverse orientations, are created and displayed.

5) Bring the forward clone into focus by selecting **Window > Untitled TA Clone (forward orientation)**. Notice that the primers, PCR product, and reading frame are all displayed in the Sequence View as features. These features can also be seen in the Feature List view, as well as all graphical views.

6) If desired, save the project and the information in its subfolders by selecting **File > Save Project As**. Otherwise, choose **File > Exit** to close the project and exit SeqBuilder.

Note: To create a cloning insert from the PCR product of the current primer pair using a cloning method other than TA Cloning, simply select the range of sequence for the PCR product you wish to amplify, and then select the appropriate command from the **Cloning** menu, depending on the type of cloning you wish to do. For information on other cloning methods, see the section *Working with SeqBuilder Cloning Projects*.

Creating primer pairs and manually introducing restriction sites

Objective: To create primer pairs for a gene of interest, and then manually introduce restriction sites to prepare for restriction cloning.

1) Launch SeqBuilder.

2) Select **File > Open** and choose *N.thermophilus_his.sbd* from the **Demo Data** folder. The sequence opens in the Linear Map view.

3) Select the range of sequence for the CDS feature named **PAS/PAC sensor signal transduction histidine kinase** by single-clicking on it. Notice in the header that the range **10086..11369** is selected.

4) Select **Priming > Create Primer Pairs**. The Create Primer Pairs dialog appears.

5) Select the radio button next to **Choose optimal primer location** and leave the other values as they are. This will allow SeqBuilder to choose the best primer locations from within 50 bp of the surrounding sequence.
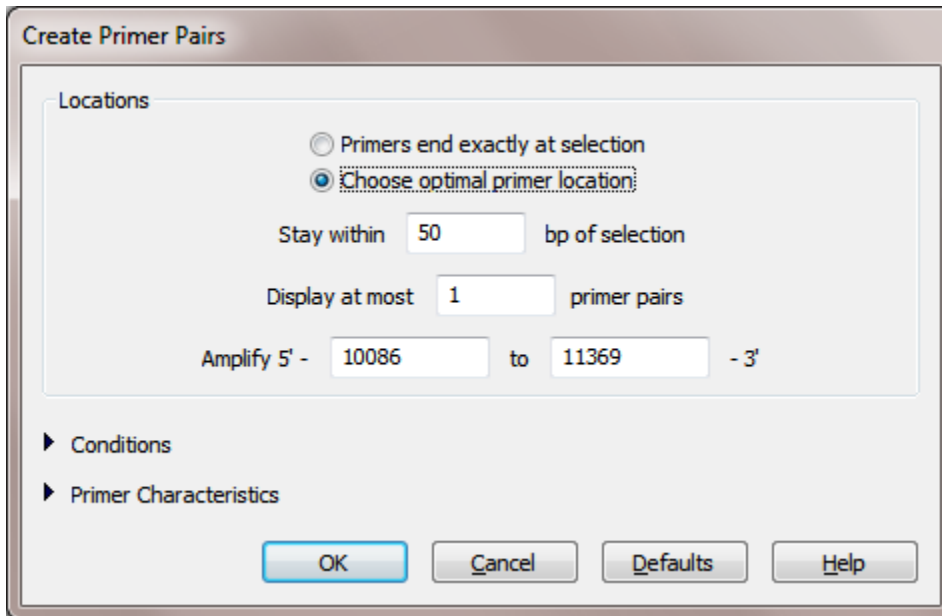
6) Click **OK**. SeqBuilder locates the best primer pair and displays the top strand primer in the Primer Design view.

7) The top strand primer is 24 bp in length. Extend the primer 8 bp by clicking and dragging the triangle shown to the left of the primer sequence. Notice that the length in the header now shows that the top strand primer is 32 bp in length.

8) Select the button from the Primer Design toolbar to enter into the mode for introducing restriction sites. Notice that your cursor changes to look like scissors.

9) Place your cursor to the immediate left (5' end) of the extended primer and **click**. From the resulting menu, select **Any Enzyme**, and then select the enzyme **SacII**. (*Hint:* To navigate quickly through the alphabetical list of enzymes, type "S" to skip directly to that letter in the list). The **SacII** enzyme is introduced.

```
                        Tsp509I
                         BtgI
                         BsaJI
                         AciI
                         PacI
        MseI             MseI
        Tsp509I    MspA1I                                    HpyCH4III
         BarI'     BstUI                                      TspRI
         PacI      AciI   NlaIV      CviJI     HpyCH4III       TspRI
        AseI       SacII   BsrI      AluI   Tsp509I            MseI
        MseI              SacII  BsrI    SetI  TsoI   MseI

        aagcgtttaattaattaactaaccagtagctattacacagtaattaacagtgattaag

        +++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
              ▶ccgcggAACCAGTAGCTATTACACAGTAATTAACAG◀
                ▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
        ttcgcaaattaattaattgattggtcatcgataatgtgtcattaattgtcactaattc
        K   A   F   N   .   L   T   N   Q   .   L   L   H   S   N   .   Q   .   L   S
                            P   R   N   Q   .   L   L   H   S   N   .   Q
        K   R   L   I   N   .   L   T   S   S   Y   Y   T   V   I   N   S   D   .   '
                N   R   G   T   S   S   Y   Y   T   V   I   N   S
        S   V   .   L   I   N   .   P   V   A   I   T   Q   .   L   T   V   I   K
                    T   A   E   P   V   A   I   T   Q   .   L   T   V
```

Note: Notice that the new restriction sites are displayed in magenta, as well as nucleotides in the primer sequence and their translations that differ from the template. Restriction enzymes whose recognition sites are eliminated by the change are shown in blue.

10) Click ▢ to display the bottom strand primer. Notice that the bottom strand primer is also 24 bp in length, as shown in the Primer Design toolbar.

11) Extend the bottom strand primer 6 bp to increase its length to 30 bp. Do this by clicking and dragging the triangle shown to the right of the primer sequence.

```
   Position: 10151
   Working Pair "Pair 1"    Top: Length=36 bp, Tm=59.4°C,   Bottom: Length=30 bp, Tm=56.0°C
   11370      11380       11390      11400      11410      11420      11430

        CviJI
        AluI
        TseI                              CviJI        CviJI
          Fnu4HI          BbvI            TseI         BfaI
          SetI            Tsp509I         Fnu4HI       SetI          PleI
          BlsI            XmnI   Tsp509I  BlsI       AluI   HinfI    MlyI   Mn

   5'  agctgcgattgatataggaagcaattcaattaggctgctaatagctagagtcaataaaaatgg
   B                           ▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪
                               ▶GTTAATCCGACGATTATCGATCTCAGTTAT◀
   o   ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
   T
```
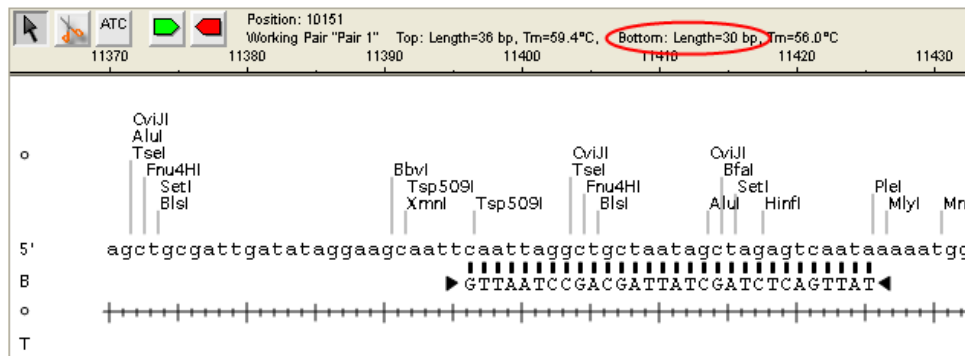
12) Select the ✂ button again, and click immediately to the right (5' end) of the extended bottom strand primer sequence. Select **Any Enzyme**, and then select the enzyme **AscI**. The **AscI** enzyme is introduced into your sequence.

Proceed to the next tutorial.

Creating a PCR insert for restriction cloning

Objective: To extract the PCR insert and use restriction cloning to insert it into a Gateway Entry clone.

Now that the appropriate restriction enzymes have been added to the sequence, we can cut and clone the histidine kinase gene into the desired vector.

1) Select **Priming > Create Insert by PCR**. A new cloning project is created and the PCR amplified insert appears in the **Inserts** folder.

   Note: This menu item extracts the sequence of the PCR product and simulates its amplification with Taq polymerase including the addition of 3' A overhangs to each end. In this example, we will use restriction cloning, which cleaves the sequence at the designated restriction sites. Thus, the 3' A overhangs will not affect the resulting clone.

2) Select **File > Open** and choose the cloning vectors catalog, **CloningVectors.sbp**, from the **Lasergene 11 Data** folder.

3) In the Cloning Vector catalog, expand the **Gateway Vectors** folder, and then expand the **Entry Clones** folder.

4) Find **pENTR™/D-TOPO** and move it into the **Vectors** folder within your cloning project, either by dragging and dropping, or by using the copy/paste options from the **Edit** menu.
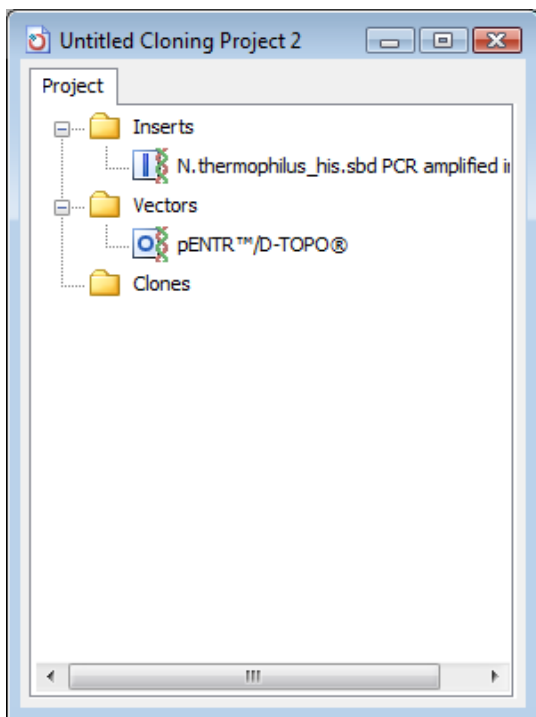
5) Open the PCR amplified insert from the **Inserts** folder by double-clicking on it. The sequence opens in the Sequence view.

6) Switch to the **Linear Map** view by selecting it from the curtain.

7) Display unique enzymes by expanding the **Enzymes Displayed** folder, then expanding **Filter by Selectors** and checking the box next to **Unique Sites**. A confirmation dialog may appear. Click **Yes**. All of the unique restriction enzymes will be displayed on your sequence.

8) Click on **SacII** on the far left (5') end of the sequence. Then, Shift-click on **AscI** on the far right (3') end of the sequence. Notice in the header that the enzymes and the sequence between them are selected.
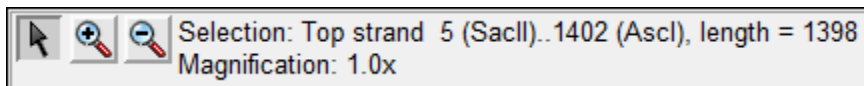
9) Select **Cloning > Copy Restriction Insert**. The insert is copied to the clipboard.

10) Return to the cloning project window by selecting **Window > Untitled Cloning Project 1**.

11) Double-click on **pENTR™/D-TOPO** from the **Vectors** folder. The sequence opens in the Circular Map view.

12) Select **Edit > Go to Position** and type **669..704**. Click **OK**. The region between the two recombination sites, shown in orange, is selected.

13) Switch to the **Sequence** view by selecting it from the curtain. Notice that there are SacII and AscI sites spanning the TOPO recognition sites.
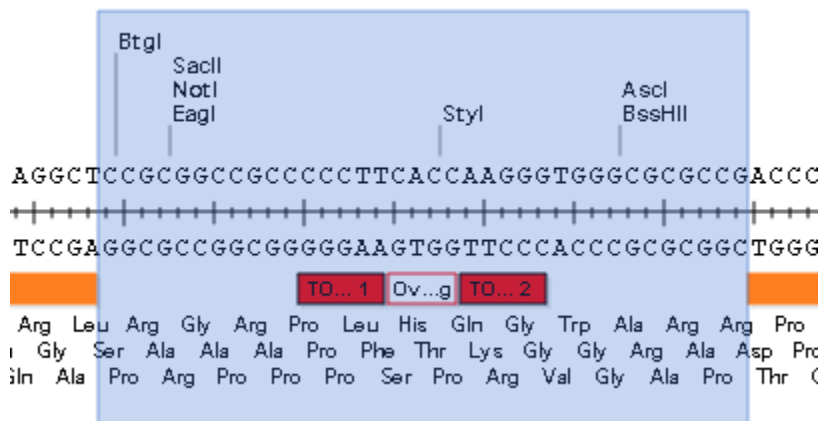
14) Click on **SacII** and then Shift-click on **AscI**. The enzymes and the sequence between them are selected.

15) Select **Cloning > Clone Restriction Insert**. The Clone Insert dialog appears.

---

16) Since both ends are already compatible, click the **Clone** button. A new Untitled Clone is created and displayed. You now have a thermostable kinase within a Gateway entry clone, ready to clone into a Destination Vector.

17) If desired, save the project and the information in its subfolders by selecting **File > Save Project As**. Otherwise, choose **File > Exit** to close the project and exit SeqBuilder.

For more information on Gateway Cloning, see the section *Working with SeqBuilder Cloning Projects*.

Further Exploration

Creating and removing panes and curtains

The default SeqBuilder document window contains two *panes* (one above the other). Each pane in turn contains a *curtain* on the left and a *view* on the right. *Curtains* contain options that provide rapid access to the most common elements that you may want to display in a given view. *Views* display data, as modified by the options selected in a curtain. By default, the upper pane displays the Sequence view and the lower pane displays the Feature List view for the same file. The SeqBuilder document window can be divided into any number of panes. Create a new pane by clicking and dragging the small box directly above a pane's vertical scroll bar on the far right, by clicking and dragging the small box to the left of a pane's horizontal scrollbar, or by using the **Vertical Split** or **Horizontal Split** options from the **Window** menu. Remove a pane by double-clicking the divider between two panes or by dragging the divider between any two panes all the way to the divider of another pane.

Searching the sequence

To locate an ORF, a particular portion of sequence, or a specific restriction site or feature, select **Edit** > **Find**. To find the next instance, click **Find Next** or select **Edit** > **Find Again**.

Using the nine SeqBuilder views

The **Views** folder in a curtain permits you to choose from the following views by clicking the icon next to the view name or by clicking on the name. The options available in a curtain vary with the view selected. If a portion of a sequence is selected (highlighted) in one view, the Circular Map, Linear Map, Minimap, Sequence, and Primer Design view in another pane will likewise display the same highlighting.

Views available for nucleotide sequences only

	Linear Map	Click to display a linear map of the current DNA sequence. This view displays on one page by default, but you can change the number of pages by going to **File > Page Layout**. This view shows selected restriction sites, ORFs, and features along the linear map.
	Primer Design	Click to design and analyze primers for your sequence. The Primer Design view is divided into 3 panes. The top pane displays the base level nucleotide sequence, restriction sites, translations, and features, as well as the currently selected primer sequence. The middle pane shows primer binding sites, and the most stable conformation of dimers, pair dimers, and hairpins, if they exist. The bottom pane displays the list of primer pairs discovered by your search, along with additional information for each pair, including product length, melting temperatures, and a quality score.
	Primer List	Click to view the results from your primer searches. The Primer List displays the primer sequence, the length of the primer and the product (for primer pairs), melting temperature, dG values, and is formatted like a spreadsheet to allow for easy selection and sorting.
	Circular Map	Click to display a circular map of the current DNA sequence. This view shows selected restriction enzymes and features. The Circular Map view is scaled to fit onto one printed page by default, but you can change the number of pages by going to **File > Page Layout**.
	Minimap	Click to display a smaller version of the current DNA sequence. The Minimap is scaled to show the whole sequence in a single line. Each row in this view shows where a particular restriction site falls along the sequence.
	Site Summary	Click to display a table of restriction sites and their locations on a DNA sequence. This view displays restriction sites that are selected in this pane or in another pane.

Views available for both nucleotide and protein sequences

ACGT	**Sequence**	Click to display the base level DNA sequence. By default, this view also shows translations and ORFs. You can also display restriction sites, partial translations, and features in this view by using options from the curtain. For protein sequences, this view displays the ruler and protein sequence. This view is in the upper pane in the initial default setting.
	Feature List	Click to display features that have previously been identified in the DNA or protein sequence. Features are annotations that provide additional information about a particular section of DNA or protein. Features may be user-defined, or they may be part of the input file from GenBank or from other Lasergene applications. By default, this view is in the lower pane.
	Comment	Click to display a comment that is associated with this DNA or protein sequence. Comments are optional and not all sequences have them. This is where the text of a GenBank entry appears. You may enter any free-form text that you wish in this view.

Viewing and editing restriction site information

- To view and edit restriction sites in the restriction enzyme library, select **Enzymes** > **Enzyme Manager**.

- To add a new restriction enzyme to the library, select **Enzymes** > **New Enzyme**.

Exporting to GenVision

- To export your SeqBuilder document as a GenVision project, select **File > Export as GenVision Project** from the Linear Map, Circular Map, Sequence View, or from the Project Window. (For protein sequences, this option is available from the Sequence view only).

- To create images in GenVision that compare the features from multiple sequences, first load and select the desired sequences into a **Project Window** (**File > New Project**). Then, choose **File > Export as GenVision Project**. The resulting dialog finds matching features in the selected sequences and allows you to select which types of features to include in the export. One GenVision project will be created for each of your selected sequences, and will display the features from one sequence connecting to the same features in all of the other sequences where it occurs.

SeqMan

Overview

SeqMan is both a sequence assembler for small scale sequence assemblies (plasmids, PCR products, etc.), and a viewer for all assemblies, including BAM alignment file based projects. For larger sequencing projects and those using next-gen data, DNASTAR's SeqMan NGen is used for the assembly process, while SeqMan is used to view and analyze the finished assembly.

Using SeqMan as an Assembler:

SeqMan enables you to assemble fragment data from sequencing projects up to tens of thousands of fragments. (Larger projects—up to and including human genome size—can be assembled with SeqMan NGen). Sequences can be assembled from numerous file formats, including trace data from automated sequencers, DNA Multi-seq (*.mseq) files with features, and data output created using 454™ technology. SeqMan removes unreliable data, including poor quality ends, sub-minimal length reads, and vector and contamination host sequences in a single pass, then assembles the trimmed data and calls the consensus.

SeqMan provides two different assembly methods: the Classic assembler, and the Pro assembler. In general, the Pro assembler should be used when your data: 1) are medium in size; 2) contain repeated sequences; 3) have noisy ends; or 4) are being used for SNP analysis. The Classic assembler should be used when: 1) your data do not contain repeated sequences; 2) you do not use vector trimming; or 3) you want to reproduce an assembly made from a previous version of SeqMan. As with all SeqMan parameters, you may select one of these assemblers as your default assembly method. You may also switch between the Pro and Classic assemblers with each SeqMan project.

For projects that utilize trace data from automated sequencers, SeqMan includes DNASTAR's unique Trace Quality Evaluation system. This system evaluates the quality of the underlying trace data, and then generates the most accurate consensus sequence possible. This method is so effective that it reduces the depth of coverage needed for accurate sequence determination, yielding spectacular savings in time and effort.

Before assembling the sequences, be sure to review SeqMan's preassembly options. These options include trimming poor-quality data manually or automatically, removing specific vector or contaminant sequences, and identifying repetitive sequences so that they will be added last during assembly. You may also adjust a wide variety of parameters for the assembly process, including which assembly method is used, as mentioned above. Once you have selected the options you want, click the Assemble button to activate all of the trimming, assembly and consensus calling options you selected. If you prefer to assemble certain groups of sequences separately, use the Assemble in Groups option.

Using SeqMan as a Viewer and Editor:

As a viewer, SeqMan lets you view the consensus, features, and other components via displays such as the Alignment View, Strategy View, Features Table, SNP Report, and Primer Walking Report. It also allows you to display and analyze multi-sample data, such as MID-tagged 454 samples, that were assembled using SeqMan NGen.

SeqMan allows editing of SeqMan (*.sqd) and Phrap (*.ace) assemblies, and can also perform a statistical calculation of their SNPs. By contrast, SeqMan supports viewing of BAM-based projects, but does not support editing or calculation of SNPs in these projects. Such SNPs are instead calculated during assembly with SeqMan NGen.

Note: SeqMan NGen produces an assembly project containing BAM alignment files for each of the contigs, BAM index files, template files, SNP files and coverage files. A BAM-based project may be gapped or ungapped, but always contains purely BAM contigs.

BAM alignment files are not editable. Because of this, some SeqMan procedures cannot be followed with BAM-based projects. Topics and commands that are not applicable to BAM-based projects are clearly marked in this manual.

Assembly information is summarized in SeqMan's tabular reports and graphical views. The Strategy View graphically summarizes the position and orientation of every constituent sequence in a contig or contig scaffold, and allows you easily assess the coverage in your assembly. If coverage seems unsatisfactory, you can add more sequences and reassemble, or you can use the Primer Walking feature to drive the closure of gaps or to fill in low coverage areas. The Alignment View gives you a more detailed picture of the assembly, and allows you to edit constituent and consensus sequences, override the called consensus, review underlying trace or flowgram data, add new features, restore previously trimmed data, and adjust alignments. You may also evaluate putative SNPs identified by SeqMan. SeqMan's SNP calling algorithm makes SNP calls with exceptional accuracy. You can also force contigs to join, or split a contig into two or more segments. Once you are satisfied with your assembly, you may save your project, export the data, or merge contigs with those imported from previous assembly projects.

7 Steps for Using SeqMan

1) Open an existing assembly or create a new SeqMan project.

    a) To open an existing SeqMan project or SeqMan NGen assembly, select **File > Open**.

    b) To open an existing Phrap or Sequencher assembly project, choose **File > Import**.

    c) To begin a new project, select **File > New**.

        i)   Add sequences to the project and set up the preassembly and assembly options, such as end trimming and scanning for vector. Projects may utilize any combination of [supported formats](#).

        ii) Review the project parameters and edit them, if desired.

        iii) Use the buttons at the top of the Unassembled Sequences window to prepare sequences for assembly. Preassembly options include the removal of vector sequence(s), removal of contaminant sequence(s), identification of repetitive sequence(s), automated 5' and 3' end trimming, and manual end setting. Removal of vector sequence(s) is strongly recommended when using Sanger data. We also recommend that you accept the default preassembly option allowing SeqMan to optimize the order in which it assembles sequences. To perform the preassembly processes without assembling the data, click **Options**, and then click **Scan All**. Otherwise proceed to the next step.

        iv) Click **Assemble** from the Unassembled Sequences window to begin the assembly. To assemble certain groups of sequences separately, use the **Assemble in Groups** option. Results will be displayed in both the Project Summary window ("*Untitled.sqd*") and in a new Report window. The Project Summary window shows constituent sequences in the bottom pane and contigs in the upper pane.

2) Use the Alignment View to examine detailed sequence alignments, traces, features, and individual consensus conflicts.

3) Use the Feature Table to view and manage all of the features within your current project. If desired, add features to the consensus sequence.

4) Use the Strategy View to obtain an overview of coverage, conflicts, paired end data consistency, and consensus features for each contig in your assembly.

5) View reports about your assembly, including the Structural Variation Report, the Project Statistics Report, the Trim Report, the Coverage Report, and Contig Details.

6) Utilize some of the many post-assembly options that SeqMan offers, including SNP discovery, primer walking, editing conflicts and gaps, and BLAST searching.

7) Select **File > Save** to save the results of your assembly project. You may also save various reports as separate files.

SeqMan Tutorials

This chapter contains five tutorials. The sections in each tutorial should be followed in order, as each is designed to walk you through a project from start to finish:

The data for each of these tutorials can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo SeqMan

The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Assembling trace data in SeqMan

Adding multiple sequences to a project

Objective: To create a new project and enter sequences for later assembly.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo SeqMan

1) Launch SeqMan. The Unassembled Sequences window and an empty Project Summary window will appear.

   Note: The Project Summary window will remain empty until after you assemble your sequences.

2) Click **Add Sequences** at the top of the Unassembled Sequences window. The Enter Sequences (Win) or Add to Project window (Mac) opens.

3) Locate and double-click the folder called **Demo SeqMan**. In the pane at the top of the dialog (Mac) or left of the dialog (Win, shown below), you should see a list of sequences with the names Sample 1...Sample 14. Each of the fragment sequences is a PE Applied Biosystems, Inc. automated sequencer trace file.
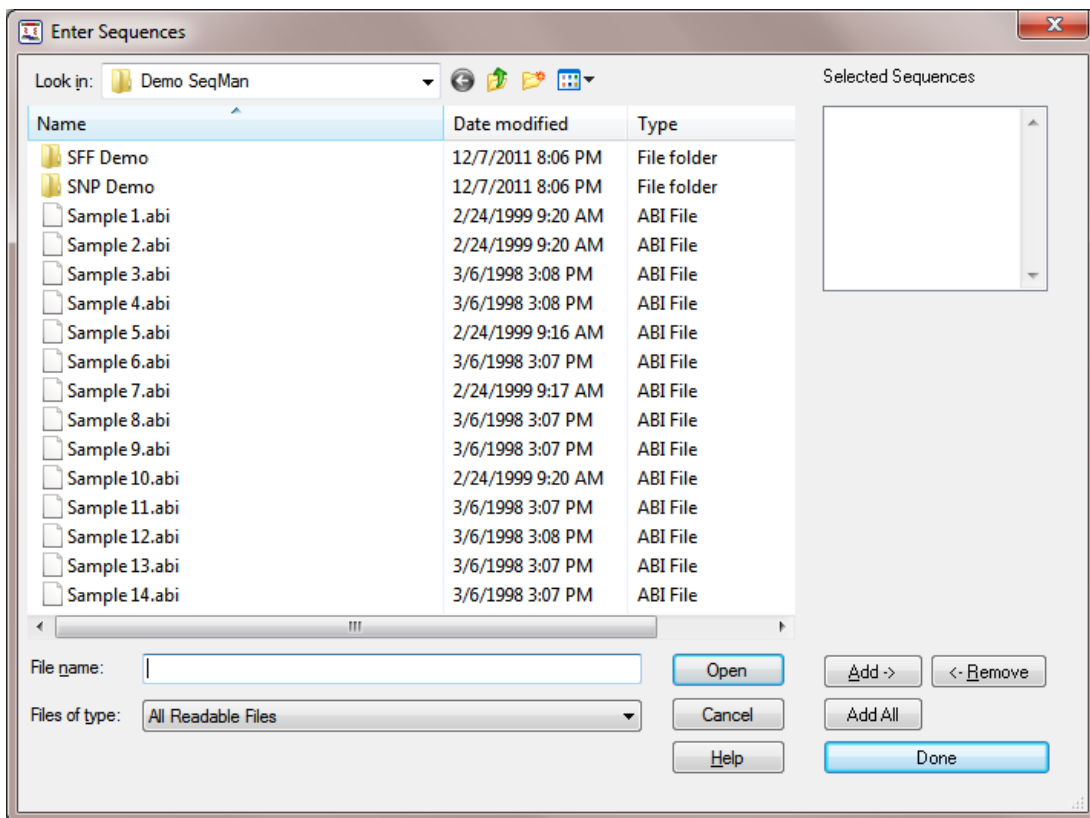
4) Add all 14 fragment sequences to the Selected Sequences (Win) or Chosen Files (Mac) pane by clicking **Add All** (Win) or **Add File** (Mac). Then click **Done**.

5) The Unassembled Sequences window 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.
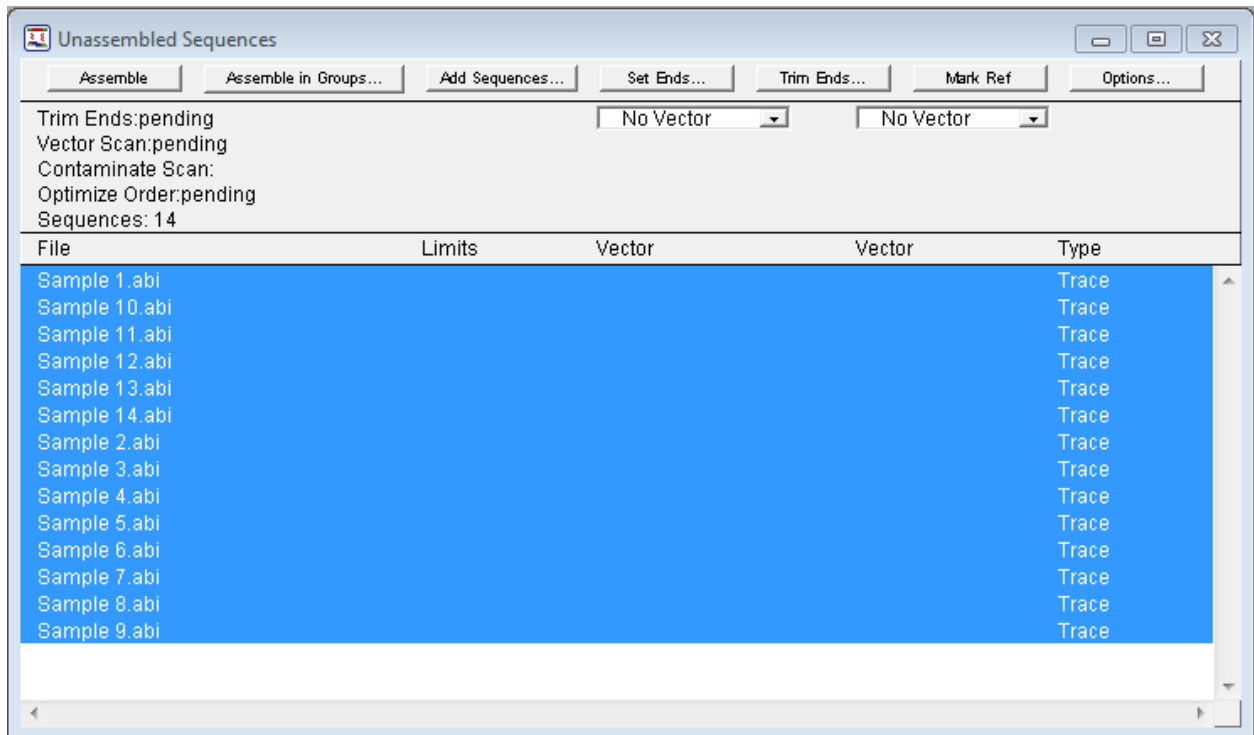
Proceed to the next tutorial.

Preparing sequences for assembly

Before assembly, it is recommended to implement pre-assembly options such as removal of vector and contaminant sequences.

Objective: To remove the Janus vector sequence from the samples and to trim sequence ends based on SeqMan's trace quality evaluation.

1) Make sure all fourteen of the sequences in the Unassembled Sequences window are highlighted by selecting **Edit > Select All.**

2) Specify the vectors to search for by choosing **Janus** from one of the vector lists and **InvJanus** from the other.

   Hint: If a vector does not appear on the list, you can add it to the vector catalog using **Project** > **Vector Catalog**.

3) Click **Options** to open the Assembly Options dialog.
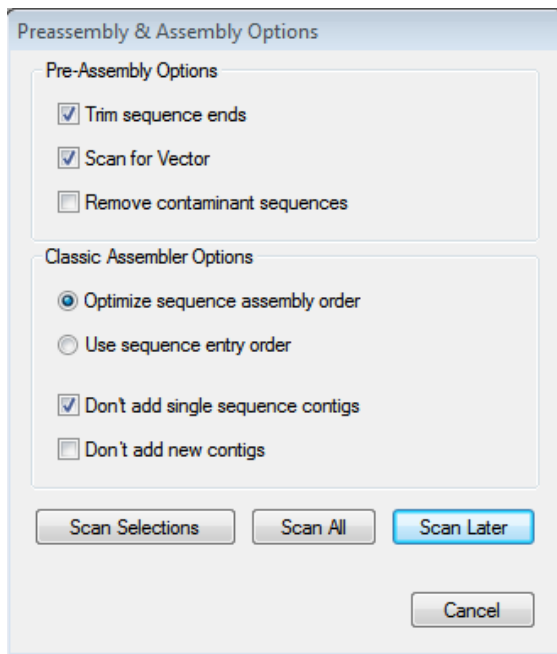
This dialog lets you choose whether to trim poor quality and vector sequence and whether to remove reads containing contaminating sequence. It also provides options related to SeqMan's "Classic" assembler.

4) Preserve the default settings and click **Scan All**. SeqMan will scan for the vector you selected, perform quality trimming on sequence ends and establish the optimal sequence assembly order.

In the Unassembled Sequences window, note that checkmarks now appear next to all of the Janus vector names in the Vector column, and upside down question marks appear next to the InvJanus names. Each check indicates that the vector was detected in that sequence. Upside down question marks indicate that the vector was not found in that sequence.

The Limits column displays the new sequence endpoints, reflecting both quality trimming and vector removal.

Proceed to the next tutorial.

Viewing trace data

Objectives: To view the Trim Report, then to examine the Trace Data view for one of the trimmed sequences.

1) Select **Project > Trim Report** to open a report showing current end-trimming parameters, followed by the name of each sequence, its average quality, amount trimmed, and both trimmed and pre-trimmed lengths. You can expand the Trim Report window by dragging its lower right corner.

2) Close the report when you are finished viewing it.

3) From the Unassembled Sequences window, double-click **Sample 8.abi** to open its chromatogram in a separate window, as shown below:
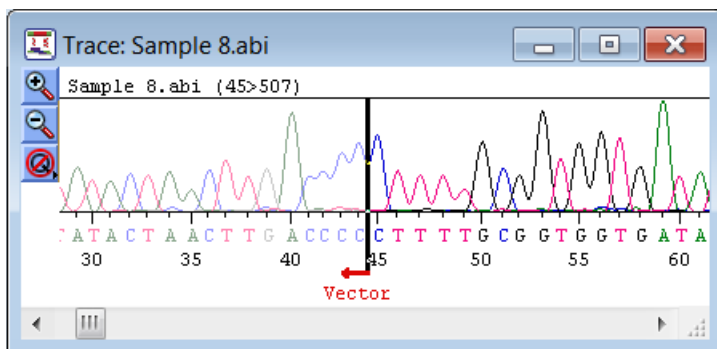
Data trimmed from the 5' end appears dimmed, denoting that it will not be included during sequence assembly. Note the good quality of most of the trimmed portion of the sequence. It was trimmed here because the sequence originates from the Janus vector. A vertical black bar appears at the intersection between trimmed and untrimmed sequence at base 45. It is possible to drag this bar left to unmask data that you feel should be included in the assembly, or right, to mask data that should not.

4) Select **Edit > Go To Position**, type **500** into the Position text box and click **OK**. Note that the peak quality falls in this region, and that beyond base 507, SeqMan determined that average peak quality falls below the medium stringency threshold. This data could be recovered, if desired, by dragging the bar at the end of the sequence.

5) Close the Trace Data window.

Proceed to the next tutorial.

Assembling and checking coverage

Objectives: This tutorial has two goals:

- To assemble the fourteen fragments.

- To use the Strategy View to examine the number of conflicts and degree of coverage for the resulting contig.

1) Click **Assemble** to assemble the 14 sequences into a single contig. The Assembly Report window appears.

2) Expand the Assembly Report window and scroll upward to review the data, which includes assembly parameters, sequences used, contigs created, preassembly time and total assembly time. When you are finished, close the window.

3) In the Project Summary window, click on "**Contig 1"** to select it and select **Contig > Strategy View** to open the following window.

   Note: Click only on the name "Contig 1" and not in the open box beside the name. Clicking on the box beside the name will change the icon to a padlock, indicating that the contig has been "locked" and preventing any further editing. To "unlock" a contig, simply click once on the padlock icon, or select **Contig > Unlock Contig**.
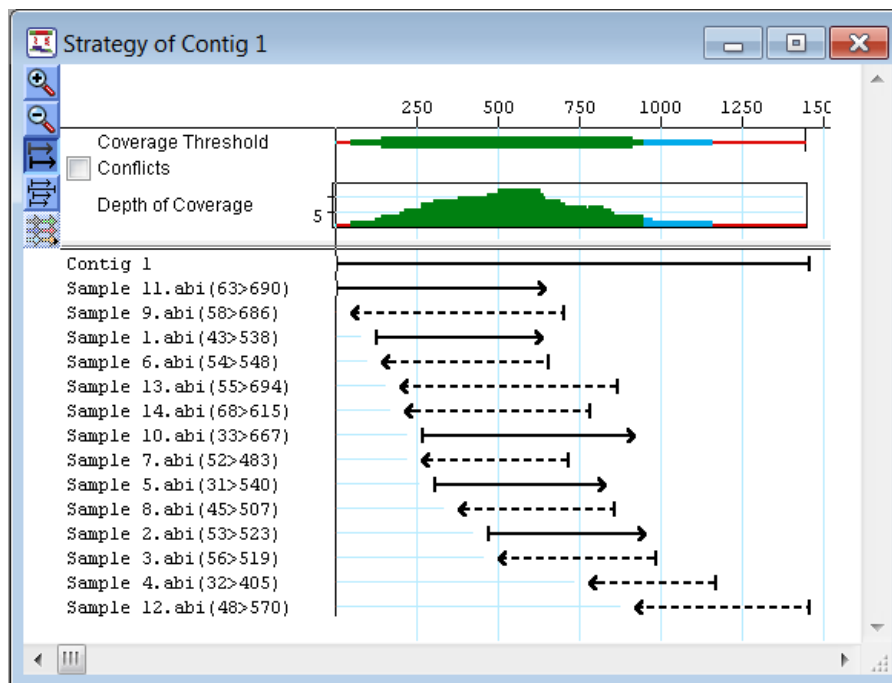
4) Enlarge the Strategy View by dragging its lower right corner down to view all of the samples.

Note the colorful Coverage Threshold graph that appears just below the ruler. The color and thickness of the band represent the coverage within the assembly as it compares to the threshold parameters defined under **Project > Parameters > Strategy Viewing & Coverage**.

The thick green band that appears between 200 and 900 denotes a region sequenced on both strands and meeting the Coverage Threshold criteria. The medium blue line to its right indicates a region sequenced in one direction only.

The thin red line appearing at each end of the contig means the region was sequenced only once.

5) Check the **Conflicts** box at the top left of the window to view a histogram showing agreement between fragment sequences. Notice that the conflict score plotted on the histogram is overall very low, indicating little or no conflict in most areas.

Proceed to the next tutorial.

Comparing consensus & restoring sequence ends

Prior to assembling the contig, SeqMan trimmed ends for the constituent sequences based on trace data quality and presence of 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 to join multiple contigs into a single contig. You may also wish to restore data in order to verify the consensus in a low-coverage area.

Objectives: To use the Alignment View to restore trimmed ends manually and to compare the results of two consensus calculation methods.

1) From the Strategy View, zoom in using until you can see and double-click on the ruler around base 50. This opens the Alignment View for that region.

The Alignment View displays the consensus sequence at the top of the window, with constituent sequences listed in the lower part.

2) Click on the **triangle** to the left of the word "Consensus" to compare two consensus sequences calculated using different methods. A line opens under the original consensus, and the names of the calculation methods are displayed. Any differences between the primary method (Trace) and comparison method (Majority) appear in red.

3) In the list of constituent sequences, note that **Sample 9** has a number of conflicts, indicated in red. View the underlying trace data for **Sample 9** by clicking on the **triangle** to the left of the sequence name.

4) Restore data to the Sample 9 sequence by dragging the **left trim bar** (the black triangle on the left end of the sequence) left to position **0** of the ruler. Note that previously trimmed sequence is easily recognized by its bright yellow background, visible as shading in the graphic below.
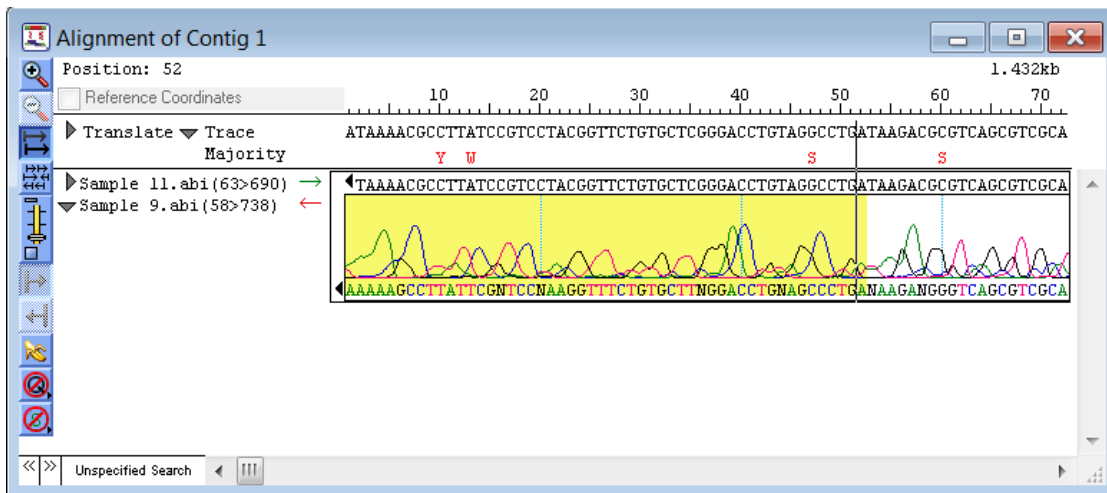
If you dragged to the correct position, the numbered range for Sample 9 will be (58>738), as in the image above.

5) Restoring trimmed sequence often reveals conflicts between the restored data and the consensus sequence. To view conflicts in the newly-restored sequence more easily, close the trace data by once again clicking on the dark **triangle** to the left of **Sample 9**.

The red letters appearing in the leftmost portion of the **Sample 9** sequence signify conflicts between the restored data and the consensus. In this case, sequence had originally been trimmed due to inferior peak quality. It is sometimes beneficial to restore such data. On the other hand, data removed due to vector contamination would have been characterized by normal peak quality in combination with a high number of conflicts. Regardless of its deceptively high peak quality, it is not recommended that you restore vector data.

Before continuing, you may wish to view an example of why the Trace consensus call is often superior to the simple Majority consensus. At **position 47**, the Trace consensus call remains **G** even though the **Sample 9** call is **C**. That's because the Trace method judges the peak quality, while the Majority method (which calls **S**) does not get the correct consensus here.

6) Open the trace data for **Sample 9** again and slide the **trim bars** toward the central sequence again until the trimmed (highlighted) data are no longer visible.

7) Click on the Project Summary window and select **File > Close** to close the current project.

8) Click **No** (Win) or **Don't Save** (Mac) when prompted to save the changes.

Assembling sequences in groups

Objective: To assemble the forward and reverse paired end reads for 3 different samples, each into its own contig, but against the same reference sequence.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo SeqMan

1) Launch **SeqMan**.

2) Click the **Add Sequences** button from the Unassembled Sequences window.
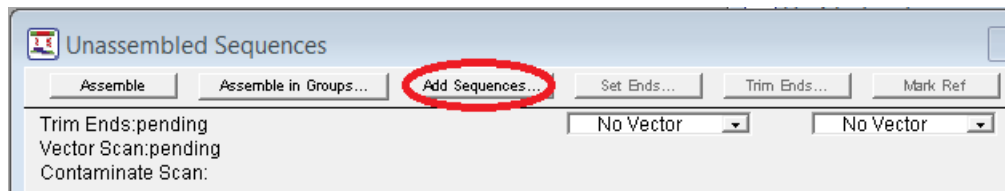
3) Locate and open the folder **Demo SeqMan**.

4) Double-click on the folder **SNP Demo**.

- **Windows:** Click **Add All** to add the seven sequences to the Selected Sequences pane.

- **Macintosh:** Click **Add Folder** to add the SNP Demo folder to the Chosen Files pane.

5) Press **Done**.

6) In the Unassembled Sequences window, select the sequence named *Reference.seq*, and click the **Mark Ref** button. The sequence is selected as the reference sequence and its name is shown in italics.

7) Click on the **Assemble in Groups** button from the top of the Unassembled Sequences window. The **Assemble in Groups** panel will be displayed in the bottom half of the window:
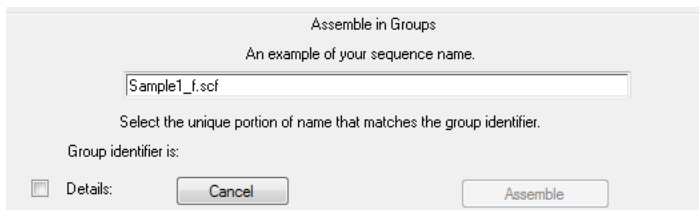
8) One of the SNP demo sequences will automatically be displayed in the field provided so that you can specify the portion of the name that SeqMan should use as a group identifier. Highlight **Sample1** so that SeqMan will use that portion of the sequence name to identify the sequence as part of a group.
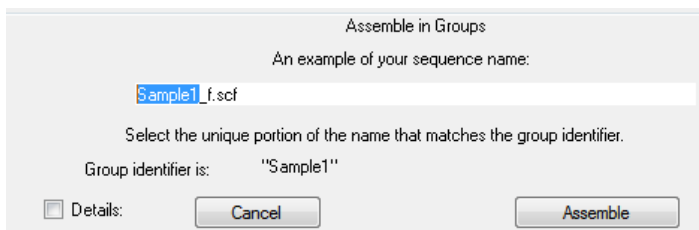
Since the SNP demo sequences have names like Sample1_f.scf, Sample1_r.scf, Sample2_f.scf, Sample2_r.scf, etc., specifying **Sample1** as the group identifier will ensure that the forward and reverse reads for the **Sample1** sequence will be assembled together; the forward and reverse reads for the **Sample2** sequence would be assembled together; and so on.

9) Click the **Assemble** button from the bottom of the Assemble in Groups panel to begin your assembly.
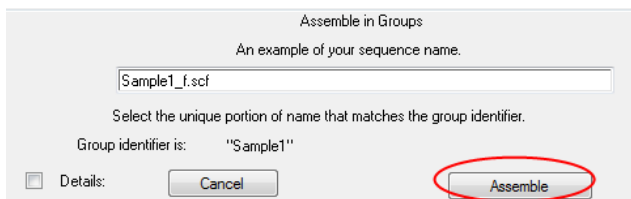
10) Once the assembly has finished, notice that in the Project Summary window, three contigs have been created. Each contig is named for its group identifier and contains three sequences: the forward and reverse reads for the sample, and the reference sequence we specified:

11) Close all open SeqMan windows using **File > Close**. When prompted to save the document, click **No** (Win) or **Don't Save** (Mac).

Designing sequencing primers and improving the coverage

The Primer Walking feature in SeqMan identifies optimal primer locations for driving the closure of gaps or for filling in areas of low coverage.

Designing sequencing primers to improve coverage

Objective: To design sequencing primers for locations surrounding areas of low coverage in an assembly.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo SeqMan

1) Launch SeqMan, if it is not already open.

2) Select **Sequence > Add**.

3) Locate and open the **Demo SeqMan** folder.

4) Do one of the following:

    - **Windows**: **Ctrl + click** on **Sample 1.abi** and **Sample 6.abi**, then click **Add**.

    - **Macintosh:** Double-click on **Sample 1.abi** and **Sample 6.abi**.

5) When both sequence names appear in the Selected Sequences pane (Win) or Chosen Files pane (Mac), click **Done**.

6) Click **Assemble** to create a contig. When the assembly is complete, the Report window appears along with the Untitled window listing the contig, the length, number of sequences and position.

7) Close the Report window.

8) In the Project Summary window, click on "**Contig 1"** to select it.
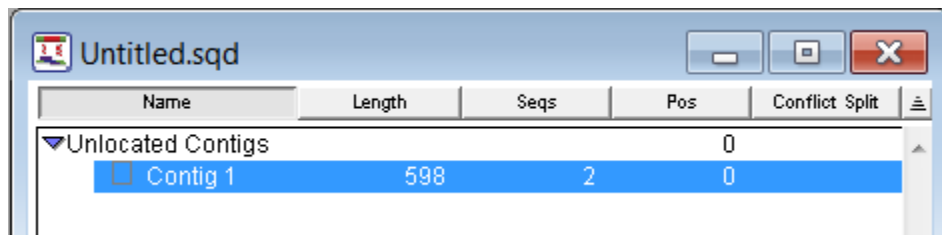
Note: Click only on the name "Contig 1" and not in the open box beside the name. Clicking on the box beside the name will change the icon to a padlock, indicating that the contig has been "locked" and preventing any further editing. To "unlock" a contig, simply click once on the padlock icon, or select **Contig > Unlock Contig**.

9) Display the Strategy View by selecting **Contig > Strategy View**.
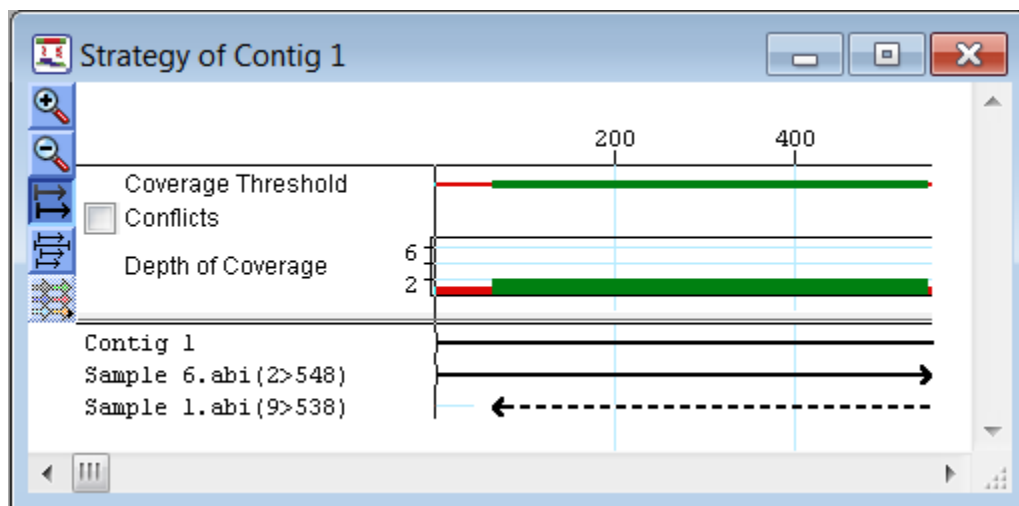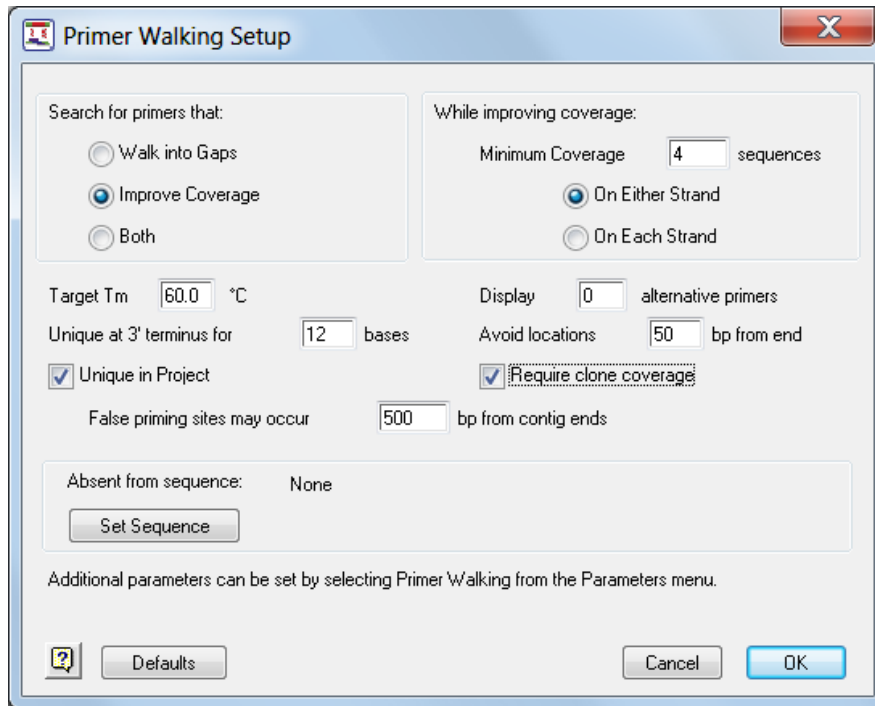
Note: If your samples are not in the order shown in the above image, choose **Contig > Complement Contig** before proceeding with this tutorial.

Observe that the two trace files that make up this contig are from upper and lower strands and overlap at all but the ends of the contig. The overlaps are green.

Click the bottom right corner of the Strategy View window and drag it to create a larger viewing window. If desired, use the zoom in tool to increase scale of the Strategy View.

10) Select **Contig > Primer Walk…** to view the Primer Walking setup parameters.

11) Click the **Improve Coverage** radio button to select a search for coverage primers.

12) Click the **Require Clone Coverage** checkbox to eliminate primers selected in areas that have no sequence coverage on the same strand as the primer.
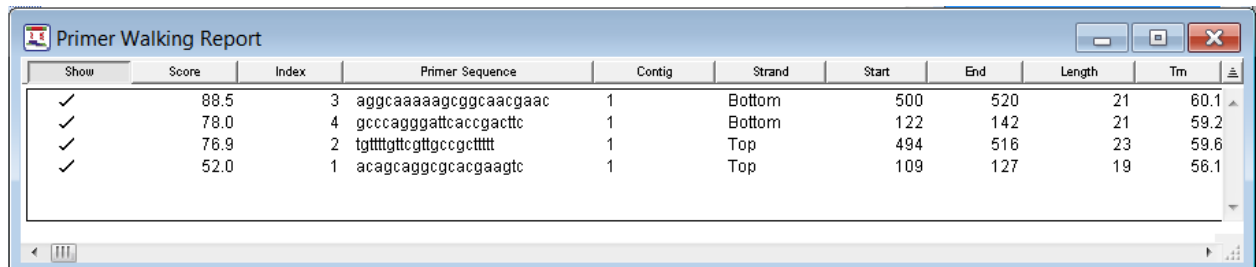
13) Click **OK** to locate sequencing primers. The Primer Walking Report appears.

14) Double-click the top-scoring primer from the Primer Walking Report to open the Alignment view displaying the primer you selected.
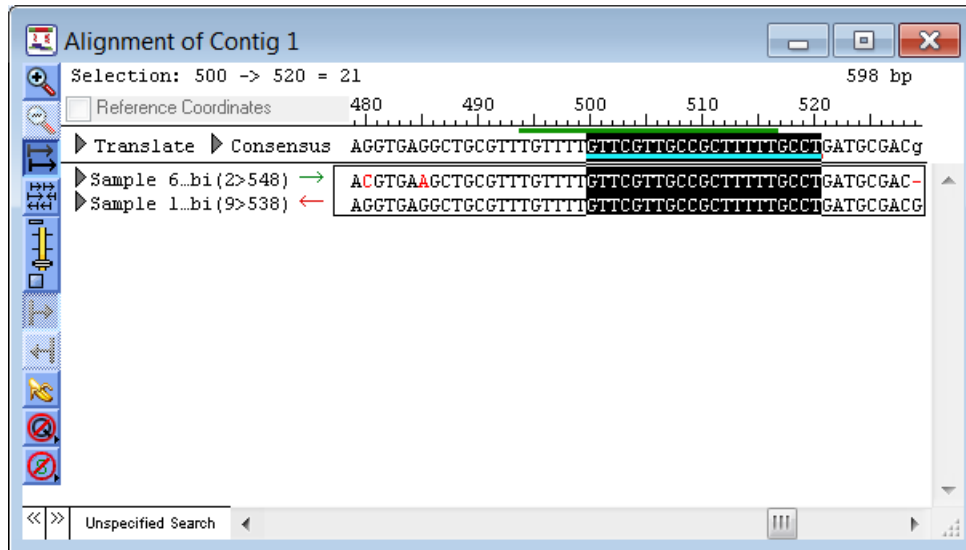
15) Select **Contig > Strategy View**. The Strategy view window shows a thin red horizontal line under the Coverage Threshold graph around base pair 100 corresponding to the bottom strand sequencing primer.
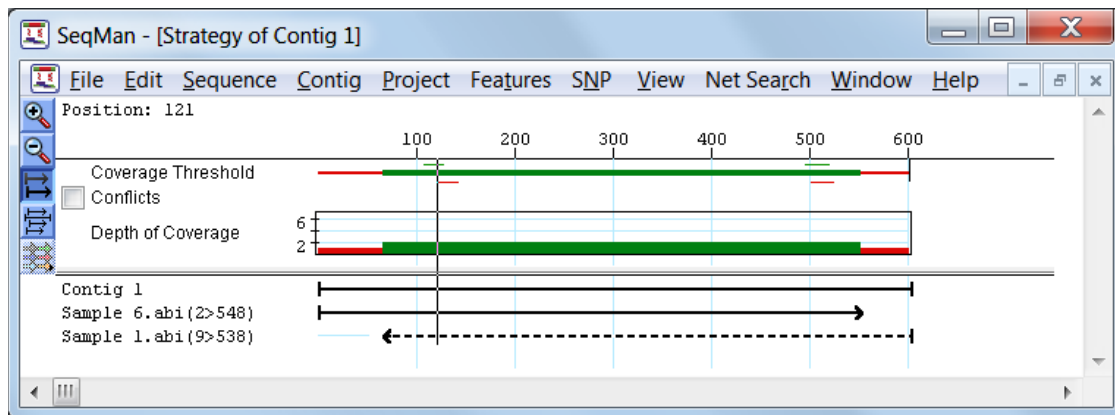
16) Double click on the primer shown around base pair 100 to open the Alignment view displaying Contig 1.
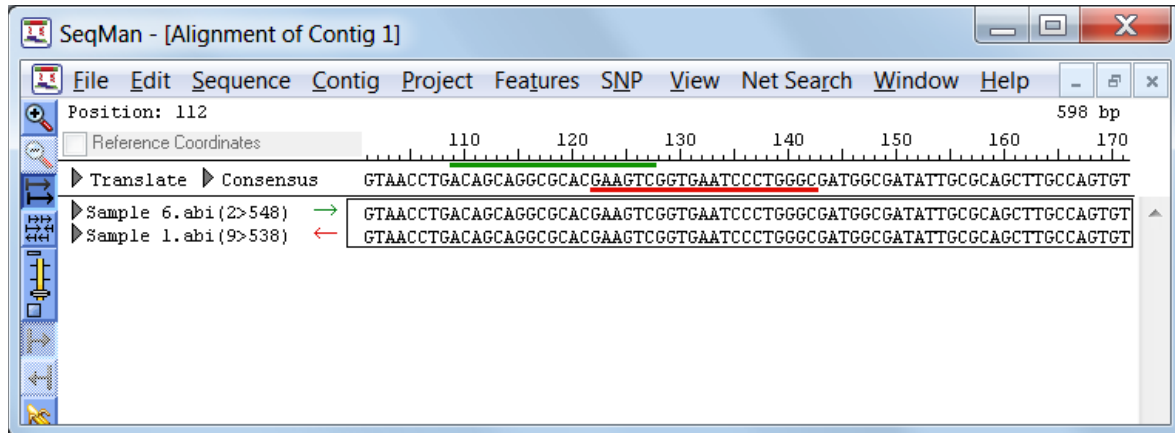
17) Click on the **Primer Walking Report** to bring it into view.

18) Select **File > Save Primer Info**.

19) Choose a location and filename. Make sure **Tab-Separated** is selected from the **Save As Type** (Win) or **Format** (Mac) pull down menu to save all of the primers within the report as a tab-separated text file.

20) Click **Save**.

21) Click on the Project Summary window and select **File > Exit** to close the current project and SeqMan.

22) Click **Don't Save** (Mac) or **No** (Win) when prompted to save changes.

Discovering SNPs

Aligning sample sequences

Objective: To align sample sequences to a reference sequence.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo SeqMan

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo SeqMan

1) Launch SeqMan, if it is not already open.

2) Select **Sequence > Add**.

3) Locate and open the folder **Demo SeqMan**.

4) Double-click on the folder **SNP Demo**.

   - **Windows:** Click **Add All** to add the seven sequences to the Selected Sequences pane.

   - **Macintosh:** Click **Add Folder** to add the SNP Demo folder to the Chosen Files pane.

5) Press **Done**.

6) In the Unassembled Sequences window, select the sequence named **Reference.seq**, and click the **Mark Ref** button. The sequence is selected as the reference sequence and its name is shown in italics.

Note: A reference sequence is a known sequence used for comparison with your sample sequences. In SNP discovery, SeqMan will determine where your sample sequences differ from the reference sequence.

7) Select **Project > Parameters > Pair Specifier**. The Pair Specifier dialog appears.

8) Select **sample_f.abi < > sample_r.abi** from the Pair Pattern drop-down menu.

9) In the Verify Pattern section, type in **Sample1_f.scf** in the Forward field and **Sample1_r.scf** in the Reverse field. The Success message indicates that the example names match the pair pattern you selected.
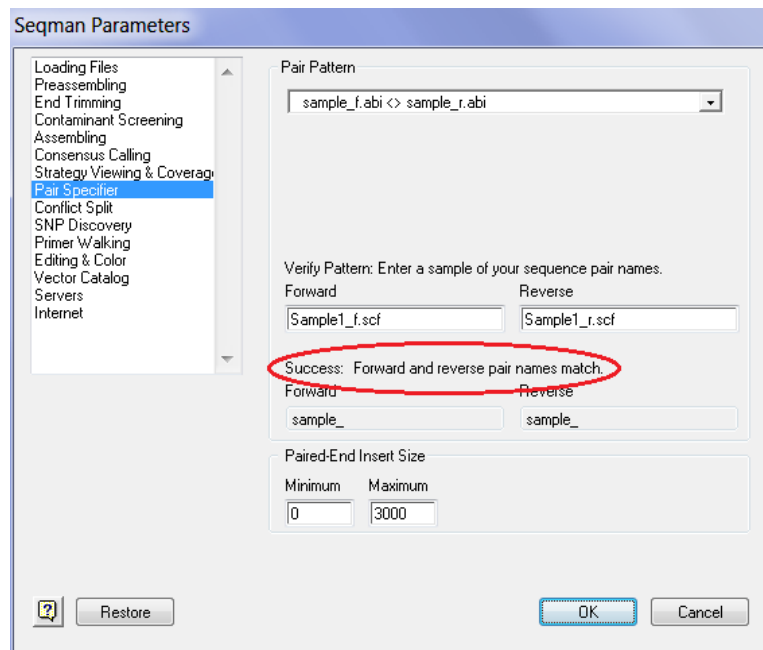
10) Click **OK** to save the changes and close the SeqMan Parameters dialog.

11) Click **Assemble**. After assembly is complete, close the Report window. The sample sequences align into a single contig with the reference sequence.

Note: After assembly, whenever the reference sequence is displayed, it appears at the top of the lower pane of the Project Summary window with its name in italics.

Please proceed to the next tutorial.

Reviewing SNPs

Objective: To review SNPs using the **Strategy** view, **Alignment** view, and **SNP Report**.

1) From the top pane of the Untitled window, click **Contig 1**, and then select **Contig > Strategy View** to display an overview of the contig. Click the **Zoom In** button and drag the lower right corner of the window to enlarge the view.

2) Select **SNP > SNP Report**.

3) Click on the Strategy View to bring it back into view:
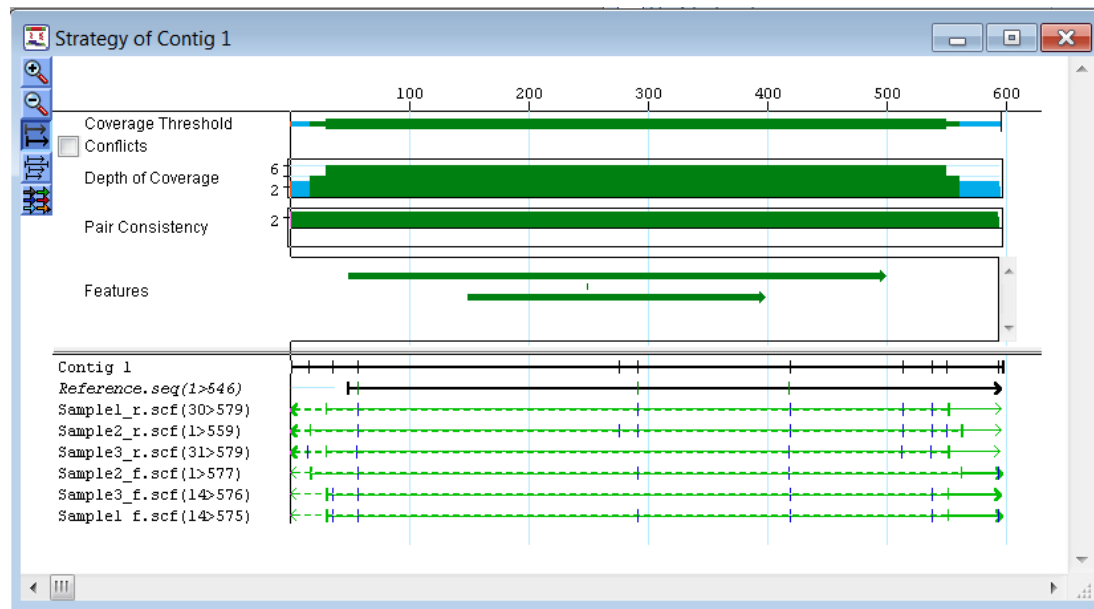
When you look at this view, notice the following:

- Bold arrows represent sample sequences. A light arrow overlapping a sample sequence arrow represents the sequence that is paired with the sample.

- Locations of bases in the contig that correspond to SNP base locations in the aligned sample sequences are indicated with hash marks on the contig goalpost.

- Locations of bases annotated as features of the type "variation" in the *Reference.seq* file are indicated with green hash marks on the reference sequence arrow.

- Locations of bases in sample sequences identified as putative SNPs by SeqMan are indicated with blue hash marks on the sequence arrows.

- Locations of bases in sample sequences identified as rejected SNPs by SeqMan are indicated with red hash marks on the sequence arrows.

4) Select **Contig > Alignment View**, and drag the lower right corner of the window to enlarge the view, so that it looks like the following window.
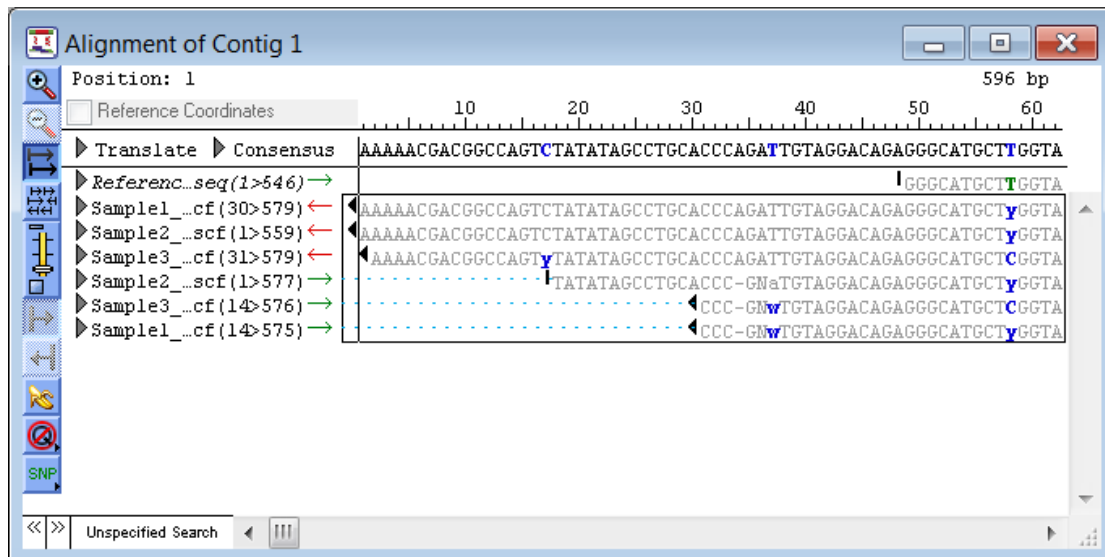
When you look at this view, notice the following:

- Bases annotated as features of type "variation" in the reference sequence file are displayed in green. The 'T' in position 58 is an example.

- Bases in sample sequences identified as putative SNPs by SeqMan are displayed in blue throughout the view.

- A base in a sample sequence that SeqMan has determined is a heterozygous putative SNP is displayed using the appropriate ambiguity code.

5) Select **SNP > SNP Report**. When you look at the window, notice the following:

- Tabs on the SNP Report allow you to change the display between two different SNP reports: **All Found SNPs** and **SNPs Summary**.

- A row in the **All Found SNPs** report lists information for a single SNP base in a sample sequence.

- A row in the **SNPs Summary** report summarizes information for all of the SNP bases in an aligned column.

Note: For further analysis, information in the SNP Report window can be copied and pasted into other documents such as MS Excel® spreadsheets. The SNP report can also be saved in tab-separated file format by selecting **File > Save SNP Report for 'Contig 1.'**

6) Select **SNP > Sort by SNP**.

7) In either the **All Found SNPs** or the **SNPs Summary** report, select a **row** and then press the **up** and **down** arrow keys to select other rows. As you do so, notice that the corresponding SNPs are selected in the Alignment View, allowing you to quickly scan through the SNP bases.

Because you selected **SNP > Sort by SNPs**, sample sequences in the Alignment View are reordered automatically according to the SNP bases in the selected column. Sample sequences with SNP bases in the aligned column are sorted to the top and samples with non-SNP bases are sorted to the bottom.

Note: A sample sequence in the Alignment View can be manually reordered by clicking on its name and dragging it up or down. Selecting **Contig > Sort > by Offset** or **Contig > Sort > by Name** automatically reorders sample sequences by offset or name.

Proceed to the next tutorial.

Analyzing putative SNP bases

Objective: To analyze putative bases and determine if you want to confirm or reject them as actual SNPs.

1) Select **Project > Parameters > SNP Discovery** to view the parameters that SeqMan uses to identify SNPs.
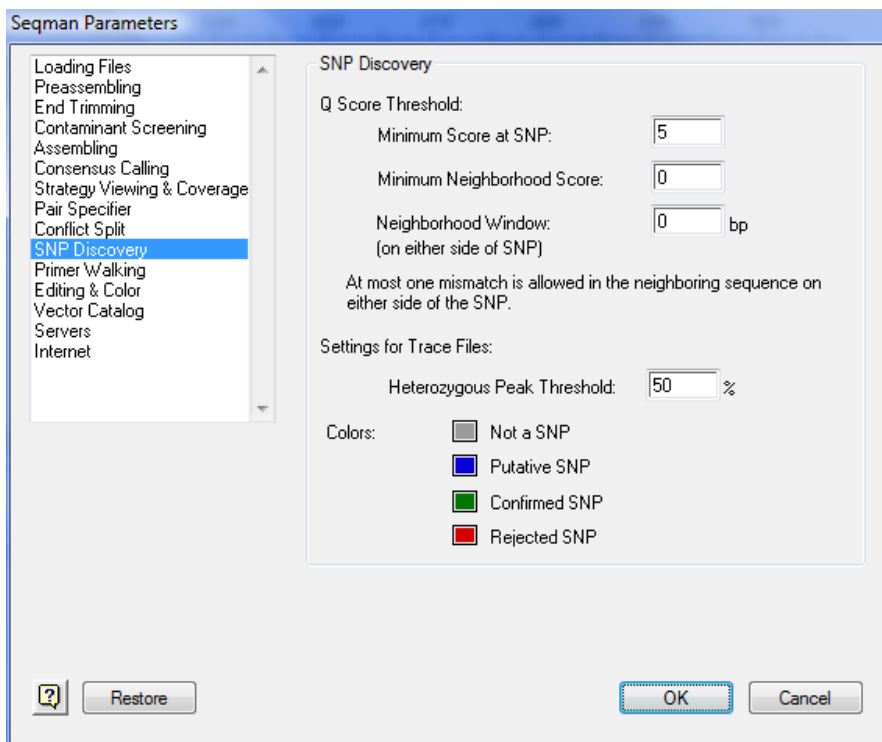
2) Notice the following default settings:

- **Minimum Score at SNP:** This parameter prevents bases with a quality score of less than the threshold value from being shown in the SNP Reports.

- **Minimum Neighborhood Score:** This parameter prevents SNPs with a *neighborhood quality score* of less than the specified value from being displayed. Note that if the specified **Neighborhood Window** (see below) contains one or more mismatches to the reference sequence, the putative SNP will be rejected, even if it meets the **Minimum Neighborhood Score** threshold.

- **Neighborhood Window:** This value specifies the size of the window to be used when calculating the *neighborhood quality score*. For example, if this value is set to 5, then the 5 bases upstream and the 5 bases downstream from the SNP base will be considered.

- **Heterozygous Peak Threshold**: By default, any base that exhibits a secondary trace peak that is at least 50% of the intensity of the primary peak will be identified as heterozygous.

3) Click **OK** to close the SeqMan Parameters dialog without making any changes.

4) In the SNP Report window, click on the **All Found SNPs** tab and look at the first column, which is labeled **SNP**. A question mark in this column indicates that the SNP base is putative, a checkmark that it is confirmed, and an 'x' indicates that it is rejected as an actual SNP.
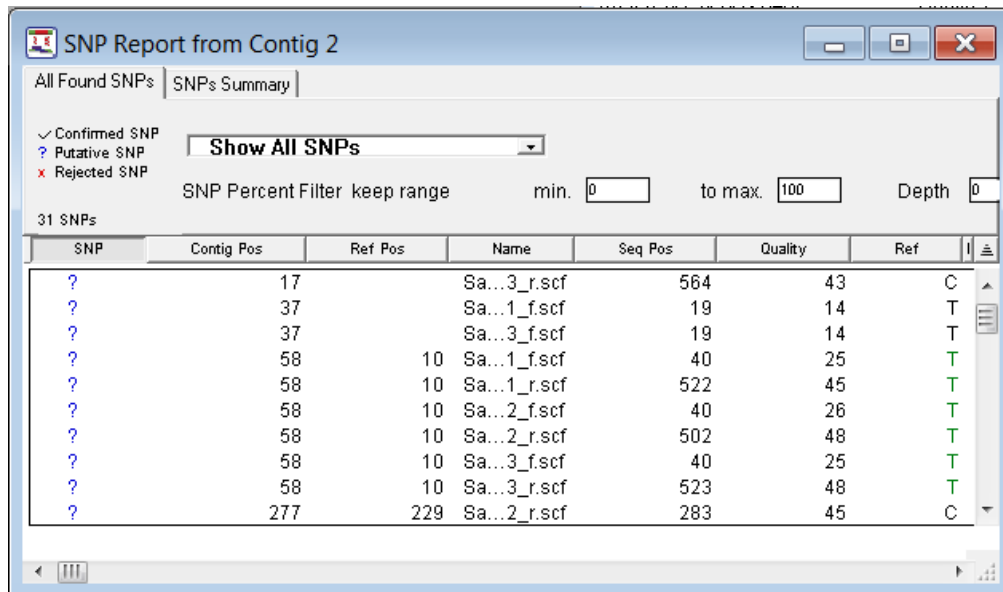
5) In the SNP Report window, click the **SNPs Summary** tab.
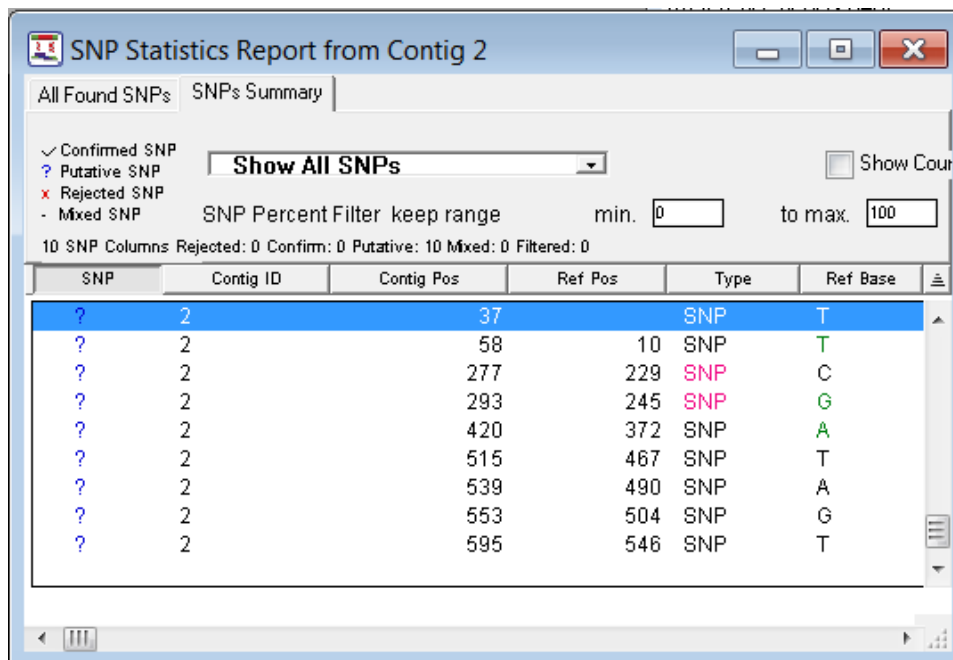
6) Look at the **SNP** column in the **SNPs Summary** report. The question marks in this column indicate putative SNP bases.

7) Locate the row in the SNP Summary report for contig position 58.

8) Double click on the contig position in the row. The Alignment View is displayed with the column at position 58 selected.

9) Display the traces for all of the sequences in the Alignment View by holding down the **Alt/Option** key (Win/Mac) while clicking on any of the triangles next to a sequence name.
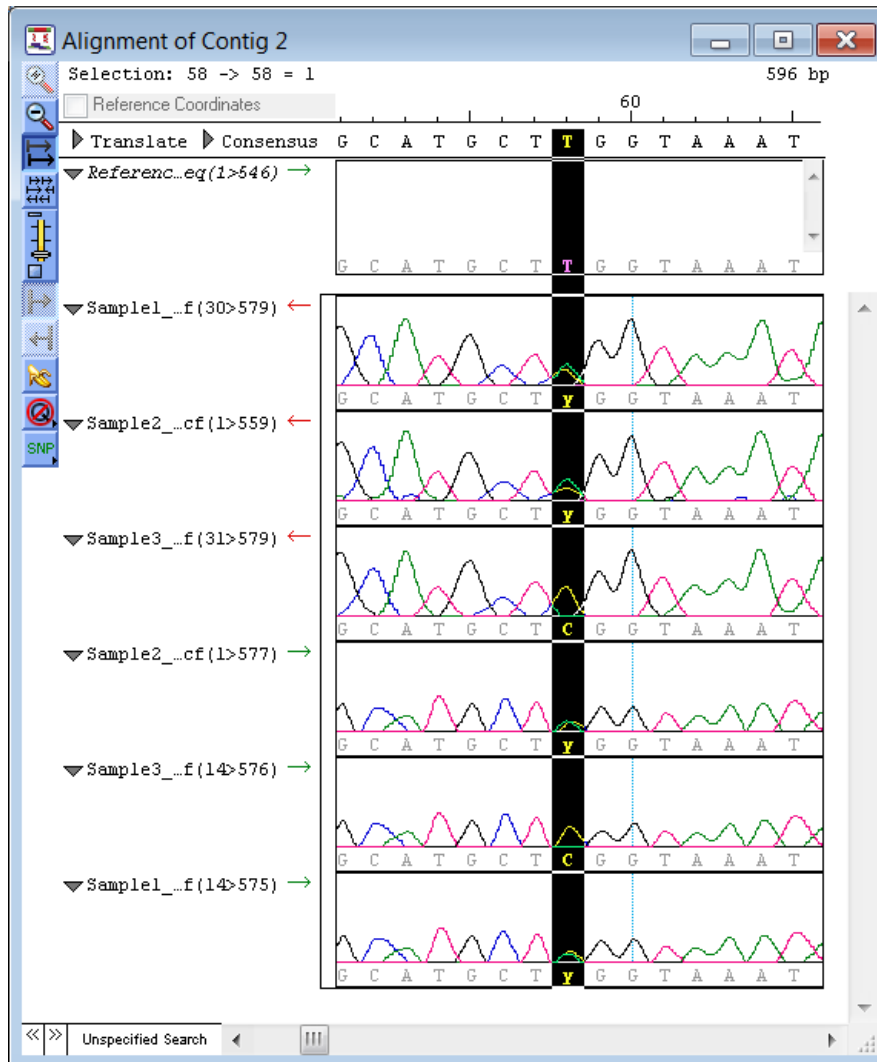
10) Examine the trace data for the bases aligned in position 58. Notice that 4 sequences show a "y" (the ambiguity code for C or T). Upon examining the trace data for these sequences, we can see that all 4 sequences show peaks for both C and T at position 58. The forward and reverse sequences for **Sample 3** clearly show a C peak. Comparing this information to the reference sequence, which is a T at position 58, we can find that the putative SNP bases in column 58 can be confirmed as actual SNPs.

11) In the SNPs Summary report, click on the row for contig position 58 to highlight it, then select **SNP > Confirm SNP**. The mark changes to a checkmark to indicate that you are confirming that the putative SNP bases in the column are actual SNPs. In the Alignment View, the color of the SNP bases in the column changes to green. This is the default color for confirmed SNP bases.

12) Click on the Project Summary window and select **File > Close**. When prompted to save the document, click **Don't Save** (Mac) or **No** (Win).

Note: The default colors for display of putative, confirmed, and rejected SNP bases, as well as non-SNP bases, can be changed in the dialog opened by selecting **Project > Parameters > SNP Discovery.**

Further Exploration

Working with features

SeqMan enables you to display, edit, and create new features for the sequences in your project. Features are read from *.seq and *.sbd files, as well as existing SeqMan projects. (*.sqd). You can also import features for your sequences through BLAST searching.

- Any sequences in your project that are not already associated with trace or flowgram data can be annotated with features within the Alignment View. To create a new feature, first select the range for the new feature on the sequence where you wish to create it, and then select **Features > New Feature**.

- The Feature Table allows you to view and manage all of the features within your current project. Display the Feature Table by selecting **Features > Show Feature Table**. To display the Feature Table for the consensus sequence, first open the Strategy View by selecting **Contig > Strategy View**, and then select **Features > Show Consensus Feature Table**.

- View features in the Alignment View by clicking the triangle shown to the left of the sequence name.

- Features that have been added to the consensus sequence are displayed in the top section of the Strategy View beneath the Depth of Coverage Graph.

- To edit a feature, first select it by clicking on it in the Alignment View or in the Feature Table. Then, select **Features > Edit Feature**. Alternately, double-click on the feature you wish to edit from the Alignment View or Strategy View.

- To collect features for sequences in your project through BLAST searching, first select the range of sequence you wish to BLAST from the Alignment View. Then, select **Net Search > BLAST selection**. After the search is complete, click the **Collect Features** button at the top of your BLAST results window. (Note that the **Collect Features** button will not be available if the sequence you selected to BLAST is already associated with trace or flowgram data).

- When exporting the consensus sequence to SeqBuilder by using **File > Send Consensus To > SeqBuilder**, SeqMan automatically annotates the sequence with features for areas of high and low coverage, as well as confirmed SNPs.

Importing Phrap, Newbler, and Sequencher data

Phrap, Newbler, or SeqMan NGen assemblies with the *.ace extension and Sequencher assemblies with the *.spf extension may be opened in SeqMan using **File > Import**.

The Alignment View for an imported Phrap project displays Phred base calls and Phrap consensus calls. By default, the top consensus (next to the word *Translate*) is the Imported Phrap consensus. Just below this is the SeqMan *Trace* or *Majority* consensus. These consensus only show differences from the Phrap consensus based on the original Phrap assembly.

Using the vector catalog

SeqMan contains a catalog for a number of common vectors. To add, delete, or modify entries in the catalog, select **Project** > **Vector Catalog.**

Handling repetitive and contaminant sequences

To remove contaminant sequences, use the following procedure.

1) Store sequence files of known contaminants in the **Contaminant Seqs** folder located:

   - **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data

   - **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data

   - **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data

2) Prior to sequence assembly, select **Project > Contaminant Seqs** and select which to screen for from the list of available files.

3) From the Unassembled Sequences window, click **Options** and check **Remove Contaminant Sequences**.

To add known repetitive sequences last to an assembly, use the following method.

1) Store sequence files of known repetitive sequences in the **Repetitive Seqs** folder located:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data

2) Prior to sequence assembly, select **Project > Repetitive Sequences** and select which files you wish to screen for in the current project.

3) From the Unassembled Sequences window, click **Options** and check **Optimize Sequence Assembly Order**.

Note: Optimizing assembly order is a parameter under the Classic assembler only, and is recommended for most assemblies using the Classic assembler whether or not repetitive sequences are specified.

Negating weights

To cancel the internal weight for a selected portion of a sequence so that it does not contribute to the consensus, highlight the region of sequence in the Alignment View and select **Edit** > **Negate Weights**. This command is particularly useful when building the assembly around a "backbone" sequence from a closely related species. Use **Edit** > **Restore Weights** to restore.

Exporting data

- To save an assembly project, select **File** > **Save**.

- To export all the constituent sequences for one or more contigs, select **Contig** > **Export Sequences**.

- To export the consensus sequence for one or more contigs, select **Contig** > **Save Consensus**.

- To export your assembly as a [GenVision](#) project, select **File > Export as GenVision Project** from the Strategy View.

- To export assembly projects as a Phrap project (*.ace), select **File** > **Export Project**.

SeqNinja

Overview

A component of DNASTAR's [Lasergene Core Suite](), SeqNinja is a utility for advanced editing of genome sequences and annotations.

SeqNinja allows you to edit an annotated sequence and examine the consequence of mutations, facilitating the process of adjusting sequence coordinates between assemblies. Other uses for SeqNinja include file format conversion; batch editing, export and translation of annotations; and merging (concatenation) and splitting of sequence data.

Each SeqNinja program consists of a user-created program, a list of operations that SeqNinja will execute in the order in which they are written.

- <u>Beginning</u> users can easily create a program in the Visual View by clicking on a pre-defined Template and completing the required fields.

- <u>Intermediate</u> users can create a program in the Visual View by "stacking" Toolkit items.

- <u>Advanced</u> users may instead choose to type operations manually into the Text View. The Text View has some functionality that is not covered by the Templates and Toolkit.

Once the program has been created, simply click on the Run button (Run) to run the program. The progress of the run can be monitored in the Console. At the conclusion of the run, the output will be saved in the file and format that was specified in the program, and the Run Results dialog will appear.

7 Quick-Start Steps for Using SeqNinja

The following steps demonstrate how to design a new program in the Visual View using SeqNinja's Templates and Toolkit.

1) Launch SeqNinja. Notice that the Templates and Toolkit panels appear on the right. The Visual View, on the left, initially displays the SeqNinja Template Gallery, a list of shortcuts to items in the Templates panel.

2) Choose a template from the SeqNinja Template Gallery or from the Templates panel. The Visual View will open with a copy of the template loaded as a new program.

To do this…	… choose this template:
Create a new, empty SeqNinja (*.star) file.	Blank Program
Create the reverse complement of one or more sequences in a file.	Reverse Complement Sequence(s)
Combine two or more sequences into a single, new sequence. (e.g. merge contigs into a pseudo-molecule with a linker).	Concatenate Sequences
Reset the origin of a circular sequence.	Change Sequence Origin
Extract a set of annotated features from a genome and save them as separate sequences.	Extract Features as Sequences
Translate one or more DNA sequences into protein.	Translate Sequence(s)
Create a modified version of a sequence.	Modify Sequence(s)
Modify a sequence by inserting new sequence characters.	Create Insertion
Convert a sequence from one type to another. (See our list of supported file formats).	Convert File Type

3) Enter the required information. For example, you may be asked to choose a sequence, specify output format and location, or to select a sequence range

4) If desired, add additional steps and/or operations to the program using items from the Templates and Toolkit panels. These items can be placed in any order, and some types of operations can be nested inside other operations or steps using drag & drop.

5) Enter the required information for the newly-added steps.

6) If desired, save the SeqNinja program as a *.star file.

7) Run the program. SeqNinja will execute the entire program from top to bottom, outputting the results in the file(s) you chose earlier. At the conclusion of the run, the output will be saved in the file(s) and format(s) specified in the program, and the Run Results dialog will appear. This dialog contains links to the folder(s) containing the output file(s).

SeqNinja Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through several SeqNinja projects from start to finish. Some tutorials end with an optional "further exploration" section highlighting additional features.

Setting Up Preferences

In this tutorial, you will use the Preferences dialogs to select a default file location, and to specify how SeqNinja should handle "file save" prompts.

1) Launch SeqNinja.

To select a default directory for file input and output:

2) Choose **Edit > Preferences** (Win) or **SeqNinja > Preferences** (Mac) and select **Default Run Options** from the list at left.

3) In the Default Run Options dialog, click the **Browse** button and navigate to the *Demo Data* (not *Demo SeqNinja*) directory for your operating system:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

4) Click **OK** (Win) or **Open** (Mac) to confirm the selected path.

5) After returning to the Default Run Options dialog, click **Apply**. Do not click **OK** at this point.

Until otherwise directed, SeqNinja will now assume that any input file is located in the Demo Data folder, and will send any output to the same directory.

6) Next, select **Editor** from the list at left.

7) In the Editor dialog locate the checkbox next to **Ask to save before running any program**. By default, this box is checked, causing SeqNinja to prompt you to save each unsaved program before running it. You may choose whether to leave the box checked or unchecked.

Note: In subsequent tutorials, you will be asked to save each program manually before running. Therefore, both options will allow you to follow the tutorials as written.

8) Click **Apply** and then **OK**.

Please proceed to the next tutorial.

Inserting Bases into an Existing Sequence

In this tutorial, you will use a template to insert three missing bases ("AAA") between positions 24 and 25 of the sequence *pbr322.seq*.

1) Choose **Create Insertion** from the New Program window, on the left, or from the Templates panel, on the right.

   Tips for entering your data appear in the yellow "comments" section above the template.

2) In the Sequence File area, press the **Choose Sequence(s)** button.

3) From the *Demo Data* folder (specified as the default data folder in *Setting Up Preferences*), double-click on *pbr322.seq*.

4) In the "Type sequence" area, enter the three missing bases. To do this, type *AAA* over the sample bases in the **Type sequence** row. You may use small or capital letters, but using capitals in this tutorial will allow you to more easily see the changes in the output file.

5) Note the two sets of range boxes above and below the row where you just typed the missing bases.

- The <u>upper range boxes</u> need to show the sequence segment that occurs *before* the insertion point. In other words, "from the left end of the sequence until base 24." In the left-hand box, ensure that one of the following is entered: "1," "Left end," or "lend" (an abbreviation for "left end). In the right-hand box, type in the number *24*.

- The <u>lower range boxes</u> need to show the sequence segment *after* the insertion point. In other words, "from 25 until the right end of the sequence." In the lower pair of range boxes, type in 25 on the left. Acceptable choices for the right box are "Right end" or "rend." If you knew the length of the sequence, you could also enter that number.

6) To the right of the **Save Results As** button, type in *pbr322-ammended*. This output file will automatically be saved in the default data folder, *Demo Data*.

7) In the **Format** drop-down menu, choose the same format as the source file, **Genbank**.

8) Save the finished SeqNinja program using the **File > Save** command. Navigate to the *Demo Data* folder (the path is provided in Step 3 of *Setting Up Preferences*). In the **File name** box, type *Insertion.star*. Then click **Save**.

The final program (below the comment) should now appear as follows:
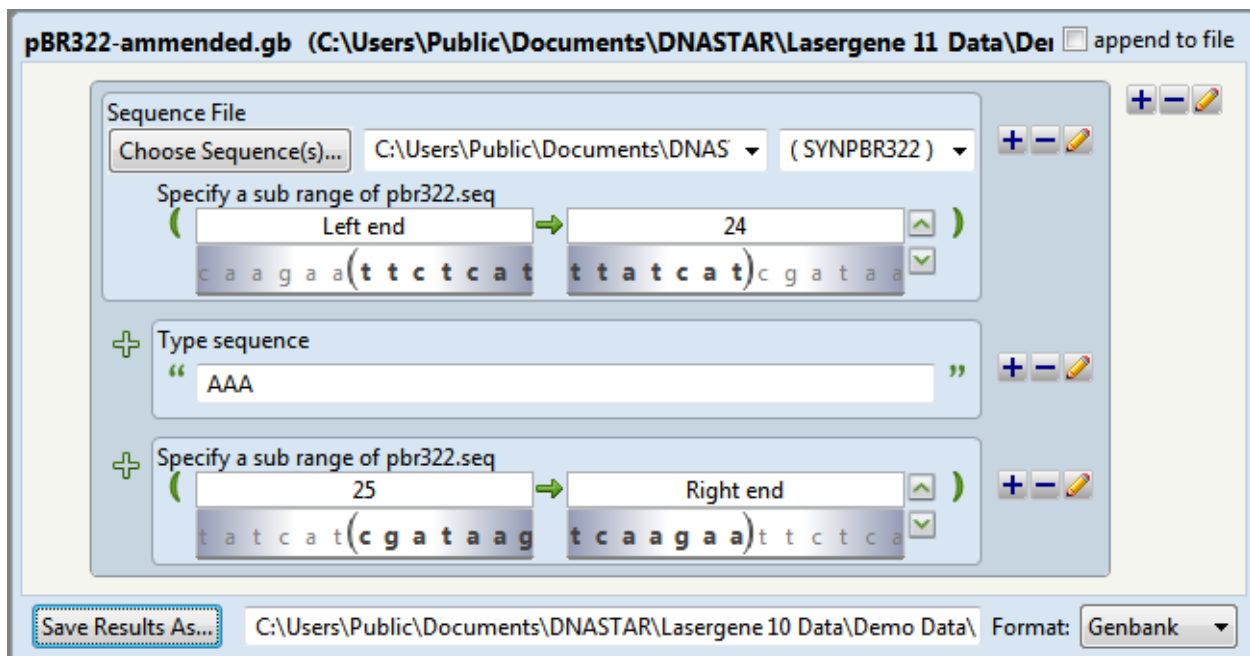
You are now ready to run the program. Please proceed to the next tutorial.

Running a Program in SeqNinja

This tutorial will describe several ways to run a SeqNinja program. When you "run" a program, you tell SeqNinja to execute all of the steps in order from top to bottom. The results are then output into the data file and format that you specified in the program.

There are several ways to initiate a run.

- Choosing **Run > Run** from the menu or clicking the **Run** tool (▶ Run) initiates a run without allowing you to choose any options.

- Selecting **Run > Run with Options** instead launches the following dialog.
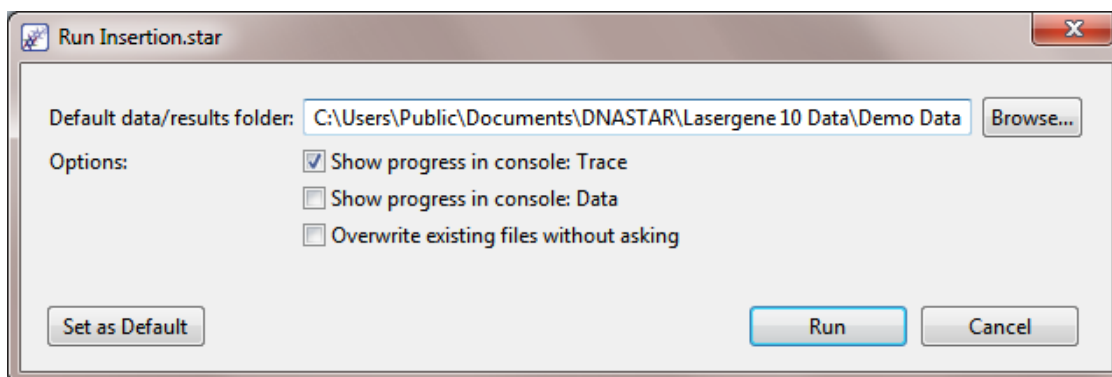
This dialog allows you to over-ride some of the global preferences from the **Edit > Preferences** dialogs, but for this program only. A useful feature in the Run with Options dialog is the **Overwrite existing files without asking** checkbox. You may check or uncheck this box to hide/show warnings about over-writing an existing data file.

1) For this tutorial, simply press the Run tool (▶ Run). Alternatively, if you opened the Run with Options dialog, keep the settings shown in the image above and click the **Run** button.

   When the run is complete, the Run Completed dialog is displayed:
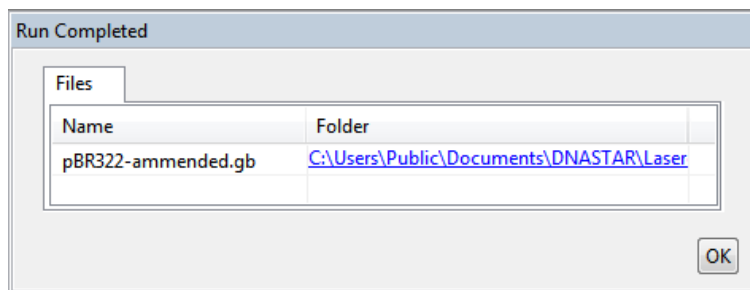
   If the run was successful, there will be a single tab, called **Files**. The **Files** tab contains a table listing the names of any output files, with links to their directories.

Note: A second tab, entitled **Warnings**, may appear after some runs. This tab provides additional information about the run, including warnings about errors (e.g., syntax errors) or about the possibility of unexpected results. For example, if you attempt to reverse complement a protein, the **Warnings** tab will advise you that the sequence being reverse complemented appears to be a protein.
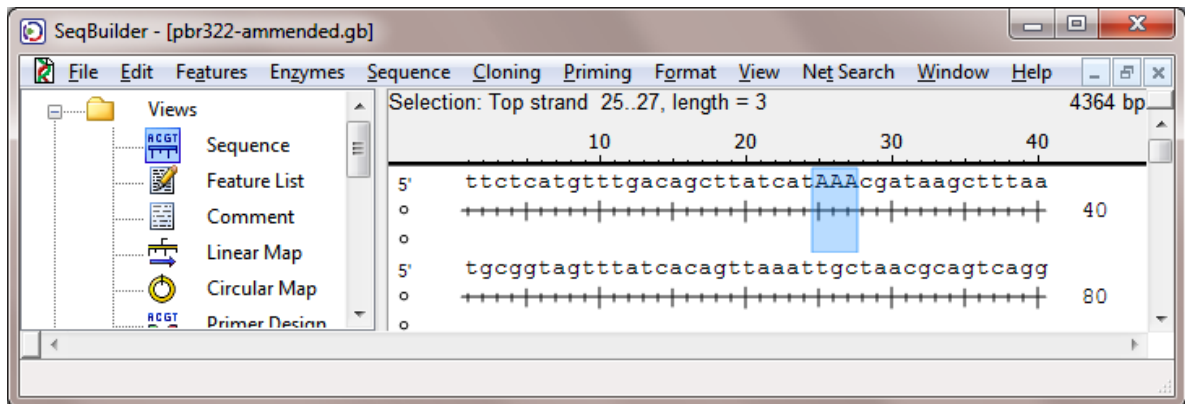
2) In the **Files** tab, click the link to go to the *Demo Data* folder.

3) Open the output file (*pBR322-ammended.gb*) in DNASTAR's SeqBuilder application or in any text editor. Note that the three 'A' bases appear, as specified, between base 24 and the base that was formerly base 25.

4) Close SeqBuilder or the text editor and return to SeqNinja.

Please proceed to the next tutorial.

Defining a Sequence Variable

In SeqNinja, a variable can be used to represent one or more files, variables or functions; or a string of text. This tutorial shows how to define a set of three sequence files (*LAMCG.seq*, *pbr322.seq* and *tethis21.seq*) as the variable "A." After doing this, you can simply enter the letter "A" whenever you need to represent those files later in the program. This variable will be used in the next tutorial, *Extracting Features as Sequences and Translating Them*.

Note: The three sequence files used in this tutorial were chosen solely for convenience, and are scientifically unrelated.

1) From the **Templates** panel, choose **Blank Program**.

2) Click on the **Toolkit** tab on the right of the SeqNinja window.

3) From the **Sequences** bar, select **Define Sequence Variable** to add the following step to the blank program:
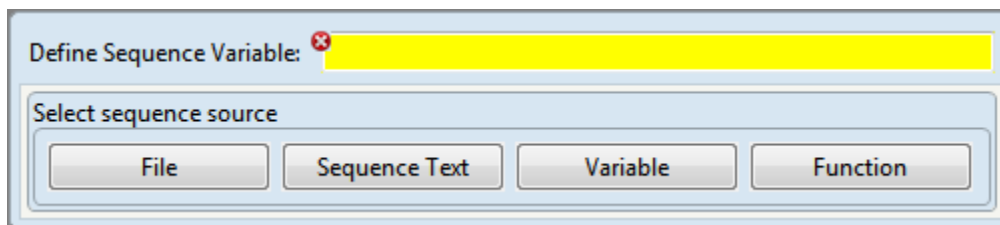
4) Enter the letter "*A*" (without quotation marks) in the text box to the right of **Define Sequence Variable**.

5) Since the variable will be defined as being equal to three sequence files, select **File** as the Sequence Source. The lower half of the Define Sequence Variable step changes into a file chooser.
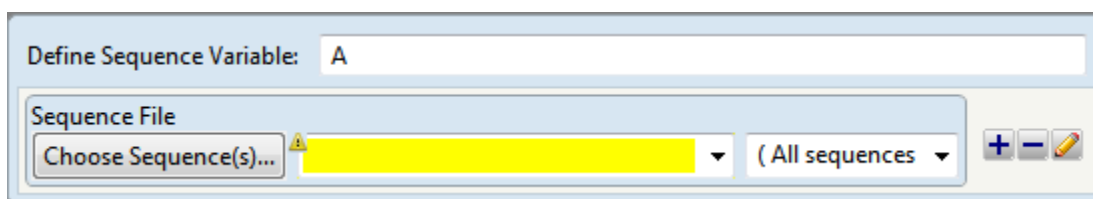
6) To choose the first sequence file, click the **Choose Sequence(s)** button. From the *Demo Data* folder, double-click on *LAMCG.seq*.

7) To add the next sequence, click the blue plus icon (✚) and select **Add Sequence File**.
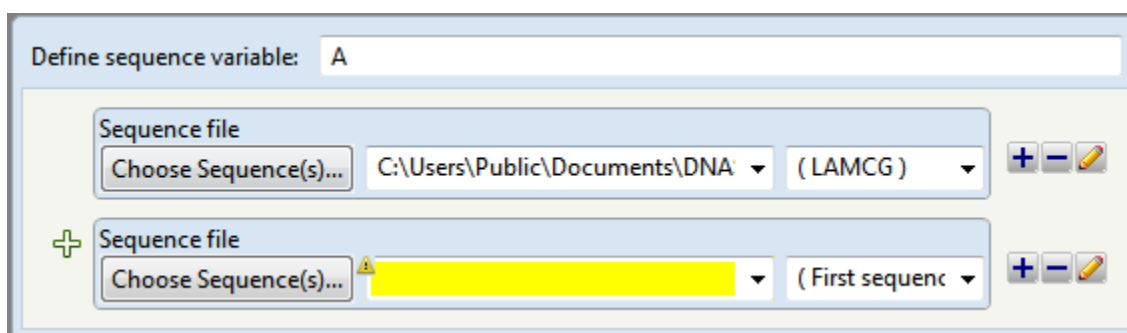
8) Press the new **Choose Sequence(s)** button**,** and then double-click on the next file, *pbr322.seq*.

9) Repeat Steps 8-9 to add the final file, *tethis21.seq*. The Define Sequence Variable step should now appear as follows:
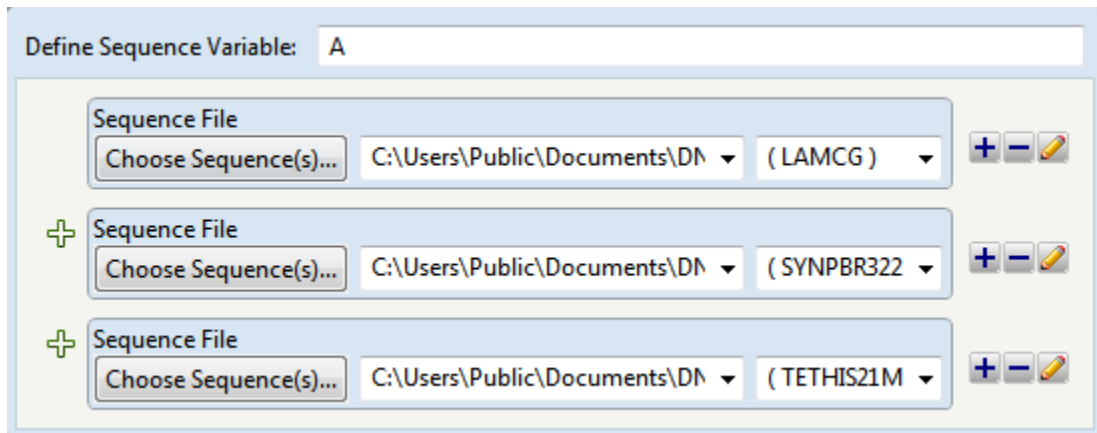
Please proceed to the next tutorial.

Extracting Features as Sequences and Translating Them

This tutorial is a continuation of the program begun in *Defining a Sequence Variable*. In the current tutorial, you will perform the following procedures:

- Extract the genes from the files represented by variable "A" and save them to a single output file.

- Translate the output sequence and save the results as a second output file.

To do these steps, you will use a combination of items from the Templates and Toolkit panels.

To add the extraction step:

1) Select the **Templates** tab on the right, then drag **Extract Features as Sequences** onto the Visual View and drop it below the Define Sequence Variable step.

Note: Be sure to use drag & drop instead of clicking on the template. If you click the template name, a new program will open instead.

The program now appears as below:
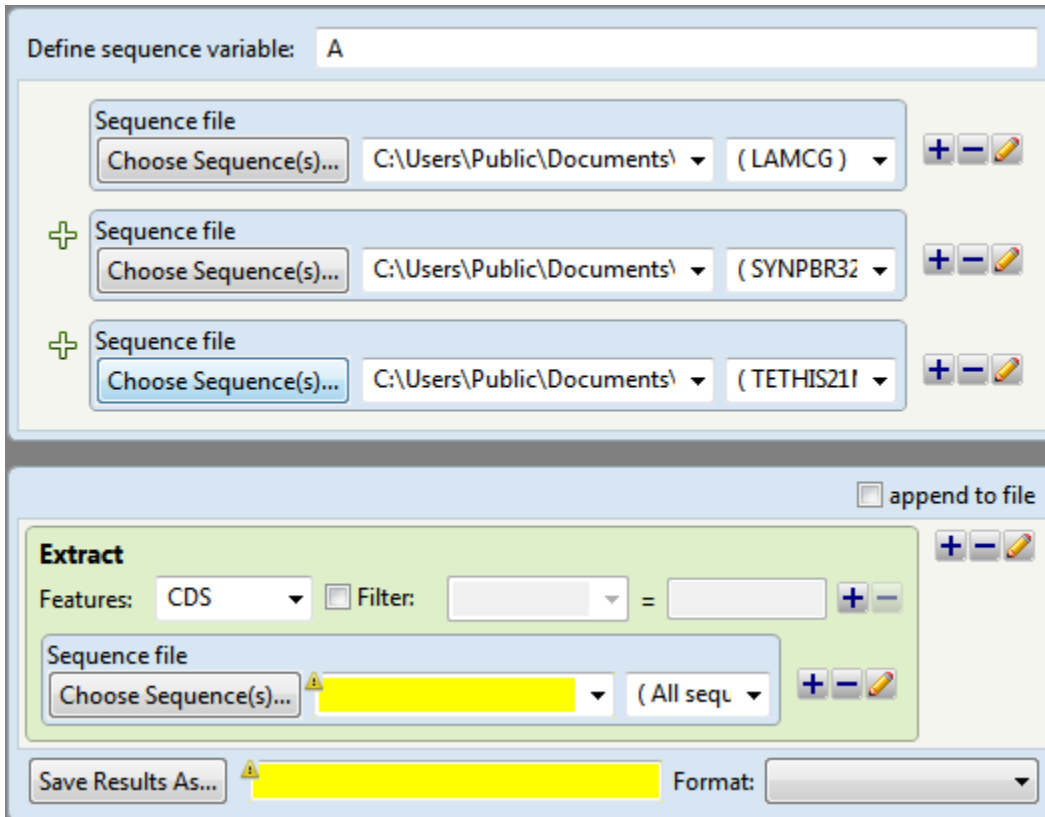
2) In the Extract step, choose **gene** from the Features drop-down menu.

Note that the next row, Sequence File, contains the standard **Choose Sequence(s)** button. However, it doesn't have an option for typing in your variable.

3) Since you want to enter the sequence variable ("*A*") to represent the input files, right-click on the words "Sequence File" and choose **Delete**. The row disappears and is replaced by a row of buttons. Choose **Variable**, then enter the letter "*A*."
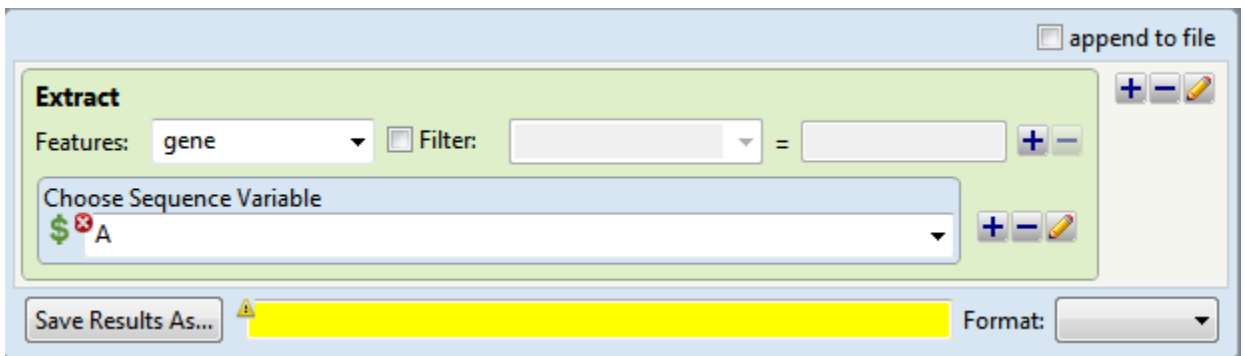
4) To the right of the **Save Results As** button, type in the filename *Extracted genes*.

5) From the **Format** drop-down menu, choose **Fasta**.

To add the translation step:

6) Drag the **Translate Sequence(s)** from the Templates panel and drop it under the existing steps.
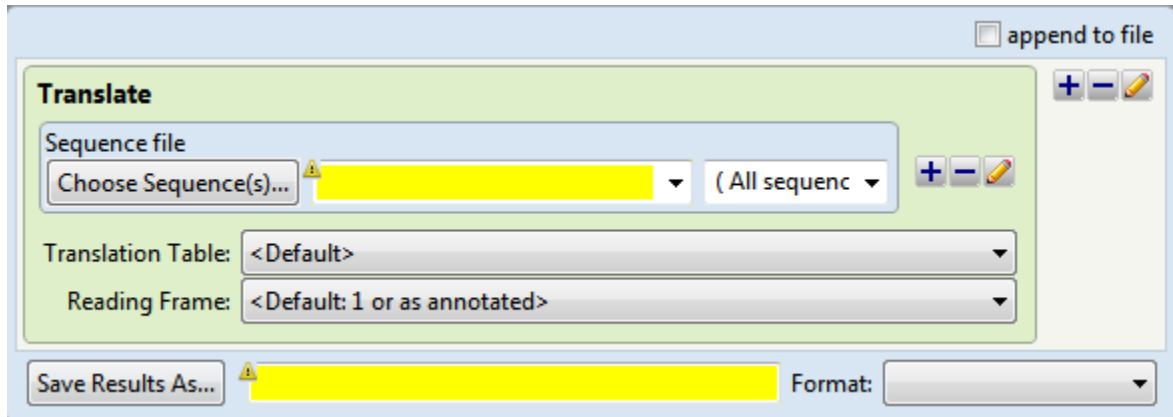
7) The input file for the translation step is the same as the output file of the extraction step. Use commands from the **Edit** menu (or standard keyboard shortcuts) to select and copy the filename from the bottom row of the Extract step and paste it into the first row of the Translate step.

8) From the **Translation Table** drop-down menu, choose **NCBI: 11**. Leave the **Reading Frame** drop-down menu at the default setting.

9) To the right of the **Save Results As** button, type the name *Translated genes*. This output file will automatically be saved in the default data folder, *Demo Data*.

10) From the **Type** drop-down menu, choose **Fasta**. The program now appears as follows:

11) To save the program, select **File > Save** and enter the name *Extract-translate.star*. Click **Save**.

12) Run the program by clicking the Run tool (Run).

When the run finishes, the Run Completed dialog will appear:

13) Click on either link to open the Demo Data folder.

14) To view the translation results, open *Translated genes.fasta* in DNASTAR's Protean 3D or in any text editor. The table below shows the genes from the three files that were located, extracted and translated.
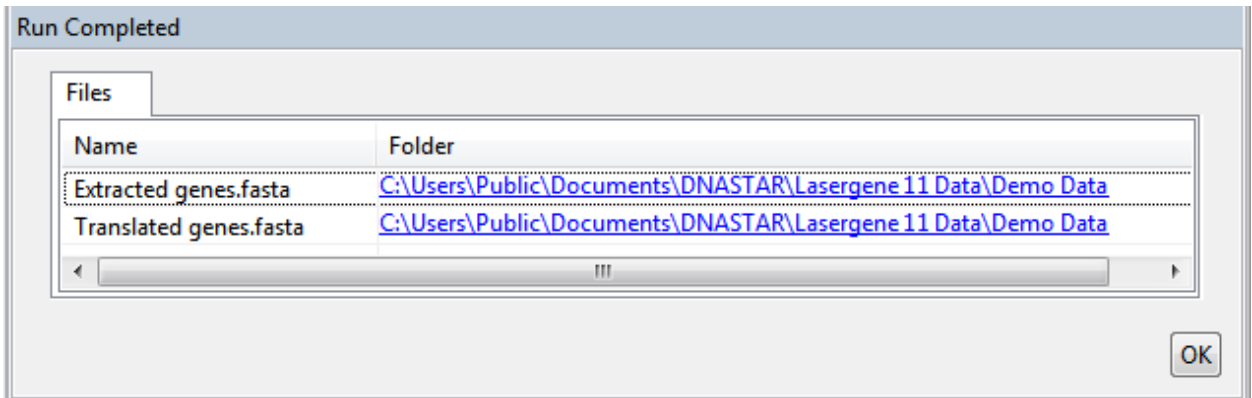
Filename	Extracted and translated genes
LAMCG.seq	*int*
	exo
pbr322.seq	*tet*
	bla
tethis21.seq	(none found)

For further exploration:

A) As you can see from the table, there were no annotated genes in the *tethis21.seq* file. However, all three files have annotated CDS features. Select **CDS** from the **Features** drop-down menu (in the Extract step) then click **Run** again to see how this affects the output files.

Note: None of the output files for this tutorial will be needed later. Therefore, you may allow SeqNinja to overwrite the earlier output files.

B) When translating a *.gbk or *.seq file that contains CDS features, the extraction step is redundant. The reason for this is that, during translation, SeqNinja automatically translates CDS features rather than the entire sequence. To see a more efficient method of extracting CDS features than was used in Part A:

1) Open the SeqNinja Text View by clicking the **Text** tab near the bottom left corner of the SeqNinja window or by choosing **View > Text View**.

The Text View contains the same information as the Visual View, but in a more concise format. With experience, you may be able to write a program more quickly in

the Text View than the Visual View. The Text View is also more efficient than the Visual View when adding large amounts of data, such as lists of sequence sub-ranges, or when copying and pasting, as in our next step.

2) Replace the existing text by pasting the following over it:

```
$A = "LAMCG.seq"+"pbr322.seq"+"tethis21.seq"
"translated genes.fasta" =
translate($A,'/transl_table=11')
```

3) To view the revised program in the Visual View, select the **Visual** tab or use the **View > Visual View** command.
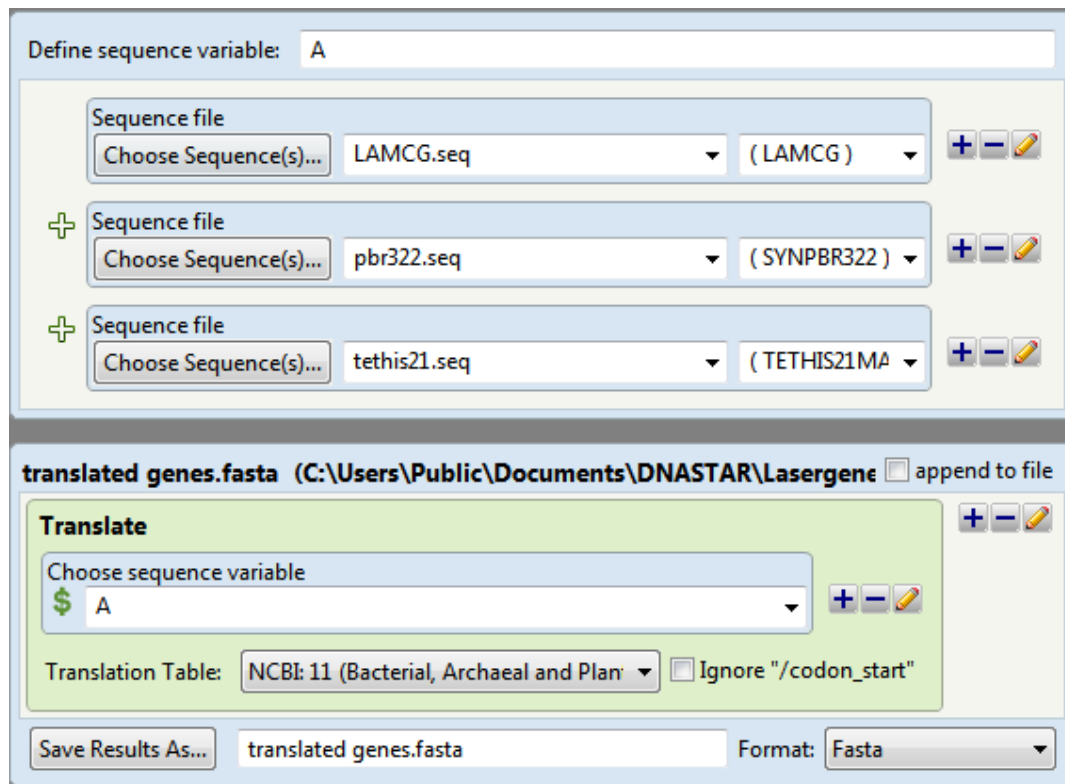
Note that the extraction step has been removed, and the sequence variable is now utilized in the translation step.

Saving a Sub-Range of a Sequence as a Separate File

This tutorial explores how to choose a sub-range for a sequence. In this example, you will extract the *int* gene and seven codons on either side of it from the file *LAMCG.seq*.

Sub-ranges can be set using the Modify Sequence(s) template. However, this tutorial demonstrates how to set up the same steps using the Toolkit only.

1) From the **Templates** panel, choose **Blank Program**.

2) From the **Toolkit** panel's **Sequences** bar, click once on **Create New Sequence** file.
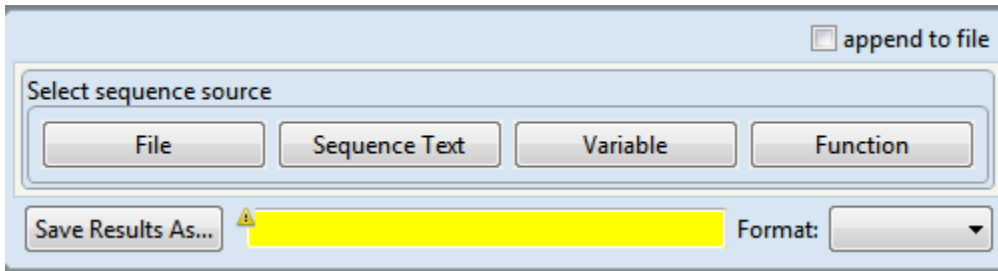
3) Since the source of the example sequence is a file (*LAMCG.seq*), click **File**.

4) Click the **Choose Sequence(s)** button, then double-click on the file *LAMCG.seq* to add it.

5) Add a step for setting the ends of the sequence. To do this, return to the **Sequences** bar of the **Toolkit** panel and click **Add Sub-Range**.

Note: Alternative ways of adding a sub-range include clicking on the ✚ icon to the right of the **Choose Sequences** row and selecting **Add Sub-Range**; or right-clicking near the **Choose Sequences** button and choosing **Add Sub-Range**.

The program now appears as follows:
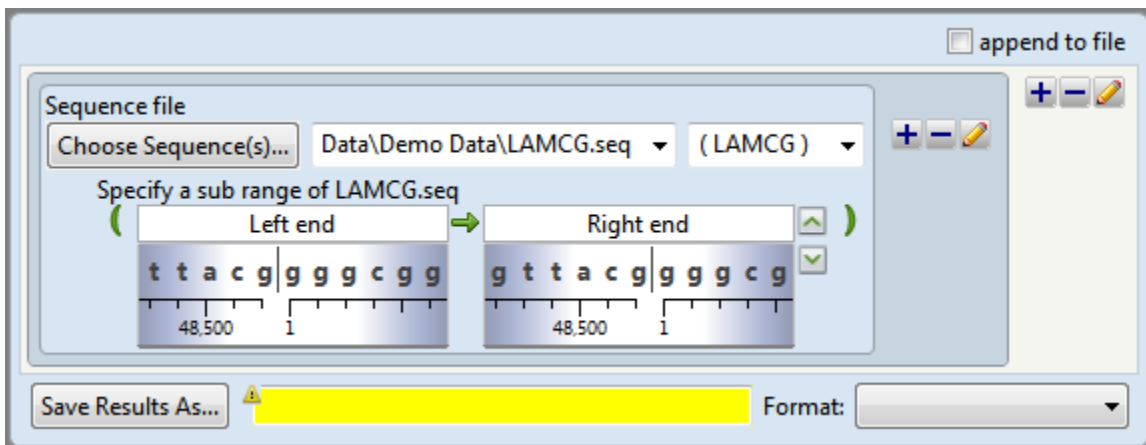
Two shaded "wheels" appear under the range boxes showing the sequence surrounding each coordinate. Note the similarities between the sub-range boxes and the double set of range boxes in the tutorial *Inserting Bases into an Existing Sequence*.

6) The *int* gene occupies a range from 27812 to 28882 bp. To scroll to the gene, type over the words "*Left end*" with "*27812*" and the words "*Right end*" with *28882*. Observe that the nucleotides and ruler under the text boxes change to reflect the new positions.

7) To view the sequence annotations for this region, click the downward-pointing green arrow on the right (⌄). The *int* gene graphic, which appears as a much larger green arrow, verifies that the end coordinates do indeed bracket the correct gene.

The arrow for the *int* gene is pointing left, indicating that it is on the opposite strand. Because we will be dealing with the isolated gene sequence, we will take the reverse complement of the sequence.

8) Click on the pencil icon (✎) in the outer box and select **Reverse Complement**. The features display will close. Re-open the features display by clicking on the green arrow (⌄), as in the previous step.
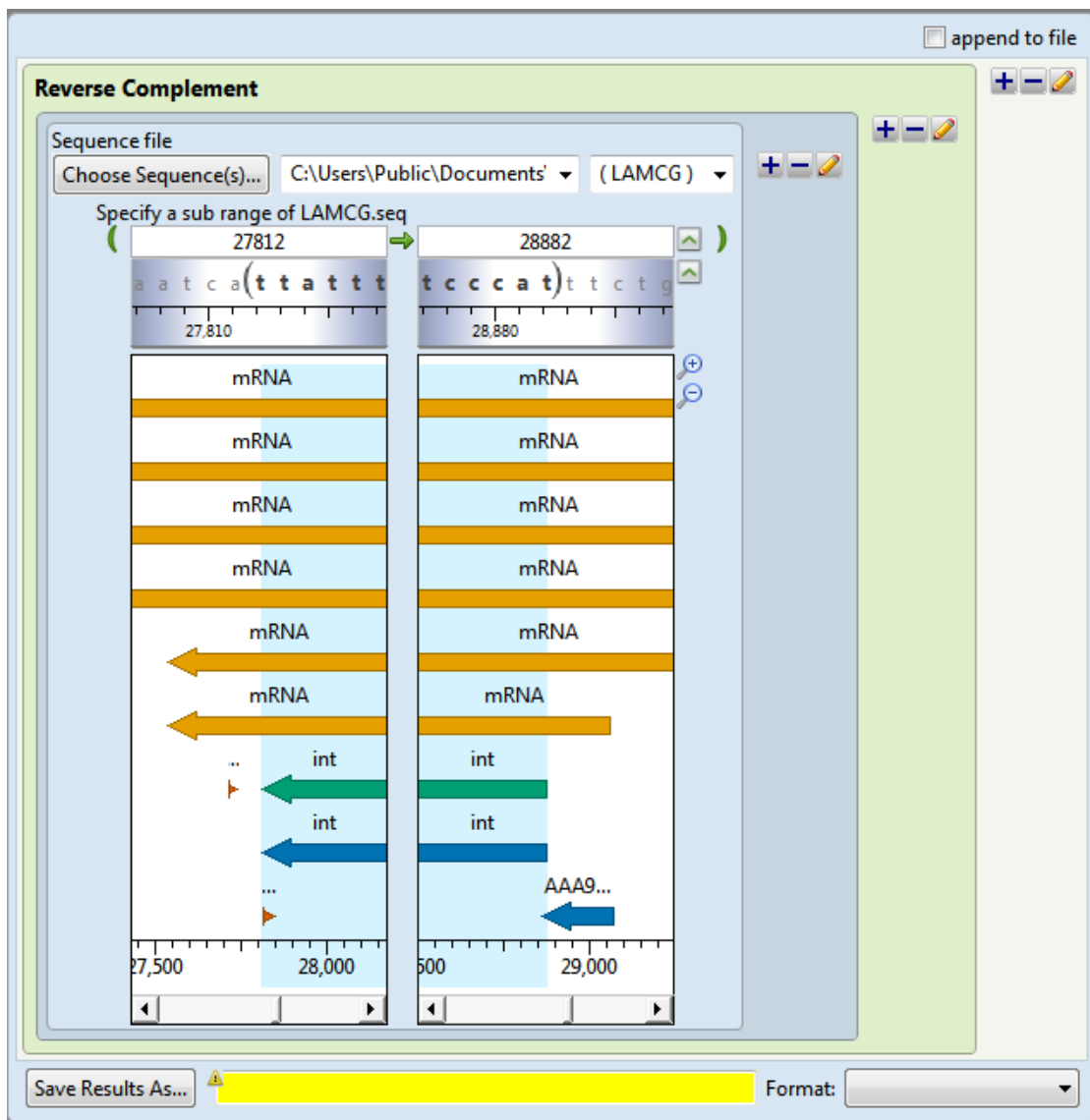
9) In this example, we want to extract not only the gene, but also seven codons on either side of it. Subtract 21 from the value on the left by typing "-21" after the existing

coordinate. Add 21 to the value on the right by typing "*+21*" after that coordinate. That is, **27812 – 21** in the left box and **28882 + 21** in the right box.

Note: Since SeqNinja will accept integer expressions as coordinates, it is not necessary to do the calculations and enter the values **27791** and **28903**. However, that would work equally well.

---

10) To the right of **Save Results As**, type the name *int Gene*.

11) From the **Format** drop-down menu, choose **Fasta**. The program now appears as shown:
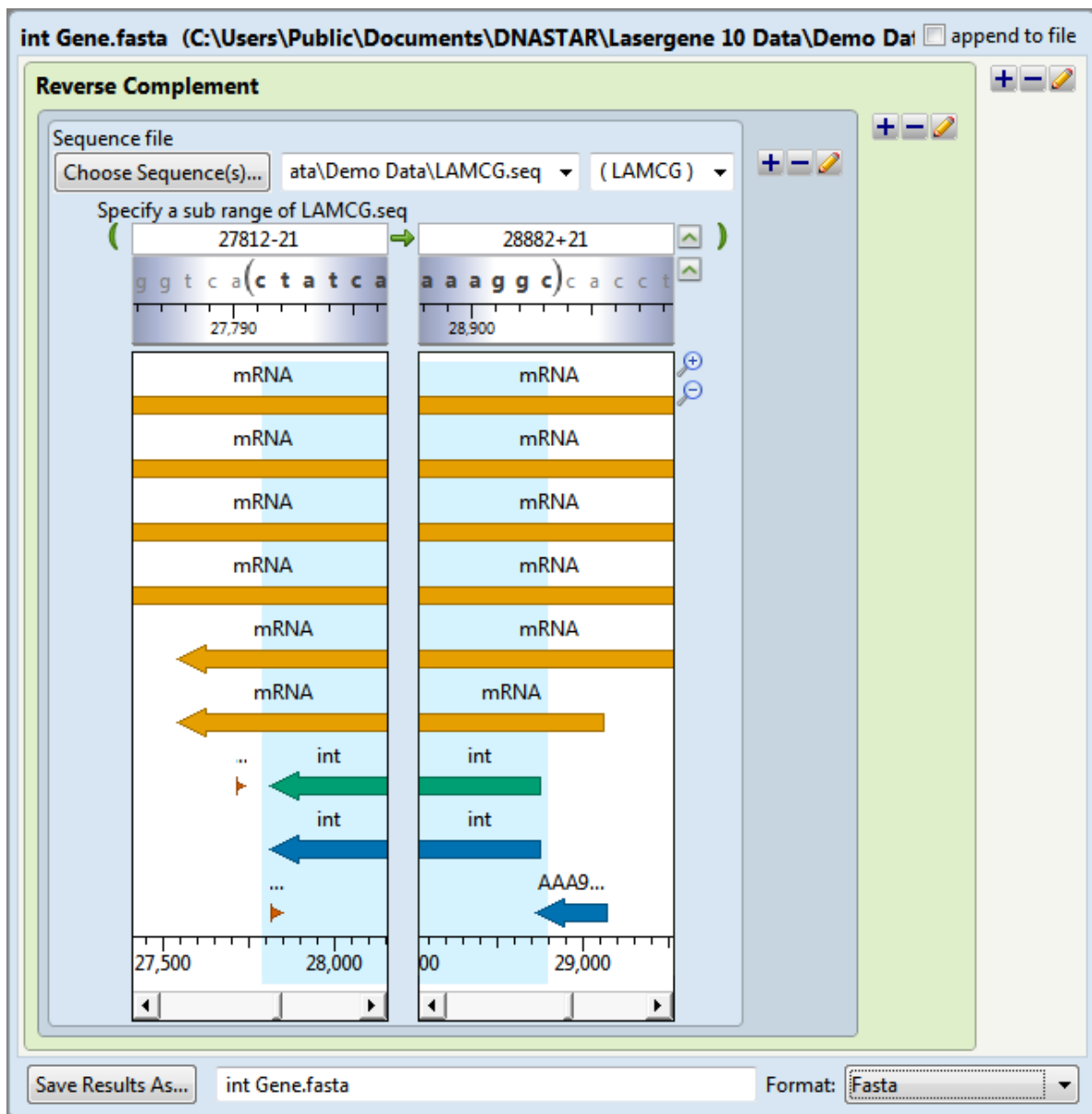
12) Use **File > Save** to save the program under the name *Sub-Range.star*.

13) Click the **Run** button (▶ Run).

14) If you now open the results file (*int Gene.fasta*) in DNASTAR's SeqBuilder or any text editor, it will appear as follows:

```
>int Gene:1
gcctttttgaagaggatcagaaatgggaagaaggcgaagtcatgagcgccgggatttaccc
cctaacctttatataagaaacaatggatattactgctacagggacccaaggacgggtaaa
gagtttggattaggcagagacaggcgaatcgcaatcactgaagctatacaggccaacatt
gagttattttcaggacacaaacacaagcctctgacagcgagaatcaacagtgataattcc
gttacgttacattcatggcttgatcgctacgaaaaaatcctggccagcagaggaatcaag
cagaagacactcataaattacatgagcaaaattaaagcaataaggagggggtctgcctgat
gctccacttgaagacatcaccacaaaagaaattgcggcaatgctcaatggatacatagac
gagggcaaggcggcgtcagccaagttaatcagatcaacactgagcgatgcattccgagag
gcaatagctgaaggccatataacaacaaaccatgtcgctgccactcgcgcagcaaaatca
gaggtaaggagatcaagacttacggctgacgaatacctgaaaatttatcaagcagcagaa
tcatcaccatgttggctcagacttgcaatggaactggctgttgttaccgggcaacgagtt
ggtgatttatgcgaaatgaagtggtctgatatcgtagatggatatctttatgtcgagcaa
agcaaaacaggcgtaaaaattgccatcccaacagcattgcatattgatgctctcggaata
tcaatgaaggaaacacttgataaatgcaaagagattcttggcggagaaaccataattgca
tctactcgtcgcgaaccgctttcatccggcacagtatcaaggtattttatgcgcgcacga
aaagcatcaggtctttccttcgaaggggatccgcctacctttcacgagttgcgcagtttg
tctgcaagactctatgagaagcagataagcgataagtttgctcaacatcttctcgggcat
aagtcggacaccatggcatcacagtatcgtgatgacagaggcagggagtgggacaaaatt
gaaatcaaataatgattttattttgactgatag
```

For further exploration: Add a translation step by dragging the template **Translate Sequence(s)** to the bottom of the existing program. After entering the required information (choose NCBI translation table 11) and clicking **Run**, the output should now appear as below. For clarity, the extra 7 codons on each side of the feature are shown boxed in red in the image below.

```
>int-transl
af.rgse ngrrrsherrdlppnlyirnngyycyrdprtgkefglgrdrriaiteaiqani
elfsghkhkpltarinsdnsvtlhswldryekilasrgikqktlinymskikairrglpd
apledittkeiaamlngyidegkaasaklirstlsdafreaiaeghittnhvaatraaks
evrrsrltadeylkiyqaaesspcwlrlamelavvtgqrvgdlcemkwsdivdgylyveq
sktgvkiaiptalhidalgismketldkckeilggetiiastrreplssgtvsryfmrar
kasglsfegdpptfhelrslsarlyekqisdkfaqhllghksdtmasqyrddrgrewdki
eik .fyfd..
```

MegAlign

Overview

MegAlign generates pairwise and multiple sequence alignments of DNA and/or protein. MegAlign can also create phylogenetic trees, as well as reports and tables showing the numerical data underlying the comparisons.

MegAlign can utilize sequence files from a wide variety of file types or from NCBI's BLAST or Entrez databases. If a sequence file is longer than you need, you may specify a portion of the sequence by coordinates, by selecting a feature from the feature table, or by highlighting a sub-region of a previously aligned project.

The following four algorithms are available for **pairwise** alignments:

- The **Wilbur-Lipman Method** constructs tables of K-tuples to find regions of similarity between two or more DNA sequence pairs using the method of Wilbur and Lipman (1983). The method reads the sequences, builds case structures of the K-tuples, finds the diagonals and matches, and creates the finished alignment.

- The **Martinez-NW Method** uses two alignment methods in succession. An approach described by Martinez (1983) identifies regions of perfect match. The Needleman-Wunsch (1970) method then optimizes the fit in between perfect matches.

- The **Lipman-Pearson Method** constructs tables of K-tuples to find regions of similarity. The alignment is then optimized using a trimmed-down version of the Needleman-Wunsch method developed by Lipman and Pearson (1985).

- The **DotPlot Method** compares sequences by overlapping them, counting mismatches, then shifting the sequences one residue relative to one another and counting mismatches again, iteratively, for every possible position. While this method is slower than other pairwise methods, it is exhaustive and can detect long regions of slight similarity between evolutionarily distant divergent genes or proteins. Parameters may also be adjusted to limit matches to shorter regions of higher similarity. Unlike other pairwise alignments, DotPlot results are displayed as customizable graphs rather than textual alignments.

The following three additional algorithms are available for **multiple** alignments:

- The **Jotun Hein Method** aligns all sequences in the Worktable using the method of Hein (1990). The Jotun Hein method builds a phylogenetic tree by examining sequence pairs and creating the best possible arrangement of ancestral branches. The method is most useful when aligned sequences are known to be related by descent.

- The **Clustal V Method** aligns all the sequences in the Worktable using the Clustal V method developed by Higgins and Sharp (1989). The Clustal Method groups sequences into clusters by examining sequence distances between all pairs. Clusters are aligned as pairs, then collectively as sequence groups, to produce the overall alignment. After a multiple alignment is completed, a neighbor-joining method is employed to reconstruct phylogeny for the putative alignment.

- The **Clustal W Method** aligns sequences using the method of Thompson et al. (1994). Clustal W was designed to create more accurate alignments than Clustal V when alignments include highly diverged sequences. However, Clustal W does not always handle end-gaps ideally, so some regions may be aligned better using the Clustal V method. Also, note that true Clustal W performance is only achieved when you choose the default "Slow-Accurate" option in the Method Parameters rather than the "Fast-Approximate" method.

Sequences may be realigned using different alignment parameters, and manual adjustments may be made using the Worktable palette tools. Once you are satisfied with the alignment, the versatile Alignment Report allows you to create a customized display of results. Your alignments may also be displayed in the form of a phylogenetic tree. Bootstrapping analysis may be applied to the tree when one of the Clustal algorithms is used to align your sequences.

10 Quick-Start Steps for Using MegAlign

Here are ten basic steps for using MegAlign:

1) Select **File > New** to open a new project or **File > Open** to open a previously saved project.

2) Add sequences to the Worktable as directed in the tutorial. You can use a sub-range of a sequence, and if necessary, you can select **Options > Font** and **Size** to reformat characters until they are clearly legible.

3) View and edit weight tables using **Align > Set Residue Weight Table**.

4) If you do not want to use the Standard Genetic Code, select **Options > Genetic Codes**. Codes may be edited using **Options > Edit Genetic Code**. Genetic codes are only important if you will be aligning DNA and protein sequences together, or translating DNA.

5) Choose a pairwise or multiple alignment method. For pairwise methods, first highlight two or more sequence names by using **Ctrl/Cmd+click** (Win/Mac) or **Shift+click**. For multiple alignments, all sequences in the Worktable are aligned whether or not they are highlighted. You may want to edit the **Align > Method Parameters** before performing the alignment. To undo a multiple alignment, select **Align > Unalign All**.

6) If you have performed a multiple alignment, you may edit gaps in the alignment from the Worktable or realign residues using the Worktable palette tools. To locate a particular position in a sequence, select **Edit > Go To Position** or **Edit > Find Disagreement**. Search commands can be used for unaligned or multiple aligned sequences, but not for pairwise alignments.

7) To examine numerical data about completed alignments, use the **Alignment Report**, **Sequence Distances**, **Residue Substitutions** or **Phylogenetic Tree** commands from the **View** menu.

8) If desired, customize the **Options > Alignment Report Contents**. You can select **Options > New Consensus** and **Options > New Decoration** to change the appearance of the report and the definition for the consensus.

9) Select **File > Print** your results, if desired.

10) Select **File > Exit** (Win) or **MegAlign > Quit MegAlign** (Mac) to exit MegAlign.

MegAlign Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example MegAlign project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Entering sequences using drag and drop

Drag and drop is a feature available in all Lasergene applications, and is very useful for entering large numbers of sequences.

Objective: To add two sequences to the MegAlign Worktable using the drag and drop procedure.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign\Histone Sequences

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign\Histone Sequences

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo MegAlign:Histone Sequences

1) Launch MegAlign.

2) Using Windows Explorer or Macintosh Finder, locate and open the **Histone Sequences** folder.

3) Highlight both **Tethis21.seq** and **Tethis22.seq** using **Shift**+**click** or a similar method.

4) Use the mouse to drag the selected files onto the MegAlign Worktable (Win) or MegAlign program icon in the dock (Mac) and drop them there. The MegAlign Worktable will appear as follows:
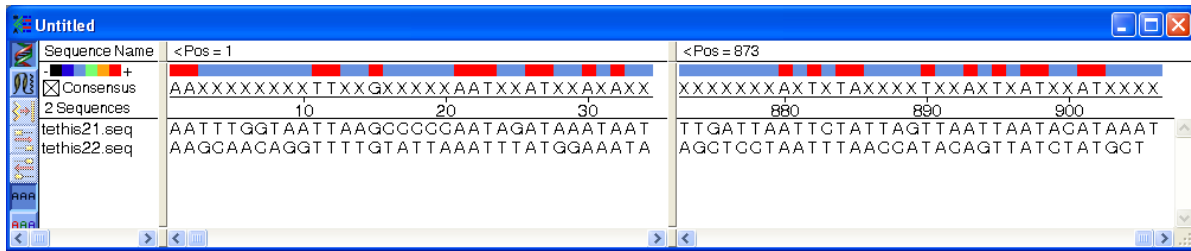
Note: If **tethis22.seq** appears before **tethis21.seq**, rearrange the sequences so that **tethis21.seq** appears first. You can do this by first selecting the sequence you want to move, and then dragging it to the desired position in the list.

- You may select **Options** > **Size** to increase or decrease the font size.

- The left pane of the worktable displays the first (starting) base pairs of both sequences. The right pane displays the final (ending) base pair of the longest sequence. In this case, because the sequences are almost the same length, the final base pairs of both sequences appear in the right pane.

Setting ends using features

In MegAlign, you may set sequence ends by typing in the desired coordinates or by selecting a feature.

Objective: To set sequence ends using one of the GenBank features imported automatically when the sequences were entered.

Note: The lengths of these sequences are 906 bp for Tethis21 and 905 bp for Tethis22.

1) Select **tethis21.seq** by clicking its name in the left pane of the Worktable.

2) Select **Options > Set Sequence Limits > From Feature Table**. The Feature List appears.

3) You can use this dialog to see the length of a sequence in the "Segments" section of the dialog.

4) Highlight the **histone H2B-1—CDS** feature in the upper pane of the dialog. The lower pane now shows the feature name and range (232 > 600) of the highlighted feature.
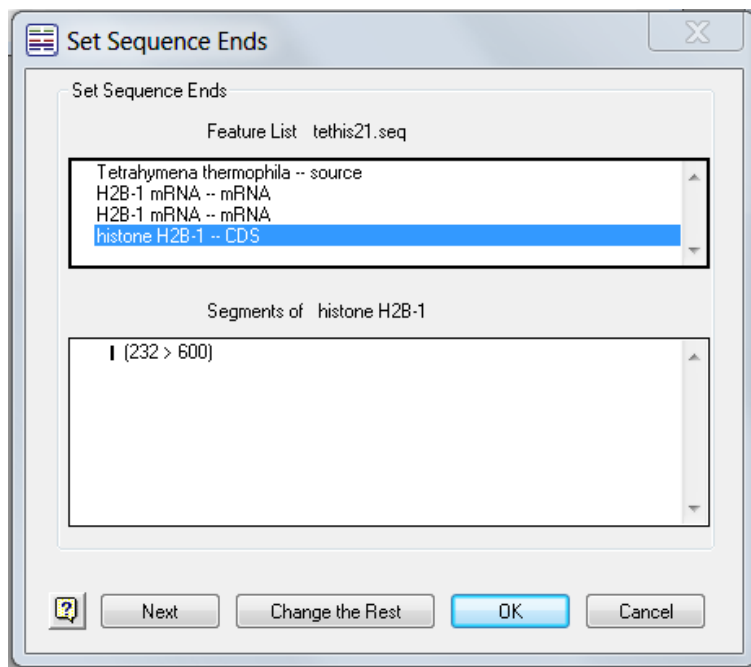
5) Click **Change the Rest** to limit the **tethis22** sequence using the same feature.

6) Click **OK** to return to the MegAlign Worktable.

Note that the selected range of the tethis21 sequence has limited the range of both original sequences on the worktable. Notice that the worktable now reflects that both sequences end at bp 369.

Performing a pairwise alignment

Since the current sequences are DNA, the pairwise methods available are Wilbur-Lipman, Martinez-Needleman-Wunsch and DotPlot.

Objective: Use the Wilbur-Lipman method to align two sequences.

1) Hold down the **Shift** key while you click each of the two sequence names to highlight them. Because there may be many sequences, MegAlign always needs to know which pairs you wish to align.

2) Select **Align > One Pair > By Wilbur-Lipman Method**. The method parameter dialog opens.

3) Click **OK** to use the default parameters for the alignment. MegAlign calculates the alignment and displays results in the Alignment View, shown below.

In the lower portion of the Alignment View, the aligned pair of sequences flanks a central consensus that shows mismatches as spaces. The window header displays the similarity index (percent of all residues that are matching), gap number, total gap length and consensus length.

4) Change the color scheme of the window by clicking ⬛. The Alignment View Options dialog appears.
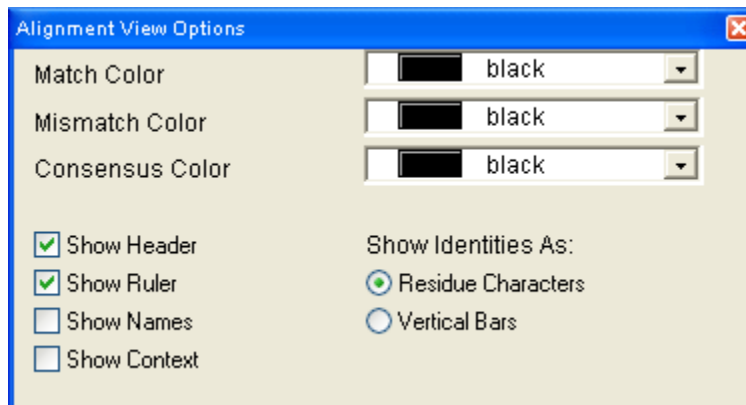
5) From the Match Color list, select **Blue**. All matching residues immediately turn blue on the Alignment View.

6) Repeat step 5, choosing **Red** for Mismatch Color and **Green** for Consensus Color.

7) Check and uncheck each of the four boxes just under the color selection boxes, and note how the Alignment View changes.

8) Click **Vertical Bars** to show agreement between the two sequences as vertical lines instead of letters.

9) Once you have finished experimenting with the Alignment Color dialog, close it using the close icon in the title bar.

10) Close the Alignment View to return to the Worktable.

Performing a Dot Plot pairwise alignment

The Dot Plot method compares sequences by plotting matches between two sequences on a chart, with one sequence on each axis.

Objective: To use the DotPlot method to align two sequences.

1) Highlight the two sequence names (if not still highlighted from the previous tutorial).

2) Select **Align > One Pair > Dot Plot**. The Filtered DotPlot dialog opens, allowing you to set the parameters of the alignment.

3) Leave the parameters set to their defaults, and click **OK**. The DotPlot opens in a separate window.

- Diagonal lines indicate regions of the two sequences that meet the threshold for similarity specified in the Dot Plot parameters.

- Dark blue diagonals show weaker similarities, with progressively stronger similarities indicated by pale blue, green, yellow, orange and red.

- The long red diagonal line indicates that these two sequences align well over most of their length.

4) Double-click on any diagonal line to open a window that shows the underlying alignment of that diagonal.
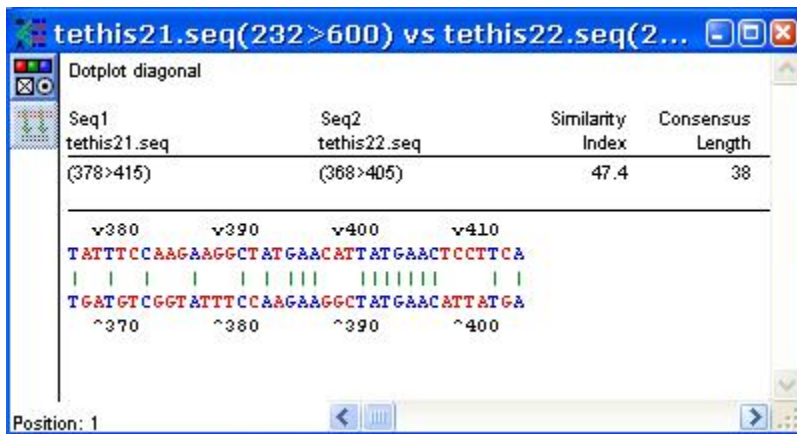
5) Once you have finished viewing the window, click on the Worktable and close the current MegAlign project using **File > Close**. When prompted to save the document, click **No** (Win) or **Don't Save** (Mac).

Performing a multiple alignment

Objective: To align fourteen calmodulin sequences using the Clustal V method.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign\

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo MegAlign

1) Select **File > Enter Sequences**.

2) Locate and open the **Demo MegAlign** folder.

3) Do one of the following:

   - **Windows**: Double-click on the folder "Calmodulin Sequences" and then click **Add All**.

   - **Macintosh**: Double-click on the "Calmodulin Sequences" folder to open it. Use **Cmd+A** to select all the sequences, then click **Add File**.

4) When all fourteen sequence names appear in the Selected Sequences pane (Win) or the folder appears in the Chosen Files pane (Mac), click **Done**.

5) Select **Align > Set Residue Weight Table**. The Residue Weight's table appears.

   - 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.

   - For the Clustal V method, the PAM100 table should be used when aligning conserved sequences, and the PAM250 table should be used for divergent sequences. Since calmodulins are highly conserved, the PAM100 table should be used for this alignment.

6) Select **PAM100** from the pull-down menu. The title bar for the current window reflects MegaAlign's default settings. The title will later change to reflect the weight table and alignment method you choose.

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	
C	14																				C
S	-1	6																			S
T	-5	2	7																		T
P	-6	1	-1	10																	P
A	-5	2	2	1	6																A
G	-8	1	-3	-3	1	8															G
N	-8	2	0	-3	-1	-1	7														N
D	-11	-1	-2	-4	-1	-1	4	8													D
E	-11	-2	-3	-3	0	-2	1	5	8												E
Q	-11	-3	-3	-1	-2	-5	-1	1	4	9											Q
H	-6	-4	-5	-2	-5	-7	2	-1	-2	4	11										H
R	-6	-1	-4	-2	-5	-8	-3	-6	-5	1	1	10									R
K	-11	-2	-1	-4	-4	-5	1	-2	-2	-1	-3	3	8								K
M	-11	-4	-2	-6	-3	-8	-5	-8	-6	-2	-7	-2	1	13							M
I	-5	-4	-1	-6	-3	-7	-4	-6	-5	-5	-7	-4	-4	2	9						I
L	-12	-7	-5	-5	-5	-8	-6	-9	-7	-3	-5	-7	-6	4	2	9					L
V	-4	-4	-1	-4	0	-4	-5	-6	-5	-5	-6	-6	1	5	1	8					V
F	-10	-5	-6	-9	-7	-8	-6	-11	-11	-10	-4	-7	-11	-2	0	0	-5	12			F
Y	-2	-6	-6	-11	-6	-11	-3	-9	-7	-9	-1	-10	-10	-8	-4	-5	-6	6	13		Y
W	-13	-4	-10	-11	-11	-13	-8	-13	-14	-11	-7	1	-9	-11	-12	-7	-14	-2	-2	19	W
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

7) Click **OK** to return to the Worktable.

8) To view the current Clustal V settings, select **Align > Method Parameters** and select the **Clustal V** tab.

9) Review the settings, leave the default values, and then click **Later**.

10) To align the sequences, select **Align > By Clustal V Method**. The Alignment Progress window shows what percentage of the project has been aligned. When the alignment is complete, the Worktable reappears with the aligned sequences.

11) To view a table showing the divergence and percent identity of the sequences, select **View > Sequence Distances**:

Pair Distances of Untitled ClustalV (PAM100)

Percent Identity (columns) / Divergence (rows)

Divergence \	1	2	3	4	5	6	7	8
1		81.9	90.6	90.6	89.9	90.6	99.3	86.6
2	20.8		84.6	83.2	83.9	84.6	82.6	81.2
3	10.1	17.3		90.6	99.3	100.0	89.9	88.6
4	10.1	19.0	10.1		89.9	90.6	89.9	92.6
5	10.8	18.2	0.7	10.8		99.3	89.3	87.9
6	10.1	17.3	0.0	10.1	0.7		89.9	88.6
7	0.7	19.9	10.8	10.8	11.6	10.8		85.9

12) To view a table showing residue substitutions, select **View > Residue Substitutions**:

Residue Substitutions of Untitled ClustalV (PAM100)

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	
C		0	2	0	0	0	0	0	0	0	0	0	0	0	0	C
S	0		4	0	8	0	4	0	0	1	0	0	0	0	0	S
T	2	4		0	0	0	0	0	0	2	0	0	0	2	2	T
P	0	0	0		0	0	0	0	0	0	0	0	0	0	0	P
A	0	8	0	0		0	0	0	0	1	0	0	0	0	0	A
G	0	0	0	0	0		2	0	0	2	0	0	0	0	0	G
N	0	4	0	0	0	2		1	1	2	0	0	0	0	0	N
D	0	0	0	0	0	0	1		12	0	0	0	0	0	0	D
E	0	0	0	0	0	0	1	12		0	0	0	0	0	0	E
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	

13) To view the phylogenetic tree, select **View** > **Phylogenetic Tree**.

14) To apply bootstrapping analysis to the tree, select **Align** > **Perform Bootstrapping**. A separate window containing the bootstrap analysis settings appears.

15) Review the settings and then click **OK**. Bootstrapping analysis begins, and bootstrapping values are displayed on the phylogenetic tree. The numbers displayed refer to the percentage of times the same branches appear when a number of sample trees (in this case, 1000) are generated.
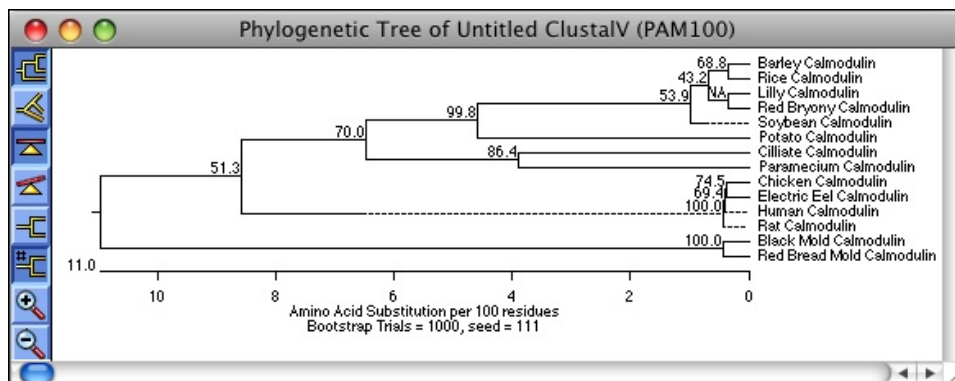
Note: MegAlign's bootstrapping algorithm uses random seeding. Therefore, the result displayed in your phylogenetic tree may differ from the image above.

16) Close all windows except for the Worktable.

Creating an alignment report

Alignments can be modified to show, hide and highlight various components and consensuses. Graphics in the Alignment Report that box, shade, or hide residues are referred to as "decorations". On the Alignment Report you may also define and display one or more custom consensuses built upon your own criteria.

Objectives: To modify the contents of the Alignment Report and to create a decoration that shades those amino acids that disagree with the consensus sequence.

1) Select **View > Alignment Report** to open a separate window containing the results of the multiple alignment. Take a moment to review its contents.

2) To modify the contents of the report, select **Options > Alignment Report Contents**.

3) Check the box next to **Show Consensus Strength** and leave all other settings at their defaults:
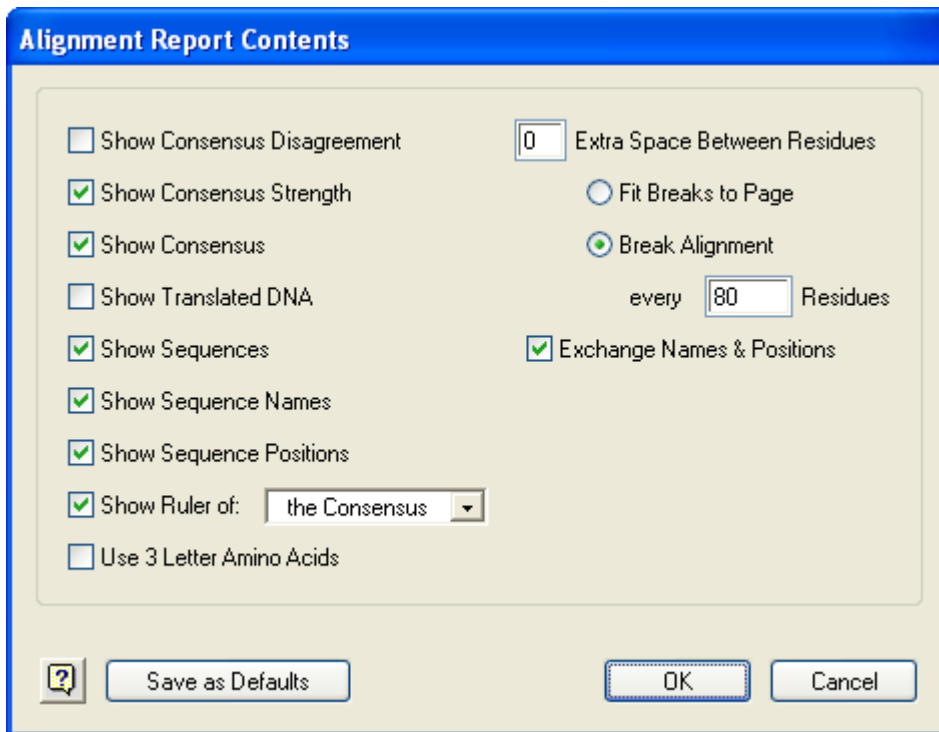
4) Click **OK** to update the already open Alignment Report. Note the color-coded histograms showing the strength of the consensus at each base.

5) Select **Options > New Decoration**. The New Decoration dialog appears.

6) In the title box, type "Shade disagreements with consensus".

7) 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**.

8) Choosing **Shade** caused two additional pull-down menus to open below it. Leave the color as black from the first menu and choose the darkest shading scheme from the second menu. Leave the distance units box at "0". The dialog should now look like the one below:
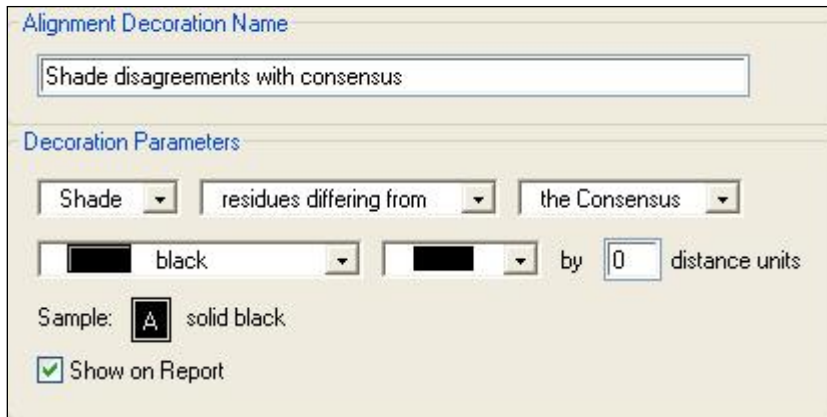
9) Click **OK**. In the Alignment Report window, any constituent sequence residues that do not match those in the consensus sequence now appear shaded. (The sequences may be listed in a different order in your project).
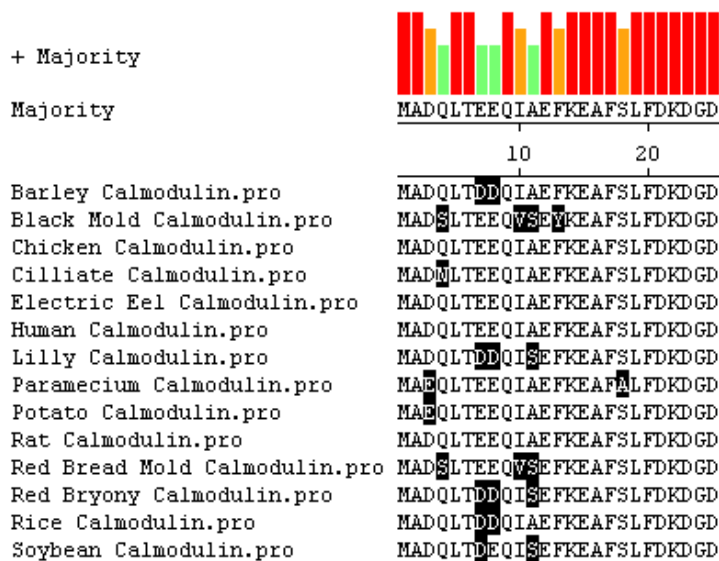
Further Exploration

Creating an alignment using a portion of the sequences

To create a new alignment from a subsection of the sequences, highlight a portion of the consensus sequence in the Worktable and select **Align** > **Create Alignment From Selection**. When prompted, choose a name and location for the new project and click **Save**. A new Worktable will open containing only the highlighted portion of sequences.

Realigning residues

After completing a multiple alignment, you may manually adjust selected portions of the alignment using the palette tools to the left of the Worktable.

 Straighten Columns inserts gaps preceding selected residues and aligns residues along their left sides.

 Shuffle Right moves gaps within selected regions so they precede the selected residues.

 Shuffle Left moves gaps within selected regions so they follow the selected residues.

Working with the Alignment Report

- To add a decoration to the Alignment Report use **Options > New Decoration**.

- To add a consensus to the Alignment Report, select **Options > New Consensus**. If desired, you can change the standard amino acid categories from the same dialog by clicking **Set Groups** then **Standards**.

- To add, remove or change decorations and consensuses, select **Options > Decorations > Decoration Manager** or **Options > Consensi > Consensus Manager**.

Exporting data

- To export sequences as separate Lasergene DNA or protein files, highlight the names of the sequences in the Worktable and select **File > Export Sequences**.

- To export the consensus as a Lasergene DNA or protein file, select **File > Export Consensus**. Gaps in the sequence are represented by X's.

- To export your alignment as a [GenVision](#) project, select **File > Export as GenVision Project(s)**. You will be given the option to export a single GenVision project, showing all of the sequences in your project as histograms, or to export one GenVision project for every sequence in the alignment. The latter option will produce a circular image for each sequence, with all of the sequences represented around the perimeter, and chords connecting the aligned, matching segments between the one single sequence and all others.

  Note: Before you can export data from the Worktable to GenVision, you must first do an alignment using **Align > By Jotun Hein Method**, **Align > By Clustal V Method** or **Align > By Clustal W Method**.

- To export the alignment for further analysis in PAUP, GCG Pileup or as a Phylip tree file, select **File > Save As** and select the format you want.

GeneQuest

Overview

GeneQuest aids you in identifying and annotating coding regions and other features of interest in your DNA sequence. The project begins when you enter a DNA sequence. The next step is to apply some of GeneQuest's many analysis methods to the sequence. These methods include:

- **Title**—Adds a title to the document.

- **Ruler**—Adds a ruler to the document.

- **Sequence**—Displays the sequence on the document.

- **Patterns - Matrix**— Patterns - Matrix - Locates Start and Stop sites, Intron and Exon splice sites using 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**—Identifies regions that match those of 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**—Uses Borodovsky's Markov method to identify potential coding regions.

- **Coding Prediction - Starts Stops ORFs**—Locates and 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 regions rich with information.

- **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.

- **Features**—Not a true "method," but a list of known features for the sequence.

Once the methods have been applied, the results are displayed graphically on a common horizontal scale. In addition, menu commands let you display tabular data summaries (e.g., base content, codon usage), fold a portion of the sequence as RNA, or view the predicted electrophoretic separation of restriction fragments through an agarose gel.

GeneQuest also reads and interprets features from the comment or feature panes of GenBank-formatted sequence files. Once you have discovered areas of interest in the sequence, you may create your own annotations and link several features together under a common title and description.

10 Quick-Start Steps for Using GeneQuest

Here are ten basic steps for using GeneQuest:

1) Create a new project or open an existing GeneQuest project.

2) Select **Analysis > Show Available Methods** to display a list of methods, and choose which of these you would like to apply to the sequence. You may wish to follow the formal strategy for locating potential genes outlined later in this chapter.

3) Move the **methods** you want from the More Methods menu to the Method Curtain. Apply methods to the sequence by using the mouse to drag them from the Method Curtain to the assay surface, if they are not already there.

4) View the results in graphical or tabular form.

5) Edit method parameters at any time. Then add, delete, rearrange and "decorate" methods until you are satisfied with the look of the document.

6) Use the graphical and tabular results to form hypotheses about your sequence, and then annotate any interesting features.

7) Search BLAST or Entrez databases for matches to your sequence, or other sequences of interest.

8) Select **File > Save** to save your results.

9) Select **File > Print** to print your results, if desired.

10) Select either **File > Exit** (Win) or **GeneQuest > Quit** (Mac) to exit GeneQuest.

GeneQuest Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example GeneQuest project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Opening an existing project and organizing methods

All folders referenced in this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:

1) Launch GeneQuest.

2) Select **File > Open**. Locate the file *WormProblem_FinalResult.dad* in the **Demo Data** folder and double-click it to open.
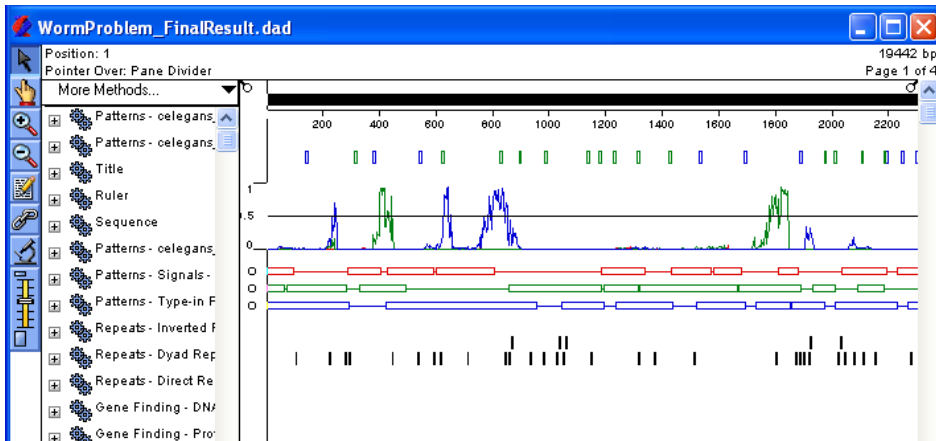
The main portion of the window is the assay surface. It contains graphic displays of analysis methods that have already been used to analyze the sequence. The Method Curtain at left displays a list of available analysis methods.

3) Select **File > Close** to close the **WormProblem_FinalResult.dad** project.

4) Select **File > New** and open the *WormProblem.seq* file located in the **Demo Data** folder.

5) Select **Analysis > Show Legend**. The Legend Curtain is displayed to the right of the assay surface. This curtain displays the names of the methods that have been applied to the assay surface.

Note: At this point, only the scale, sequence, as well as ORFs and Starts for the first 3 reading frames are displayed.

Proceed to the next tutorial.

Applying methods to the sequence

1) In the Method Curtain (displayed to the left of the assay surface), double-click on the method **Coding Prediction-Borodovsky- volvox_0.mat** so that the method parameters dialog is displayed.
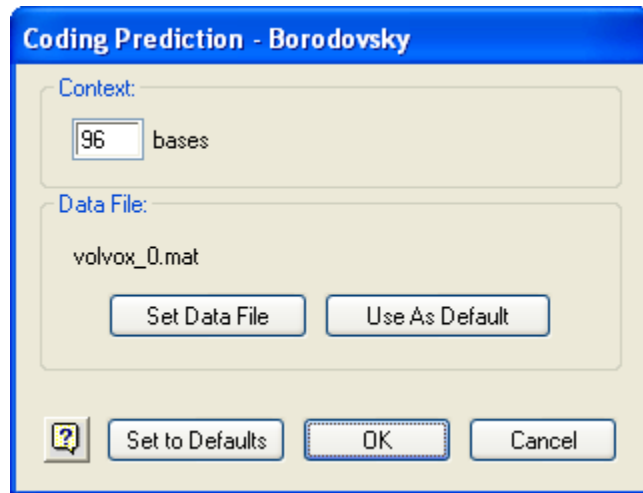
2) Click **Set Data File** and select the **ce_lo_4.mat** file from the **Borodovsky Matrices** folder. Click **Open**, then click **OK** to close the dialog.

3) Click on the **plus sign** (Win) or the **triangle** (Mac) to the left of the method **Coding Prediction – Borodovsky – ce_lo_4.mat** to reveal the list of possible displays.
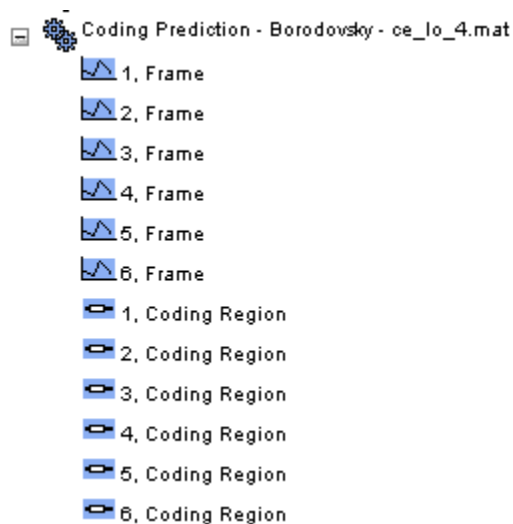
4) Click the **white space** to the left of the displays to ungroup them.

5) **Shift**+**click** each of the **graph symbols** next to Frames **1**, **2**, and **3** to select them.

6) Drag the selected displays onto the assay surface and drop them under the existing displays.

Note: GeneQuest calculates potential coding regions and displays them as peaks. The higher the Borodovsky peak, the higher the probability that the area is a coding region.

7) You can use colors to differentiate each frame of the newly applied Borodovsky displays from each other.

   - First, click the **white space** underneath the Borodovsky displays to ungroup them.

   - Then, click the Frame 1 Borodovsky display to select it.

   - Select **Options > Line Color > Red**.

   - Repeat the above steps for the Frame 2 and Frame 3 displays, but select **Green** for Frame 2 and **Blue** for Frame 3.

8) **Shift**+**click** frames **1**, **2**, and **3** of the Borodovsky methods, then select **Options > Superimpose Objects**.

9) In the Method Curtain, click on the **plus sign** (Win) or the **triangle** (Mac) to the left of the method **Patterns - Matrix**. Locate and select the file **celegans_ds_2.mat** from the **Matrix Patterns** folder and click **Open**. This method will now be listed in the Method Curtain. Select the **Top Strand** of this method and drag it to the bottom of your assay surface.

10) At the top of the Method Curtain, select **More Methods > Patterns > Matrix**. A new entry called **Patterns - Matrix** will appear at the top of the Method Curtain.

11) Repeat steps 9-10 to add two matrix patterns **celegans_as_2.mat** and **celegans_atg_2.mat** to this new method. For each method, drag the **Top Strand** to the bottom of your assay surface.

Note: GeneQuest calculates potential translation start, donor and acceptor sites and displays them as vertical lines. Each apparent line is really a narrow box containing a numeric score denoting the likelihood of that area being a site.

12) To view the scores only, select the **Zoom In** tool , then click and drag diagonally to draw a box around the area you would like to see in more detail.

13) To view the scores and other information, select the **Range Selector** tool , then click and hold on a site until an information box is displayed.

- The cursor snaps to a particular point on the site. The exact base to which the cursor snaps is shown on the header at the top left of the window next to the word "Position."

- The pop-up information box displays the bases spanned by the pattern and a candidate ranking score.

14) Close the method curtain by selecting **Analysis > Hide Available Methods**.

Proceed to the next tutorial.

Predicting coding regions and creating features

1) To make it easier to locate potential genes, click the **Object Selector** tool . Drag and drop the method names so they appear in the Legend Curtain in the order shown:

```
□ Scale

□ Sequence

□ Top Strand, caenorhabditis & elegans acceptor - Matrix - celegans_as_2.mat

□ Top Strand, caenorhabditis & elegans donor - Matrix - celegans_ds_2.mat

□ 1, ORFs - Starts Stops ORFs
□ 2, ORFs - Starts Stops ORFs
□ 3, ORFs - Starts Stops ORFs
□ 1, Starts - Starts Stops ORFs
□ 2, Starts - Starts Stops ORFs
□ 3, Starts - Starts Stops ORFs

□ 1, Frame - Borodovsky - ce_lo_4.mat
  □ 2, Frame - Borodovsky - ce_lo_4.mat
    □ 3, Frame - Borodovsky - ce_lo_4.mat

□ Top Strand, caenorhabditis & elegans start - Matrix - celegans_atg_2.mat
```

2) Use the **Object Selector** tool to select each matrix pattern result in turn, then use the **Options** menu to change the **Line Color**. For example, choose green for the donor sites, blue for the acceptor sites, and black for the start sites.

3) Use the **scroll bar** at the bottom of the window to locate the area of the sequence from **9,000-12,000** bp. Use the **Zoom** tools or to enlarge or reduce the view until you can see the entire sequence range.

4) Click the **Range Selector** tool and then click on the frame **3** ORF centered on base 10,000.
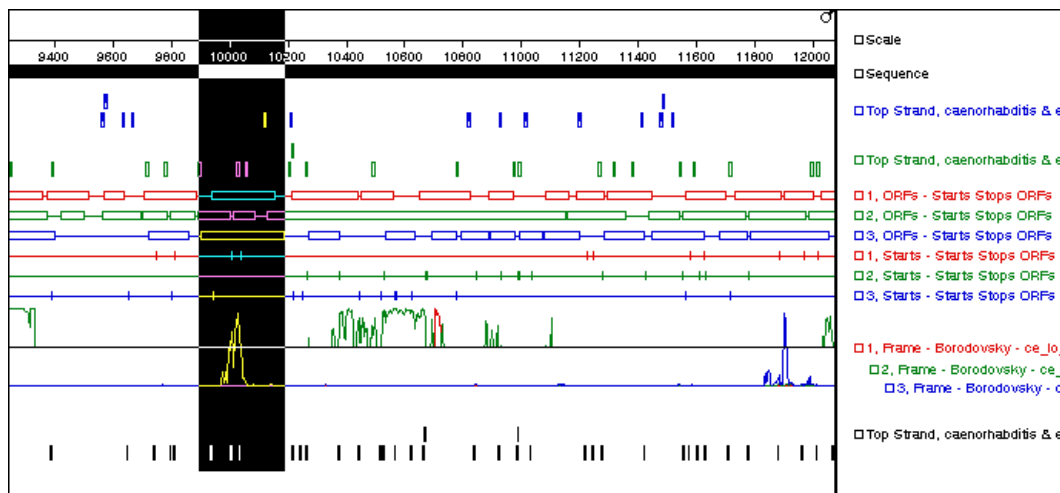
Note that a large peak in the frame 3 Borodovsky graph lies entirely within this highlighted ORF. It is likely that this area represents an exon.

You are now ready to find the beginning of the potential gene.

5) In the **Top Strand…atg_2.mat** display, click and hold on the candidate start site at the far left end of the ORF. This site is shown enlarged below.

Pattern: caenorhabditis & elegans start
9.0 match from 9935 to 9943 on top strand.
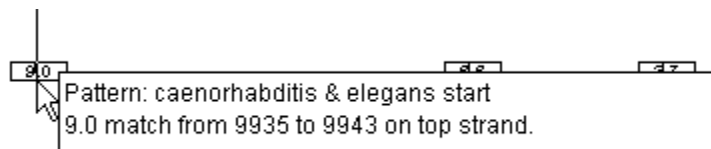
- The header shows the position of the candidate, while the pop-up information box shows the score.

- If this is the first exon of a multi-exon gene, it must end with a donor splice site that precedes the stop codon for the ORF in the same frame. In the Top Strand…ds_2.mat display, observe that there are two candidate donor splice sites (shown circled in the following figure) beyond the Borodovsky peak and still within the corresponding frame 3 ORF.
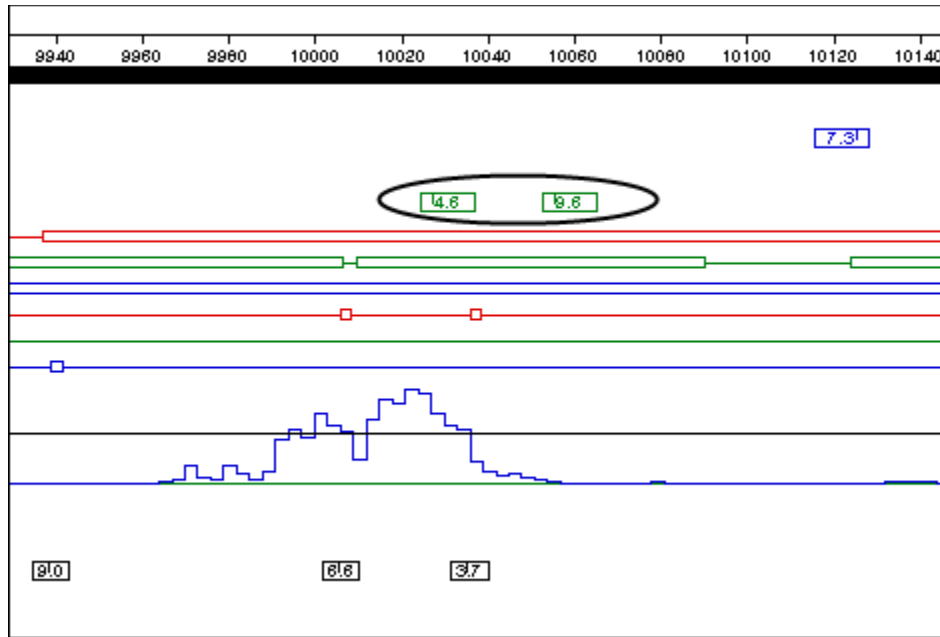
✓ Exon 1 Donor: Base 10028; score = 4.6

✓ Exon 1 Donor: Base 10056; score = 9.6

You are now ready to annotate the first exon.

6) Select **Edit > Go To Location**. Enter the range 9939,10055 and then click **OK**.

The position 10056 indicates the beginning of the donor splice site; therefore, the actual range of the exon extends from 9939 to 10,055.

7) Click the Annotation tool ![icon] to open the Feature Editor. The location of the gene segment appears in the Segment box.
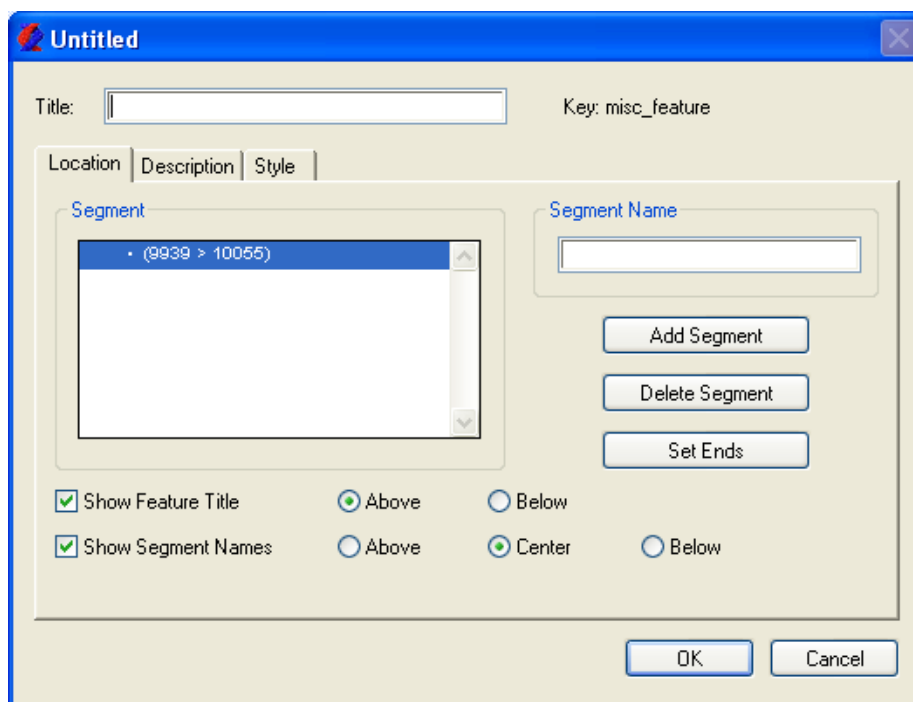
- Enter **genE** into the **Title** box.

- Enter **exon1** into the **Segment Name** box, then click **OK**. A new feature is created and displayed on your assay surface.

8) Now that the first exon has been annotated, you may continue searching for additional exons.

- Any further exons must begin with an acceptor splice site, and that site should be in an ORF for which there is Borodovsky evidence for coding potential.

- Observe that the next significant Borodovsky peak is in frame 2, centered near base 10,600.

- There is one very likely acceptor candidate in the Top strand…as_2.mat display, between bases 10,207 and 10,219: Exon 2 Acceptor: Base 10,217; score = 12.0

- If there are more than two exons to this gene, Exon 2 must terminate with a donor splice site. There is a strong candidate in the Top strand...ds_2.mat display at position 10,778, some distance to the right of where the frame-2 Borodovsky peaks fall to the baseline: Exon 2 Donor: Base 10,778; score = 12.6.

- Now you may annotate this exon and join it to the first gene segment.

9) Select **Edit > Go To Location**.

10) Enter the range **10217,10777** and then click **OK**.

11) Click the Joining Tool .

12) Select **genE - misc_feature** from the list and then click **Join**.

13) The next Borodovsky peak is not only in the same frame as exon 2, but also in the same ORF. The best candidate acceptor splice site to start exon 3 has a likelihood value of 14.1 out of a maximum 16.7—an excellent candidate. Two candidate donor splice sites are available to define the 3′ end of this exon, at 10,973 and 10,993. The first site scores higher at 6.4.

   - Exon 3 Acceptor: Base 10,826; score = 14.1

   - Exon 3 Donor: Base 10,973; score = 6.4

14) The next Borodovsky peak is once more in frame 2, but there is no candidate acceptor site (at the default threshold) under this peak. However, there is a site to the left of the peak, at position 11,021, that scores an impressive 15.1; this is the best candidate start for exon 4. You will note that there is no candidate donor splice site within the corresponding frame 2 ORF.

15) This indicates that Exon 4 is probably the final exon in the gene. To locate the stop codon, select **Analysis > Show Available Methods** and expand **Coding Prediction – Start Stops ORFs** by clicking on the **plus sign** (Win) or the **triangle** (Mac) next to the method name.

16) Select the **Stops** method for **frame 2** and drag and drop it onto the assay surface. Notice that there is a stop codon at the end of this ORF (base 11,155). Therefore, use the splice site at base 11,021 to start this exon and the stop codon at base 11,155 to terminate this exon.

   - Exon 4 Acceptor: Base 11,021; score = 15.1

   - Exon 4 Donor: No likely candidates

   - Stop Codon: Base 11,155

17) After locating the endpoints for Exons 3 and 4, annotate them by repeating steps 9-12 for each exon and entering the appropriate ranges.

18) Now that all four gene segments have been annotated, click on either the **Object Selector** tool or the **Range Selector** tool and then **double-click on the feature** to reopen the Feature Editor.

19) In turn, click on the coordinates for the three unnamed segments and type their names into the empty **Segment Name** boxes (**exon2**, **exon3**, **exon4**).
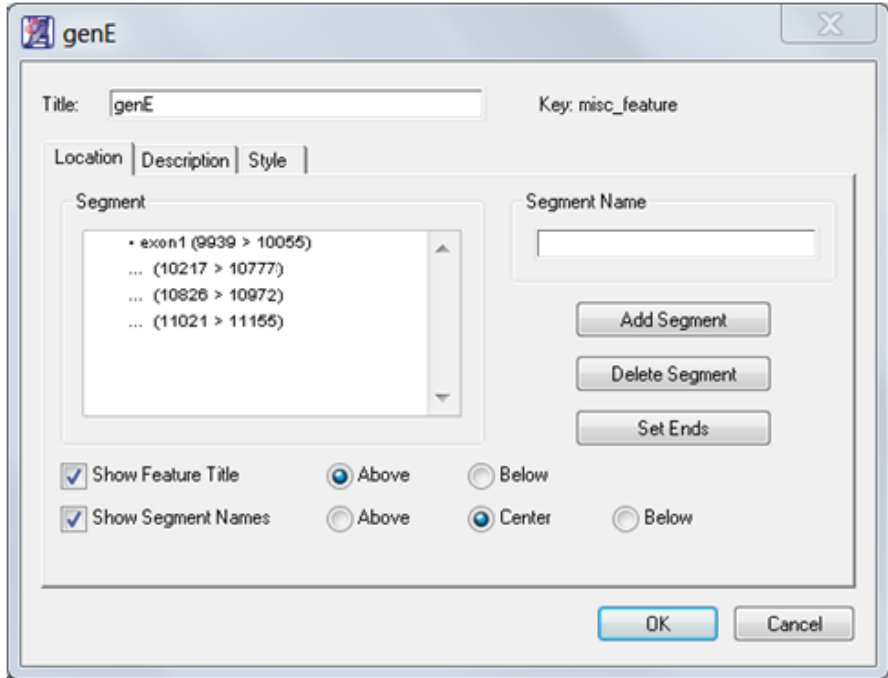
20) Click **OK** to finish. The feature should resemble the figure below:
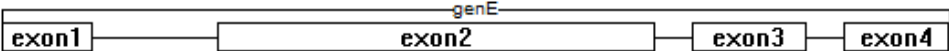
Proceed to the next tutorial.

Corroborating predictions by BLAST search

1) Click the **Range Selector** tool and then click anywhere in the **genE** feature.

    Note: The entire sequence spanning the feature is highlighted. However, the **Selection** line in the header above the assay surface indicates that only the annotated segments of the feature are selected.

2) Select **Net Search > BLAST ORF**. Leave **nr** selected in the Database drop-down menu, then click **OK**.

    - This sends the translation of the selected DNA sequence to NCBI's BLAST server as a blastp query (protein vs. protein).

    - GeneQuest displays search results when they are returned from the server—this may take several seconds.

3) Highlight the match with Identifier **17553974** (this will likely be the first entry) and click the **Put In Document** icon.

4) In the ensuing dialog, **keep the defaults** and click **OK** to save copy of the highlighted sequence. It now appears on the assay surface under the **genE** feature.

    It also appears as a new **Gene Finding - Protein Finder** method at the bottom of your Method Curtain. You can apply it or remove it from the assay surface as you would any other method. This method aligns the protein sequence with the corresponding DNA region(s) in your sequence where the protein matches the sequence translation, as pictured below.
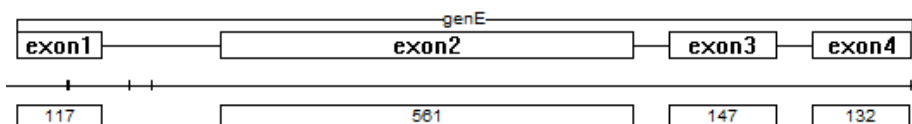
    Note: Exon4 is actually 135 bp in length (range=11155-11021), but the feature obtained via BLAST searching omits the stop codon.

5) To view the sequence up close, select the **Zoom In** tool , then click and drag diagonally to draw a box around the area you would like to see in more detail.

Proceed to the next section.

Checking predictions against external data

In this section, we will be comparing our results with the results the authors of this sequence obtained.

1) Select **File > Open Entrez Sequence**. Enter **Z92803** in the dialog box, select the **nucleotide** database to search and click **OK**. When the Entrez result is returned, click **Save** to save a local copy of the sequence and open it in GeneQuest, applying the default method set.

> **Note:** The Z92803 sequence is exactly the same length as your problem sequence—19,442 bases. To see all the annotations made by the authors, go to the next step.

2) In the Method Curtain for **Z92803**, expand the **Features** method. Drag the **K01G5.3:CDS** feature to the Assay Surface.

3) Examine the annotated gene in the region 9-12 kb. How does it match up to yours?
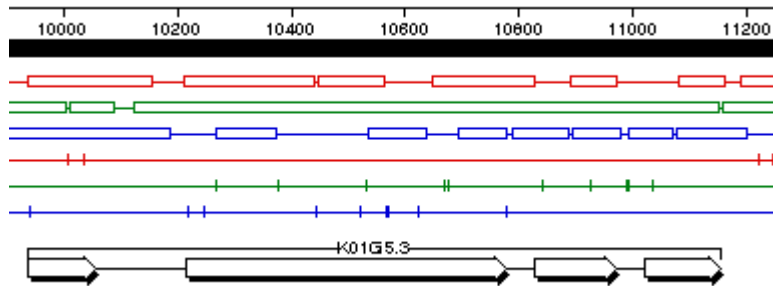
4) Select **File > Save** to save the project if you wish and close GeneQuest.

Further Exploration

Additional analyses

- To view a tabular summary of restriction fragments, apply one or more Enzymes - Restriction Map methods, then select **Sites & Features > Restriction Fragment Summary**.

- GeneQuest assumes that sequences are circular unless told otherwise. To have the application treat a sequence as linear, select **Sites & Features > Linearize Sequence**.

- To simulate multiple digestion experiments and examine fragment sizes in predicted agarose gel migration patterns, apply one or more Enzymes - Restriction Map methods, then select **Sites & Features > Agarose Gel Simulation**. The following window opens, showing the predicted electrophoretic separation of the restriction fragments:
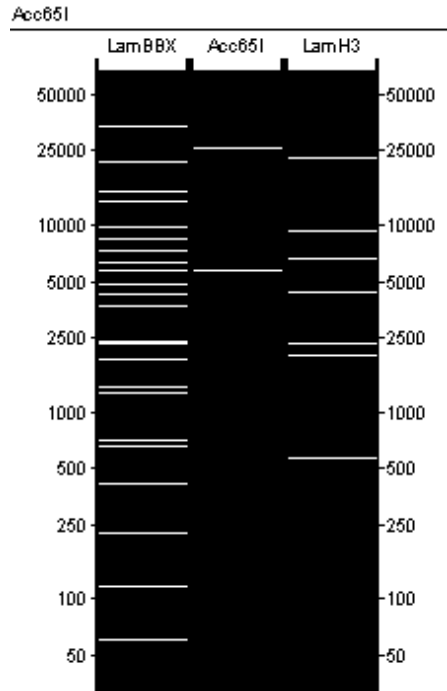
Note: Select a portion of sequence prior to using the following commands.

- To view tabular summaries showing method results, nucleotide composition or codon distribution, select **Analysis > Tabular Data**, **Analysis > Base Composition or Analysis > Codon Usage**, respectively.

- To view repetitive portions of a sequence, apply one or more Repeats methods and select **Analysis > Repeat Summary**.

- To predict how a nucleotide sequence might fold as RNA, select a <1500 bp portion of sequence and select **Analysis > Fold as RNA**:
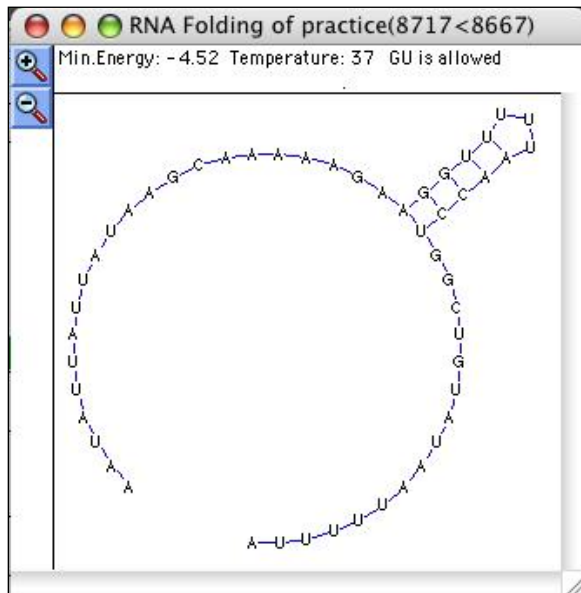
Using the microscope view

To view the sequence for any region under the mouse pointer, click ![microscope icon] move the pointer to a location. The microscope tool lets you stay zoomed out for the big picture, but still read the sequence.

Changing method parameters

Access method parameters by double-clicking on a method name or display anywhere in the GeneQuest window. Note that parameter changes made to a method in the method curtain affect all current and future displays created from that copy of the method. Similarly, parameter changes made to a display on the assay surface are applied to all displays originating from the same method in the method curtain.

To view multiple displays of a method using different parameters, add another copy of the method to the method curtain, then change method parameters separately for the added copy of the method.

Changing formatting options

Note: Click 🖐 and select one or more displays prior to using the following commands.

- To optimize the look of a selected display(s), select **Options > Line Color, Line Weight**, **Fill Pattern**, or **Fill Color**.

- To change the font and point size of text other than feature annotations, select **Options > Font or Size**.

Using method outlines

When you create a new assay document, GeneQuest uses a file called the default method outline ("*default.dao*" on Windows, "Default DNA Outline" on Macintosh) to determine which methods to place in the method curtain and which of those to apply to the assay surface. In the course of revising your assay document, you may have deleted some of the default methods, added new methods of your own, and optimized method parameters and display options. You may wish to use a similar setup in other assay documents. GeneQuest allows you to save the method outline from the current assay document and apply it to selected assay documents or to all new assay documents.

- To save the current outline as the default method outline, select **Analysis > Save As Default Method Outline**.

- To save the current method outline as a file without saving it as the default, select **Analysis > Save Method Outline**.

- To apply an outline from another document or saved as a separate file, select **Analysis > Apply Method Outline**.

Exporting to GenVision

- To export a GeneQuest assay document as a GenVision project, select **File > Export as GenVision Project**.

- A dialog will prompt you to choose the content you wish to export: **Applied Method Displays** exports all analysis methods currently applied to your assay surface; **Selected Method Displays** exports only those analysis methods currently selected on your assay surface; **Scored ORFs** exports ORFs that correspond to a Borodovsky method applied on your assay surface for the same frame.

Protean 3D

Overview

A component of DNASTAR's [Lasergene Core Suite](), Protean 3D is an integrated visualization framework designed for the needs of today's molecular biologists and geneticists. With Protean 3D's user-friendly interface, you can explore macromolecular structure and function with ease.

Protean 3D gives you the ability to work with a protein sequence and its molecular structure simultaneously on one screen. Structural features such as binding sites, disulfide bonds, helices and beta-sheets are color-coded for effortless identification. You can zoom in, move or rotate the structure, change its rendering, turn on and off the display of structural components or features, and re-color items of interest. One of Protean 3D's highlights is its ability to synchronize the selection made in one view with all other views.

Protean 3D also features an Analysis View displaying graphical views of selected "prediction methods" on a common horizontal access. These methods allow you to find PROSITE and protease pattern matches, predict B-cell epitopes, and to evaluate secondary structural characteristics such as amphilicity, charge density, disorder, flexibility and hydropathy.

The Details panel allows you to view details about the structure, sequence or selection, including 2D structure graphics. Information in this section is updated continuously to reflect the currently selected data.

Protean 3D supports PDB formatted structure files for protein, DNA and RNA. The application also lets you view multiple model PDB files from NMR experiments as individual or overlapped structures. Protean 3D also allows you to view and apply methods to sequence-only files.

In addition, Protean 3D offers a browseable and searchable Motion Library. This unique Motion Library contains over 300 annotated macromolecular structure files that use animation to show conformational changes.

10 Quick-Start Steps for Using Protean 3D

1) Launch Protean 3D and use **File > Open** to open a protein file (*.pro, *.gbk, etc.) or structure file (*.structure, *.pdb, etc.). For information on locating and downloading structure files from the Protein Data Bank (http://www.pdb.org/pdb/home/home.do), see the tutorial entitled *Downloading a file from the Protein Data Bank*.

2) Use the Molecules (**View > Explorer > Molecules**) and Features (**View > Explorer > Features**) bars to show, hide or select various structural components and features. If your document only contains a sequence, proceed to step 8.

3) If your document is a structure file, select one of the first five toolbar icons

   (), then use your mouse to rotate, move, zoom in/out on the structure, or center on a selection

4) Customize the Structure View by using the Color bar to change the color of the background or of any selection. To access the Color bar, choose **View > Style > Color** from the menu. Selections can be made from the Structure and Sequence views or the Molecules and Features bars, and are synchronized between all views.

5) Change the appearance of the structure using the Rendering bar.

6) Customize the Sequence View using the Sequence bar.

7) If you are working with a multi-model document, use the Models panel (**View > Models**) to control model visibility and animation.

8) Evaluate secondary structural characteristics of the sequence by applying prediction methods to the Analysis View.

9) Use **File > Save** to save the Protean 3D document.

10) Use **File > Export Image** to export graphics, if desired.

Protean 3D Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example Protean 3D project from start to finish. During this workflow, observe that any selections you make in a panel or view are simultaneously highlighted in the Structure View and Sequence View.

Downloading a file from the Protein Data Bank

Most of the following Protean 3D tutorials use a PDB structure file named *7TIM*. This tutorial demonstrates how to locate and download this file from the RCSB Protein Data Bank.

Note: If you would like to skip this section, you may instead proceed to the next topic and use the *7TIM.structure* file located in the following folder:

Windows 7 & Windows 8: C:\Users\Public\Documents\DNASTAR\Lasergene 11 Data\Demo Data

Windows XP: C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

Macintosh OS X: Applications:DNASTAR:Lasergene 11 Data:Demo Data

To download the file from the Protein Data Bank:

1) Using your web browser, go to the Protein Data Bank home page: http://www.pdb.org/pdb/home/home.do.

2) Enter the term *7TIM* into the search bar near the top of the window. Then click the blue **Search** icon on the right.
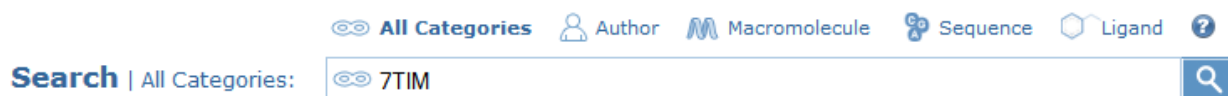
3) In the results page, locate the **Download Files** menu in the top right corner.
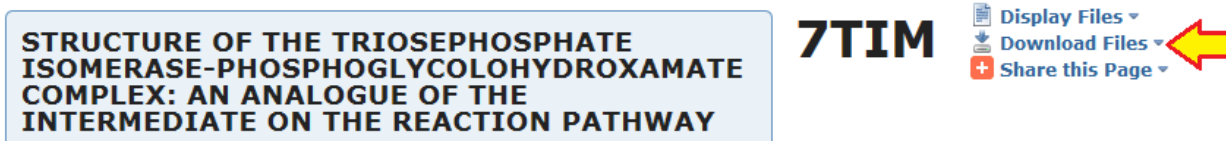
4) Click on the triangle to open the menu, and choose one of the PDB file formats ("Text" or "gz"). Once the file has been downloaded to your computer, proceed to the next tutorial.

Using the Structure View

This tutorial highlights usage of the toolbar and demonstrates how to make a selection in the Structure View.

1) Launch Protean 3D.

2) Choose **File > Open** and select the *7TIM* file you saved in *Downloading a file from the Protein Data Bank*. Depending upon the browser used and the file format chosen in that tutorial, the file may have any of several names: *7TIM.pdb, 7TIM.pdb.gz* or *pdb.7tim.ent*.

Note: If you skipped the previous tutorial, instead open *7TIM.structure* from the following folder.

Windows 7 & Windows 8: C:\Users\Public\Documents\DNASTAR\Lasergene 11 Data\Demo Data

Windows XP: C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

Macintosh OS X: Applications:DNASTAR:Lasergene 11 Data:Demo Data
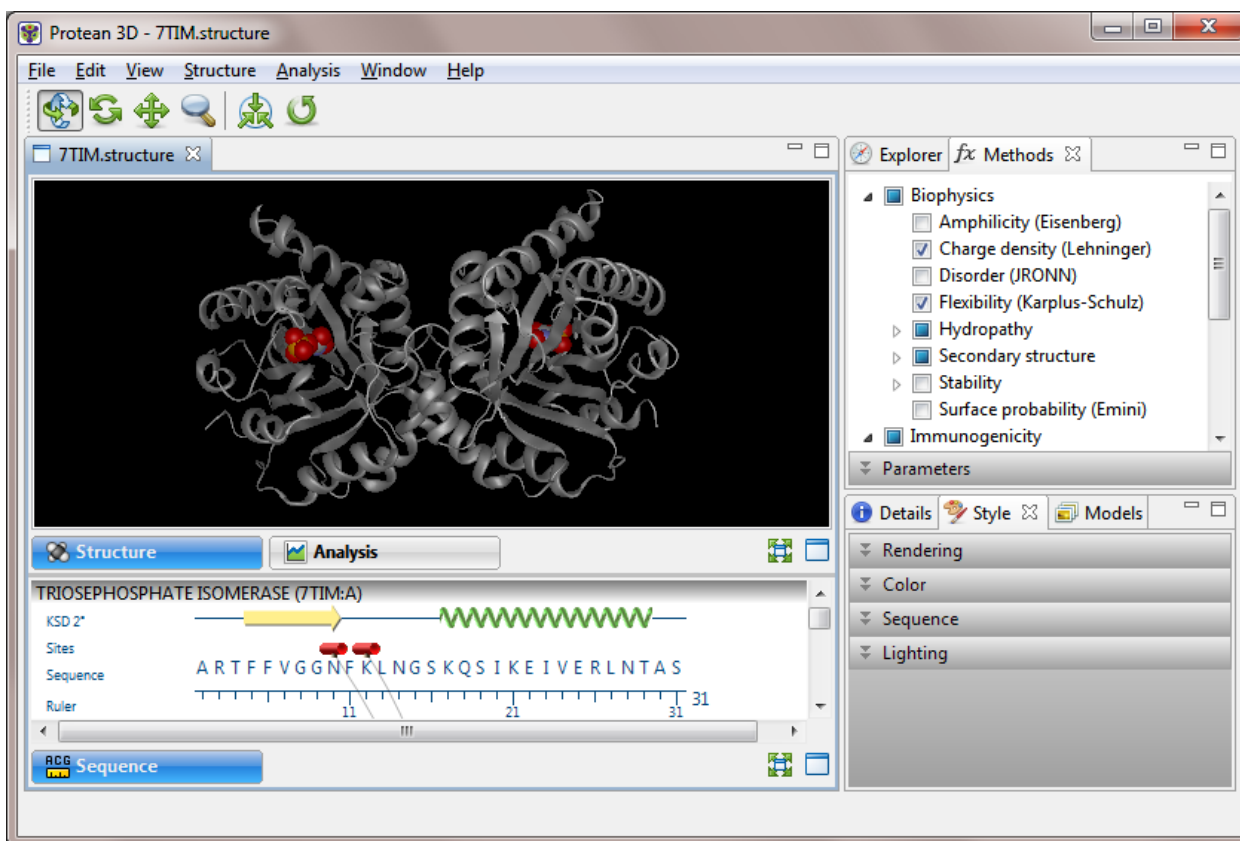
3) Zoom in and out on the structure by clicking the Zoom tool ( ) and dragging the mouse up and down.

> **Note:** If you don't see the zoom icon, use the **View > Show Toolbar** command to restore the toolbar. Another alternative is to use the commands in the **Structure** menu instead of the toolbar icons.

4) Slide the molecule partially out of view and back again by clicking the Move tool ( ) and dragging the mouse in any direction.

5) Rotate the molecule by clicking the Rotate tool ( ) and dragging the mouse in any direction.

6) Double-click on any residue in the molecule to select it. Selection is shown as a light blue highlight. De-select by double-clicking on the black background.
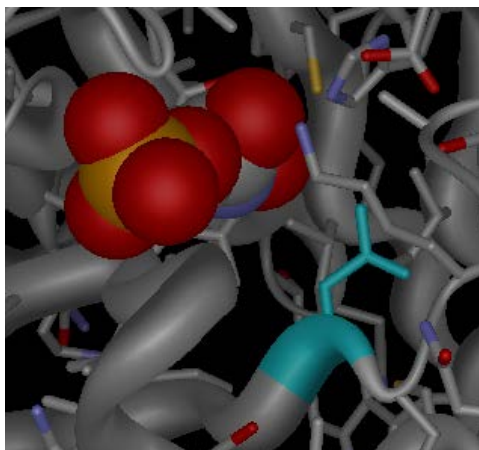
7) Return the structure to its original orientation with the Reset Structure View tool (⟳).

Changing the rendering of the structure

This tutorial shows how to change the rendering of the structure backbone.

1) Open the Rendering expand bar by clicking anywhere on the bar that says "Rendering." If the Rendering bar isn't visible, use the **View > Style > Rendering** command to reveal it.

Note: Throughout these tutorials, you may open a closed expand bar (or vice versa) by clicking on the bar bearing its name. If a panel or expand bar is not visible, reveal it by selecting the corresponding command from the **View** menu.

2) Select "Ribbon" from the **Backbone** drop-down menu.
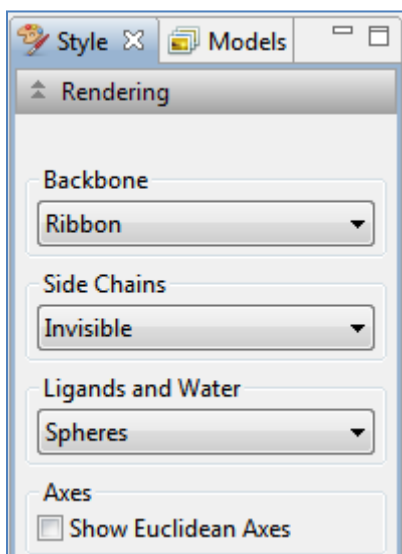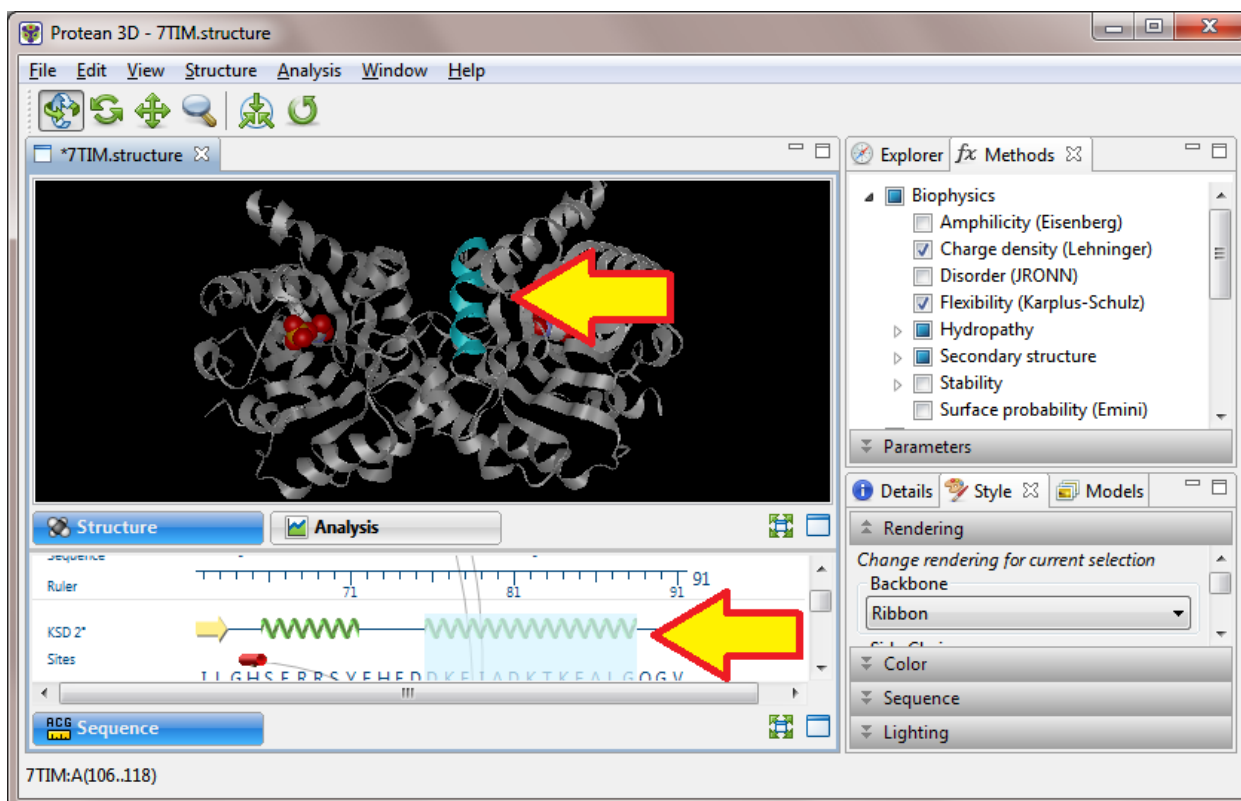
The structure now appears as follows:

Using the Sequence View

This tutorial demonstrates how to make selections from the Sequence View.

1) In the Sequence View, select any helix (represented by a green spiral) by clicking on it. The selected helix will turn light blue. Note that the corresponding helical element also appears highlighted in the Structure View. (If you don't see a highlighted element in the Structure View, try zooming in on the molecule or rotating it for a better view).

2) Click on any red tube icon in the Sequence View to simultaneously select all residues involved in a binding site. Note that the corresponding binding site also appears highlighted in the Structure View.
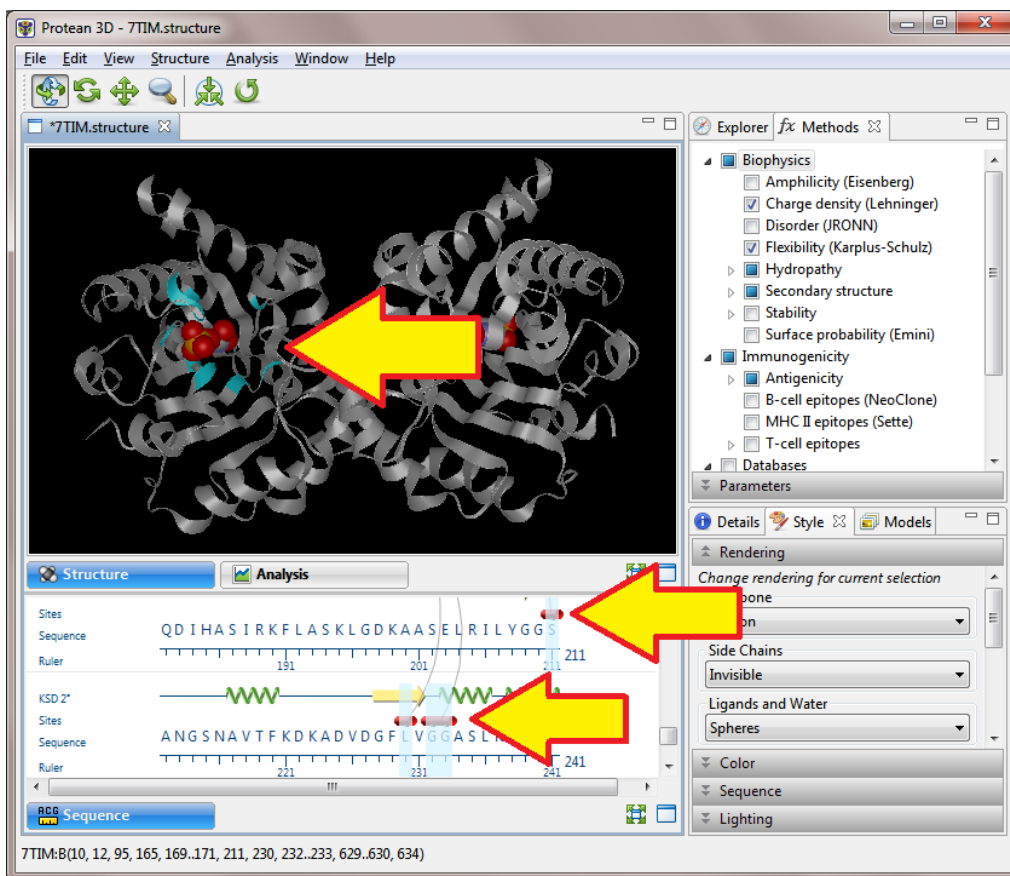
Selecting and hiding a structural component

This tutorial shows how to select a particular structural element from the Molecules bar.

1) Open the Molecules bar using the command **View > Explorer > Molecules**.
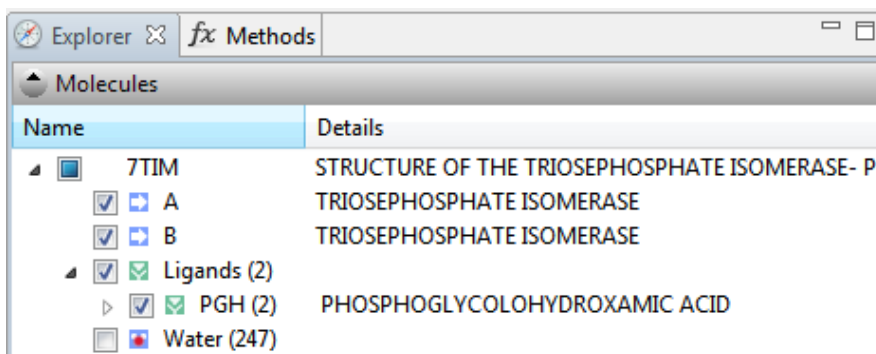
Notice that, by default, all boxes in the Molecules bar are checked except for **Water**.

2) Select the first *Triosephosphate Isomerase* subunit by clicking on the letter "*A*" (not on its checkbox). Observe the corresponding selections in the Structure and Sequence Views.
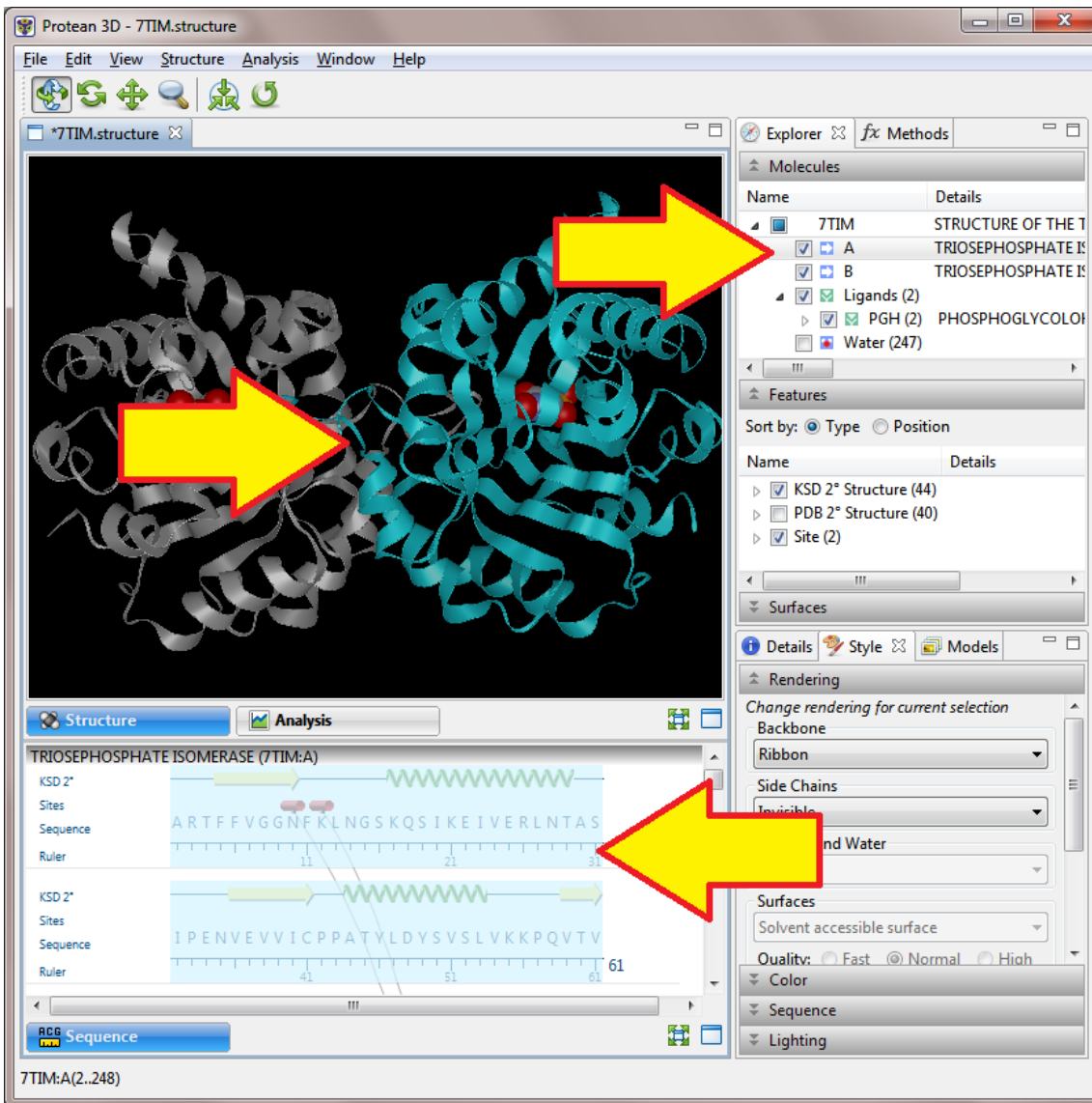
3) Next, uncheck the box to the left of "*A*." Observe that the analogous *Triosephosphate Isomerase* is now hidden in both Structure and Sequence Views.
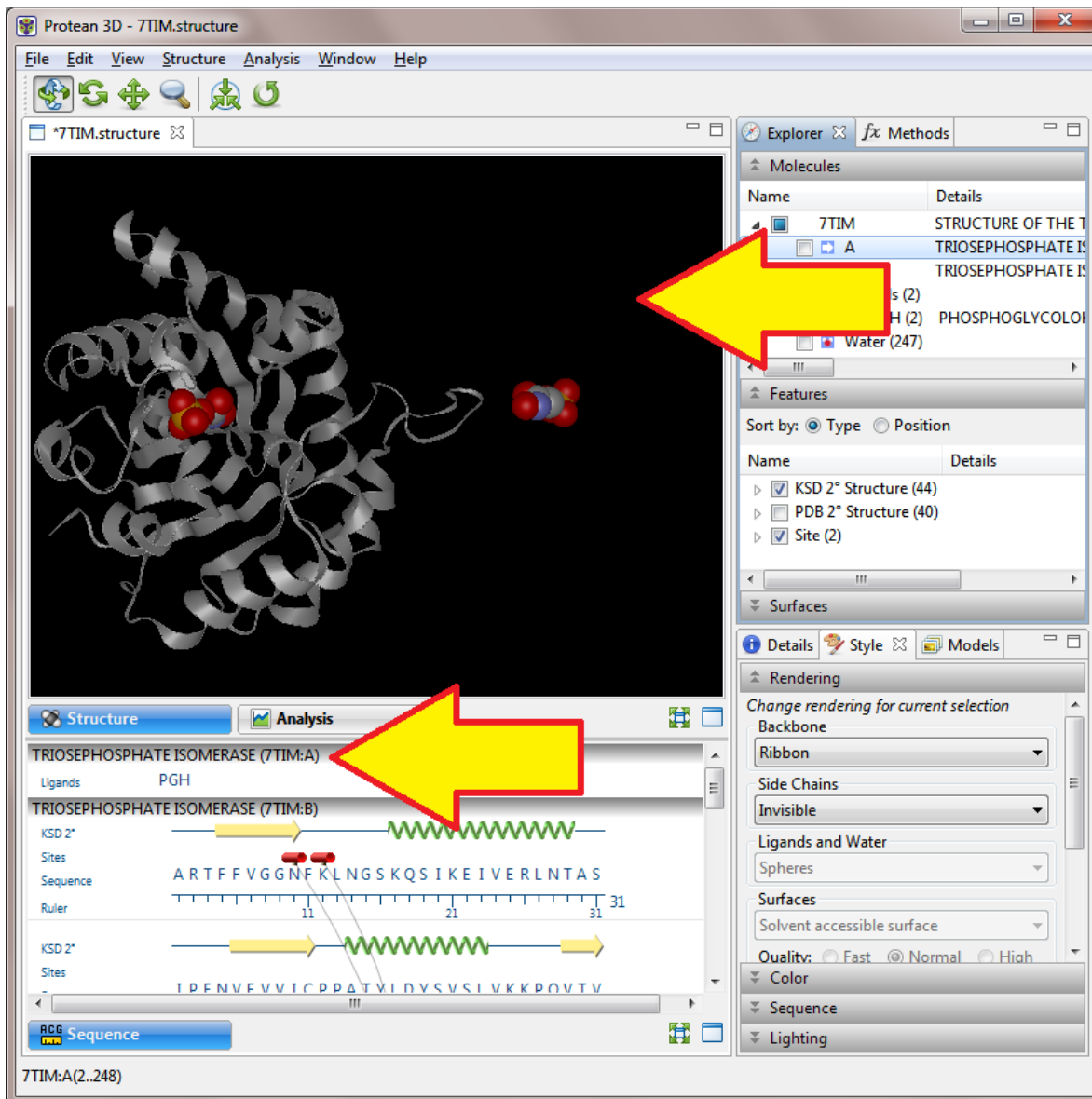
4) Recheck the box so that all boxes except **Water** are checked.

Recoloring a feature

This tutorial demonstrates how to color a structural element and its related feature light green.

1) Maximize the Protean 3D window and open both the Molecules and Features expand bars so that you can see their contents at the same time. To reveal more of the Features bar contents, you may wish to drag the Style panel downwards. You can also use the vertical scrollbar to view more of the expand bar's contents.
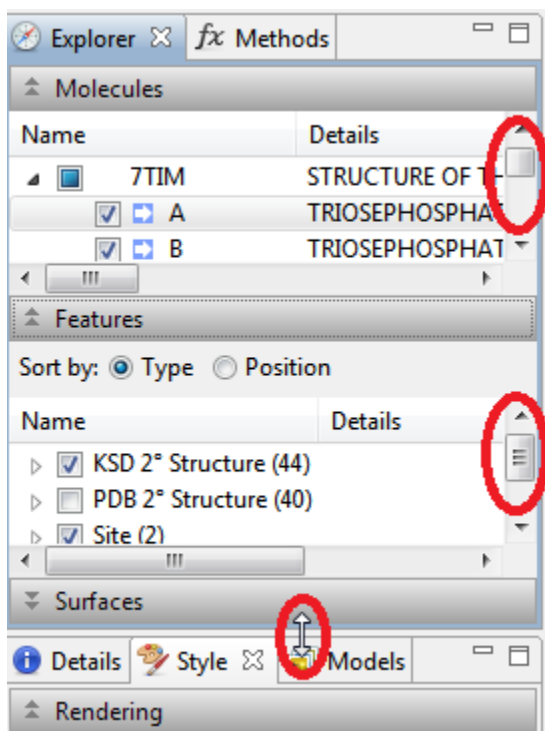
2) In the Molecules bar, click on the expand arrow to the left of the **Ligands** *PGH* checkbox to reveal *PGH [A:249]* and *PGH [B:249]*.

Note: The expand arrow may be hidden until you hover over it with the mouse.
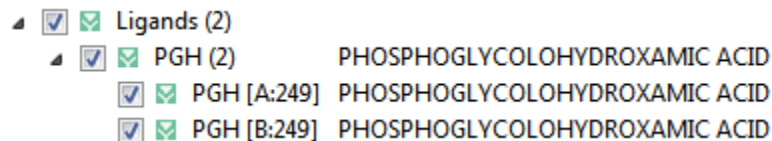
3) In the Features bar, click the expand arrow to the left of the **Site** checkbox, as well as the expand arrows one level down, to reveal features *AC1* (binding site for residue *PGH A:249*) and *AC2* (binding site for residue *PGH B:249*).

4) Now you will select the *B:249* ligand and binding site simultaneously. In the Molecules bar, click on the text "*PGH [B:249]*" to select the ligand. Hold down the **Ctrl** key (Win) or ⌘ key (Mac) and click on the text "*AC2*" in the Features bar to select the related binding site.

5) To color the site and feature, open the Color bar by selecting **View > Style > Color** from the menu. In the Color bar, click on the box under the word **Fill** to open the color picker for your operating system.
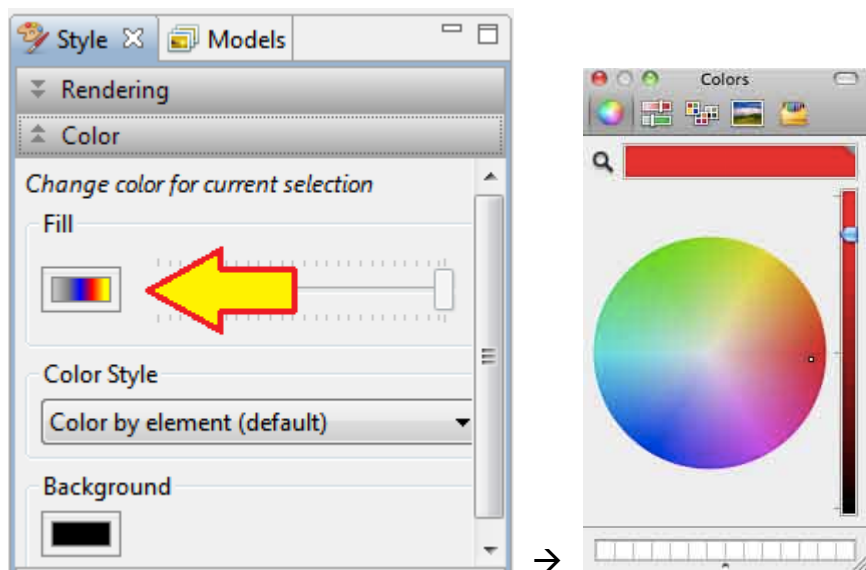
On Windows, choose any shade of light green and click **OK** to close the window and return to Protean 3D. On Macintosh, you must close the color picker after selecting the light green color.

6) In the Structure View, the selection initially remains highlighted in blue. De-select it by double-clicking on the black background. The *AC2* ligand and the associated binding site now appear light green.
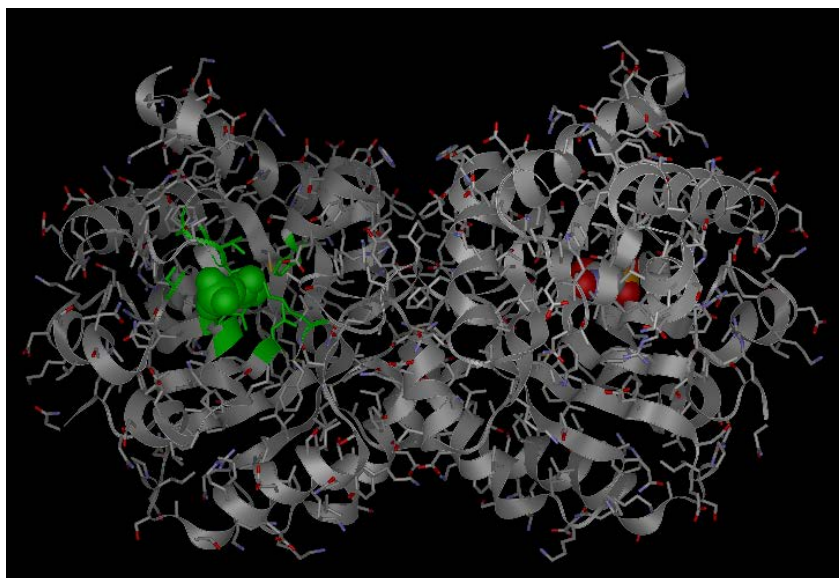
Viewing structure and sequence details

This tutorial demonstrates how to view detailed information about a structure or sequence.

1) If you have not already double-clicked on the background of the Structure View, do so now.

2) Open the Details panel by clicking on it, or by choosing **View > Details**.
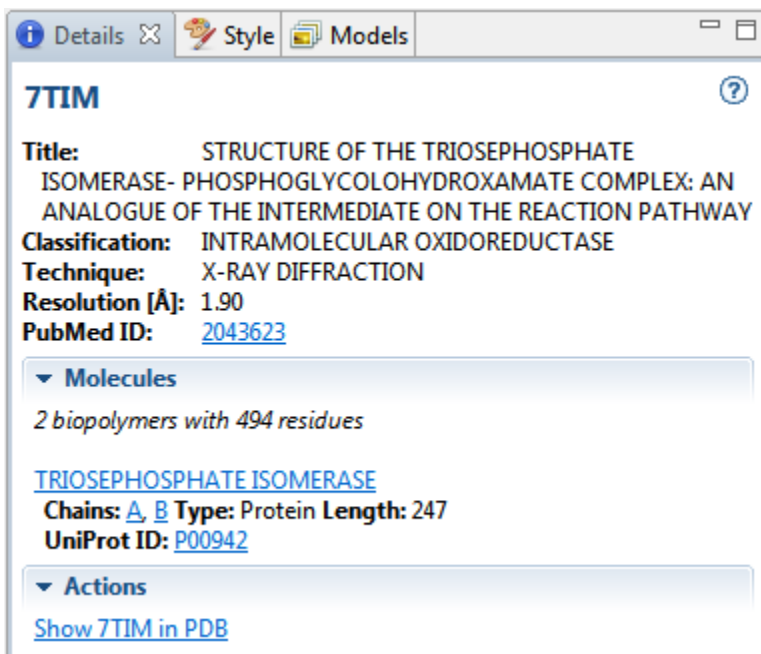
The Details panel displays a different subset of information depending on your selection. Because no specific selection has been made here, the panel displays information about the structure as a whole.

Note the hyperlink in the Actions bar that allows you to view the structure in the PDB website.

3) Click on any hyperlink to open a web page relevant to the structure. After viewing the web page, return to Protean 3D.

4) In the Molecules bar, click on ⬚ A TRIOSEPHOSPHATE ISOMERASE to select that molecule in all views. Note that the Details panel now contains different expand bars from before, including bars for Biophysical Properties, Structural Properties, and Features in Selection. The hyperlinked task in the Actions bar has also changed, and now allows you to export the selection details to a file.
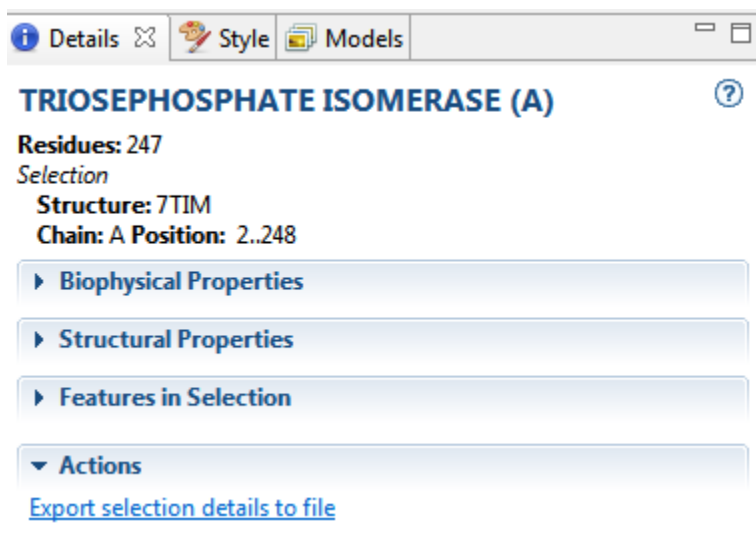
5) Click on the triangles to the left of each bar to reveal more details.

You have now observed what the Details panel looks like when you do not make any selection, and when you select a molecule, feature or region. If you wish, try making the following selections and note how the panel changes:
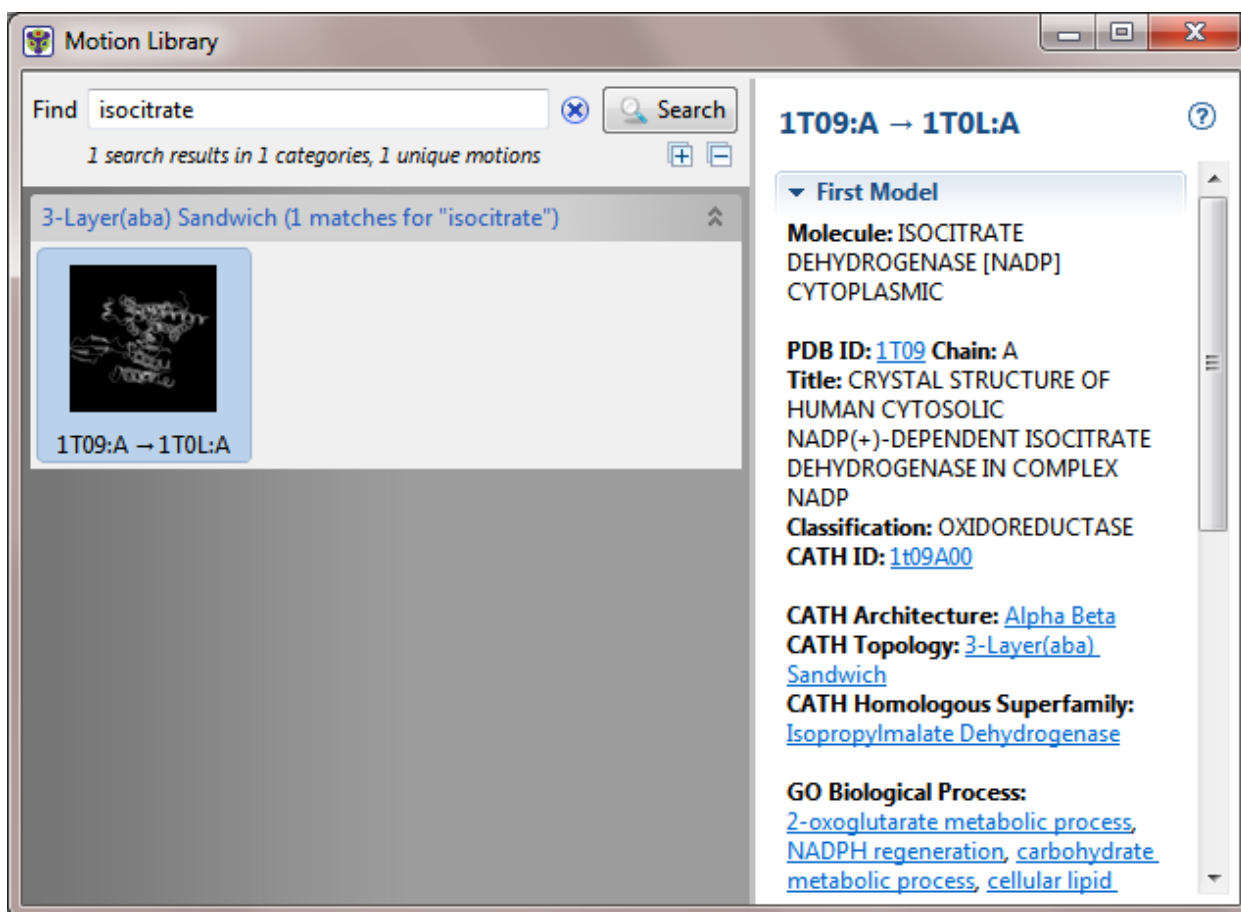
- An entire structure

- A group of structures (see *Opening and aligning multiple structures*)

- A single residue or ligand

- A method (from the Methods panel)

- A PROSITE or protease pattern or profile (from the Methods panel)

- An individual PROSITE or protease match (from the Analysis View)

Working with a multi-model file

This tutorial shows how to open and work with a multi-model file from the Motion Library.

1) Choose **File > Open from Motion Library** to open the floating Motion Library window.

2) In the Find bar, enter *isocitrate* and click **Search**. One match will be shown. Notice that the animated thumbnails that did not match the search term are now hidden.

3) The panel on the right displays detailed information about the first and last models in the animation. Hyperlinks lead to additional data about the molecule in the PDB, CATH, and GO databases.

Double-click on the thumbnail to open the animation as a new document. If you have not previously downloaded the Motion Library, you will be prompted to download the file you selected:
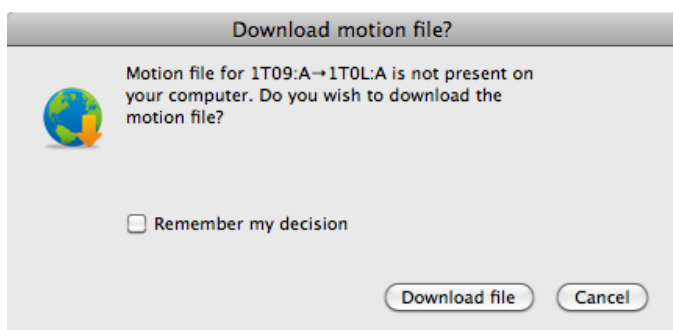

Check the **Remember my decision** box if you would like to download individual Motion Library files in the future without being prompted. Then click **Download file**.

4) In the Models panel, move the slider to the left to slow the animation.
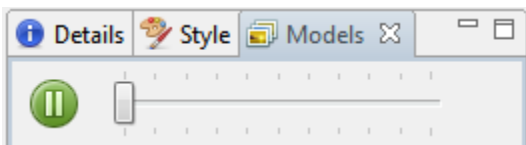
5) Click the **Pause** button ( ) to stop the animation. The **Model** last displayed in the Structure View as a "movie frame" remains checked and in view.

6) Check a second **Model** box to overlay the two models in the Structure View.
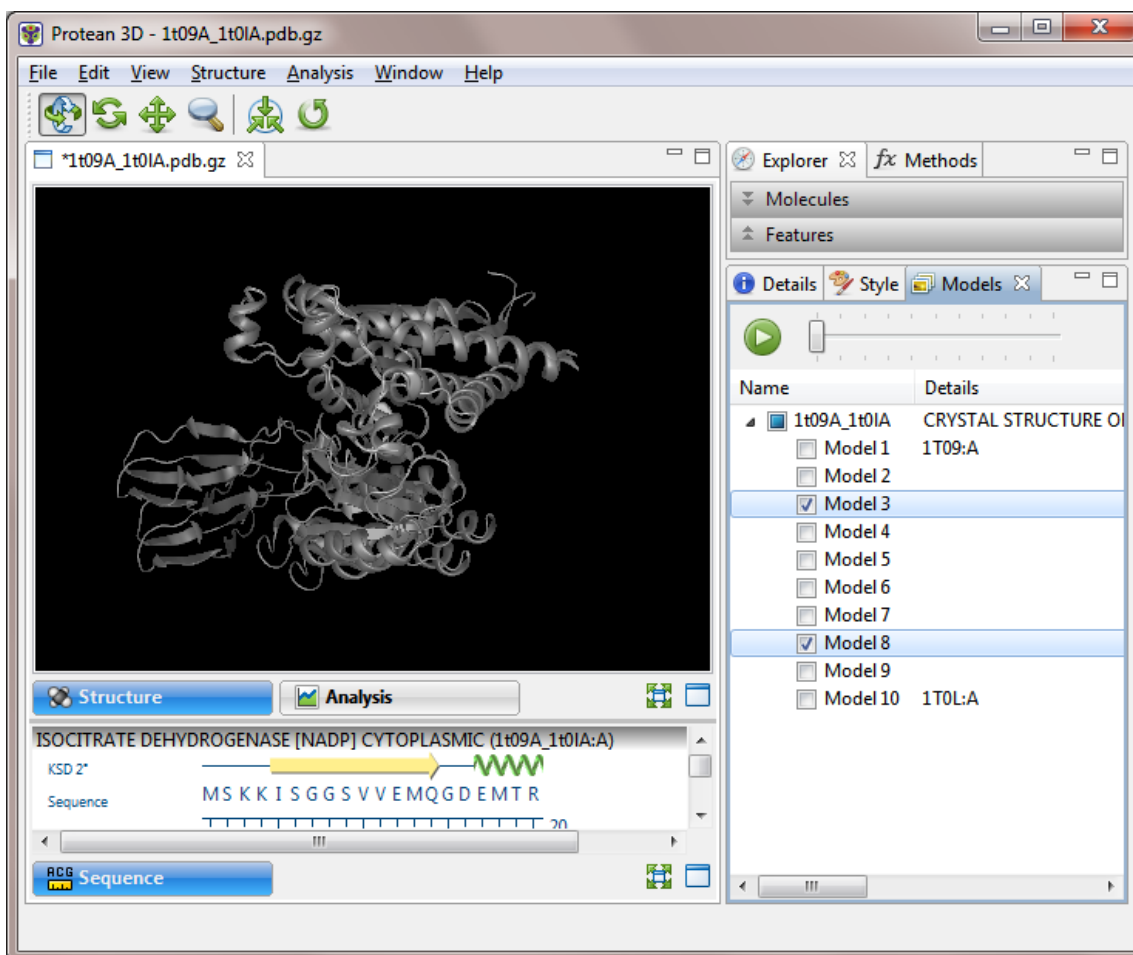
Overlays are helpful for checking where multiple models in NMR files agree and disagree. They can also be used as presentation graphics to show the path of motion in a movie file.

You can also perform the same tasks with multi-model files as you did in earlier tutorials with the single-model file *7TIM*, including selecting, recoloring or hiding various structural elements and features, and changing the structure rendering.
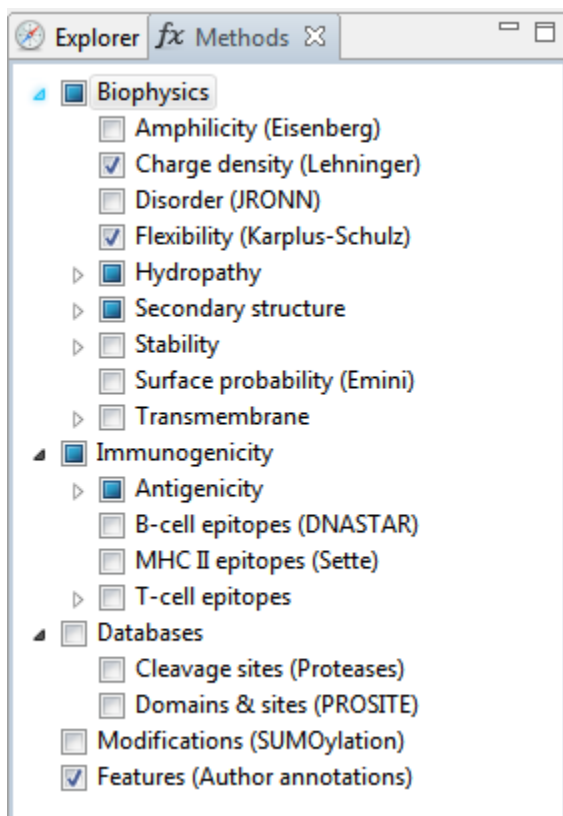
Applying methods to the Analysis View

Protean 3D's Analysis View allows you to apply a variety of "prediction methods" to your protein sequence. In this section and the following ones, we will locate potential calcium-binding regions for a human calmodulin sequence. We know from the literature that these regions are correlated with predicted turn regions and with flexible regions.

1) Launch Protean 3D if it is not already open.

2) Choose **File > Open** and select the protein sequence file *Human Calmodulin.pro* from the demo data folder:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OS X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

The Analysis View opens with some methods already applied by default.

3) To see a list of the applied methods, select **View > Methods > Methods** from the menu.
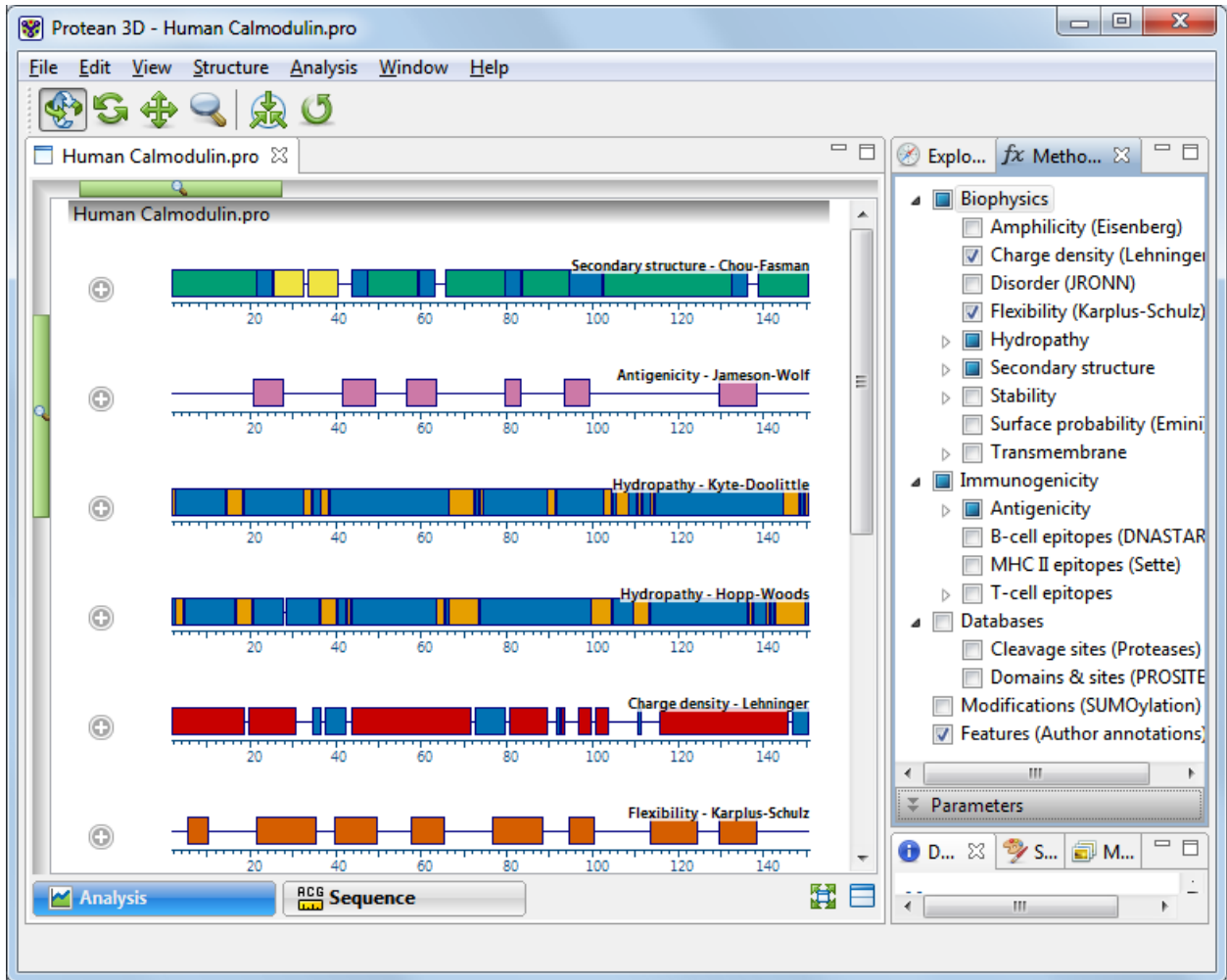
Maximize the Analysis View by clicking its Remove Split tool (⬜).
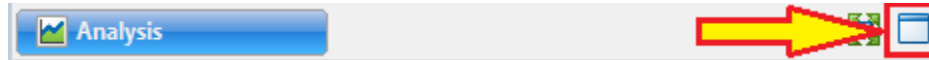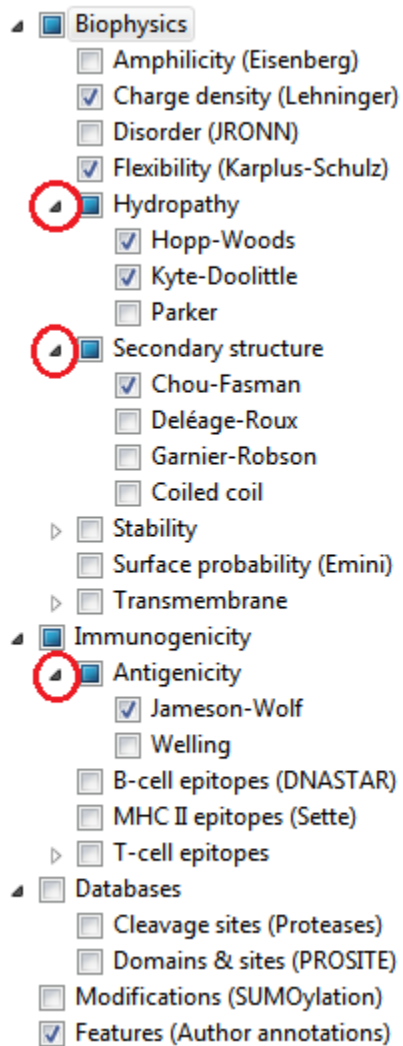

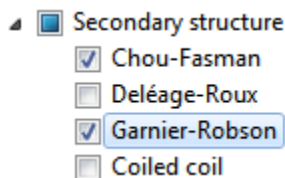If the NovaFold panel is open, close it by clicking on its 'x.'

The Protean 3D document now appears as follows:

4) In the Methods tab, click the expand arrows to the left of some methods to reveal their sub-methods:

5) Add the Garnier-Robson method to the Analysis View by checking the box to its left.

6) Double-click on **Garnier-Robson** in the Methods bar to open its parameters. Change the **Method version** from **GOR II** to **GOR I**.
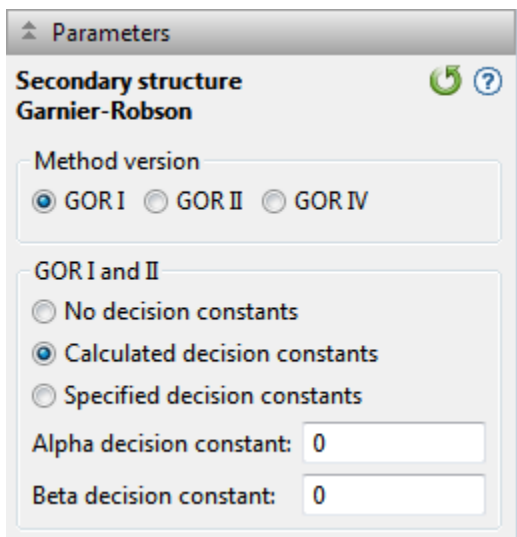
7) In the Analysis View, reorder the methods so the first three methods are Garnier-Robson, Chou-Fasman and Flexibility-Karplus-Schulz. To do this, double-click on each method synopsis, and drag and drop it in the desired location.
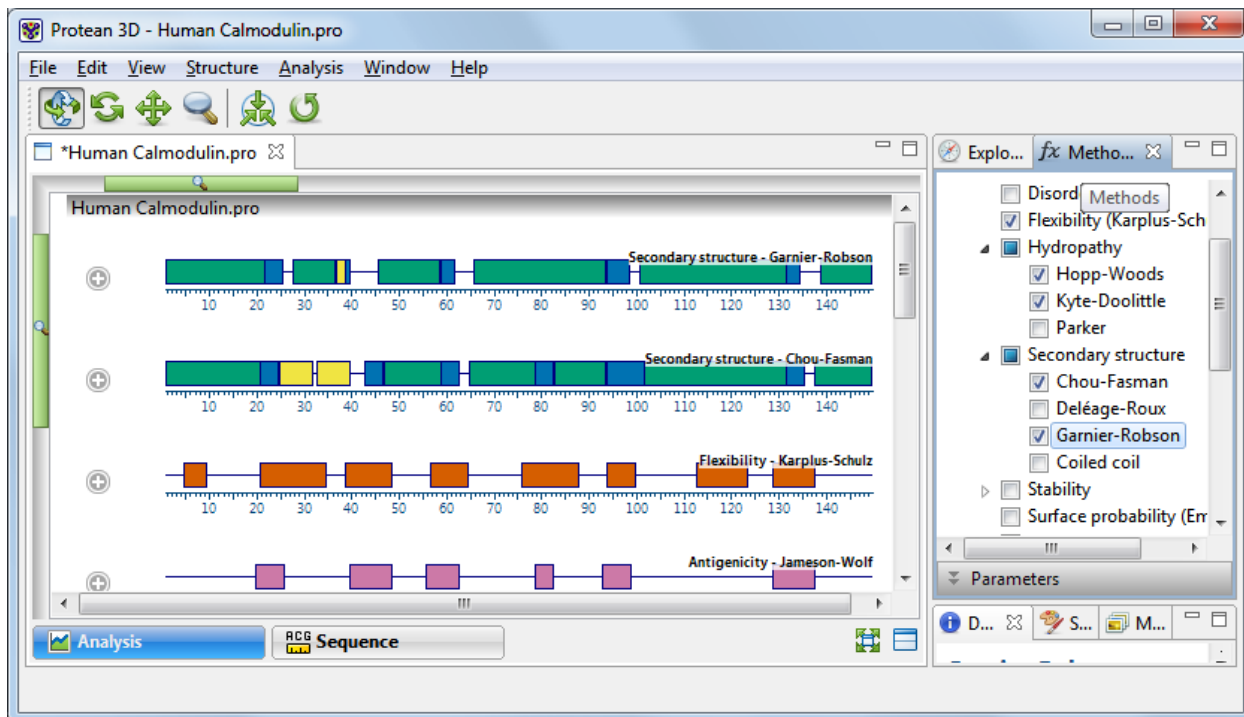
Comparing results from two analysis methods

This tutorial demonstrates how the results of one method can be corroborated by looking at a similar graph from another method.

1) So far, only the synopsis row for each method is visible in the Analysis View.
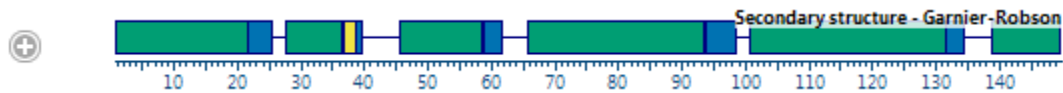
To view the individual graphs, click the plus signs (⊕) to the left of the first three synopses. The individual method graphs are now displayed below each synopsis. The name of each graph appears on the right of the Analysis View.
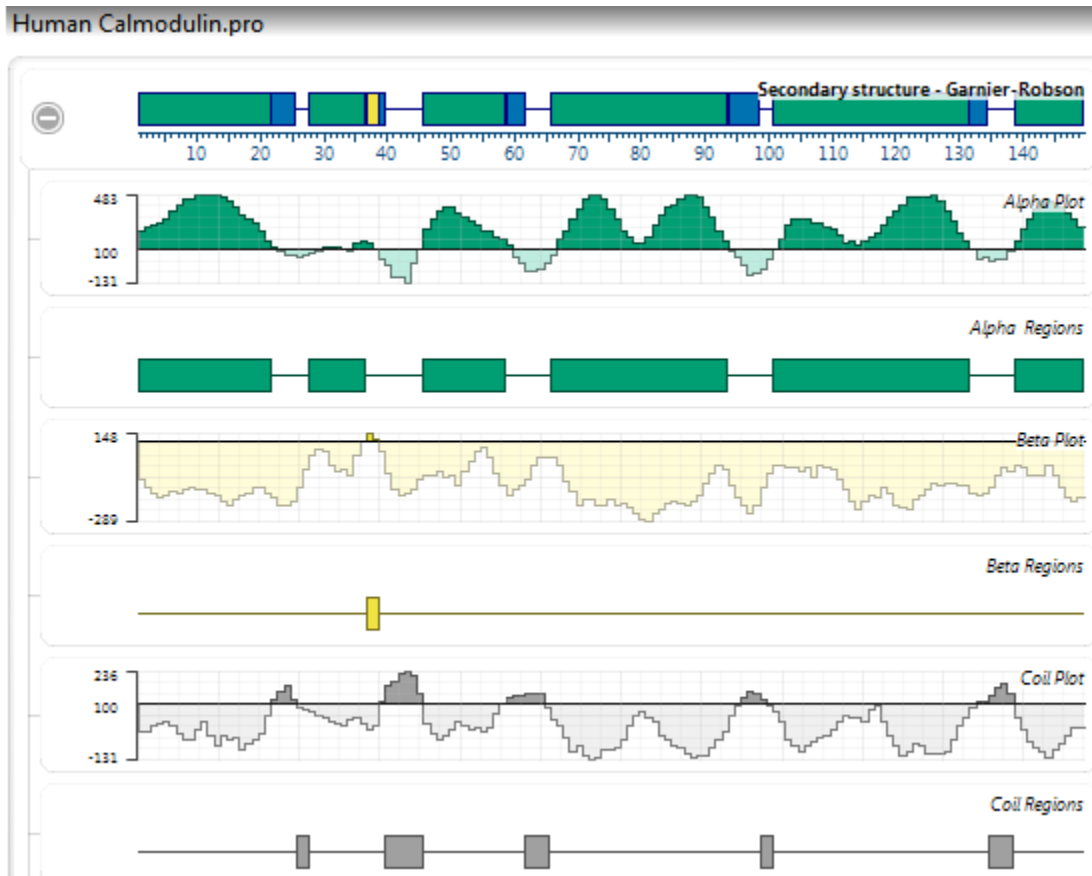
2) To rescale the display so that you can view it more comfortably, use the zoom sliders located on the top and left sides of the Analysis View.
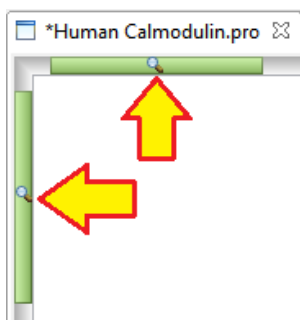
3) Examine the Garnier-Robson and Chou-Fasman "Turn Region" plots. These plots show the predicted turn regions, and thus, the probable calcium binding regions for the protein sequence. (The plots are shown as adjacent in the graphic below, but they are not actually adjacent in the Analysis View).

(Garnier-Robson)
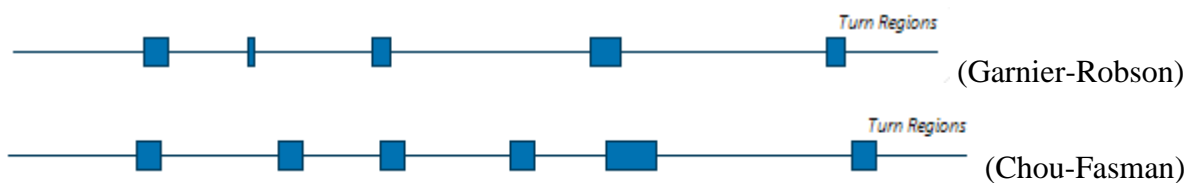
(Chou-Fasman)

4) To verify which turn regions are the same in both methods, select the Chou Fasman "Turn Regions" by holding down the **Ctrl** or **Cmd** key while clicking once on each of the six blue boxes.

5) Visually follow the highlighting upwards to see corresponding turn regions highlighted in the Garnier-Robson method.

Of the six turn regions found in Chou-Fasman, four are nearly exact matches (shown by red boxes, below) with the turn regions in Garnier-Robson, and another is a close match (shown by red circles, below). Thus, the Garnier-Robson method provides partial corroboration of the turn regions from the Chou-Fasman method.
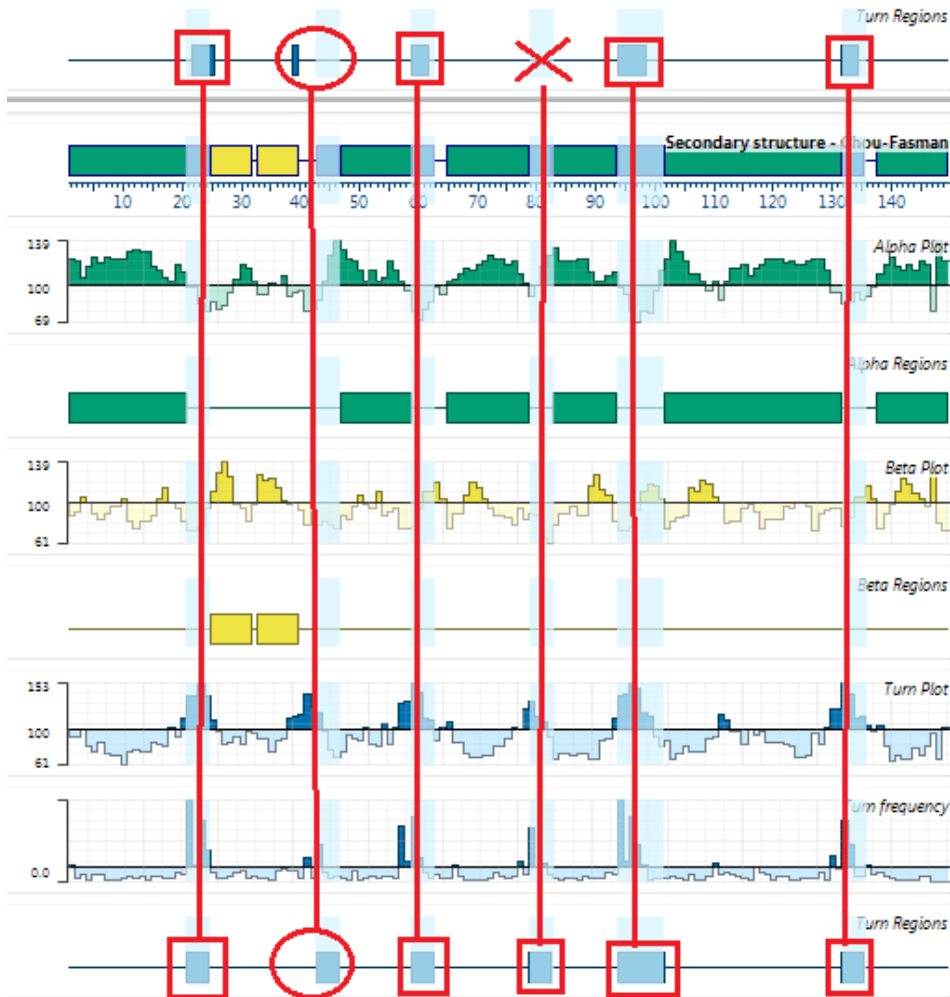
Changing analysis method parameters

Scroll down to the Flexibility-Karplus-Schulz graph named "Flexibility Regions." Note that under the current (default) parameters, there are 8 boxes in this graph compared to 6 for Chou-Fasman "Turn Regions":
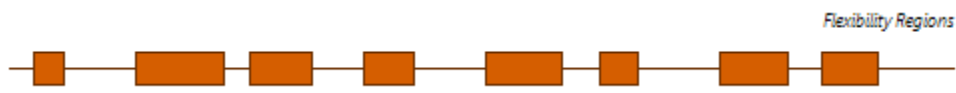
By making the parameters more stringent, we can find which of these areas are most likely to be flexibility regions.

1) Open the parameters for the Flexibility-Karplus-Schulz method by going to the Method Bar and clicking on the words "Flexibility (Karplus-Schulz)." The parameters are displayed in the Parameters bar:
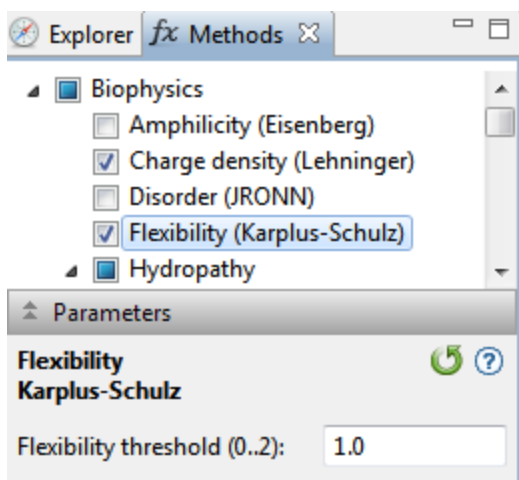
2) Change the **Flexibility threshold** to "1.07."

Note that the Flexibility Regions graph now corresponds almost exactly with Chou-Fasman Turn Regions selected in the previous tutorial. This provides additional confirmation that these regions may correspond to calcium binding regions.

3) Return to the default parameters by clicking the **Restore Default Values** button ( ).

4) Remove the Flexibility (Karplus-Schulz) method from the Analysis view by removing its checkmark in the Methods panel.

Recoloring graphs in the Analysis View

We will now recolor an Analysis View graph.

1) Double-click on any green **Alpha Regions** graph to select it.

2) If it is not already open, open the Color bar by clicking on it. (If the Color bar is not visible, first choose **View > Style > Color** to reveal it).

3) Click on the orange box under **Fill**, and choose bright purple from the color picker.
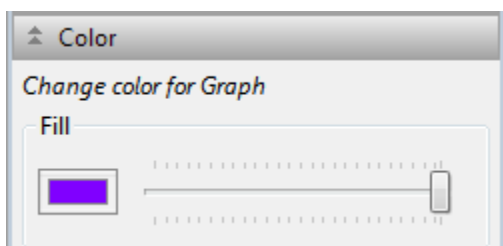
Since the selected graph is of the "alpha helix" type, all graphs of the same type will be simultaneously re-colored bright purple, including graphs in other methods.
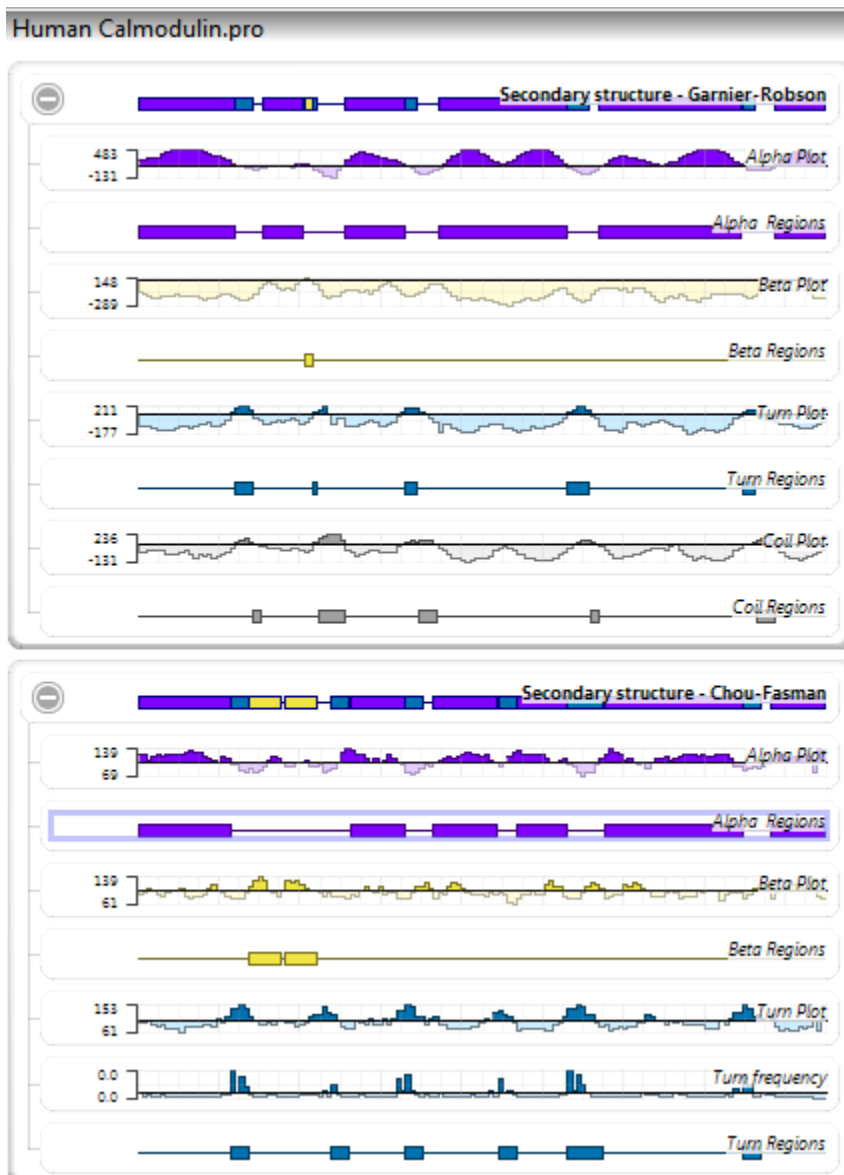
Collapsing rows in the Analysis View

If many individual method graphs are displayed at the same time, the Analysis View can become crowded. To instantly close all of the individual graphs, and display only the synopsis rows, use the **Analysis > Collapse All Rows** command.

Opening and aligning multiple structures

This tutorial demonstrates how to align two structurally homologous protein files within a single document. This technique is useful for discovering similarities and differences between related structures.

1) Using the method described in *Downloading a file from the Protein Data Bank*, download and save the following two structures as PDB files. ("Text" or "gz" versions are equally suitable). We suggest saving the files to your desktop for ease of locating them in Step 3.

   - *1BFD* – The crystal structure of benzoylformate decarboxylase

   - *1PVD* – The crystal structure of pyruvate decarboxylase

   These two proteins exhibit structural homology, although they have only a 19% sequence identity. Their structural homology may indicate either distant sequence homology or convergent evolution.

2) In Protean 3D, choose **File > Open Group** to launch the Open Group dialog.

3) Keep the default setting of **Center on-screen to align similar structures**.

4) Click the **Add** button. Navigate to the *1BFD* file that you saved and double-click on it to add it to the Files section of the dialog.

5) Repeat the previous step for the *1PVD* file.
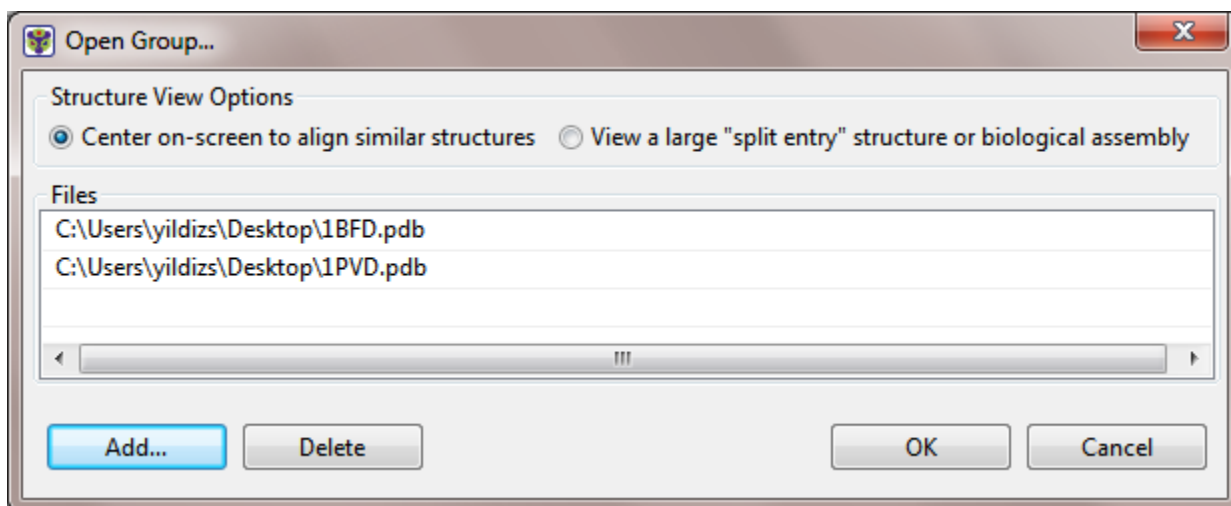
6) Click the **OK** button.

7) Now that both structures are open in a single document, we will color them by chain to facilitate their comparison. First, double-click on the black background of the Structure View to remove any existing selection. Then choose **View > Style > Color** to open the Color bar. From the Color Style drop-down menu, select **Color by chain**.
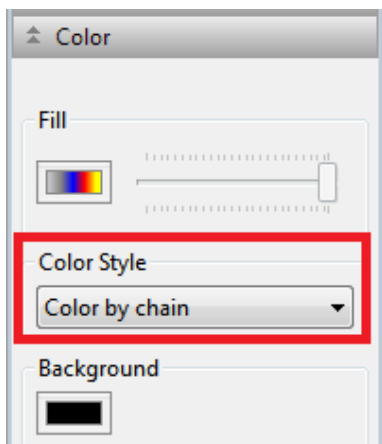
The *1BFD* structure now consists of a single green chain, while the *1PVD* structure consists of two chains: one blue and the other yellow. Since we have not yet aligned the structures, the three chains overlap randomly.
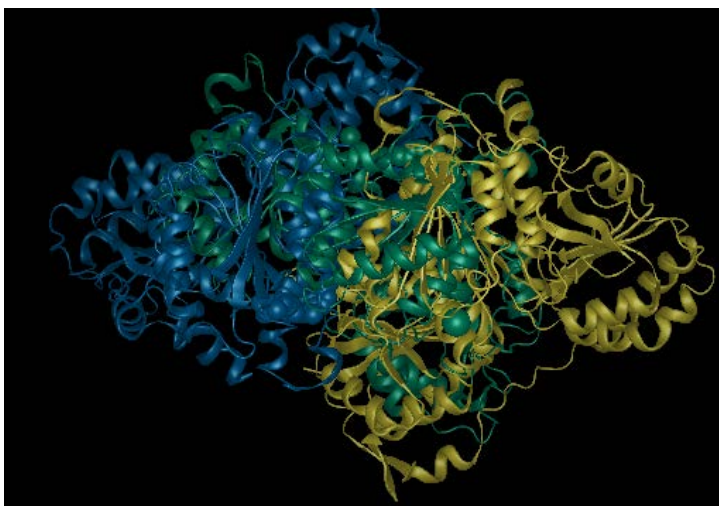
8) To select the chains you wish to align, open the Molecules panel using **View > Explorer > Molecules**. Using **Ctrl+click** or **Cmd+click**, select both the "A" chain from *1BFD* and the "B" chain from *1PVD*.

9) To align the selected chains, choose **Structure > Align Structures** > **Rigid-body Alignment for Selected Regions**. During alignment, the following graphic will appear in the lower left corner of the Structure View.
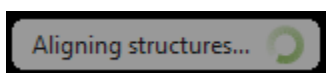
10) When alignment is complete, double-click on the black background to clear the selection. Observe that chain A of protein *1BFD* (green) is now aligned to chain B of protein *1PVD* (blue).

11) Open the Details panel using the **View > Details** command, and look for the Root Mean Square Deviation (RMSD). Note that the RMSD value for the two proteins is 3.079 Å.

RMSD quantifies the degree of similarity by describing the average distance, in Angstroms (Å), between the atoms of two superimposed structures. The RMSD value is inversely proportional to the degree of similarity between the structures.

As a very general rule of thumb:

- RMSD ≤ 0.5 Å indicates near-identical structural similarity.

- RMSD between 0.5-2.0 Å corresponds to a high degree of similarity.

- RMSD between 2.0-5.0 Å indicates a lower, but still noticeable, degree of similarity.

The RMSD in this case is 3.079, which is compatible with our knowledge that the two sequences have a 19% sequence identity.

Further Exploration

Detaching and reattaching a document view

To detach the Structure, Sequence or Analysis Views and turn them into free-floating windows, use the **View > Detach Document View** command, or right-click on the name of the view and choose **Detach View**.
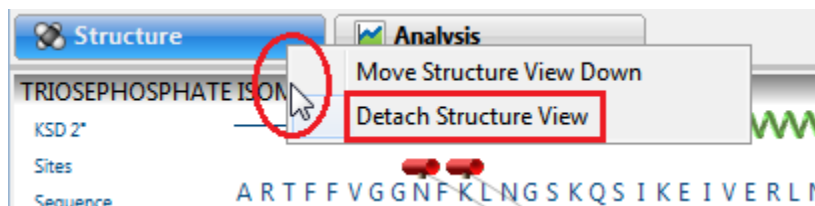
To reattach a detached view to its original location, click the 'X' in the top corner of its window. To reattach *all* detached views, choose **View > Reattach All Views**.

Exporting an image

When working with a structure file, you can export the Structure View image as follows:

- To save an image from the Structure View as a still image, choose **File > Export Image**. Still images can be exported as high-resolution *.png, *.jpg or *.gif files. This method can also be used to export a still image from an animated sequence.

- To save an animated sequence in QuickTime Movie (*.mov) or AVI format, choose **File > Export Movie**. Both formats are saved to play back at 6 frames per second (166 ms/frame), which is one-quarter the speed of standard television programs and films. This speed was chosen because it is optimal for viewing Motion Library files, most of which are 10-30 frames in length.

- To save as an animated *.gif choose **File > Export Image**. Animated *.gifs are exported with a frame rate of 10 milliseconds/frame.

GenVision Standalone

Overview

GenVision™ is a powerful tool for creating publication-quality images depicting large and complex data sets. The tutorials in this document are intended to familiarize you with much of the functionality GenVision has to offer.

To learn more about specific features, please consult the GenVision help that was installed with your software by selecting **Help > Contents**.

Note: In addition to being a Standalone application, GenVision is supplied as a Plug-In for Adobe® Illustrator® on both Windows and Macintosh platforms. The tutorials in this document are intended to be used with the GenVision Standalone application only. For additional information about the GenVision Plug-In, or to access the GenVision Plug-In tutorials, please visit the DNASTAR website.

GenVision Standalone Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example GenVision project from start to finish.

In order to complete the two tutorials in this document, *Designing a Circular Map* and *Designing a Linear Map*, you will need the sample data that was installed automatically with your software in the following directory:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo GenVision

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\ Demo GenVision

- **Macintosh OS X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data: Demo GenVision

Note: Throughout the GenVision Standalone tutorials, to see the most up-to-date image preview after editing a value or reordering panels, click the **Run GenVision** button ( ).

Designing a Circular Map

Objective: The purpose of this tutorial is to familiarize you with building a circular map image in GenVision by using the 4,639,221 bp *E. coli* genome.

Setting up a new circular map image

Objective: To set up a circular map figure for the *E. coli* genome.

1) Launch GenVision if it is not already open. The **Map Data** pane appears, with the **Circular Map** radio button selected by default.
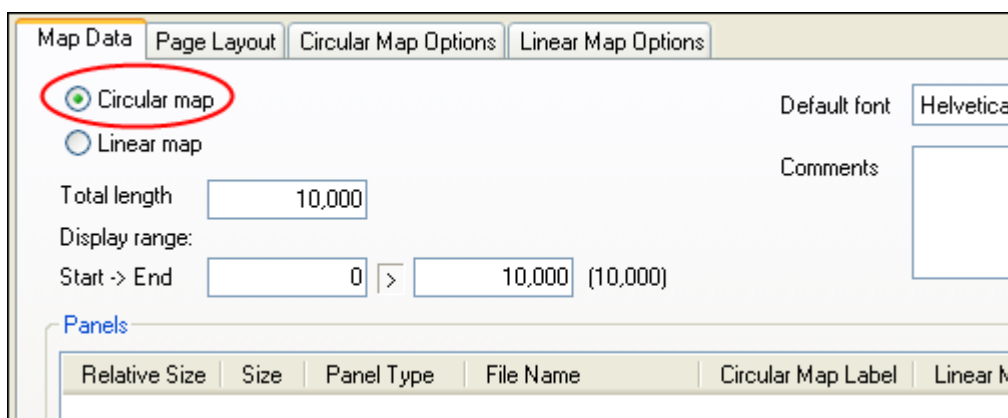
2) Enter **4,639,221** (without commas) into the **Total length** field and then click in the Display range **End** field. A message will appear, prompting you to save the new project you created:
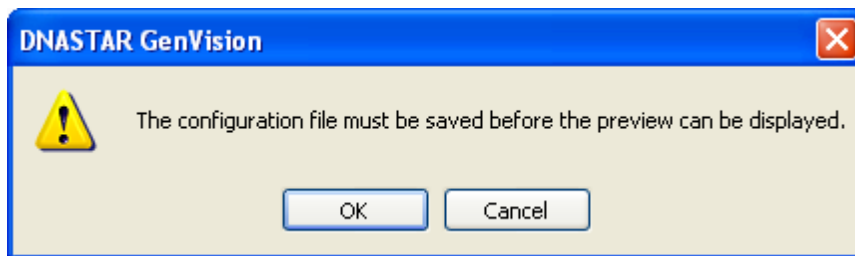
3) Click **OK** in the message dialog.

4) Choose the directory where you want your project to be saved, enter the filename **Ecoli Circular Genome**, and click **Save**. A blank PDF document with the center title "Untitled Genome" will be displayed.

5) From the Map Data pane, enter **4,639,221** (without commas) in the **End** field so that the entire length of the genome will be displayed in the image. Then click on the **Circular Map Options** tab. The PDF image will be updated, although no visible changes will be shown at this point.

6) From the **Circular Map Options** pane, change the **Inside radius** value to **1.25**, but leave the **Outside radius** at **3.5** inches. This will expand the map so that it nearly fills an 8.5" x 11" page.
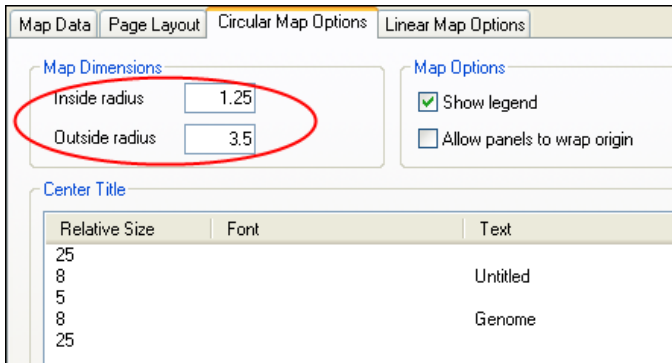
7) Also, make sure that the **Show legend** box is checked. This will ensure that a legend defining each panel in your image will be created and displayed in the bottom right corner of the circular map. The PDF image will be updated again to include the changes you just made, although you will not be able to see any visible difference since we have not yet added any data to the project.

Defining foreground and background colors

The **Page Layout** pane allows you to control many aspects of how your image will be displayed, including the foreground and background colors of the image. In circular maps, the foreground color defines the color used for page numbers, the text for page legends, and also sets the initial default color for each panel type. The background color is the color that will be used in areas where no other image elements are specified.

Objective: To define the foreground and background colors that will be used for the image.

1) Click the **Page Layout** tab to open the **Page Layout** pane:
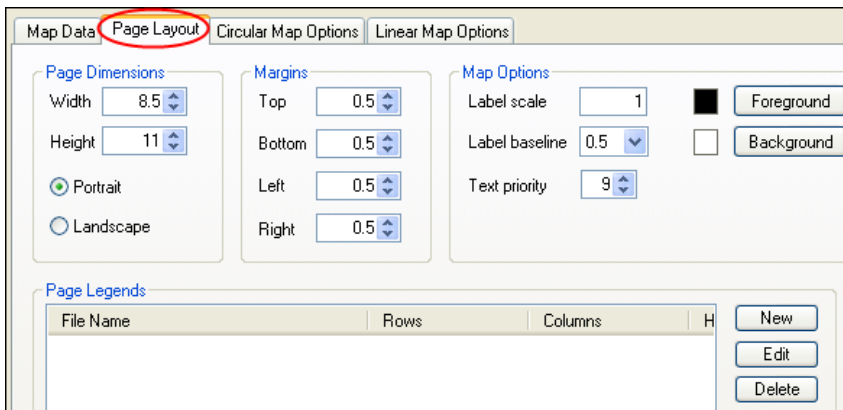
2) Click on the **Foreground** button to display the Color dialog.

3) Select a **bright green** color, like the one circled above, and then click **OK**. The preview of what will become your circular map image will be updated and displayed, although still no visible changes can be seen at this point.

4) Click on the **Background** button to display the Color dialog.

5) Select a dark blue color, like the one circled above and then click **OK**. The preview of what will become the circular map image is updated to show a dark blue background. Notice that changing the background color to dark blue has made the center title, displayed in black, difficult to see. We will change this in the next section when we edit the title of the map.

Adding a title to the circular map

Objective: To create a title that will appear in the center of the image.

1) Return to the **Circular Map Options** pane by clicking on its tab:
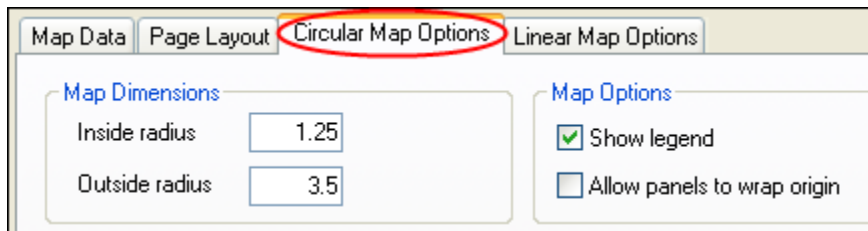
2) Notice the **Center Title** section of this pane, which allows you to control the title that will be displayed in the center of your image. (This is the text that is currently being shown as "Untitled Genome" in the PDF document).
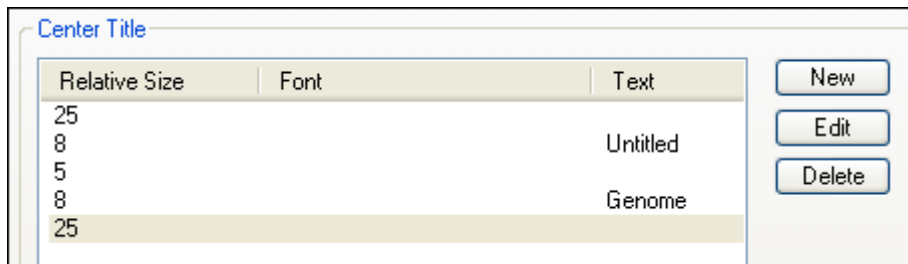
3) Notice that besides the title text, blank spaces are also specified between, above and below the text. Double-click on the row containing the word "Untitled" to open the Edit Title Panel dialog.
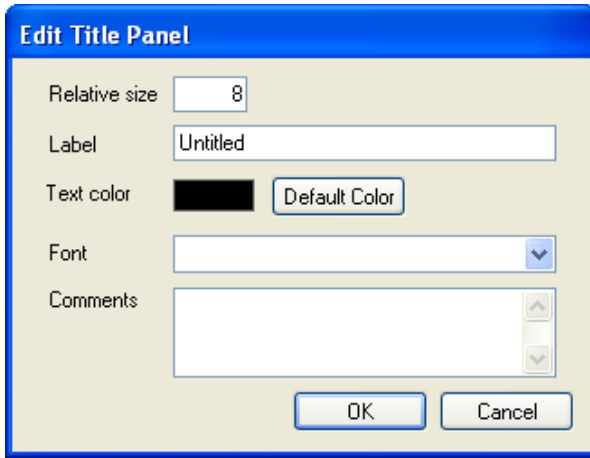
4) Change the **Relative size** value to **6**. The **Relative Size** values for title panels define the size of each text label *in relation to* all of the other text and blank spaces within the center title area. (For more information, see the section "Panel Size" within the GenVision Help).

5) Type **E. coli K-12** in the **Label** field. This will appear as the first line of text in the center title section.

6) Click on the **Default Color** button to open the Color dialog.
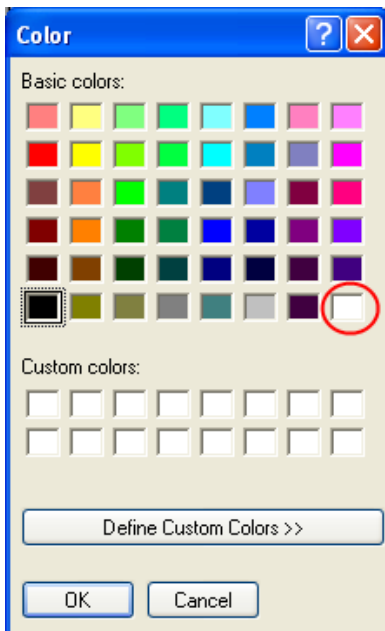
7) Select the color **white** from the Color dialog, as circled above, and then click **OK**.

8) After making these changes, click **OK** to close the Edit Title Panel dialog. The preview of the circular map image updates, and the text "E. coli K-12" now replaces the previous "Untitled" text and is displayed in white.

9) Double-click on the row containing the word "Genome" to open the Edit Title Panel dialog for this panel.

10) Change the **Relative size** value to **4** and type **4,639,221 bp** into the **Label** field.

11) Click on the **Default Color** button to open the Color dialog. Select **white** again, and then click **OK**.

12) After making these changes, click **OK** to close the Edit Title Panel dialog. The preview of the circular map image updates to show the new text.

13) Finally, edit the **Relative size** values for each of the blank panels by double-clicking on them one at a time. Doing this will change the size of the blank space above, below, and in between the two lines of text in the center title. Change their **Relative size** values to **14**, **3** and **14**, respectively. The circular map preview will update with each change. When you are finished, the Center Title section of the **Circular Map Options** pane should appear as follows:

Center Title		
Relative Size	Font	Text
14		
6		E. coli K-12
3		
4		4,639,221 bp
14		

Adding panels to build the circular image

GenVision illustrations are built from one or more *panels*, each drawn using information from an associated *data file*. Data files contain the annotation data you wish to display in a tab-delimited format, and may be created manually, in a program such as Microsoft Excel, or automatically by exporting from a Lasergene application. There are several different types of panels that you may use to visualize your annotation data, including arrows, histograms, graphs, and text.

For more information on creating your own data files, please see the section "Working with Data Files" in the GenVision Help which can be viewed by selecting **Help > Contents**.

Objective: To build a circular map image of the E. coli genome by adding 16 different panels.

Adding a scale to the image

Scales allow you to see exact coordinate positions on your image.

Objective: To add a scale to the image.

1) Click on the **Map Data** tab to return to the **Map Data** pane.
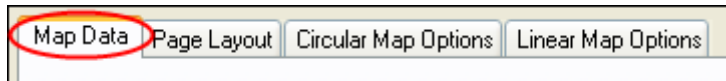
2) From the **New panel type** drop-down list, select **Scale**. The New Scale dialog appears:
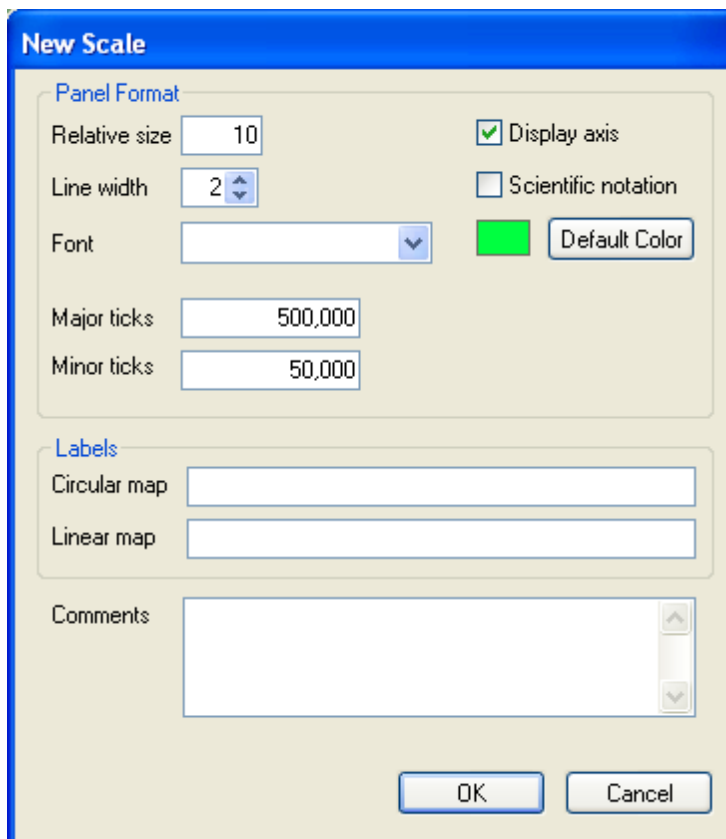
3) Increase the **Major ticks** interval to **1,000,000** (without commas) and the **Minor ticks** interval to **100,000**. (Major tick marks in the scale will be numbered, while the minor ticks will remain unnumbered).

4) Leave the **Relative size** value at **10**, but increase the **Line width** value to **3**. This will provide a slightly thicker scale.

5) Select **Helvetica** from the **Font** menu. This specifies the font for the numbers displayed on the scale.

6) Click **OK**. A preview of the circular map image will be displayed with the scale added, but the scale will appear extremely large. This is because the scale is the *only* panel in our image so far, and each panel's size is relative to *all* of the panels in the image. As panels are added to the project, each panel, including the scale, will get smaller.

7) To add some space between the scale and the next data panel, choose **Blank** from the **New panel type** drop-down list. The New Blank dialog appears:
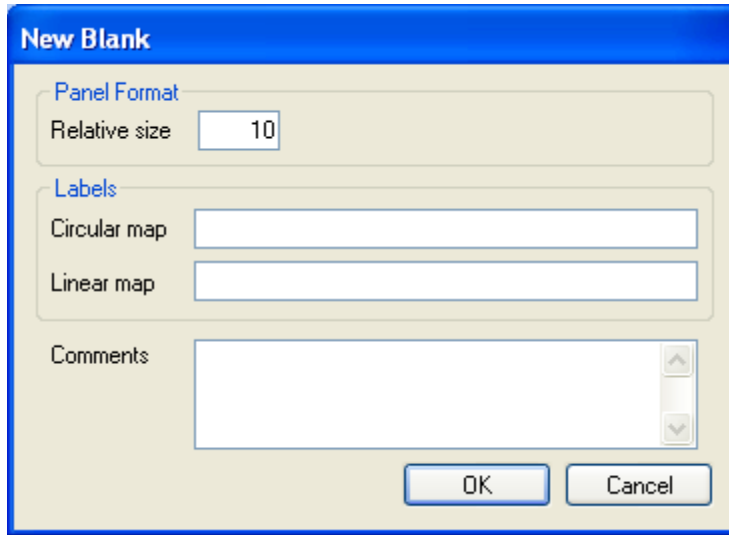
8) Change the **Relative size** value to **7**. Remember, the scale panel's relative size was 10, so this blank space will be slightly smaller than the scale panel.

9) Click **OK**. A preview of your circular map will be displayed. You will notice that the size of the scale has been reduced slightly due to the addition of another panel, but is still very large. It will continue to decrease in size (as will all of the panels) as each panel is added to the project.

Adding replichore arrows

Objective: To add an arrow panel representing replichore arrows to the image.

1) Select **Arrow** from the **New panel type** drop-down list. The New Arrow dialog will appear:

2) Click the **Browse** button, and then navigate to and select **ReplicationArrow.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of the two *E. coli* replichores, as well as the thickness and color for the arrows depicting them.

3) Type "Replichores" into the **Circular map** field. This provides the text that will be used in the circular map legend.

4) Change the **Relative size** value to **8**. This will make this arrow panel smaller than the scale panel, but slightly larger than the blank panel.

5) Click **OK**. A preview of your circular map will be displayed, including the two replichore arrows. Notice that the two arrows are shown in a turquoise color even though the **Default Color** in the New Arrow dialog shows bright green. This is because the color for the arrows was specified in the data file we selected. Colors specified within data files always take precedence over the color settings in any panel dialog.

6) To add some space between the arrow panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **7**, and then click **OK**. The preview of the circular map image will be updated.

Adding tRNA genes

Objective: To add an arrow panel representing tRNA genes to the image.

1) Select **Arrow** from the **New panel type** drop-down list. The New Arrow dialog will appear:

2) Click the **Browse** button, and then navigate to and select **tRNAArrows.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of each of the *E. coli* tRNA genes, as well as the thickness and color for the arrows representing them.

3) Increase the **Relative size** value to **12** to make this arrow panel the largest panel in our project so far.

4) Increase the **Minimum length** value to **40,000**. This value specifies that arrows will be drawn at a length of 40,000 nucleotides, even though the actual genomic coordinates for each tRNA gene will be much smaller.

5) Type "tRNA" into the **Circular map** field. This provides the text that will be used in the circular map legend.

6) Leave all of the other settings as they are, and then click **OK**. A preview of the circular map will be displayed, including the tRNA arrows, drawn in red and yellow. As with the replichore arrow panel, the colors for the tRNA arrows were specified in the data file we selected.

7) To add some space between this panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **7**, and then click **OK**. The preview of the circular map image will be updated.

Adding rRNA genes

Objective: To add an arrow panel representing rRNA genes to the image.

1) Select **Arrow** again from the **New panel type** drop-down list. The New Arrow dialog will appear:

2) Click the **Browse** button, and then navigate to and select **rRNAArrows.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of each of the *E. coli* rRNA genes, as well as the thickness and color for the arrows representing them.

3) Increase the **Relative size** value to **12** to make this panel the same size as the tRNA genes panel.

4) Increase the **Minimum length** value to **50,000**. This value specifies that arrows will be drawn at a length of 50,000 nucleotides, even though the actual genomic coordinates for each rRNA gene will be much smaller.

5) Type "rRNA" into the **Circular map** field. This provides the text that will be used in the circular map legend.

6) Click **OK**. A preview of the circular map will be displayed, including the rRNA genes, drawn as yellow and red closed arrows. Note that the style of the arrow head, as well as the color, was specified within the data file we selected.

7) To add some space between this panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **7**, and then click **OK**. The preview of the circular map image will be updated.

Adding ORFs

Objective: To add a histogram representing ORFs to the image.

1) Select **Histogram** from the **New panel type** drop-down list. The New Histogram dialog will appear:

2) Click the **Browse** button, and then navigate to and select **ORFsHistogram.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the locations of the *E. coli* genes, as well as the color for the histograms representing them.

3) Leave the default **Relative size** value as **30**.

4) Type "ORFs" into the **Circular map** field. This provides the text that will be used in the circular map legend.

5) Leave all of the other settings as they are, and then click **OK**. A preview of the circular map image will be displayed, including the ORF histogram you just added, drawn in yellow and red.

6) To add some space between this panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **20**, and then click **OK**. The preview of the circular map image will be updated.

Adding codon usage

Objective: To add a histogram representing the inverse codon adaptation index for *E. coli* genes. This index is used to flag genes that may have been acquired by horizontal gene transfer.

1) Select **Histogram** from the **New panel type** drop-down list. The New Histogram dialog will appear:

---

2) Click the **Browse** button, and then navigate to and select **CodonUseHistogram.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, height and color of each bar in the histogram.

3) Increase the **Relative size** value to **75** to make this panel the largest one in our image so far.

4) Remove the checkmark next to **Display axis** so that the optional axis line will not be displayed on the map.

5) Type "Codon Usage" into the **Circular map** field. This provides the text that will be used in the circular map legend.

6) Leave all of the other settings as they are, and then click **OK**. A preview of the circular map image will be displayed, including the codon usage histogram you just added, drawn in yellow, red, and orange.

Note: Unlike the ORFs histogram (red and yellow rings), the height of the Codon Use histogram varies, reflecting the "Inverse Codon Adaptation Index"—a measure of codon usage for each ORF. It is thought that the higher the inverse CAI, the more likely it is that the ORF was acquired in the *E. coli* chromosome by horizontal transmission. To emphasize the ORFs most likely to have been acquired by horizontal transmission, histogram boxes for ORFs with inverse CAI values above 0.75 are colored yellow instead of red or orange.

---

7) To add some space between this panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **7**, and then click **OK**. The preview of the circular map image will be updated.

Adding a graph panel

Objective: To add a graph to the image representing remnants of bacteriophages.

1) Select **Graph** from the **New panel type** drop-down list. The New Graph dialog will appear:

2) Click the **Browse** button, and then navigate to and select **PhagesGraph.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the x and y coordinates for each point on the graph.

3) Increase the **Relative size** value to **65** to make this panel just slightly smaller than the Codon Use histogram.

4) Type "Phages" into the **Circular map** field. This provides the text that will be used in the circular map legend.

5) Leave all of the other settings as they are, and then click **OK**. A preview of the circular map image will be displayed, including the newly added graph.

6) Notice that for this particular data set, the few high y-value data points tend to mask the variation in the majority of the data points that have much lower values. (If you wish, you can confirm this by opening **PhagesGraph.txt** in Microsoft Excel and sorting by the **Top** column, which contains the y-coordinates for each data point).

7) In order to greater emphasize the differences among the lower data points, select the "Phages" graph panel from the list panels in the **Map Data** pane by single-clicking on its row.

Relative Size	Size	Panel Type	File Name	Circular Map Label	Linear Map Label	Offset
10	0.08	Scale				---
7	0.06	Blank				---
8	0.07	Arrow	ReplicationArrow.txt	Replichores		0
7	0.06	Blank				---
12	0.10	Arrow	tRNAArrows.txt	tRNA		0
7	0.06	Blank				---
12	0.10	Arrow	rRNAArrows.txt	rRNA		0
7	0.06	Blank				---
30	0.25	Histogram	ORFsHistogram.txt	ORFs		0
20	0.17	Blank				---
75	0.63	Histogram	CodonUseHistogra...	Codon Usage		0
7	0.06	Blank				---
65	0.55	Graph	PhagesGraph.txt	Phages		0

8) Then, click the **Edit** button. The Edit Graph dialog will be displayed.
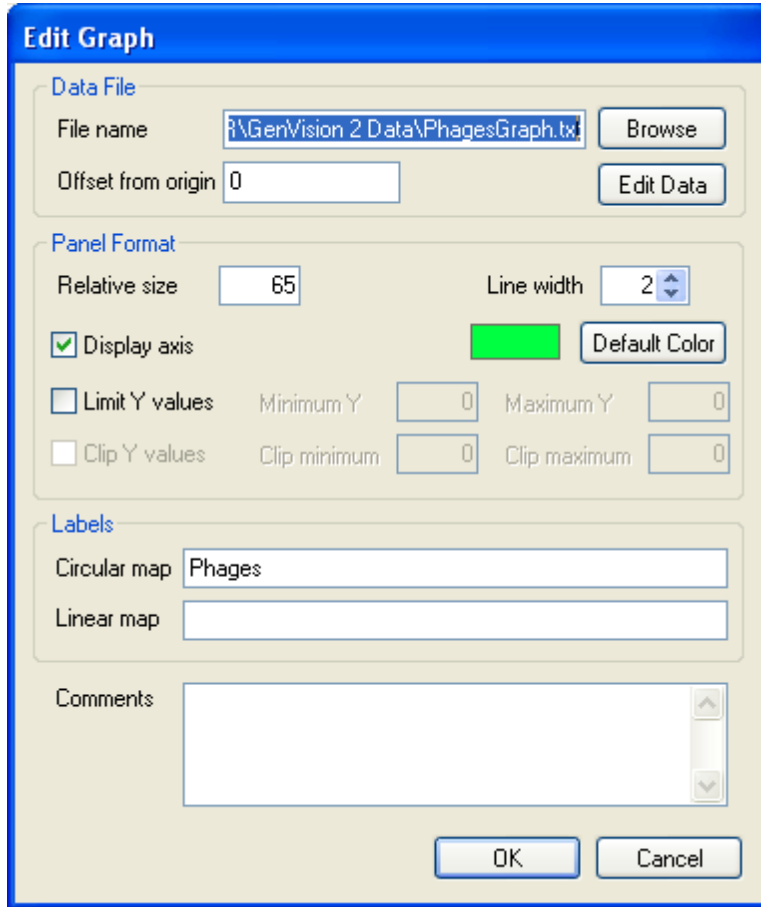
9) Check the box next to **Limit Y values**, and increase the **Maximum Y** value to **0.3**. This will cause GenVision to scale your graph so that the top of the graph panel is equal to the Y coordinate 0.3.

10) Next, check the box next to **Clip Y values** and set the **Clip maximum** value to **0.6**. This will allow data points greater than 0.3 and up to 0.6 to extend beyond the space allocated for your panel.

11) Click **OK**. The preview of the circular map image will be updated. Notice that the graph towards the center of the map now shows greater variation.

12) To add some space between this panel and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **7**, and then click **OK**. The preview of the circular map image will be updated.

Adding text to the image

Text panels allow you to place text alongside the other panels in your image with respect to map coordinates.

Objective: To add a text panel to the map.

1) Select **Text** from the **New panel type** drop-down list. The New Text dialog appears:
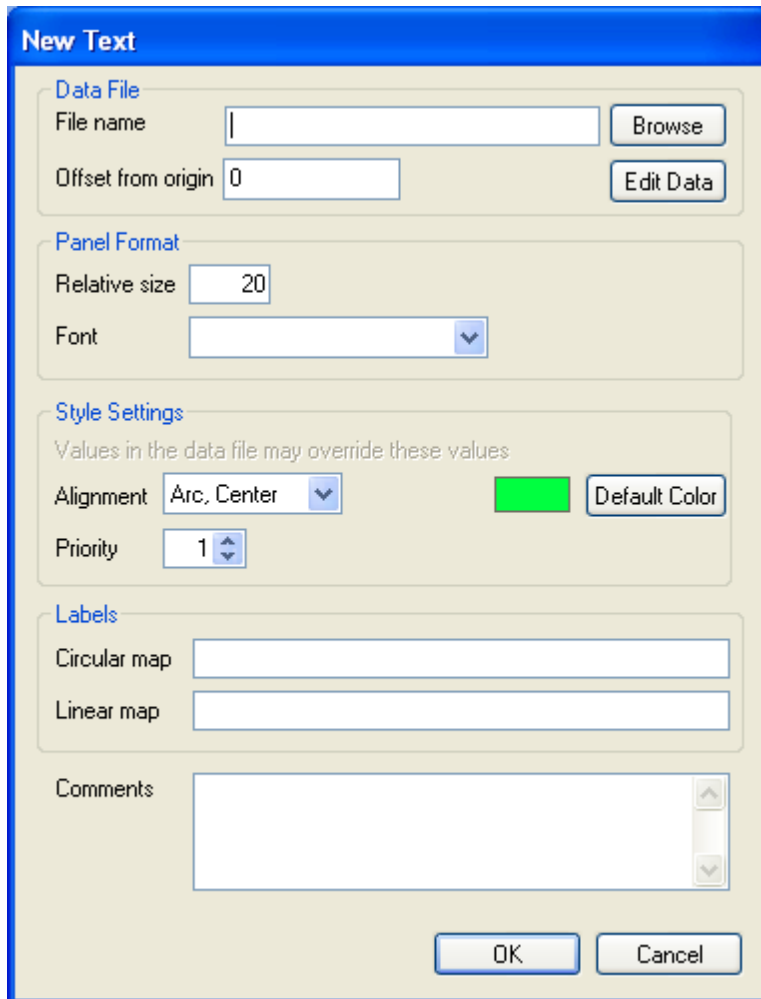
2) Click the **Browse** button, and then navigate to and select **GenVisionText.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the position and text for the label that will reside around the perimeter of the *E. coli* genome.

3) Leave the default **Relative size** value as **20**, and then click **OK**. The image will display the text "**GenVision by DNASTAR**" in two segments of the map, centered on coordinates 300,000 and 2,619,611.

4) In order to display the text so that it appears as the outer-most ring of the circular map, select the text panel from the Panels section of the **Map Data** pane by single-clicking on its row.

> **Note:** Panels in the **Map Data** pane are listed from outer- to inner- most for circular maps.

Relative Size	Size	Panel Type	File Name	Circular Map Label
7	0.05	Blank		
8	0.06	Arrow	ReplicationArrow.txt	Replichores
7	0.05	Blank		
12	0.09	Arrow	tRNAArrows.txt	tRNA
7	0.05	Blank		
12	0.09	Arrow	rRNAArrows.txt	rRNA
7	0.05	Blank		
30	0.23	Histogram	ORFsHistogram.txt	ORFs
20	0.15	Blank		
75	0.57	Histogram	CodonUseHistogra...	Codon Usage
7	0.05	Blank		
65	0.50	Graph	PhagesGraph.txt	Phages
7	0.05	Blank		
20	0.15	Text	GenVisionText.txt	

5) Then, drag the text panel row up to the top of the list of panels and drop it above the scale panel so that it appears as the first panel in the list. The preview of the circular map image will be updated to show the text on the outside of the scale.

Adding Demarcation Lines to the Image

The final panel for this project is a demarcation panel. A demarcation is a line drawn radially on a circular map, perpendicular to the map coordinate axis.

Objective: To add demarcation lines to the image defining the replication origin and terminal regions.

1) From the **Map Data** pane, first click on the last panel in the list (a Blank panel), and then select **Demarcation** from the **New panel type** drop-down list. The New Demarcation dialog appears:
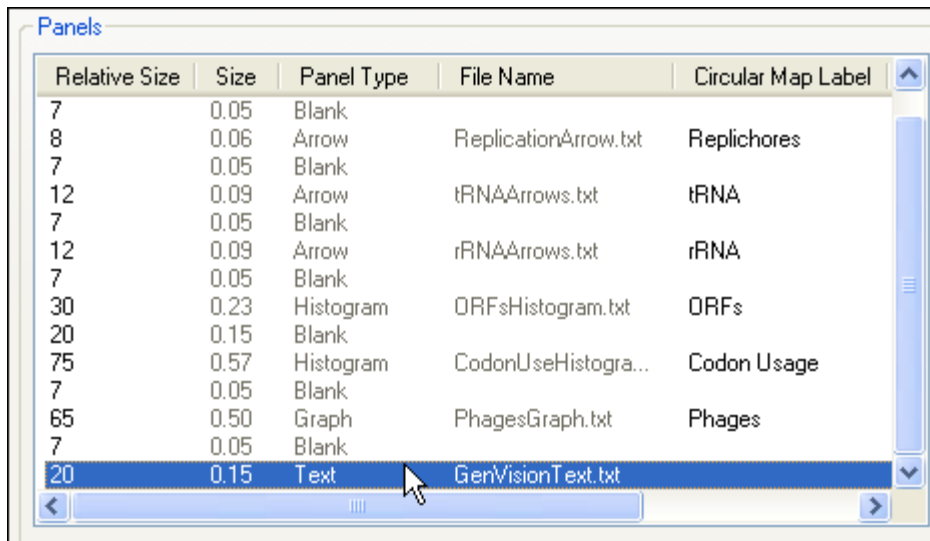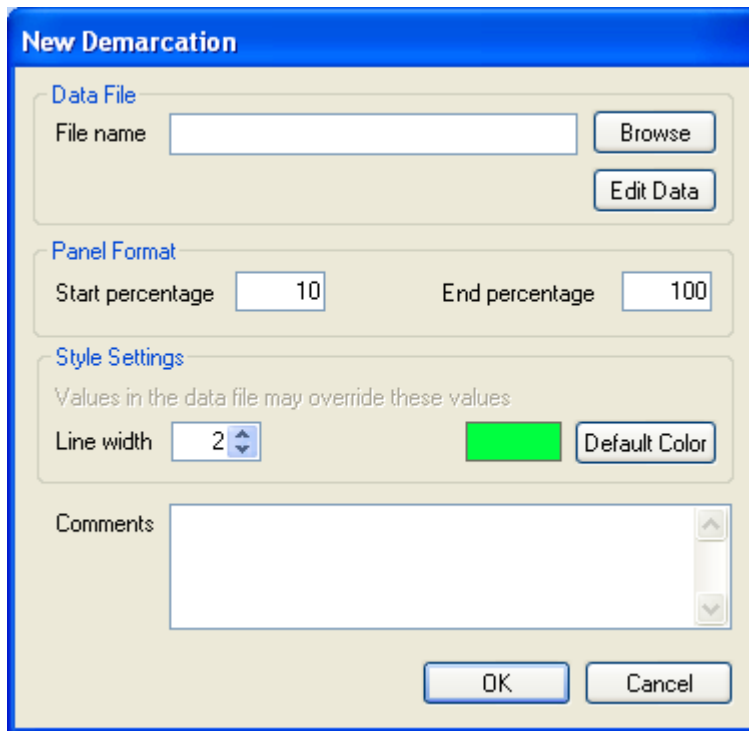

2) Click the **Browse** button, and then navigate to and select **OriTerDemarcation.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the coordinates of the two demarcations for the *E. coli* map, one at the origin, and the other at the terminus of replication.

3) Change the **Start percentage** value to **0** and leave the **End percentage** value at **100**. These values specify the portion of the figure where the demarcation lines will be drawn. In this case, the lines will span the entire panel space.

4) Click **OK**. The preview of the completed circular map image will be updated, including the demarcation lines, and should look similar to the image shown after Step 5.

5) The completed map is now ready for printing or for insertion into another document. Select **File > Save** to save the GenVision project (*.gnv) for later use.
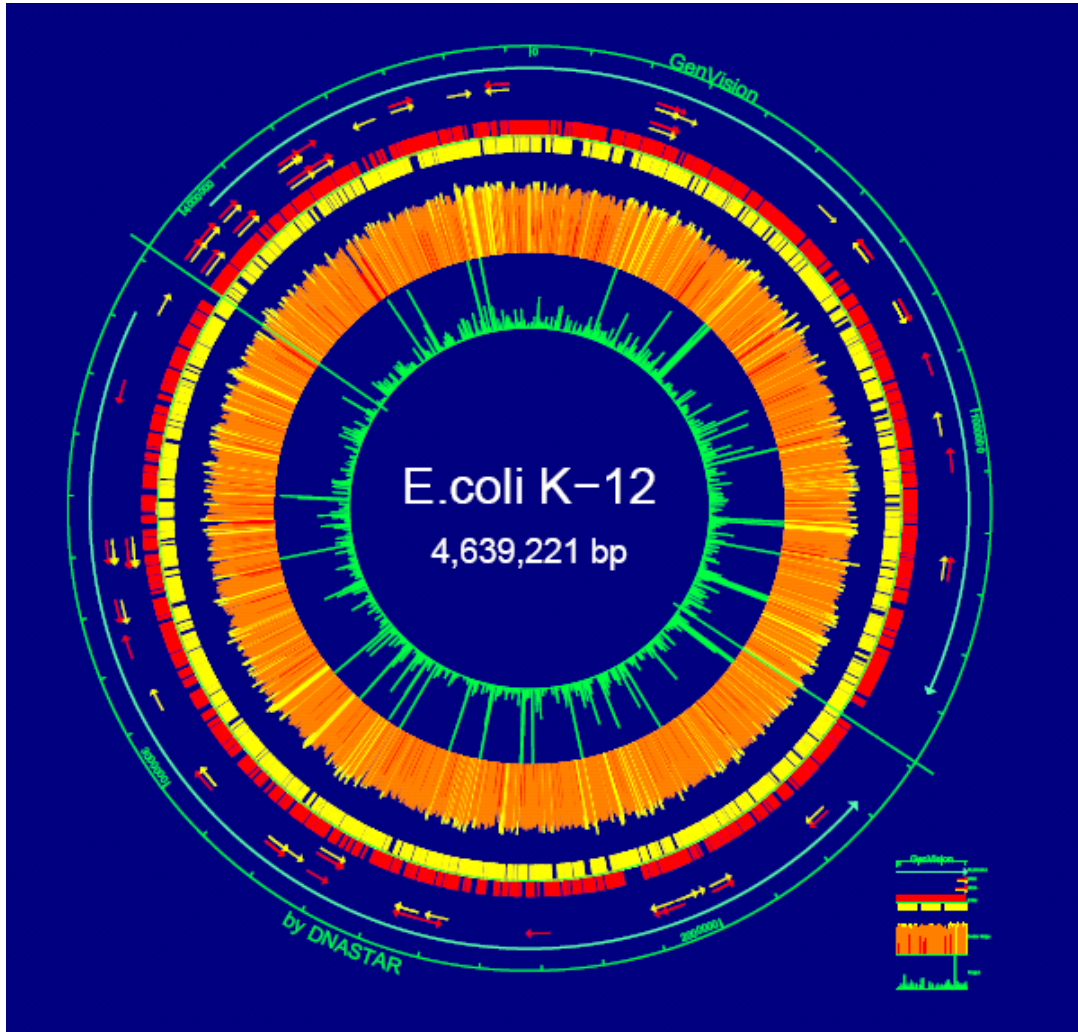
Designing a Linear Map

Objective: The purpose of this tutorial is to familiarize you with building a linear map image in GenVision using the first 600,000 bp of the *E. coli* genome.

In order to complete this tutorial, you will need the sample data that was installed automatically with your software in the following directory:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo GenVision

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\ Demo GenVision

- **Macintosh OS X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data: Demo GenVision

Creating a new linear map image

Objective: To create a new linear map figure for a portion of the *E. coli* genome.

1) Launch GenVision if it is not already open. The **Map Data** pane appears, with the **Circular Map** radio button selected by default.
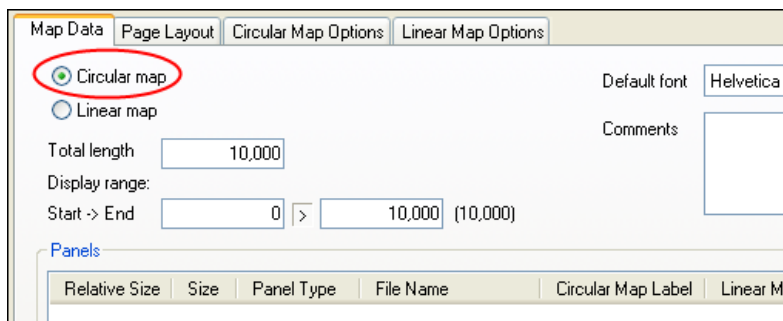
2) Select the **Linear Map** option. A message will appear, prompting you to save the new project you created:
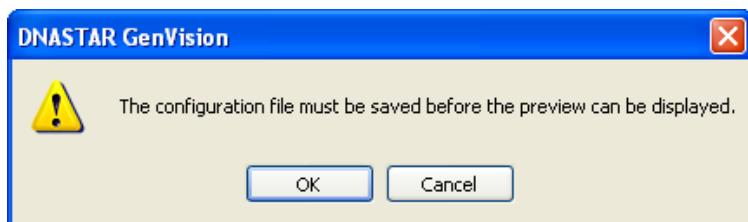
3) Click **OK** in the message dialog.

4) Choose the directory where you want to save the project, enter the filename **Ecoli Linear Map**, and click **Save**. A blank PDF document will be displayed.

Adjusting user options

GenVision offers several options for how and when the images it creates are viewed. By default, images are displayed as PDF files each time you make a change to the figure.

Objective: To turn off the auto-preview feature while we are setting up the image.

1) From the GenVision menu, select **View > Options**. The Options dialog is displayed:
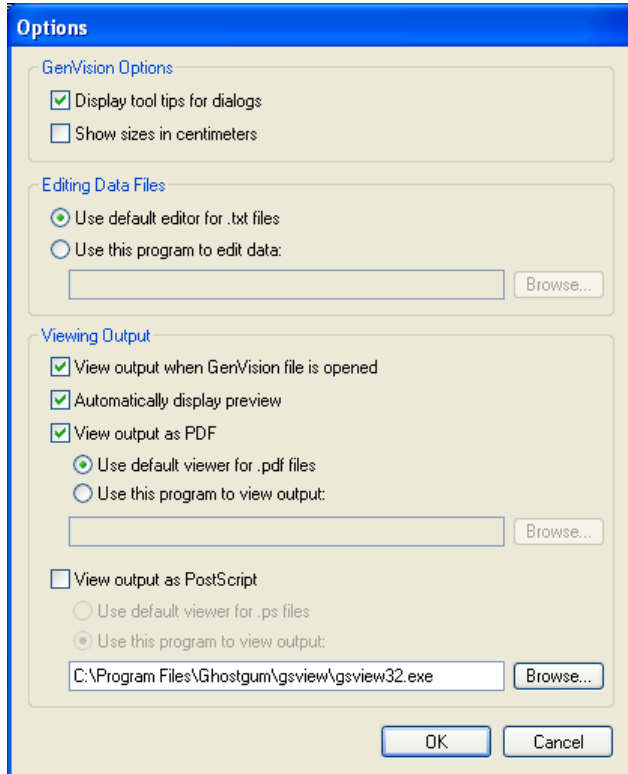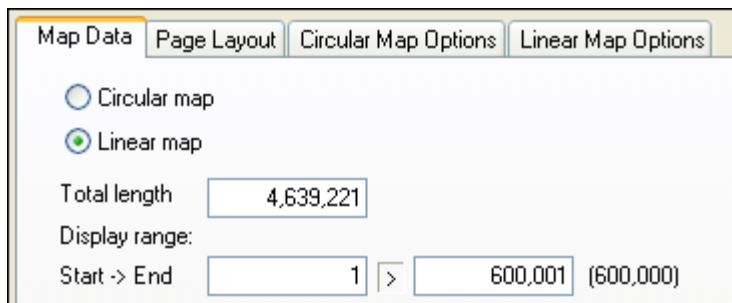
2) Notice the default settings under the "Viewing Output" section. By default, GenVision automatically creates a preview of your image as a PDF document each time you make a change to your map.

3) While we are setting up the layout of our linear map, we will be making lots of changes to the image. Turn off the auto-preview feature by removing the checkmark next to **Automatically display preview**, then click **OK**. We will turn this feature back on later when we are ready to begin adding data panels.

Defining the layout and options for the linear map

Objective: To define the layout of the linear map and limit the image to the first 600,000 nucleotides of the *E. coli* genome.

1) From the **Map Data** pane, enter **4,639,221** (without commas) into the **Total Length** field.

2) Under **Display Range**, enter **1** as the **Start** value and **600,001** (without commas) as the **End** value. Doing this will limit the map we are creating to the first 600,000 nucleotides of the genome.
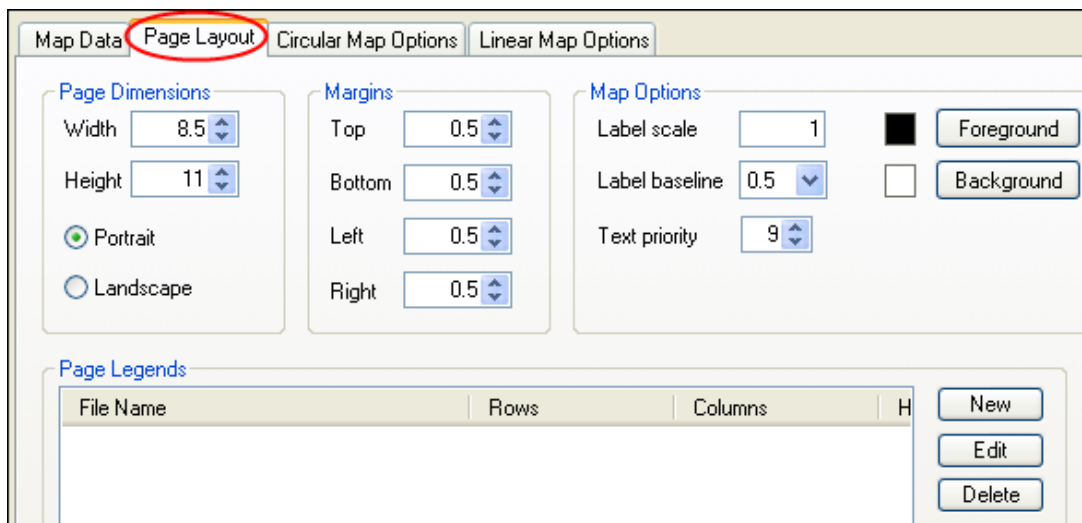
3) Click on the **Page Layout** tab. The **Page Layout** pane will be displayed:

4) In the Page Dimensions section, select **Landscape** orientation.

5) Change the **Label baseline** value to **0.0** using the pull-down menu. This will cause the linear map labels for each panel to be drawn even with the baseline of each panel.

6) Click on the **Linear Map Options** tab to display the **Linear Map Options** pane:
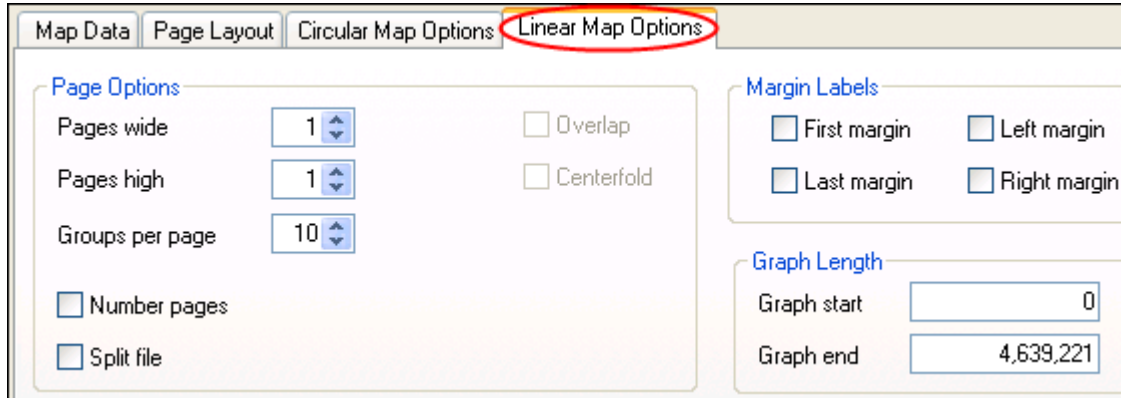
7) Change the **Pages wide** value to **3** to specify that the data will be plotted across the width of three pages.

8) Change the **Groups per page** to **12.** A *group* is a single set of all the data panels in your project. This setting specifies that only twelve groups of panels will be displayed per page.

9) Check the box next to **Number pages** to generate page numbers in the bottom left corner of each page.

10) In the Margin Labels section, check the box next to **Left margin**. This setting indicates that the labels for each panel will be displayed in the left margins of the linear map.

11) Change the **Graph start** value to **1** so that the scale will show a "1" for the first base in the map.

12) Change the **Graph end** value to **605,000** (without commas). Since we are mapping the first 600,000 bp of the genome, entering this value will create a space equivalent to 5,000 nucleotides at the end of the map to accommodate an End Legend.

13) Now that we are finished setting up our linear map, we will return to the Options dialog by selecting **View > Options**.

14) In the Options dialog, check the box next to **Automatically display preview** to turn back on the auto-preview feature, and then click **OK**. It will be beneficial to see previews of our linear map image as we begin to add data panels to the project.

Adding panels to build the linear image

GenVision illustrations are built from one or more *panels*, each drawn using information from an associated *data file*. Data files contain the annotation data you wish to display in a tab-delimited format, and may be created manually, in a program such as Microsoft Excel, or automatically by exporting from a Lasergene application. There are several different types of panels that you may use to visualize your annotation data, including arrows, histograms, graphs, and text.

For more information on creating your own data files, please see the section "Working with Data Files" in the GenVision Help which can be viewed by selecting **Help > Contents**.

Objective: To build a linear map image of a portion of the E. coli genome by adding 8 different panels.

In order to complete this section, you will need the sample data that was installed automatically with your software in the following directory:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo GenVision

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\ Demo GenVision

- **Macintosh OS X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data: Demo GenVision

Displaying transcripts and genes

Objective: To add an arrow panel representing transcripts and a histogram panel representing the *E. coli* genes.

1) Click on the **Map Data** tab to return to the **Map Data** pane.
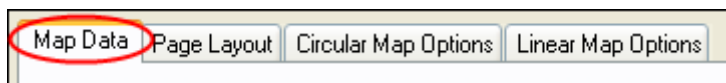
2) From the **New panel type** drop-down list, select **Arrow**. The New Arrow dialog appears:

3) Click the **Browse** button, and then navigate to and select **Transcripts.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, thickness and color for the arrows representing the transcripts.

4) Change the **Relative size** value to **8**. Relative Size values define the size of each panel *in relation to* all of the other panels within the image. (For more information, see the section "Panel Size" within the GenVision Help).

5) Change the **Minimum length** value to **100**. This specifies that any arrows less than 100 bases in length will be drawn as if they were 100 bases long, enabling you to see the arrow shafts more easily.

6) Type "Transcripts" into the **Linear map** field. This provides the text that will be used in the left margins of the linear map to identify this panel.

7) Leave all the other settings as they are, and then click **OK**. A preview of the linear map image will be displayed with the black and red transcript arrows added, but the arrows will appear fairly large. This is because the arrow panel is the *only* panel in our image so far, and each panel's size is relative to *all* of the panels in the image. As panels are added to the project, each panel, including the transcript arrows, will get smaller.

8) To add some space between the transcript arrows and the next data panel, choose **Blank** from the **New panel type** drop-down list. The New Blank dialog appears:
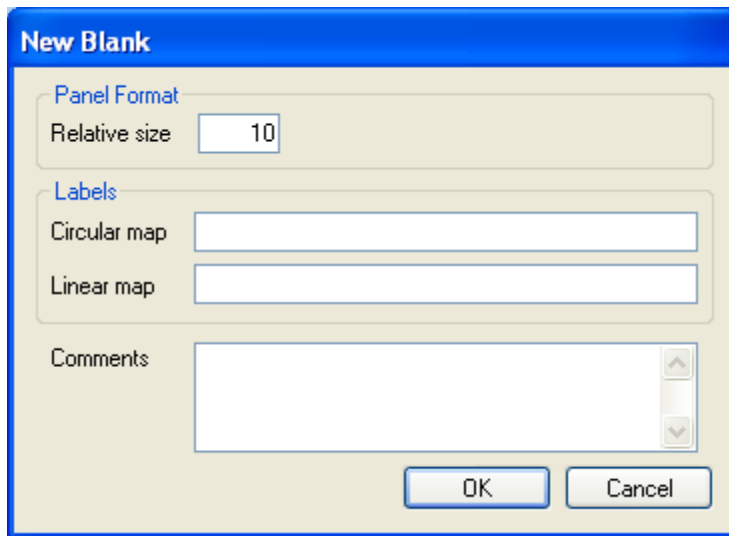
9) Change the **Relative size** value to **2**. Remember, the arrow panel's relative size was 8, so this blank space will be a quarter of the size of the arrow panel.

10) Click **OK**. A preview of your linear map will be displayed. You will notice that the size of the arrow panel has been reduced slightly due to the addition of another panel, but is still large. It will continue to decrease in size (as will all of the panels) as each panel is added to the project.

11) In order to add the genes to the image, select **Histogram** from the **New panel type** drop-down list. The New Histogram dialog appears:

12) Click the **Browse** button, and then navigate to and select **Genes.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, color and height for the histogram depicting the known and putative *E. coli* genes.

13) Change the **Relative size** value to **6**. This will make the histogram panel slightly smaller than the arrow panel.

14) Type "Genes" into the **Linear map** field. This provides the text that will be used in the left margins of the linear map to identify this panel.

15) Leave all the other settings as they are, and then click **OK**. The preview of the linear map image will be updated to include the histogram panel representing *E. coli* genes.

Note that the histogram is displayed in several different colors. These colors were defined within the data file we selected and are being used to indicate gene function. Later in this tutorial we will add an *End Legend*, which will act as a key for the color coding that is used.

16) To add some space between the genes and the next data panel, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **2**, and then click **OK**. The preview of the linear map image will be updated.

Adding ORF names

Objective: To add a text panel to the image displaying the names of the ORFs.

1) Select **Text** from the **New panel type** drop-down list. The New Text dialog appears:

2) Click the **Browse** button, and then navigate to and select **ORFNames.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the position and text for the gene labels.

3) Change the **Relative size** value to **10**, making it the largest panel in our project so far.

4) Type "ORF Names" into the **Linear map** field. This provides the text that will be used in the left margins of the linear map to identify this panel.

5) Leave all the others settings as they are, and then click **OK**. The preview of the linear map image will be updated to include the gene labels we just added.

6) Notice that the gene labels are displayed underneath the gene histogram panel. This is because the panels in the **Map Data** pane are listed from top to bottom for linear maps and the text panel is at the bottom of the list. In order to display the text so that it appears at the top of the panel space, select the text panel from the Panels section of the **Map Data** pane by single-clicking on its row.

Relative Size	Size	Panel Type	File Name	Circular Map Label	Linear Map Label
8	0.24	Arrow	Transcripts.txt		Transcripts
2	0.06	Blank			
6	0.18	Histogram	Genes.txt		Genes
2	0.06	Blank			
10	0.30	Text	ORFNames.txt		ORF Names
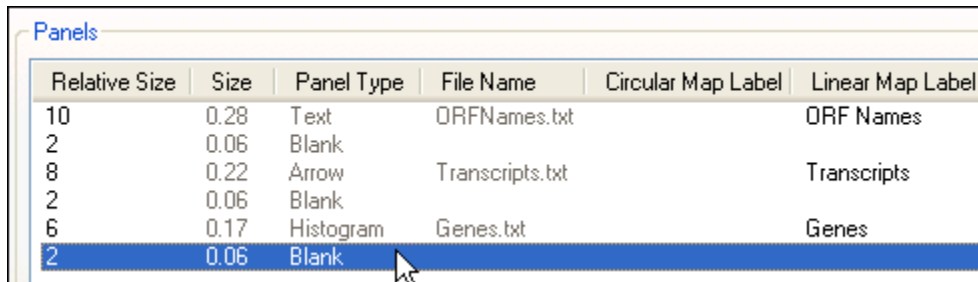
7) Then, drag the text panel row up to the top of the list of panels and drop it above the arrow panel so that it appears as the first panel in the list. The preview of the linear map image will be updated to show the text above the transcript arrows.

8) In order to add some space between the text and the transcript arrows, first select the text panel by single-clicking on its row. Then, choose **Blank** from the **New panel type** drop-down list.

9) In the resulting New Blank dialog, change the **Relative size** value to **2**, and then click **OK**. The preview of the linear map image will be updated.

Adding and optimizing a scale

Scales allow you to see exact coordinate positions on your image.
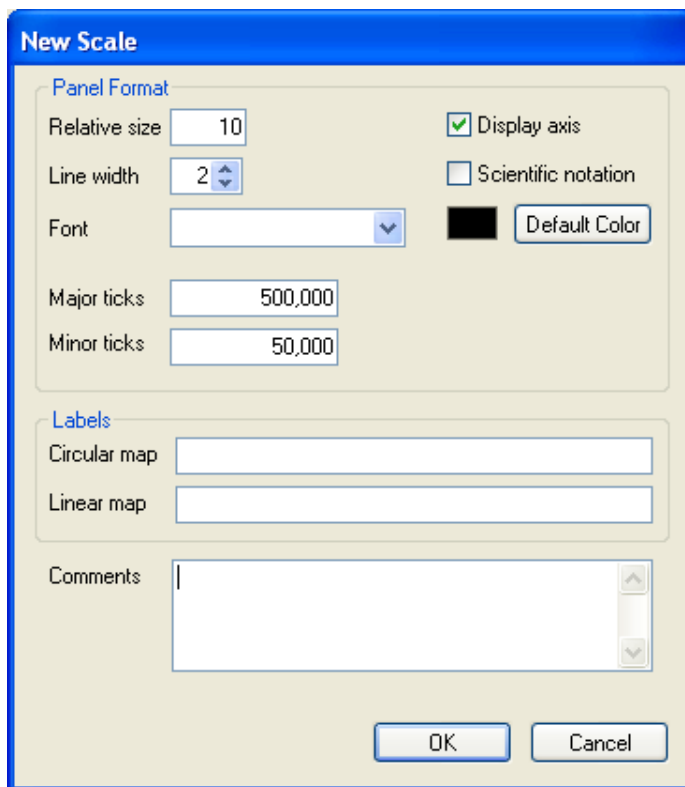
Objective: To add a scale to the image and optimize it for the size of the map.

1) In order to add the scale underneath the genes histogram in the linear map, first select the last **Blank** panel in the list by single-clicking on its row.

Relative Size	Size	Panel Type	File Name	Circular Map Label	Linear Map Label
10	0.28	Text	ORFNames.txt		ORF Names
2	0.06	Blank			
8	0.22	Arrow	Transcripts.txt		Transcripts
2	0.06	Blank			
6	0.17	Histogram	Genes.txt		Genes
2	0.06	Blank			

2) Then, select **Scale** from the **New panel type** drop-down list. The New Scale dialog appears:

3) Change the **Relative size** value to **4**, making it equal in size to the transcript arrow panel.

4) Decrease the **Major ticks** interval to **10,000** (without commas) and the **Minor ticks** interval to **1,000**. (Major tick marks in the scale will be numbered, while the minor ticks will remain unnumbered).

5) Select **Helvetica** from the **Font** menu. This specifies the font for the numbers displayed on the scale.

6) Leave all the others settings as they are, and then click **OK**. The preview of the linear map image will be updated to include the scale.

7) To add some space after the scale, choose **Blank** from the **New panel type** drop-down list. In the resulting New Blank dialog, change the **Relative size** value to **8**, and then click **OK**. The preview of the linear map image will be updated, and the Panels section of the **Map Data** pane should now appear as follows:

Relative Size	Size	Panel Type	File Name	Circular Map Label	Linear Map Label
10	0.20	Text	ORFNames.txt		ORF Names
2	0.04	Blank			
8	0.16	Arrow	Transcripts.txt		Transcripts
2	0.04	Blank			
6	0.12	Histogram	Genes.txt		Genes
2	0.04	Blank			
4	0.08	Scale			
8	0.16	Blank			

Adding an end legend

Start/End Legends are a type of legend specific to linear maps that are used as keys for panels that use color coding. As the names indicate, they are displayed either at the start or end of the linear map.

Objective: To add a legend at the end of the map with a key showing gene function.

1) Click on the **Linear Map Options** tab to return to the **Linear Map Options** pane.
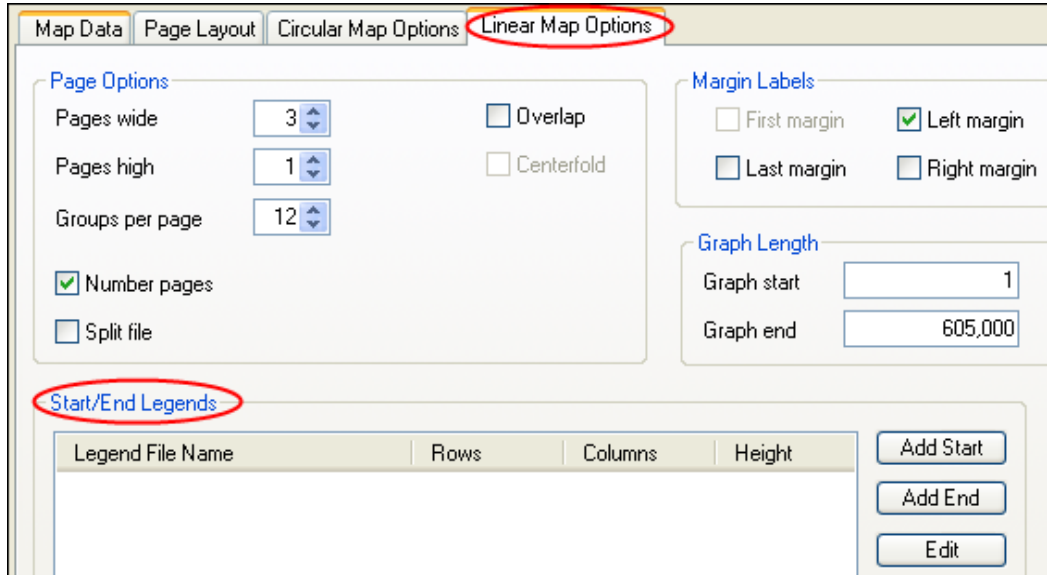
2) From the Start/End Legends section, click the **Add End** button. The End Legend dialog will appear:
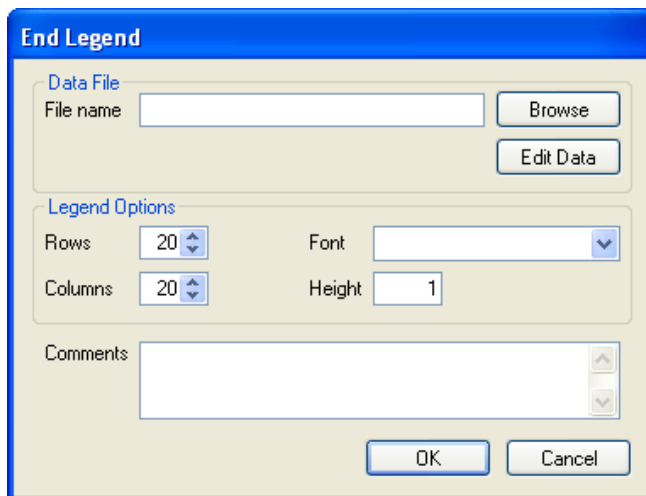
3) Click the **Browse** button, and then navigate to and select **EndLegend.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the organization and color of a legend defining gene functions.

4) Change the **Rows** value to **16** and leave the **Columns** value as **20**. These values define the size of the grid that will be used for the legend.

5) Leave the other settings as they are, and then click **OK**. The preview of the completed linear map image will be updated, including the End Legend. In order to see the End Legend, scroll to the third page of the linear map image and look in the bottom right corner. If necessary, use the zoom tool to view it:
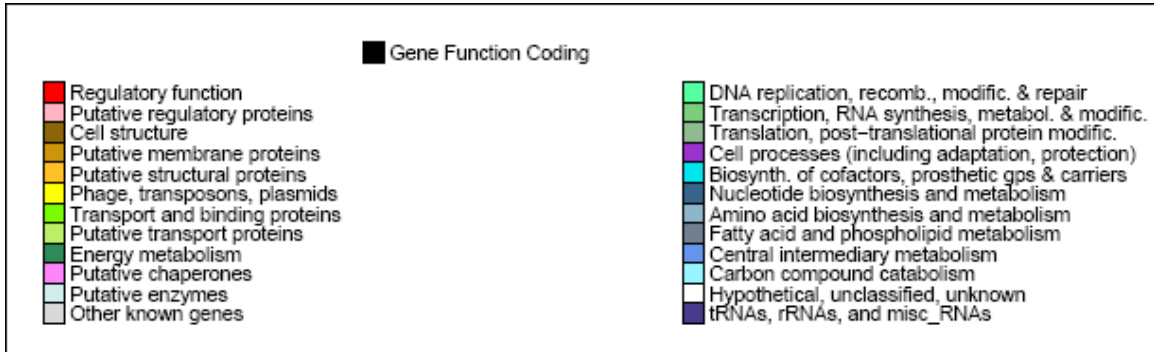
6) The completed linear map image should look similar to the small segment displayed below:
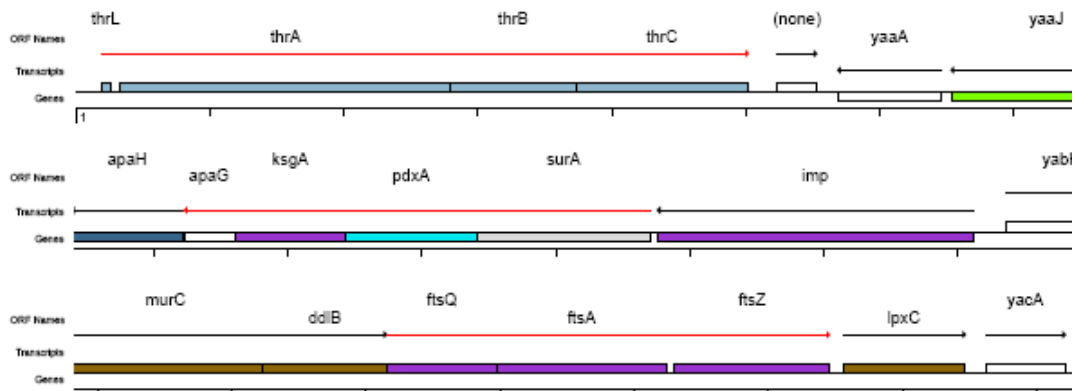
7) The completed linear map is now ready for printing or for insertion into another document. Select **File > Save** to save the GenVision project (*.gnv) for later use.

GenVision Plug-In for Adobe Illustrator

Overview

The GenVision™ Plug-In for Adobe Illustrator is a powerful tool for creating publication-quality images depicting large and complex data sets. The tutorials in this document are intended to familiarize you with much of the functionality GenVision has to offer.

To learn more about specific features, please consult the GenVision Help that was installed with your software:

- Windows: **Help > About DNASTAR Plug-ins > GenVision Help**

- Macintosh: **Illustrator > About DNASTAR Plug-ins > GenVision Help**

Note: In addition to being a Plug-in for Adobe Illustrator on both Windows and Macintosh platforms, GenVision is supplied as a Standalone application for Windows. The tutorials in this document are intended to be used with the GenVision Plug-in for Adobe Illustrator only. For additional information about the GenVision Standalone application, or to access the GenVision Standalone tutorials, please visit the [DNASTAR website](#).

GenVision Plug-In Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example GenVision project from start to finish.

In order to complete the two tutorials in this document, *Designing a Circular Map* and *Designing a Linear Map*, you will need the sample data that was installed automatically with your software in the following directory:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo GenVision

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\ Demo GenVision

- **Macintosh OS X:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data: Demo GenVision

Designing a Circular Map

Objective: The purpose of this tutorial is to familiarize you with building a circular map image using the GenVision Plug-In. For this example workflow, we will create a map of the 4,639,221 bp *E. coli* genome.

Setting up a new circular map image

Objective: To set up a circular map figure for the E. coli genome.

1) Launch Adobe Illustrator if it is not already open, and select **File > New**.

2) Enter "Ecoli Circular Genome" into the Name field, choose **portrait** orientation, select **RGB** as the Color Mode from the Advanced section, and then click **OK**.
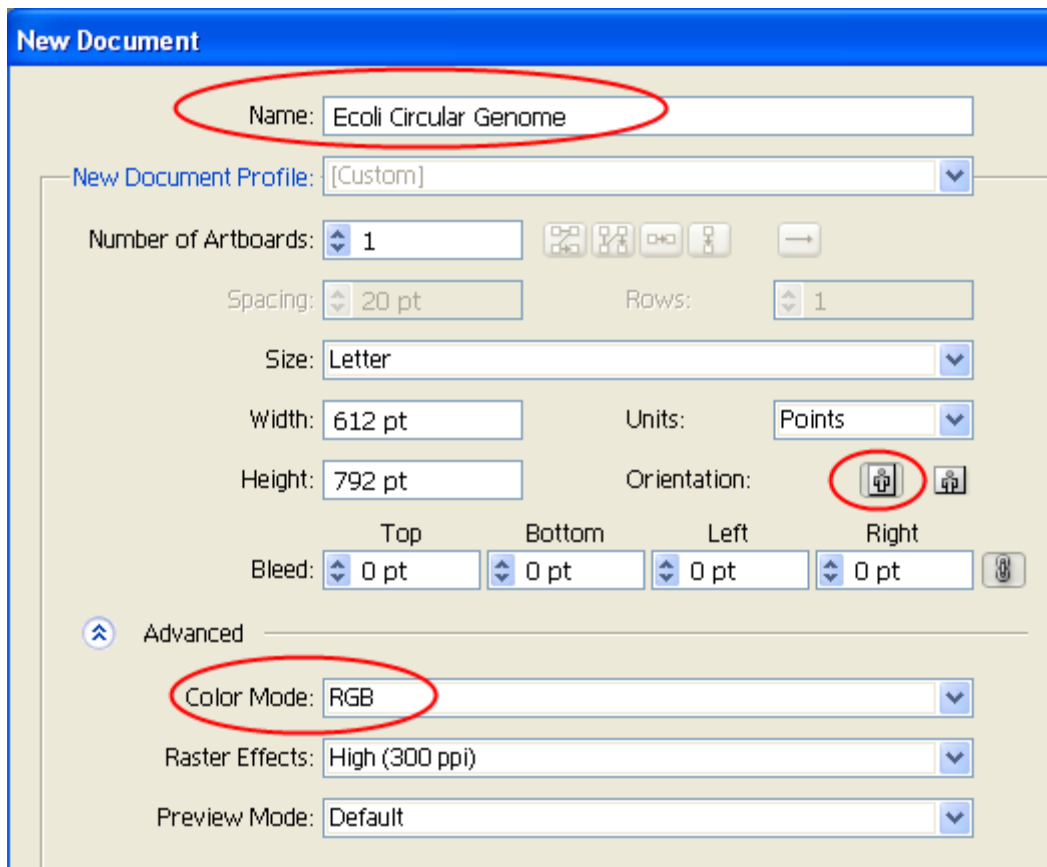
3) Open the Figure Settings dialog by selecting **Window > GenVision Dialogs > Figure Settings**.
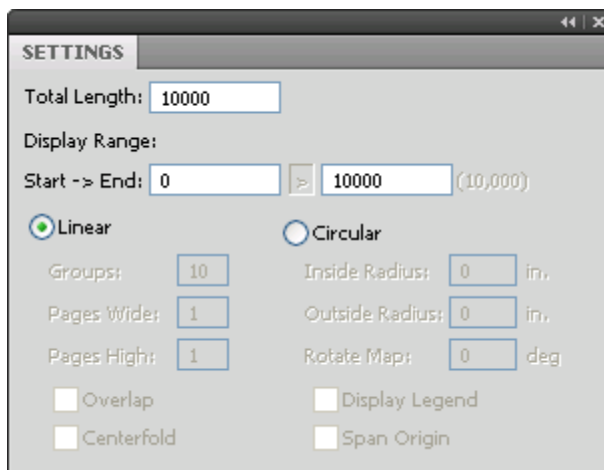
4) In order to specify the length of the genome, enter **4,639,221** (without commas) into the **Total Length** field. Since we want to display the entire genome on our map, leave the **Start** value as **0**, but change the **End** value to **4,639,221**.

5) Select the **Circular** radio button. The circular map fields will become active.
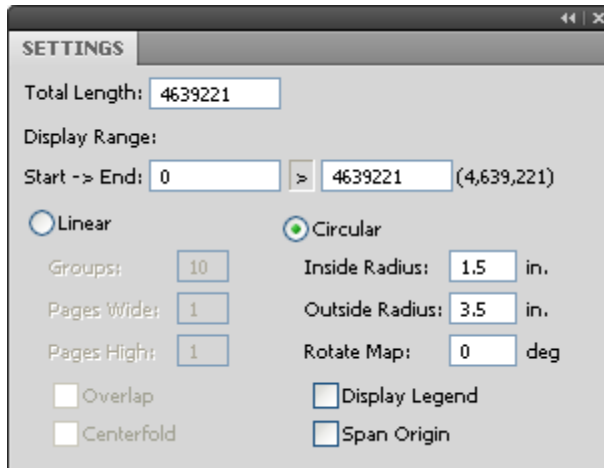
6) Change the **Inside Radius** value to **1.25**, but leave the **Outside Radius** at **3.5** inches. This will expand the map so that it nearly fills an 8.5" x 11" page.

7) Check the box next to **Display Legend** to ensure that a legend defining each panel in your image will be created and displayed in the bottom right corner of the circular map.
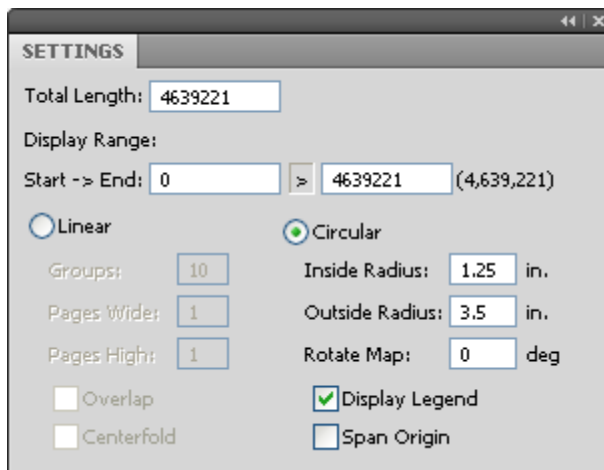
8) Close the Figure Settings dialog before proceeding to the next section.

Defining background and default colors

The Figure Defaults dialog allows you to control several aspects of how your image will be displayed, including the background and default colors of the image. In circular maps, the default color defines the color used for page numbers, the text for page legends, and also sets the initial default color for each panel type. The background color is the color that will be used in areas where no other image elements are specified.

Objective: To define the background and default colors that will be used for the image.

1) Open the Figure Defaults dialog by selecting **Window > GenVision Dialogs > Figure Defaults**.
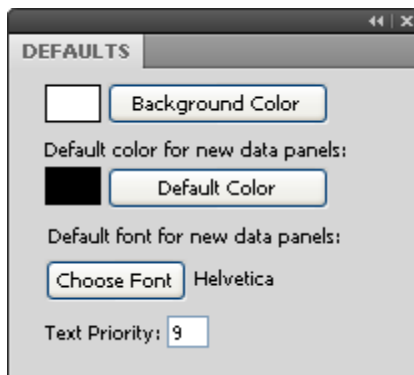
2) Click on the **Background Color** button to display the Color dialog.

3) Select a **dark blue** color from the Color dialog, either by selecting it from a list of available colors, or by entering the values **0, 0, 120** in the **Red**, **Green**, and **Blue** fields, respectively. Once the color has been selected, click **OK**. The Illustrator document will be updated to show a dark blue background. Notice that changing the background color to dark blue has made the center title, displayed in black, difficult to see. We will change that in the next section when we edit the title of the map.

Both Macintosh (top) and Windows (bottom) Color dialogs are shown here:
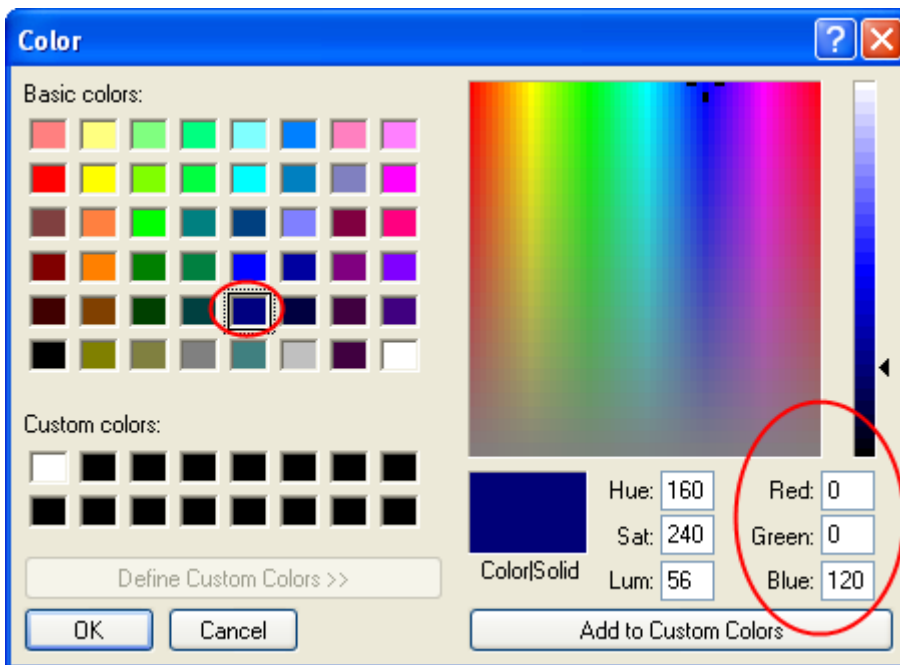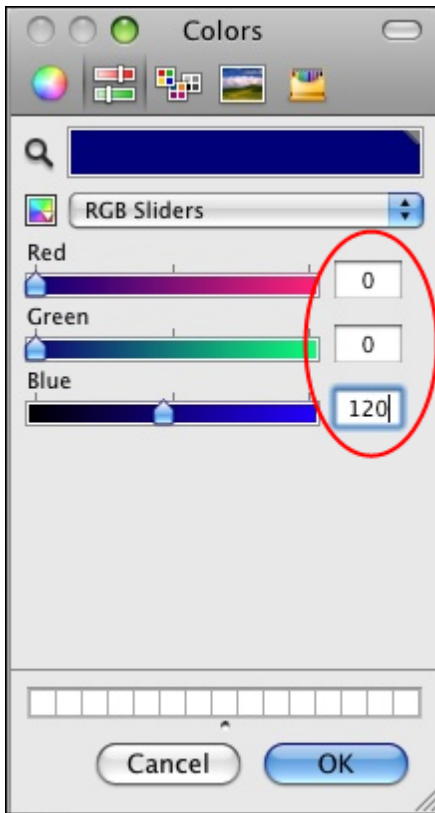
4) Click the **Default Color** button to open the Color dialog again. This time, choose a **bright green** color, either by selecting it from a list of available colors, or by entering the values **0, 200, 0** in the **Red**, **Green**, and **Blue** fields, respectively. Once the color has been selected, click **OK**. Since we have not yet added any elements that use the bright green default color, the Illustrator document will remained unchanged.

5) Close the Defaults dialog before proceeding to the next section.

Editing the circular map title

Objective: To edit the title that will appear in the center of the image.

1) To edit the title of the circular map, first select the Type Tool [T] from the Illustrator toolbar.

2) Double-click on the word "Untitled" to highlight it, and then double-click on the large **Fill** color box from the toolbar to open the Color dialog.

3) Change the color of the text to white as follows:

On Windows: In the Color Picker dialog, type the number "255" into each of the **R**, **G** and **B** text boxes. Then click **OK**.
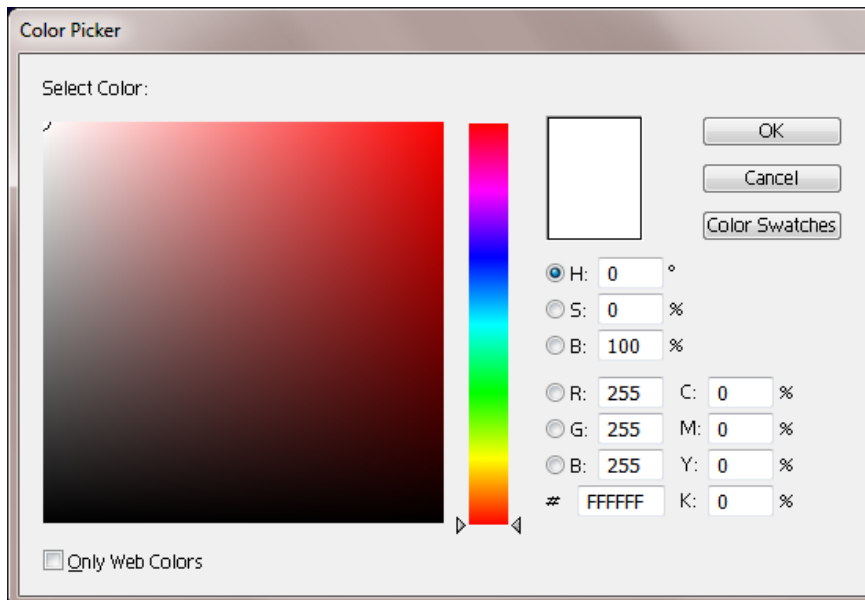
On Macintosh: In the Color dialog, use the small arrow in the top right corner to select **Grayscale**:
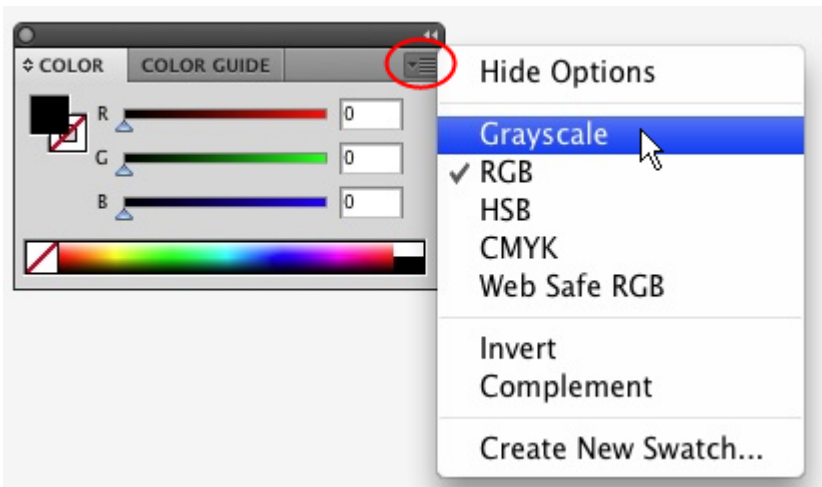
Verify that the **K%** field contains a value of **0**, and then press **Enter.**
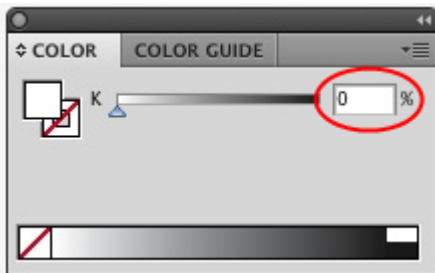
4) Select the "Untitled" text again by double-clicking on it. Then, select **Type > Size > 18 t**o slightly reduce the font size.

5) Finally, select the text a third time and replace the word "Untitled" with "**E. coli K-12**."

6) Repeat steps 2-5 to change the word "Genome" to "**4,639,221 bp**" in 18 pt white text.

7) When you are finishing making the changes, press **Enter**. The center title should now appear as follows:
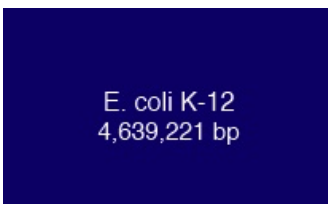
8) Close the Color dialog or Color Picker dialog before proceeding to the next section.

Adding panels to build the circular image

GenVision illustrations are built from one or more *panels*, each drawn using information from an associated *data file*. Data files contain the annotation data you wish to display in a tab-delimited format, and may be created manually, in a program such as Microsoft Excel, or automatically by exporting from a Lasergene application. There are several different types of panels that you may use to visualize your annotation data, including arrows, histograms, graphs, and text.

For more information on creating your own data files, please see the section "*Working with Data Files*" in the GenVision Help which can be viewed as follows:

- Windows: **Help > About DNASTAR Plug-ins > GenVision Help**

- Macintosh: **Illustrator > About DNASTAR Plug-ins > GenVision Help**

Objective: To build a circular map image of the E. coli genome by adding 11 different panels.

Adding a scale to the image

Scales allow you to see exact coordinate positions on your image.

Objective: To add a scale to the image.

1) Access the Scale dialog by selecting **Window > GenVision Dialogs > Scale**. In the resulting dialog, click the small arrow in the top right corner and select **Add Scale Panel**.
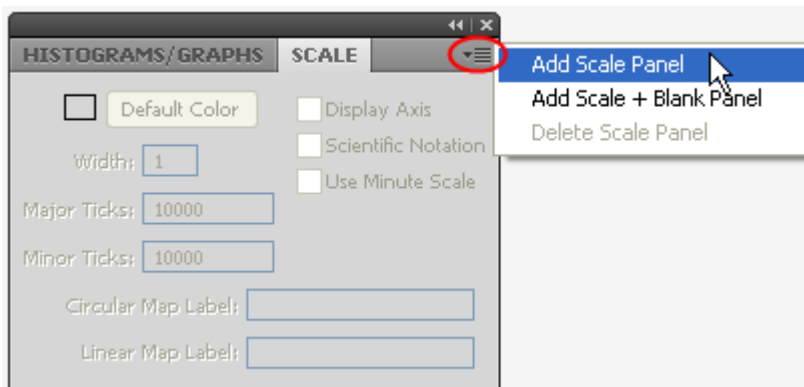
2) In the Scale dialog, increase the **Major Ticks** interval to **1,000,000** (without commas), the **Minor Ticks** interval to **100,000**, and then press **Enter.**
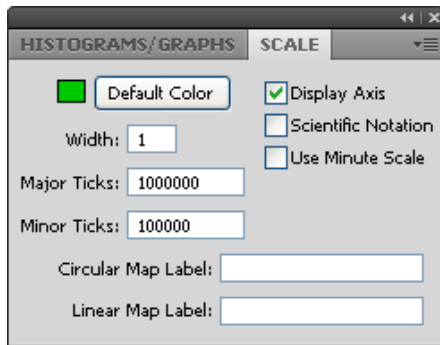
Note: Major tick marks in the scale will be numbered, while the minor ticks will remain unnumbered.

3) Notice that the circular map image will be updated to include the scale in bright green (the default color we specified earlier), but the scale will appear extremely large. This is because the scale is the *only* panel in our image so far, and each panel's size is relative to *all* of the panels in the image. As panels are added to the project, each panel, including the scale, will get smaller.

4) Close the Scale dialog before proceeding to the next section.

Adding replichore arrows

Objective: To add an arrow panel representing replichore arrows to the image.

1) Access the Arrow dialog by selecting **Window > GenVision Dialogs > Arrow**. In the resulting dialog, click the small arrow in the top right corner and select **Add Arrow Panel**.
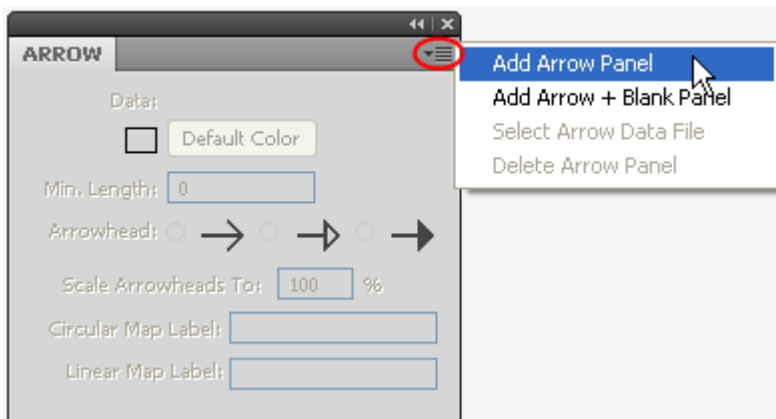
2) When prompted, navigate to and select **ReplicationArrow.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of the two *E. coli* replichores, as well as the thickness and color for the arrows depicting them.

3) The circular map image will be updated to display the two replichore arrows. As with the scale panel, the arrows currently appear very large, but will continue to decrease in size as more panels are added to the figure.

4) Also notice that the two arrows on the map are shown in a turquoise color even though the **Default Color** in the Arrow dialog shows bright green. This is because the color for the arrows was specified in the data file we selected. Colors specified within data files always take precedence over the color settings in any panel dialog.

5) In the Arrow dialog, type "Replichores" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend.
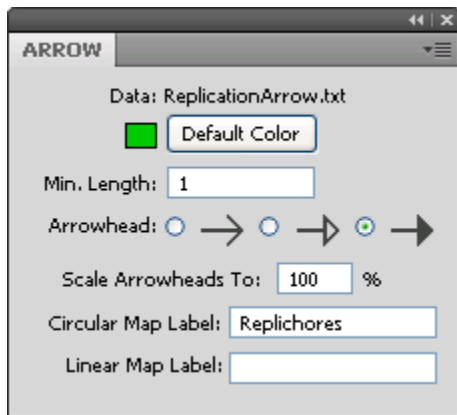
Adding tRNA genes

Objective: To add an arrow panel representing tRNA genes to the image.

1) From the open Arrow dialog, click the small arrow in the top right corner again and select **Add Arrow Panel**.
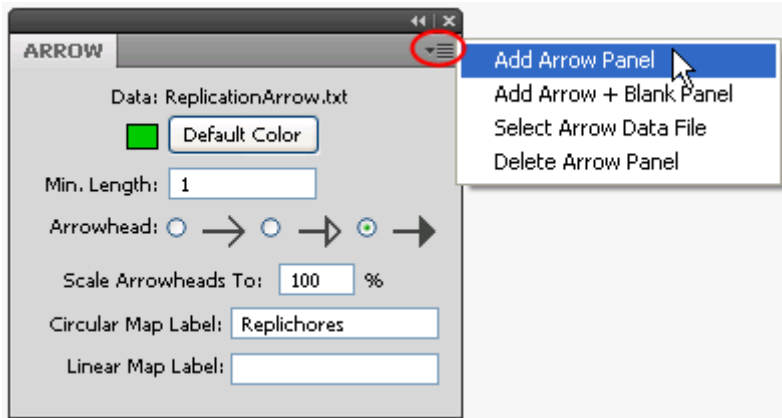
2) When prompted, navigate to and select **tRNAArrows.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of each of the *E. coli* tRNA genes, as well as the thickness, color and arrowhead style for the arrows representing them.

3) The circular map image will be updated to display the tRNA arrows, drawn in red and yellow. As with the replichore arrow panel, the colors for the tRNA arrows were specified in the data file we selected.

4) In the Arrow dialog, increase the **Min. Length** value to **40,000**. This value specifies that arrows will be drawn at a length of 40,000 nucleotides, even though the genomic coordinates for each tRNA gene will be much smaller.

5) Finally, type "tRNA" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend. The Arrow dialog should now appear as follows:

Adding rRNA genes

Objective: To add an arrow panel representing rRNA genes to the image.

1) From the open Arrow dialog, click the small arrow in the top right corner and select **Add Arrow Panel**.
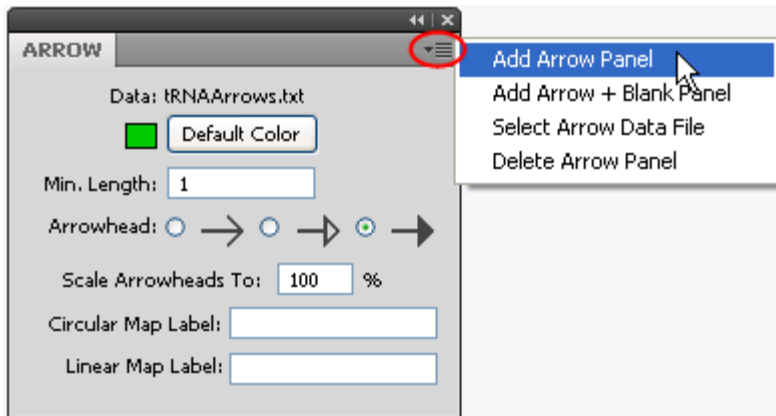
2) When prompted, navigate to and select **rRNAArrows.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file defines the genomic locations of each of the *E. coli* rRNA genes, as well as the thickness, color, and arrowhead style for the arrows representing them.

3) The circular map image will be updated to display the rRNA arrows, drawn as yellow and red closed arrows.

4) In the Arrow dialog, increase the **Min. Length** value to **50,000**. This value specifies that arrows will be drawn at a length of 50,000 nucleotides, even though the genomic coordinates for each rRNA gene will be much smaller.

5) Finally, type "rRNA" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend. The Arrow dialog should now appear as follows:
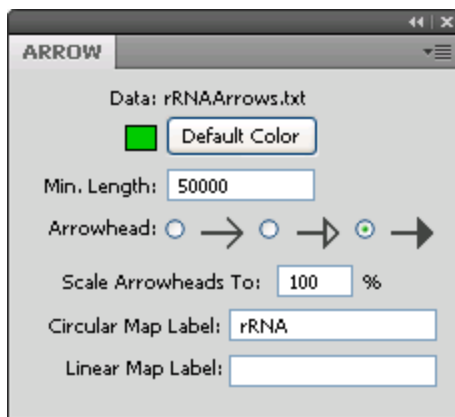
6) Close the Arrow dialog before proceeding to the next section.

Adding ORFs

Objective: To add a histogram representing ORFs to the image.

1) Access the Histogram dialog by selecting **Window > GenVision Dialogs > Histogram/Graph**. In the resulting dialog, click the small arrow in the top right corner and select **Add Histogram Panel**.
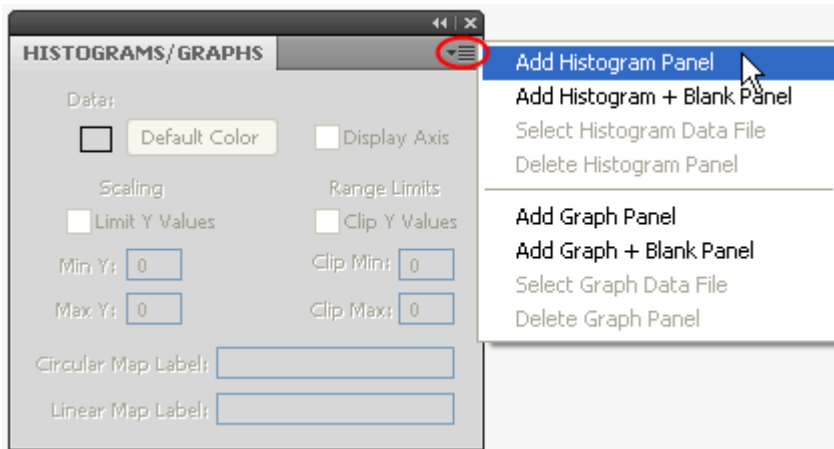
2) When prompted, navigate to and select **ORFsHistogram.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the locations of the *E. coli* genes, as well as the color for the histograms representing them.

3) The circular map image will be updated to display the ORF histogram, drawn in yellow and red.

4) Type "ORFs" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend.
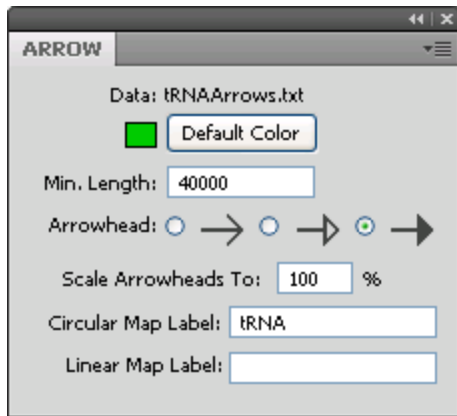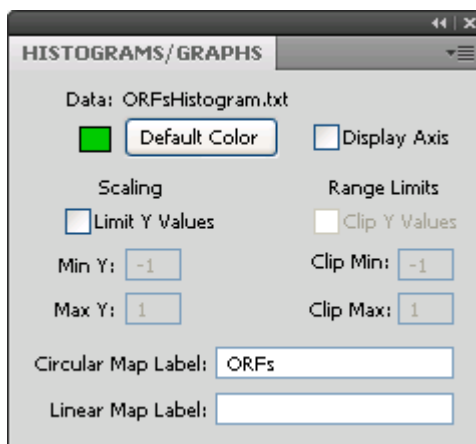
Adding codon usage

Objective: To add a histogram representing the inverse codon adaptation index for E. coli genes. This index is used to flag genes that may have been acquired by horizontal gene transfer.

1) From the open Histograms/Graphs dialog, click the small arrow in the top right corner and select **Add Histogram Panel.**
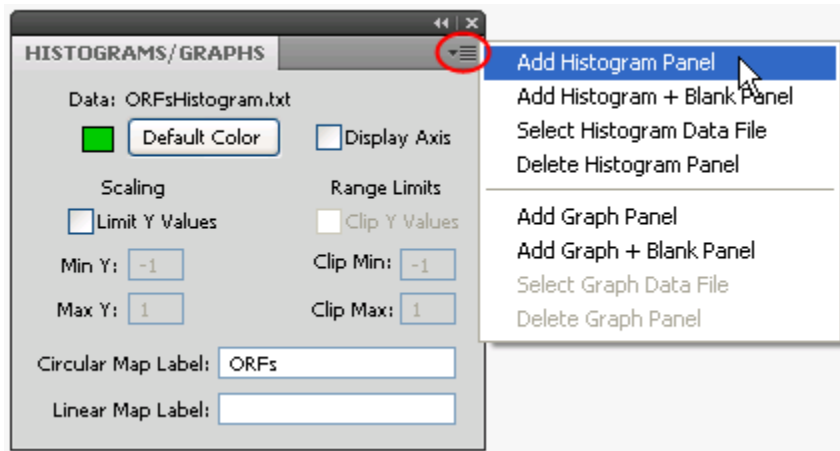
2) When prompted, navigate to and select **CodonUseHistogram.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, height and color of each bar in the histogram.

3) The circular map image will be updated to display codon usage histogram you just added, drawn in yellow, red, and orange.

   Note: Unlike the ORFs histogram (red and yellow rings), the height of the Codon Use histogram varies, reflecting the "Inverse Codon Adaptation Index"—a measure of codon usage for each ORF. It is thought that the higher the inverse CAI, the more likely it is that the ORF was acquired in the *E. coli* chromosome by horizontal transmission. To emphasize the ORFs most likely to have been acquired by horizontal transmission, histogram boxes for ORFs with inverse CAI values above 0.75 are colored yellow instead of red or orange.

4) From the Histograms/Graphs dialog, type "Codon Usage" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend.
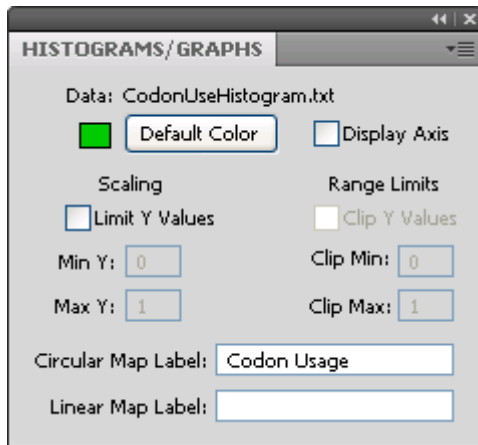
Adding a graph panel

Objective: To add a graph to the image representing remnants of bacteriophages.

1) From the open Histograms/Graphs dialog, click the small arrow in the top right corner and select **Add Graph Panel**.
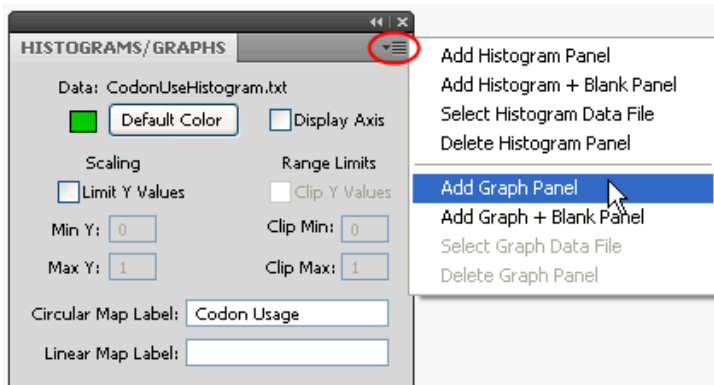
2) When prompted, navigate to and select **PhagesGraph.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the x- and y- coordinates for each point on the graph.

3) Click anywhere outside the artboard to view the newly added graph, drawn in bright green in the center of the image. Notice that for this particular data set, the few high y-value data points tend to mask the variation in the majority of the data points that have much lower values. (If you wish, you can confirm this by opening **PhagesGraph.txt** in

Microsoft Excel and sorting by the **Top** column, which contains the y-coordinates for each data point).

4) In order to greater emphasize the differences among the lower data points, first access the Panels dialog by selecting **Window > GenVision Dialogs > Panels**. (The Panels dialog allows you to edit, move, re-size or delete any of the panels in your project, as well as import panels from an existing project). Select the graph panel from the Panels dialog by clicking on **PhagesGraph.txt**.
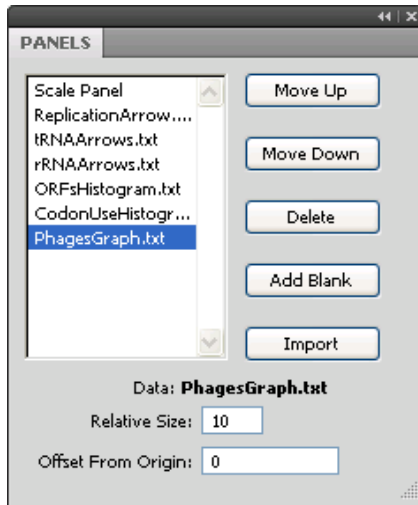
5) From the Histograms/Graphs dialog, check the box next to **Limit Y Values**, and change the **Max Y** value to **0.3**. This will cause GenVision to scale your graph so that the top of the graph panel is equal to the y- coordinate 0.3.

6) Next, check the box next to **Clip Y Values** and set the **Clip Max** value to **0.6**. This will allow data points greater than 0.3 and up to 0.6 to extend beyond the space allocated for your panel.

7) Finally, type "Phages" into the **Circular Map Label** field. This provides the text that will be used in the circular map legend.
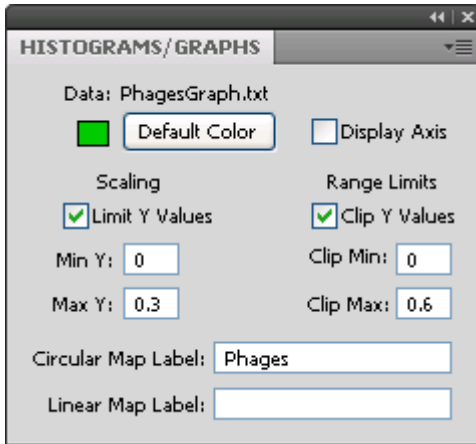
8) When you are finished, press **Enter**. Notice that the graph towards the center of the map now shows greater variation, and the highest peaks overlap the orange codon usage histogram. At the end of this tutorial, we will add some space between the two panels so that the amount of overlap is reduced.

9) Close the Histograms/Graphs dialog before proceeding to the next section.

Adding text to the image

Text panels allow you to place text alongside the other panels in your image with respect to genomic coordinates.

Objective: To add a text panel to the map.

1) Access the Text dialog by selecting **Window > GenVision Dialogs > Text**. In the resulting dialog, click the small arrow in the top right corner and select **Add Text Panel**.
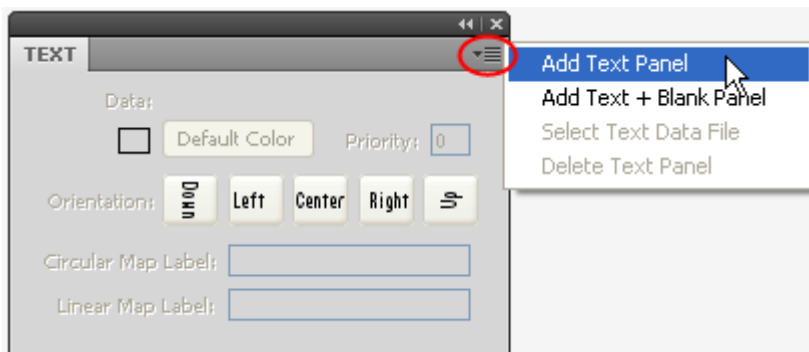
2) When prompted, navigate to and select **GenVisionText.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies

the position and text for the label that will reside around the perimeter of the *E. coli* genome.

3) Notice that the circular map image is updated to display the text "**GenVision by DNASTAR**" in two segments of the genome, centered on coordinates 300,000 and 2,619,611.

4) In order to display the text so that it appears as the outer-most ring of the circular map, first select the text panel from the Panels dialog by clicking on **GenVisionText.txt**.
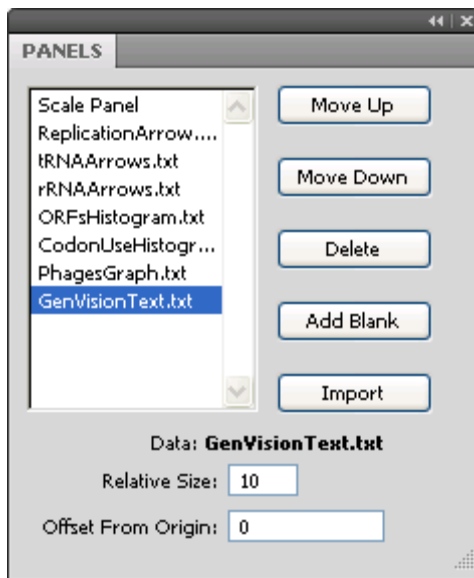
Note: Panels in the Panels dialog are listed from outer- to inner- most for circular maps.

5) Click the **Move Up** button so that the text panel appears above the phages graph in the Panels dialog.

6) Continue clicking the **Move Up** button until the GenVisionText.txt panel appears at the top of the list. The circular map image will be updated to show the text on the outside of the scale.

7) Close the Text dialog before proceeding to the next section.

Adding demarcation lines to the image

The final panel for this project is a demarcation panel. A demarcation is a line drawn radially on a circular map, perpendicular to the genomic coordinate axis.

Objective: To add demarcation lines to the image defining the replication origin and terminal regions.

1) Access the Demarcation dialog by selecting **Window > GenVision Dialogs > Dermarcation**. In the resulting dialog, click the small arrow in the top right corner and select **Add Demarcation Panel**.
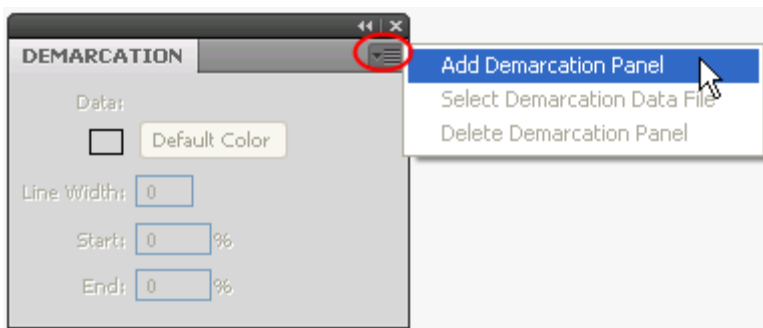
2) When prompted, navigate to and select **OriTerDemarcation.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the genomic locations of the two demarcations for the *E. coli* map, one at the origin, and the other at the terminus of replication.

3) From the Demarcation dialog, leave the **Start percentage** value at **0** and change the **End percentage** value to **100**. These values specify the portion of the figure where the demarcation lines will be drawn. In this case, the lines will span the entire panel space.

4) Press **Enter**. Notice that the circular map image has been updated to include the demarcation lines, spanning across the entire allotted panel space in your figure. The circular map should now look similar to the following image:
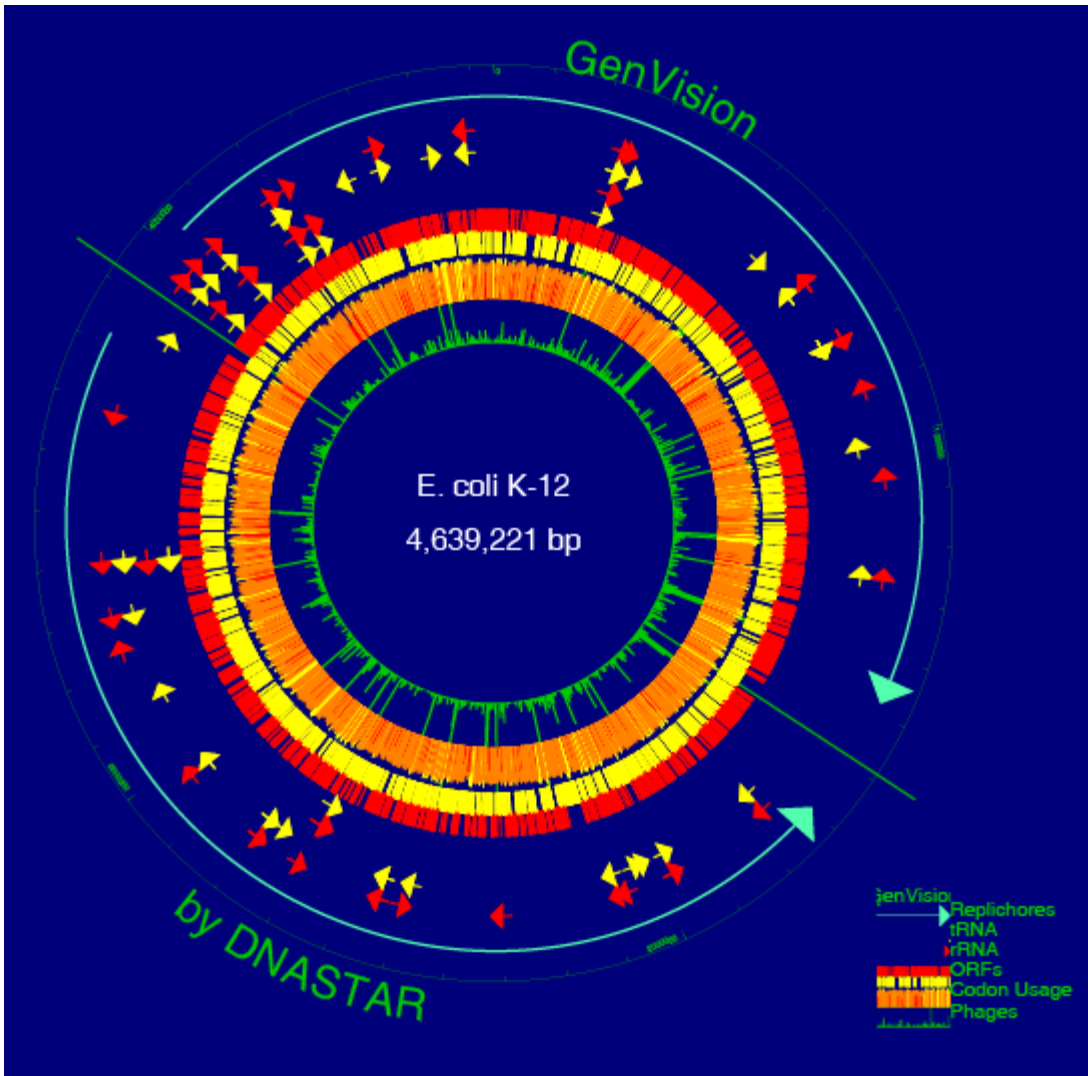
5) Close the Demarcation dialog before proceeding to the next section.

Adding blank space and resizing panels

Notice that the three histogram panels towards the center of the image appear bunched together, and the phages graph in the center of the map overlaps the orange codon usage histogram quite a bit. In addition, the text on the outside of the map is too large and the scale is too small. All of these issues can be addressed by using the Panels dialog.

Objectives: To resize panels and add blank space between panels in order to finalize the appearance of the circular map image.

1) From the Panels dialog, first select the **GenVisionText.txt panel** by clicking on its row. Note that the **Relative Size** value for the text panel is **10**. As mentioned earlier, each panel's size is relative to *all* of the panels in the image.
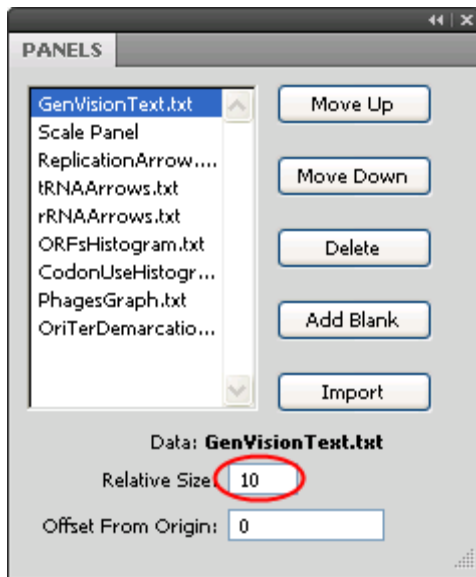
2) Decrease the **Relative Size** value for the text panel to **8**, then press **Enter**. The circular map image will be updated, and the "GenVision by DNASTAR" text on the perimeter of the map will appear slightly smaller than it did before.

3) Next, select the **Scale Panel** from the Panels dialog. Note that the **Relative Size** value for the scale panel is **2**
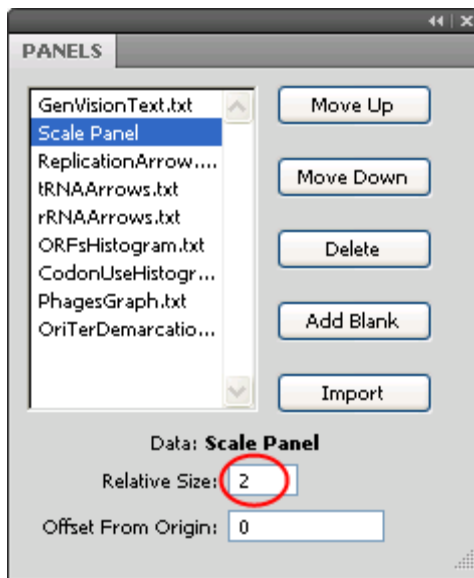
4) Increase the **Relative Size** value for the scale panel to **5**, and then press **Enter**. The circular map image will be updated to show a slightly larger scale.

5) Although the numbers on the scale are now easier to read, the tick marks and axis of the scale are still very thin. To make them bigger, first select the **Scale Panel** again from the Panels dialog. Then, open the Scale dialog by selecting **Window > GenVision Dialogs > Scale**.
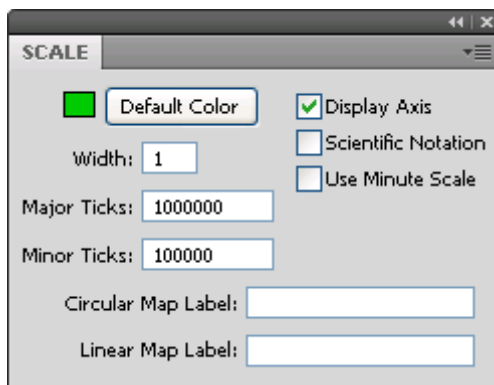
6) In the Scale dialog, change the **Width** value to **3** and then press **Enter**. The circular map image will be updated again, and the entire scale should now be easier to read.

7) In order to add some space between the histogram panels, next select **ORFsHistogram.txt** from the Panels dialog.

8) Click the **Add Blank** button. The circular map image will be updated, and a blank space will now appear between the red and yellow ORFs and the orange codon usage histograms:
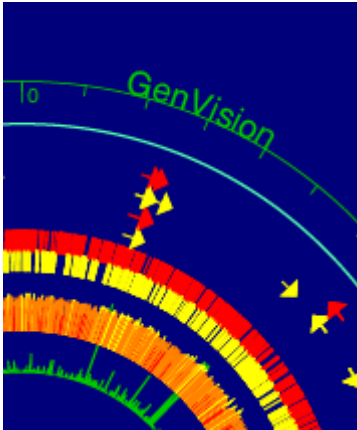
9) Next, select **CodonUseHistogram.txt** from the Panels dialog. Note that the **Relative Size** value for this panel is **10**.

10) In order to see more variation within this histogram panel, increase its **Relative Size** to **20** and then press **Enter**. The circular map image will be updated.

11) Select **CodonUseHistogram.txt** from the Panels dialog again, and then click the **Add Blank** button. The circular map image will be updated to include the new blank panel. Notice that the green bacteriophage graph no longer overlaps the orange codon usage histogram.

12) The completed circular map should now look similar to the following image:
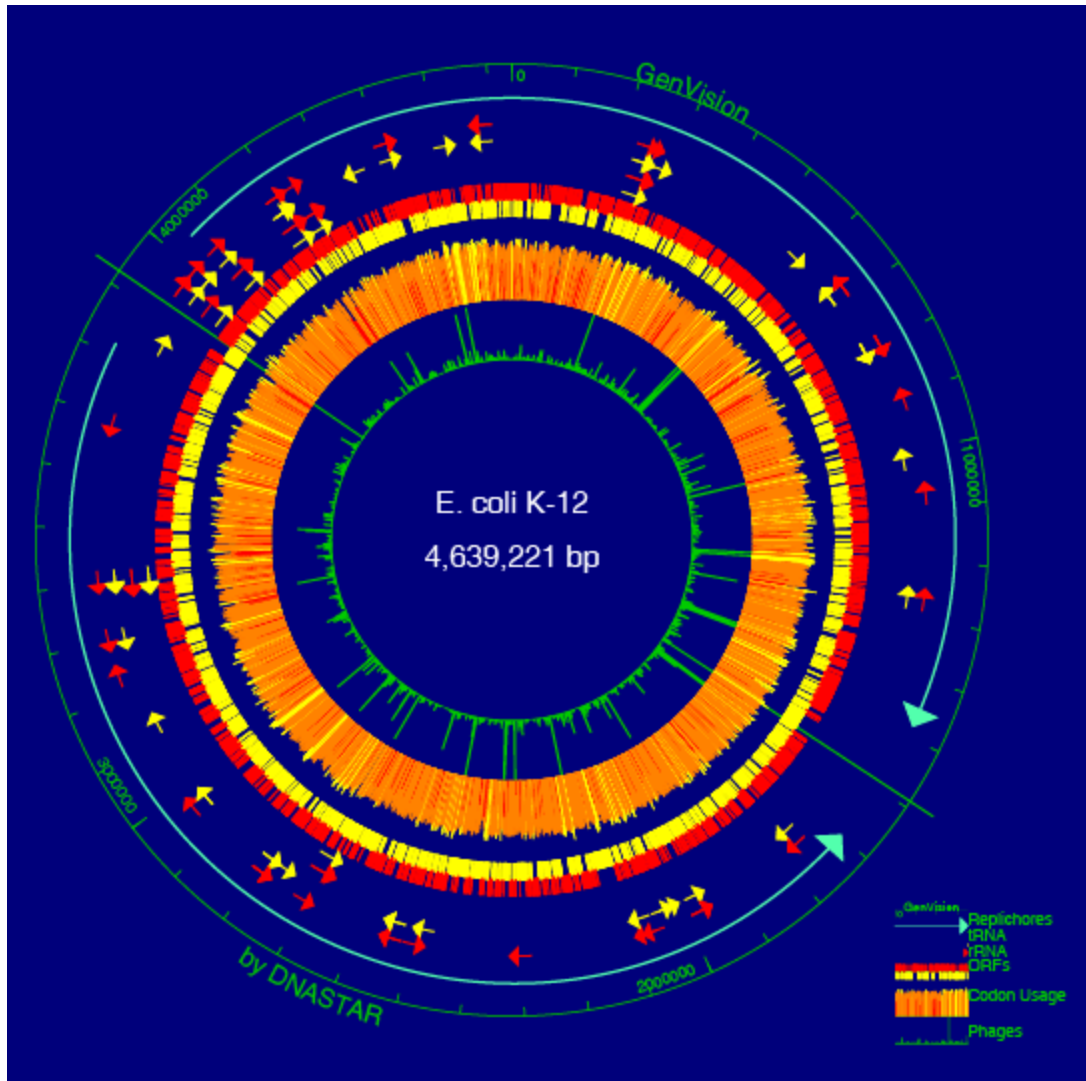
13) The completed map is now ready for printing or for insertion into another document. Select **File > Save** and select **GenVision Config File (*.gnv)** from the **Format/Save as Type** (Mac/Win) field to save the GenVision project for later use.

Designing a Linear Map

Objective: The purpose of this tutorial is to familiarize you with building a linear map image using the GenVision Plug-In. For this example workflow, we will create a map using the first 600,000 bp of the *E. coli* genome.

Creating a new linear map image

Objective: To create a new linear map figure for a portion of the E. coli genome.

1) Launch Adobe Illustrator if it is not already open, and select **File > New**.

2) Enter "Ecoli Linear Map" into the Name field, choose landscape orientation, select **RGB** as the Color Mode from the Advanced section, and then click **OK**.
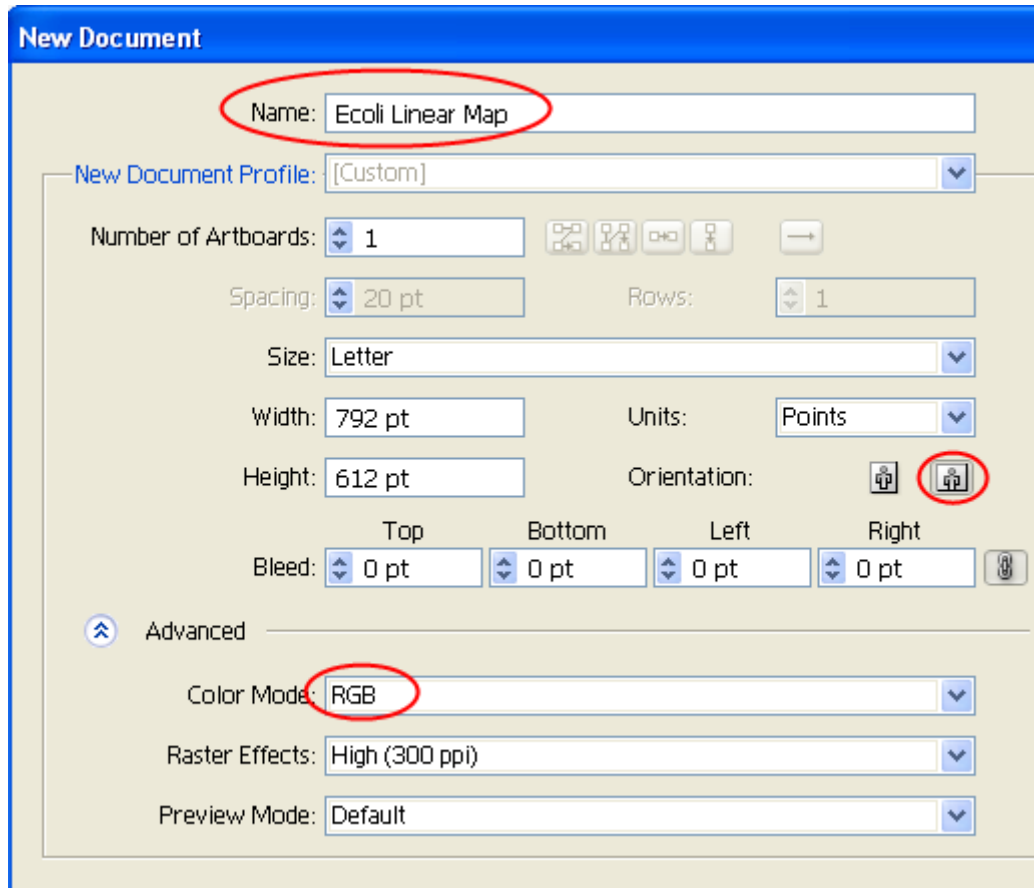
3) Open the Figure Settings dialog by selecting **Window > GenVision Dialogs > Figure Settings**.
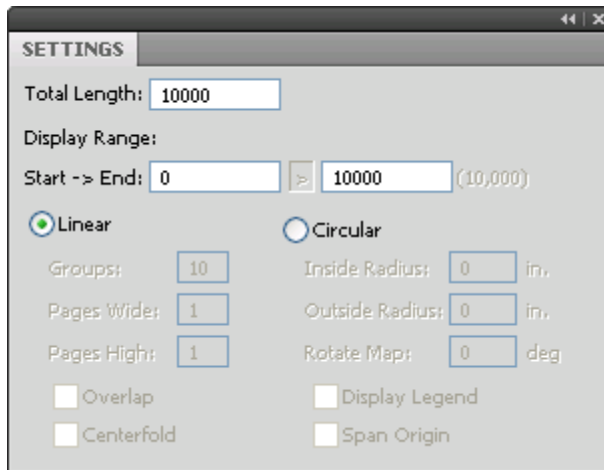
4) Enter **4,639,221** (without commas) into the **Total Length** field. This specifies the length of the entire *E. coli* genome.

5) Under **Display Range**, enter **1** as the **Start** value and **600,001** as the **End** value. Doing this will limit the map we are creating to the first 600,000 nucleotides of the genome.

6) Under the **Linear** options, Change the **Groups** value to **12.** A *group* is a single set of all the data panels in your project. This setting specifies that only twelve groups of panels will be displayed per page.

7) Change the **Pages Wide** value to **3** to specify that the data will be plotted across the width of three pages.

---
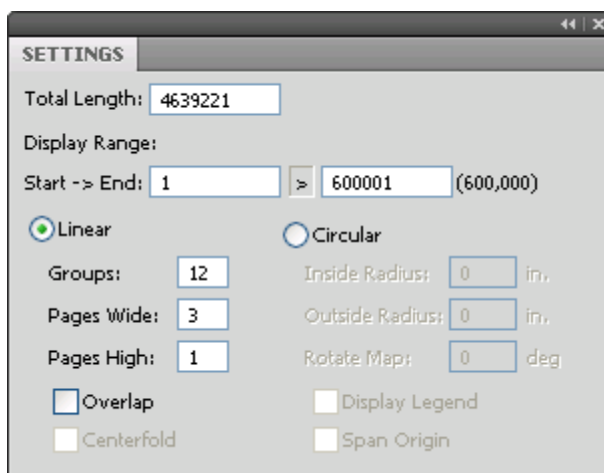
8) Press **Enter**. The following message may appear, notifying us that we may need to adjust the print settings in Illustrator to accommodate the changes we made to the layout of the map:

To correctly print the figure it may be necessary to adjust the paper size and orientation, set the tile full pages option, or adjust the page grid using the page tiling tool.

Please refer to the Illustrator help for more information.

OK
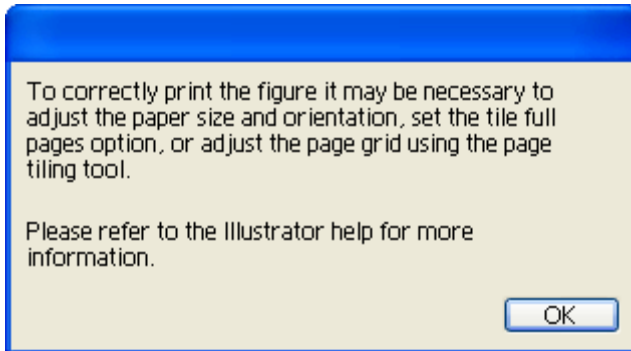
9) Click **OK** to close the notification.

10) Double-click on the Drawing Size Tool ⊢–⊣ from the Illustrator toolbar. The following dialog will appear:

Graph Start: 0
Graph End: 4639221
Cancel   OK
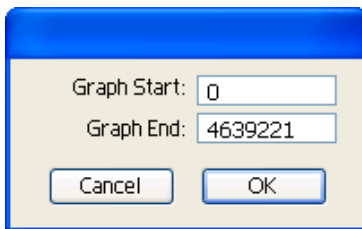
11) Change the **Graph Start** value to **1** and the **Graph End** value to **605,000** (without commas), and then click **OK**. Since we are mapping the first 600,000 bp of the genome, entering this value will create a space equivalent to 5,000 nucleotides at the end of the map to accommodate an End Legend.

Graph Start: 1
Graph End: 605000
Cancel   OK
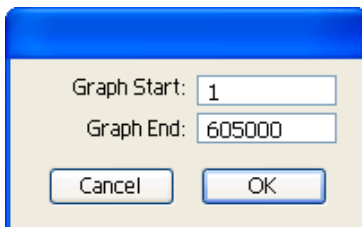
12) Close the Figure Settings dialog before proceeding to the next section.

Defining label settings

Labels are the text items that define the panels within your image. For linear maps, labels appear beside the specified panel in the margin of the map. The Labels dialog allows you to define how the linear map labels in your image are displayed.

Objective: To define the label settings for the linear map image.

1) Access the Labels dialog by selecting **Window > GenVision Dialogs > Labels**.
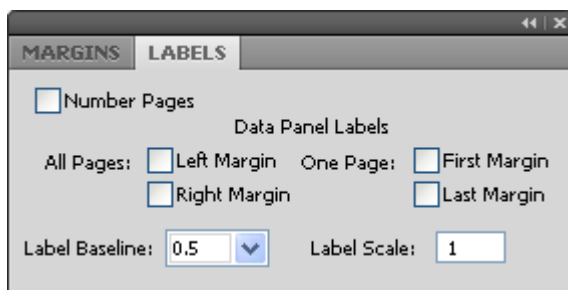
2) Check the box next to **Number Pages** to generate page numbers in the bottom left corner of each page.

3) Also check the box next to **Left Margin**. This setting indicates that the labels for each panel will be displayed in the left margins of each page of the linear map.

4) Close the Labels dialog before proceeding to the next section.

Adding panels to build the linear image

GenVision illustrations are built from one or more *panels*, each drawn using information from an associated *data file*. Data files contain the annotation data you wish to display in a tab-delimited format, and may be created manually, in a program such as Microsoft Excel, or automatically by exporting from a Lasergene application. There are several different types of panels that you may use to visualize your annotation data, including arrows, histograms, graphs, and text.

For more information on creating your own data files, please see the section "*Working with Data Files*" in the GenVision Help which can be viewed as follows:

* Windows: **Help > About DNASTAR Plug-ins > GenVision Help**

* Macintosh: **Illustrator > About DNASTAR Plug-ins > GenVision Help**

Objective: To build a linear map image of a portion of the E. coli genome by adding 6 different panels.

Displaying transcripts and genes

Objective: To add an arrow panel representing transcripts and a histogram panel representing the *E. coli* genes.

1) Access the Arrow dialog by selecting **Window > GenVision Dialogs > Arrow**. In the resulting dialog, click the small arrow in the top right corner and select **Add Arrow Panel**.
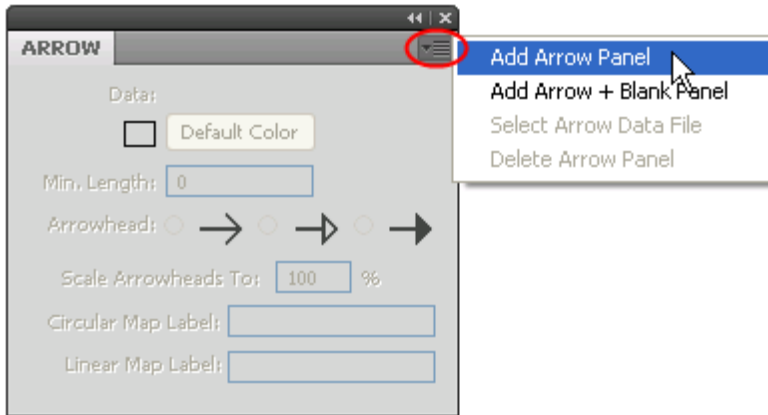
2) When prompted, navigate to and select **Transcripts.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, thickness and color for the arrows representing the transcripts.

3) The linear map image will be updated, displaying the newly added black and red transcript arrows.

4) In the Arrow dialog, change the **Min. Length** value to **100**. This specifies that any arrows less than 100 bases in length will be drawn as if they were 100 bases long, enabling you to see the arrow shafts more easily.

5) Finally, type "Transcripts" into the **Linear Map Label** field. This provides the text that will be used in the left margins of the linear map to identify this panel.
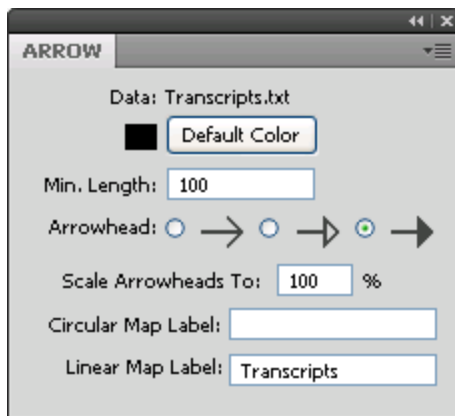
6) Next, access the Histograms/Graphs dialog by selecting **Window > GenVision Dialogs > Histogram/Graph**. In the resulting dialog, click the small arrow in the top right corner and select **Add Histogram Panel**.
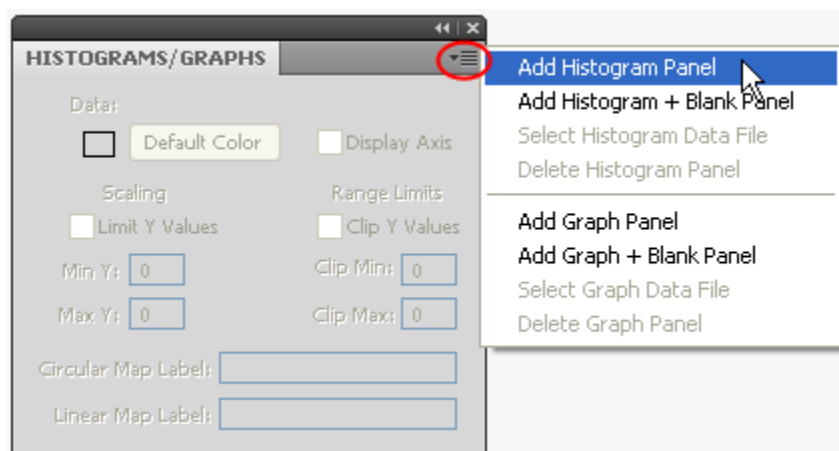
7) When prompted, navigate to and select **Genes.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the location, color and height for the histogram depicting the known and putative *E. coli* genes.

8) The linear map image is updated to include the histogram panel representing *E. coli* genes.

Note: The histogram is displayed in several different colors. These colors were defined within the data file we selected and are being used to indicate gene function. Later in this tutorial we will add an *End Legend*, which will act as a key for the color coding that is used.

---

9) In the Histograms/Graphs dialog, check the box next to **Display Axis** so that a line will be drawn across the axis, separating positive and negative values.

10) Finally, type "Genes" into the **Linear Map Label** field. This provides the text that will be used in the left margins of the linear map to identify this panel.
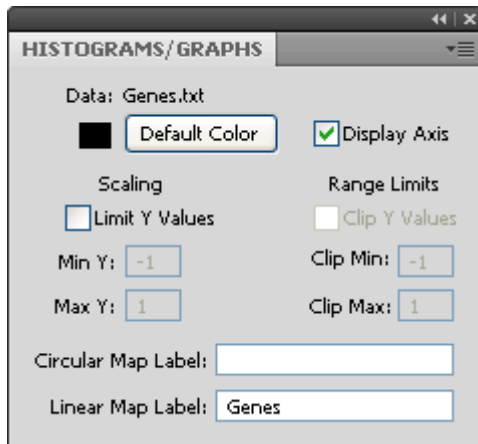
11) Close the Arrow dialog and the Histograms/Graphs dialog before proceeding to the next section.

Adding ORF names

Objective: To add a text panel to the image displaying the names of the ORFs.

1) Access the Text dialog by selecting **Window > GenVision Dialogs > Text**. In the resulting dialog, click the small arrow in the top right corner and select **Add Text Panel**.
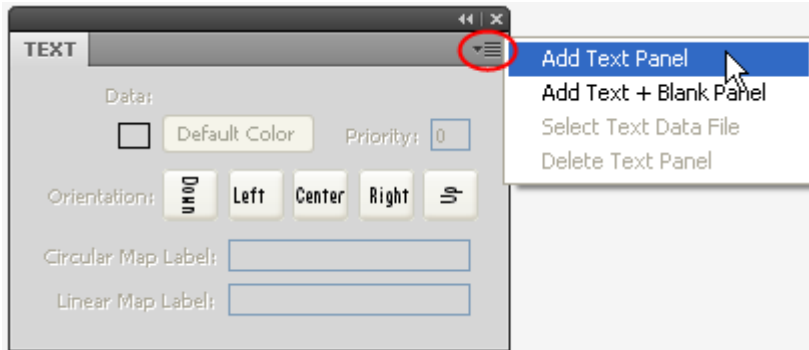
2) When prompted, navigate to and select **ORFNames.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the position and text for the gene labels.

3) The linear map image is updated to include the newly added gene labels.

4) Type "ORF Names" into the **Linear Map Label** field, and then press **Enter**. This provides the text that will be used in the left margins of the linear map to identify this panel.

5) Notice that the gene labels are displayed underneath the gene histogram panel. In order to display the text so that it appears at the top of the panel space, first access the Panels dialog by selecting **Window > GenVision Dialogs > Panels**. (The Panels dialog allows you to edit, move, re-size or delete any of the panels in your project, as well as import panels from an existing project). Select the text panel from the Panels dialog by clicking on **ORFNames.txt**.
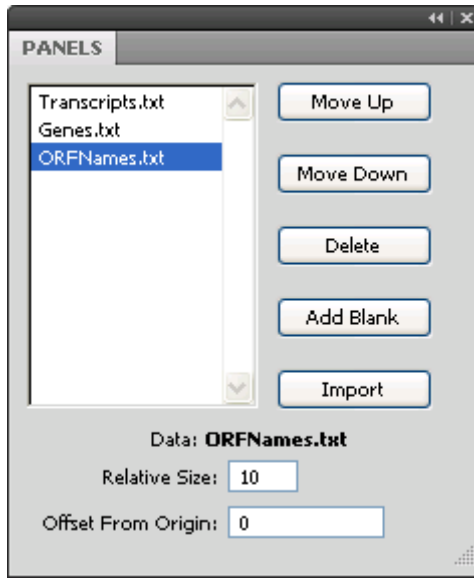
6) Click the **Move Up** button twice so that the text panel appears at the top of the list. The linear map image will be updated to show the gene labels at the top of each panel group.

7) Close the Text dialog before proceeding to the next section.

Adding and optimizing a scale

Scales allow you to see genomic coordinates on your image, and are marked in exact nucleotide positions.

Objective: To add a scale to the image and optimize it for the size of the map.

1) Access the Scale dialog by selecting **Window > GenVision Dialogs > Scale**. In the resulting dialog, click the small arrow in the top right corner and select **Add Scale Panel**.
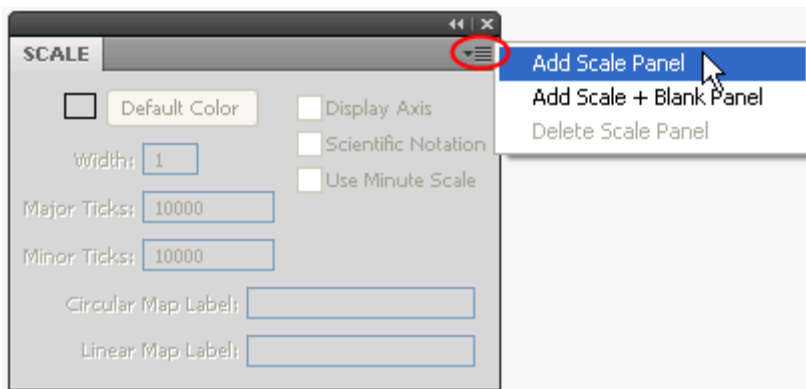

2) In the Scale dialog, decrease the **Major Ticks** interval to **10,000** (without commas), the **Minor Ticks** interval to **1,000**, and then press **Enter.**
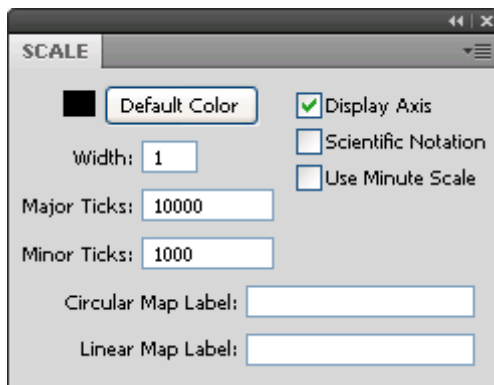
Note: Major tick marks in the scale will be numbered, while the minor ticks will remain unnumbered.

3) The linear map image will be updated to include the scale.

4) Close the Scale dialog before proceeding to the next section.

Adding blank space and resizing panels

Notice that the scale in the linear map image appears very small, and also that it appears very close to the histogram panel. Both of these issues can be addressed by using the Panels dialog.

Objectives: To resize panels and add blank space between panels in order to finalize the appearance of the linear map image.

1) From the Panels dialog, first select the **Scale Panel** by clicking on its row. Note that the **Relative Size** value for the scale panel is **2**. As mentioned earlier, each panel's size is relative to *all* of the panels in the image.
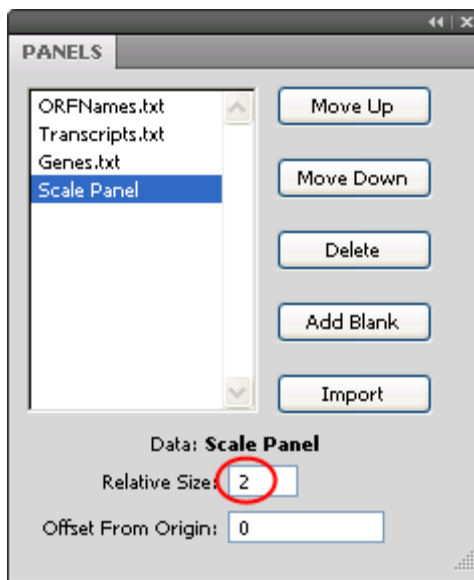
2) Increase the **Relative Size** value for the scale panel to **5**, and then press **Enter**. The linear map image will be updated to show a larger scale that is easier to read.

3) In order to add some space between the histogram panel and then scale, next select **Genes.txt** from the Panels dialog.
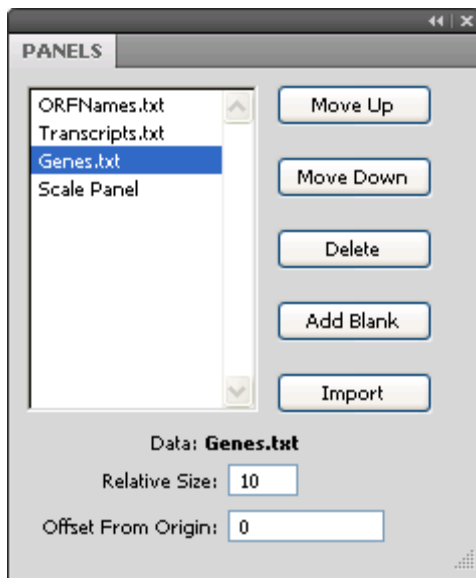
4) Click the **Add Blank** button. The linear map image will be updated, and a small blank space will now appear between the histogram and the scale:
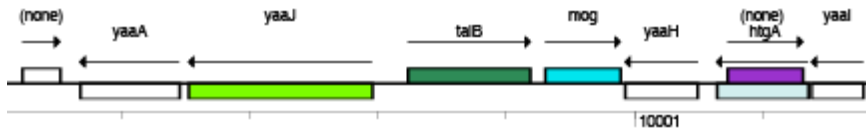
5) To increase the space between the bottom of the scale and the top of the gene labels for the panel group below, first select the Scale Panel from the Panels dialog by clicking on its row:
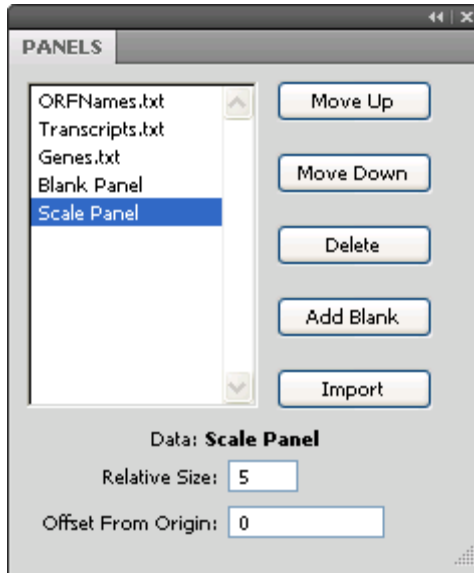
6) Then, click the **Add Blank** button. Each panel group in the linear map will now appear slightly more separated.

Adding an end legend

Start/End Legends are a type of legend specific to linear maps that are used as keys for panels that use color coding. As the names indicate, they are displayed either at the start or end of the linear map.

Objective: To add a legend at the end of the map with a key showing gene function.

1) Access the Legends dialog by selecting **Window > GenVision Dialogs > Start/End Legends**. In the resulting dialog, click the small arrow in the top right corner and select **Add End Legend**.
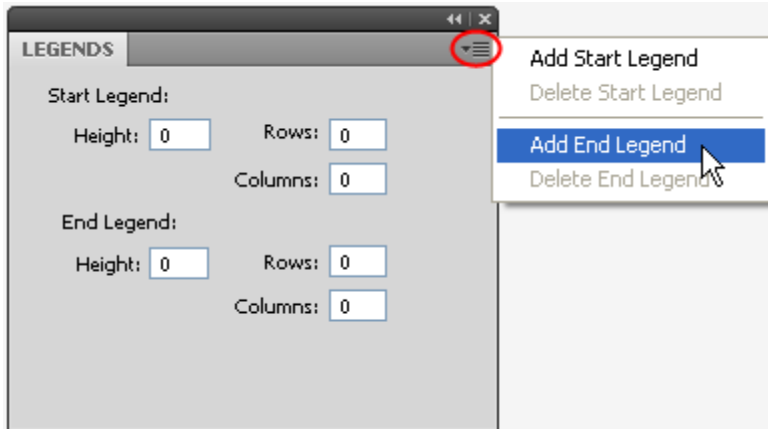
2) When prompted, navigate to and select **EndLegend.txt** from the **Demo GenVision** folder. (See the beginning of this tutorial for the full directory). This data file specifies the organization and color of a legend defining gene functions.

3) The following notification may appear:
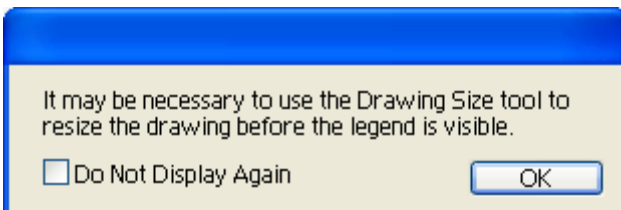
4) Because we already added space for the End Legend earlier in this tutorial, simply click **OK** in this dialog.

5) In the End Legend section of the Legends dialog, change the **Rows** value to **16**, and leave the **Columns** value as **20**. These values define the size of the grid that will be used for the legend.

6) Press **Enter**. The completed linear map image will be updated, including the end legend. In order to see the End Legend, scroll to the third page of the linear map image and look in the bottom right corner. If necessary, use the zoom tool to view it:
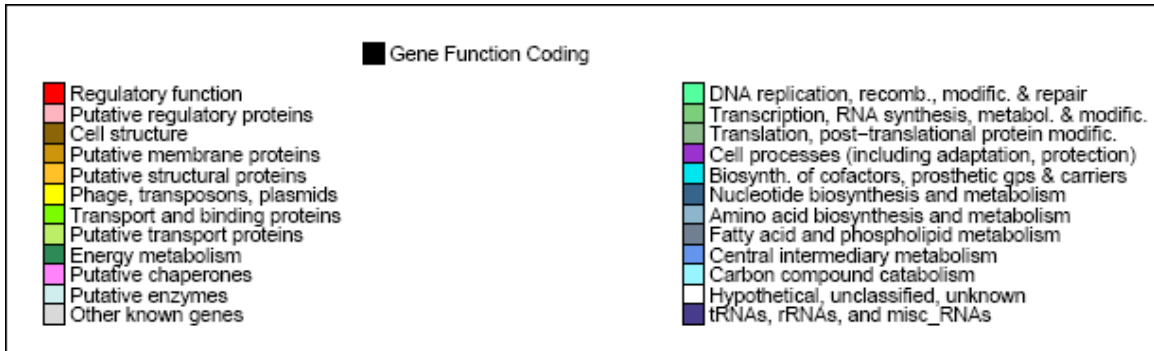
7) The completed linear map image should look similar to the small segment displayed here:
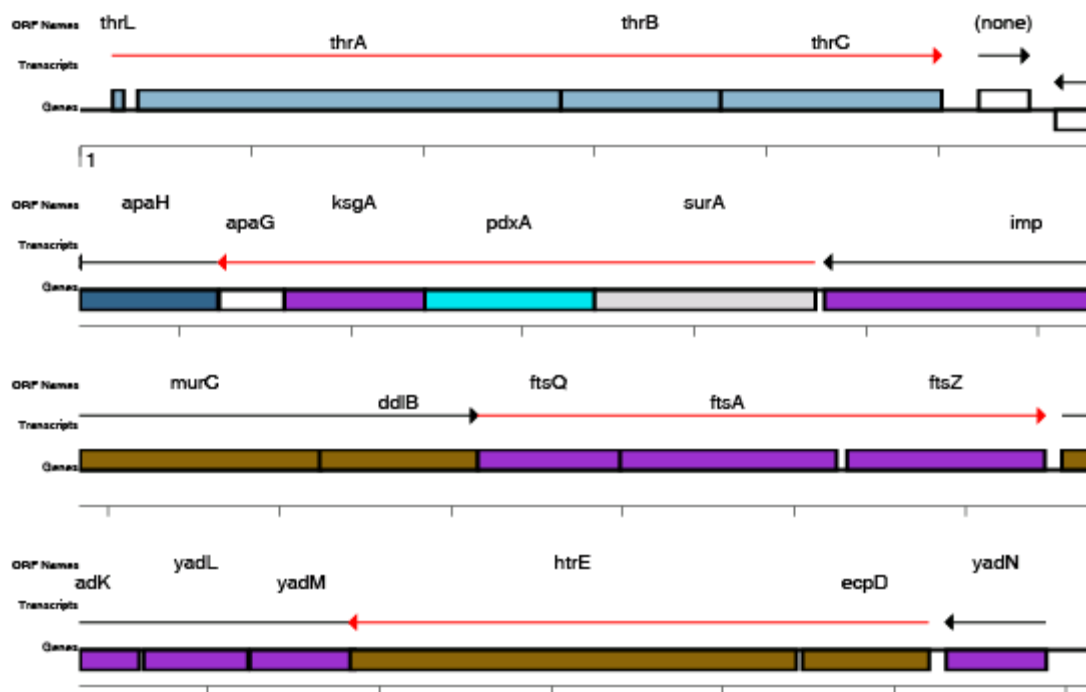
8) The completed map is now ready for printing or for insertion into another document. Select **File > Save** and select **GenVision Config File (*.gnv)** from the **Format/Save as Type** (Mac/Win) field to save the GenVision project for later use.

Protean

Overview

Protean aids you in predicting and annotating the structural character of your protein sequence and provides a graphical display and tabular summaries of sequence composition and other data. The project begins when you enter a protein sequence. The next step is to evaluate secondary structural characteristics of the polypeptide sequence by applying some of Protean's many analysis methods to the sequence. These methods include:

- **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 mini-map.

- **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 - Deléage-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.

- **Post-Translational - Sumoylation**—predicts sites that undergo the addition of small, ubiquitin-like modifiers after translation.

- **Features**—Not a true "method," but a list of known features for the sequence.

Once the methods have been applied, the results are displayed graphically on a common horizontal scale. In addition, menu commands let you display tabular data summaries, calculate titration curves, create protease digestion maps, and perform SDS PAGE simulations.

Protean reads and interprets GenBank-formatted features from the comment or features pane of sequence documents. Furthermore, you may create your own annotations for features of interest and even link several features together under a common title and description.

You can also use Protean to create images for publication. You can juxtapose and superimpose method results graphically or utilize one of Protean's tabular data summaries. You may also wish to take advantage of Protean's specialized protein fonts, which allow you to view sequences as chemical formulas or space-filling models.

10 Quick-Start Steps for Using Protean

Here are ten basic steps for using Protean:

1) Create a new project (**File > New**) or open an existing project (**File > Open**) and review the results of the current method outline.

2) Select **Analysis > Show/Hide Available Methods** to display a list of methods, and select which of these you would like to apply to the sequence.

3) Move the methods you want from the More Methods menu to the Method Curtain. Apply methods to the sequence by using the mouse to drag them from the Method Curtain to the assay surface. Continue to add, delete or rearrange method displays until you are satisfied with the look of the document.

4) View the results in tabular form, if desired, using **Analysis > Tabular Data**.

5) Edit method parameters at any time, either by double-clicking on the method display or on the method name in the Method Curtain. The method display is updated after each edit.

6) Use the graphical and tabular results to form hypotheses about your sequence, and then annotate any interesting features.

7) If you wish, create a structural model of your protein using **Analysis > Model Structure**, or calculate its titration curve using **Analysis > Titration Curve**.

8) If you have applied proteolytic digestion methods to the sequence, you may run a **Sites & Features > SDS Page Gel Simulation**.

9) Search BLAST or Entrez databases for matches to your sequence, or other sequences of interest.

10) **File > Save** your results, **File > Print** them, if desired, and quit using **File > Exit** (Win) or **PROTEAN > Quit Protean** (Mac).

Protean Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example Protean project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Correlating annotations with predicted structures

Objective: To open an existing protein assay and to view the locations of known calcium-binding sites and correlate them with predicted turns.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch Protean.

2) Select **File > Open** to open *Human Calmodulin.pad* from the **Demo Data** folder. Protean opens the project in a new window.
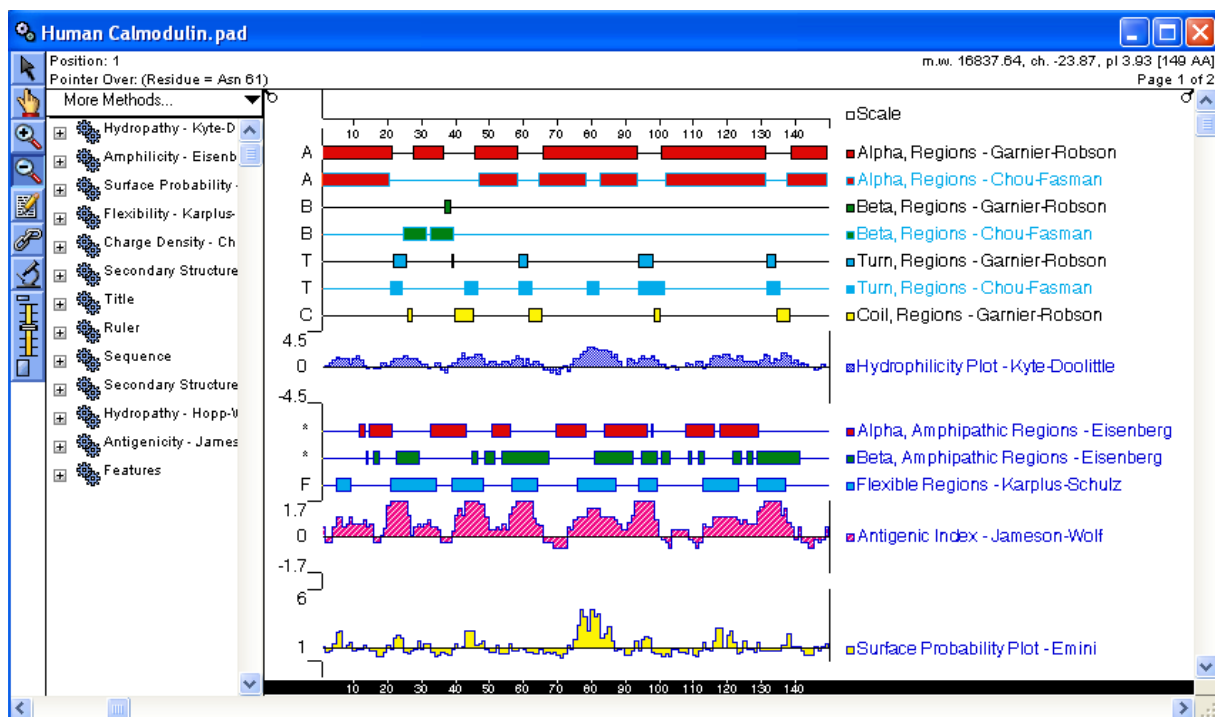
Several analytical method displays are already present on the **Assay Surface**, the main area of the Protean window. These methods are listed on the right side of the figure above.

This figure displays the **Method Curtain**, open at the left. The Method Curtain contains methods that can be applied to the assay surface using a drag and drop method.

3) In the Method Curtain, click on the **triangle** (Mac) or the **plus sign** (Win) to the left of the method **Features** to reveal a list of annotations that were imported with the sequence. If the items in the list appear highlighted, click in the white space to their left to ungroup them.

4) Sort the features by name using the command **Sites & Features > Sort Features > Name**.

5) Using **Ctrl+click** (Win) or **Cmd+click** (Mac) highlight each of the four features named **Calcium binding region:Region**. If you cannot see the complete names of the features in the Method Curtain, widen it by dragging the curtain pull (⌀) to the right.

6) Drag the highlighted features to the right and drop them on the assay surface. The four features now appear on the assay surface and remain selected.
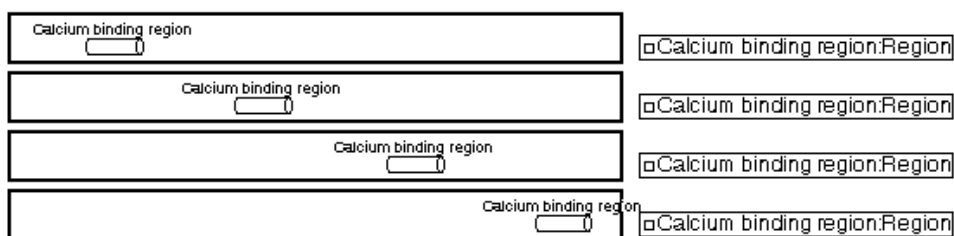
7) Select **Options > Superimpose Objects** to place all four calcium binding site locations (represented by cylinders) on the same line.
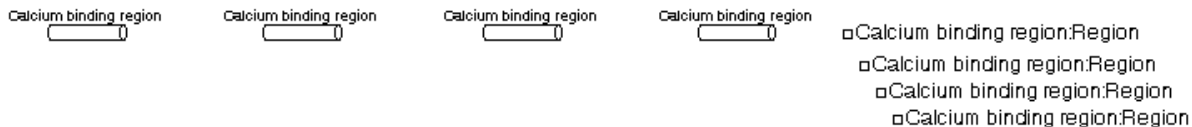
When you have predicted structures and features displayed in this way, you can begin to evaluate the structural characteristics of the protein and its corresponding features.

8) To make it easy to compare the calcium-binding site locations with predicted turns, scroll to the far right of the assay surface, select the name **Turn, Regions - Garnier-Robson** from the list, and drag it down to be just above the calcium binding site display name.
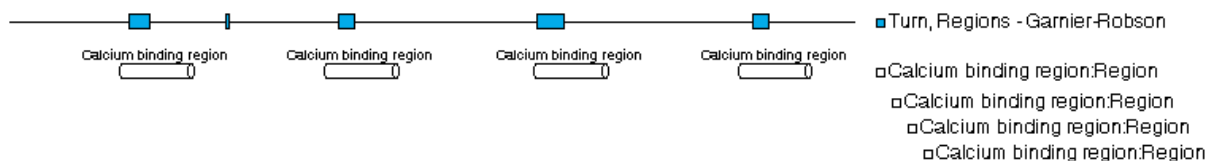
Note that the calcium binding sites are all located in predicted turn regions. This is consistent with the known structure of these binding sites, which are classical "EF Hand" calcium-binding sites wherein calcium is bound at the "elbow" between two adjacent alpha-helical regions.

Proceed to the next tutorial.

Calculating methods with different parameters

You can access method parameters by double-clicking on a method name or display anywhere in the Protean window. Note that parameter changes made to a method in the Method Curtain affect all current and future method displays created from that copy of the method. Similarly, parameter changes made to a display on the assay surface are applied to all displays originating from the same parent method in the method curtain. To view multiple displays of a method using different parameters, therefore, you can add a second (or third, etc.) copy of the method to the Method Curtain, then change method parameters separately for the added copy of the method.

Objective: To apply a new **Hydropathy – Kyte-Doolittle** method to the sequence, to change its parameters, and to compare it to the already applied **Hydropathy – Kyte-Doolittle** method.

1) Click **More Methods** at the top of the Method Curtain to open a menu containing methods. From this menu, you can select methods that aren't listed in the lower part of the method curtain or make additional copies of methods that are.

2) Select **Hydropathy – Kyte-Doolittle**. A new copy of the method now appears at the top of the method curtain list.

3) Now that the method is in the main body of the method curtain, it may be applied to the sequence. Click on the **plus sign** (Win) or the **triangle** (Mac) to the left of one of the Hydropathy – Kyte-Doolittle methods to reveal the four possible display types, shown below. If the items in the list appear highlighted, click in the white space to their left to ungroup them.
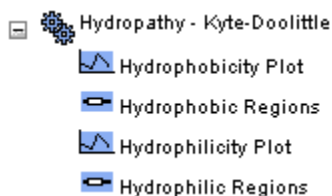
4) Click on **Hydrophilicity Plot**, drag it to the assay surface and drop it below the existing **Hydrophilicity Plot – Kyte-Doolittle** display.

- Predicted hydrophilic regions of your protein are shown as a black and white graph in the new method.

5) Double-click on the newly added display to open its method parameter dialog.

6) Change the **Residues to Average** to 3 and click **OK**.

- The original display used the default value of 9.

- The two displays now show the effects of applying the different parameter values:
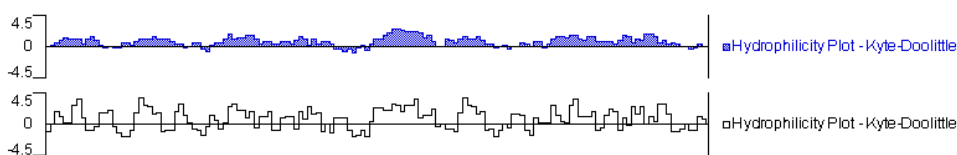
Proceed to the next tutorial.

Predicting proteolytic cleavage sites and performing SDS PAGE simulations

Objectives: To identify and display proteolytic sites, then to simulate an SDS PAGE run.

1) From **More Methods** at the top of the Method Curtain, select **Proteases – Protease Map**.

2) View the following information about the available proteases by choosing **Sites & Features > Show Protease List**.

3) When you are finished, select **File > Close** to close the Protease List.

4) Click on the **plus sign** (Win) or the **triangle** (Mac) to the left of **Proteases - Protease Map** method name to view the list of proteases. If the items on the list appear highlighted, click in the white space to their left to ungroup them.

5) Use **Ctrl+click** (Win) or **Cmd+click** (Mac) to highlight both **Chymotrypsin** and **CNBr**, and then drag them onto the assay surface. The resulting method displays show recognition sites as vertical bars.
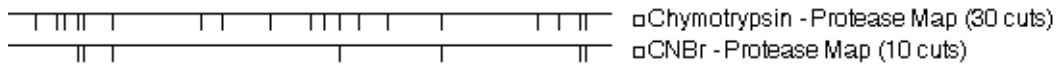
6) Leaving the displays highlighted, select **Sites & Features > SDS PAGE Gel Simulation**. An image of the simulated separation will appear in a new window. Above each column in the gel is the name of the corresponding protease or molecular weight standard.
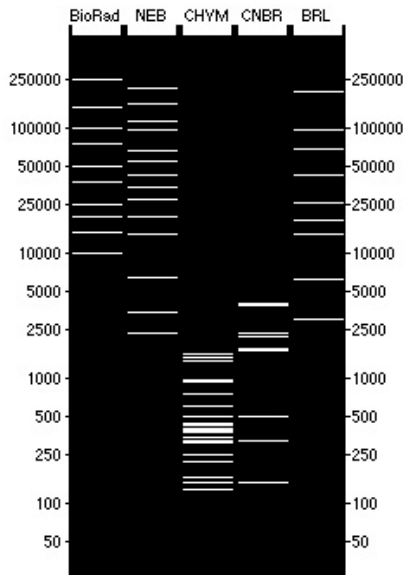
7) Move your cursor over the surface of the gel and notice how the header at the top of the window changes. When the cursor is over a protease fragment (CHYM or CNBR), the header displays its molecular weight, sequence range, HPLC retention time and predicted pI.

8) Select **Sites & Features > Protease Fragment Summary**. This opens a summary of information about the proteases applied to the sequence.

```
Chymotrypsin - 30 cuts
    1     1   149.20   5.55   >20h    1.33   M
    2     4   445.46   3.45   >20h    7.58   ADQL
    6     8   966.00   3.46   >20h   16.17   TEEQIAEF
   14     4   493.54   6.22     3m    6.61   KEAF
   18     2   218.24   5.55   >20h    5.53   SL
   20     1   165.18   5.55     3m    6.52   F
   21    13  1392.46   3.99     3m   15.36   DKDGDGTITTKEL
   34     4   406.49   5.55   >20h    5.60   GTVM
   38     3   374.43  10.04     2m    6.67   RSL
   41     9   957.96   3.57   >20h   10.90   GQNPTEAEL
   50     3   392.42   3.45    10m    4.35   QDM
   53    14  1479.51   3.07    30m   17.60   INEVDADGNGTIDF
   67     3   391.41   3.75     7m    9.64   PEF
   70     1   131.16   5.55     3m    4.52   L
   71     2   250.30   5.55   >20h    2.18   TM
   73     1   149.20   5.55   >20h    1.33   M
   74     4   504.64  11.05   >20h    4.23   ARKM
   76    13  1566.60   3.92     3m   17.58   KDTDSEEEIREAF
   91     3   420.50  10.04     2m    9.40   RVF
   94     7   767.73   3.92     3m    5.26   DKDGNGY
  101     6   602.67   3.75    30m   13.20   ISAAEL
  107     2   311.33  10.05     2m    0.00   RH
  109     2   246.33   5.55   >20h    3.67   VM
  111     3   346.37   5.55   >20h    2.45   TNL
  114     4   445.50   6.22   >20h    7.35   GEKL
  118     8   966.96   3.12   >20h    9.44   TDEEVDEM
  126    14  1564.62   3.58    30m   16.79   IREADIDGDGQVNY
  140     3   423.41   3.57    30m    6.99   EEF
  143     3   376.46   5.55   >20h    5.90   VQM
  146     1   149.20   5.55   >20h    1.33   M
  147     3   318.35   9.00   >20h    0.00   TAK

CNBr - 10 cuts
    1     1   149.20   5.55   >20h    1.33   M
    2    36  3979.37   3.64   >20h   71.35   ADQLTEEQI...TTKELGTVM
   38    15  1666.62   3.96     2m   54.38   RSLGQNPTEAELQDM
   53    20  2196.36   3.04    30m   63.77   INEVDADGNGTIDFPEFLTM
   73     1   149.20   5.55   >20h    1.33   M
   74     4   504.64  11.05   >20h    4.23   ARKM
   76    33  3829.16   4.48     3m   66.59   KDTDSEEEI...SAAELRHVM
  111    15  1722.63   3.57   >20h   54.26   TNLGEKLTDEEVDEM
  126    20  2326.49   3.46    30m   61.18   IREADIDGDGQVNYEEFVQM
  146     1   149.20   5.55   >20h    1.33   M
  147     3   318.35   9.00   >20h    0.00   TAK
```

Further Exploration

Additional analyses

- To view tabular summaries of method results and sequence composition, select **Analysis > Tabular Data** or **Composition**.

- To view the secondary structure of the sequence, select **Analysis > Model Structure** and select from **Helical Wheel**, **Helical Net**, **Beta Net**, **Linear Space Fill** or **Chemical Formula**. A helical wheel structure is shown below:
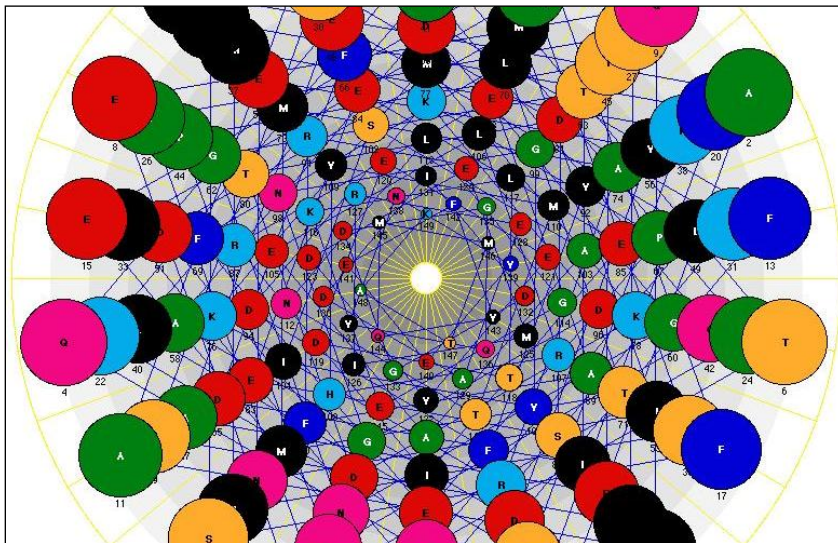
- To view a titration curve like the one below, select **Analysis > Titration Curve**:
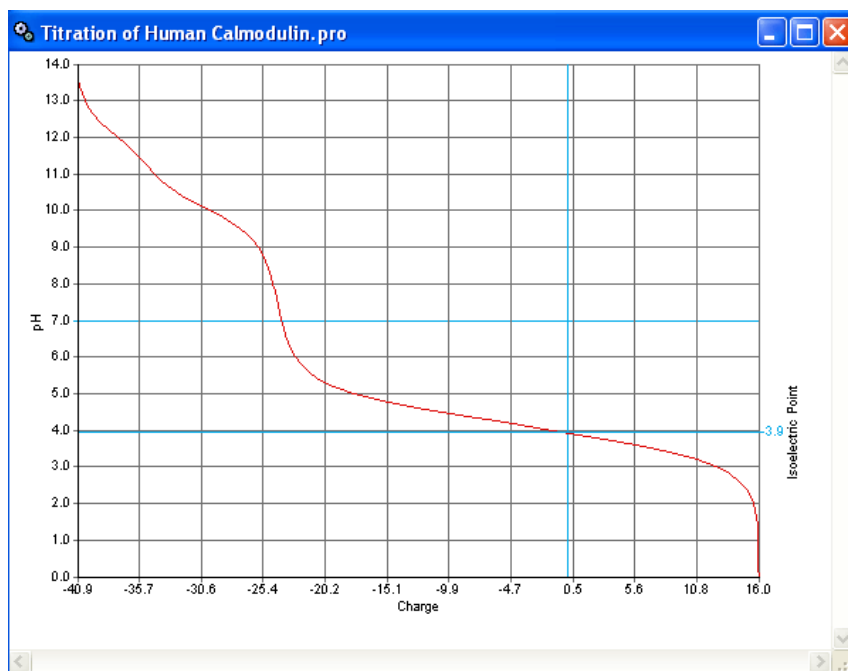
Using method outlines

When you create a new assay document, Protean uses a file called the default method outline ("*default.pao*" on Windows, "Default Outline" on Macintosh) to determine which methods to place in the method curtain and which of those to apply to the assay surface. In the course of revising your assay document, you may have deleted some of the default methods, added new methods of your own, and optimized method parameters and display options. You may want to use the same setup or a similar setup in other assay documents, as well. Protean allows you to save the method outline from the current assay document and apply it to selected assay documents or to all new assay documents.

- To save the current outline as the default method outline, select **Analysis > Save As Default Method Outline**.

- To save the current method outline as a file without saving it as the default, select **Analysis > Save Method Outline**.

- To apply an outline from another document or saved as a separate file, select **Analysis > Apply Method Outline**.

Using the microscope view

To view the sequence for any region under the mouse pointer, activate (the **Microscope** palette tool) and move the mouse pointer to the desired location. The microscope tool lets you stay zoomed out for the "big picture," but still read the sequence.

Exporting to GenVision

- To export a Protean assay document as a GenVision project, select **File > Export as GenVision Project**.

- A dialog will prompt you to choose the content you wish to export: **Applied Method Displays** exports all analysis methods currently applied to your assay surface; **Selected Method Displays** exports only those analysis methods currently selected on your assay surface.

PrimerSelect

Overview

PrimerSelect helps you design primers and probes for PCR, sequencing, and hybridization experiments.

The project begins when you enter a template sequence from any of a wide variety of sequence files or from NCBI's BLAST or Entrez databases. If a sequence file is longer than you need, you may specify a portion of the sequence by coordinates or by selecting a feature from the feature table.

Sequences added to a project are arranged in the Document Window, the principle working area in PrimerSelect. The Document Window is used to collect and inspect template sequences and to evaluate search results. The window has five views, which display templates, primers and products. Once you have entered one or more template sequences, PrimerSelect processes the template for melting temperatures, free energy, and terminal free energy in pentamer windows. These primary calculations are all controlled by initial conditions that you set yourself.

After template processing, PrimerSelect identifies and scores primers based on your specified criteria. Primer "Workbenches" allow you to modify your primer and see how any 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 amplification products.

10 Quick-Start Steps for using PrimerSelect

Here are ten basic steps for using PrimerSelect:

1) Select **File > New** to create a new project or **File > Open** to open a previously saved project. To reopen a document that has recently been opened in PrimerSelect, select it from the **File > Recent Documents** list.

2) Add one or more template sequences to the Document Window using **File > Enter Sequence**.

3) Define your starting conditions and desired primer characteristics by using the options on the **Conditions** menu.

4) If desired, save your conditions and primer characteristic settings for use with other template sequences using **Conditions > Save Conditions**.

5) Let PrimerSelect choose primers for you by using **Locate > Primers & Probes**, or enter your own primers. To use only those primers in the catalog, select **Locate > Only Catalogued Primers**. Sort the resulting primers, if desired, by using **Locate > Sort Primers**.

6) Search for primer pairs by using **Locate > PCR Primer Pairs**. Select **Locate > Adjust Scoring** to change the weight of such factors as internal stability and product length.

7) View results in tabular form by going to the **Report** menu and choosing Amplification Summary, Composition Summary, Primer Self Dimers, Primer Pair Dimers, or Primer Hairpins.

8) Activate a primer pair in the Located Primer Pairs window by double clicking on it.

9) Make any necessary modification to the active primers using the Workbenches accessed from the **Edit** menu. The mutated primers can then be saved in the primer catalog.

10) **File > Save** your results, if desired, then **File > Exit** (Win) or **PrimerSelect > Quit PrimerSelect** (Mac) to exit PrimerSelect.

PrimerSelect Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example PrimerSelect project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Defining primer design criteria

Objective: To specify primer characteristics and their locations using the cloning vector pBR322 as the template sequence.

The data for this tutorial can be found in the following location:

* **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

* **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

* **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch PrimerSelect.

2) Select **File > Enter Sequence** and highlight **pbr322.seq** from the **Demo Data** folder.

3) Click **Add** (Win) or **Add File** (Mac) to move the sequence into the Selected Sequences/Chosen Files list.

4) Click **Done** to open the selected sequence in PrimerSelect's Document Window. The window should appear as follows:
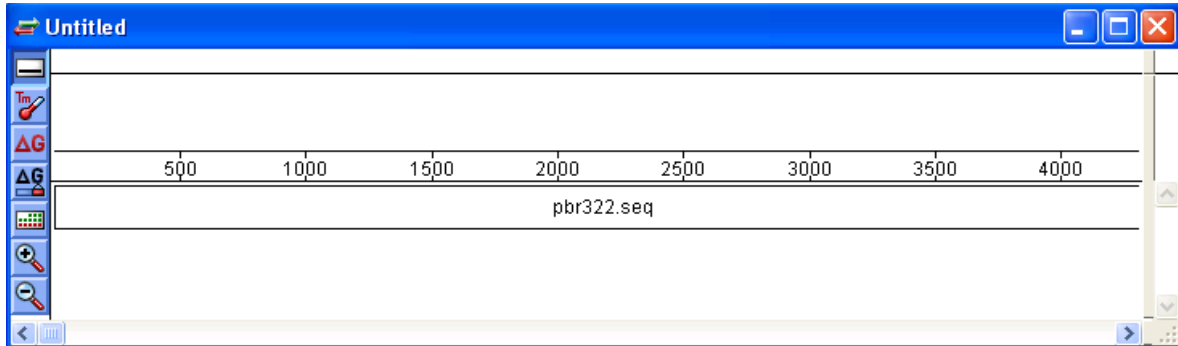
5) Select **Conditions > Primer Characteristics** to open the dialog below.

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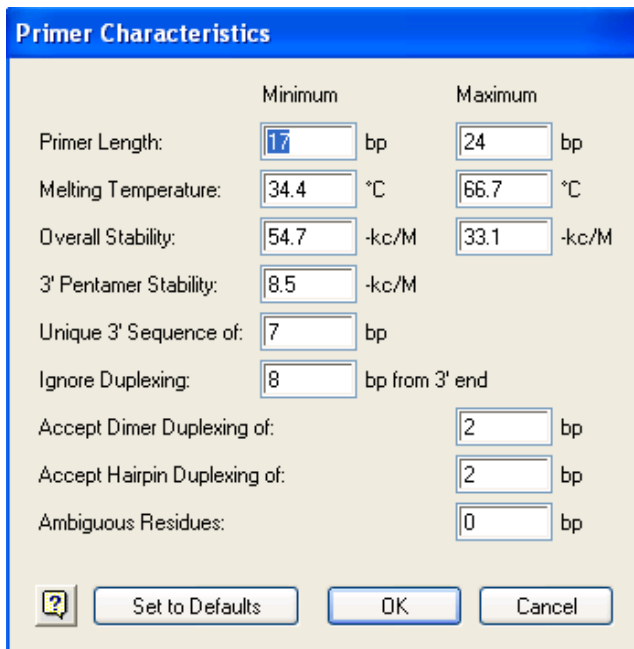
6) Leave the default settings and click **Cancel** to exit the dialog.

---

7) Select **Conditions > Primer Locations** to open a dialog for restricting primer locations based on a given product length, a range in the template, or a combination of both.
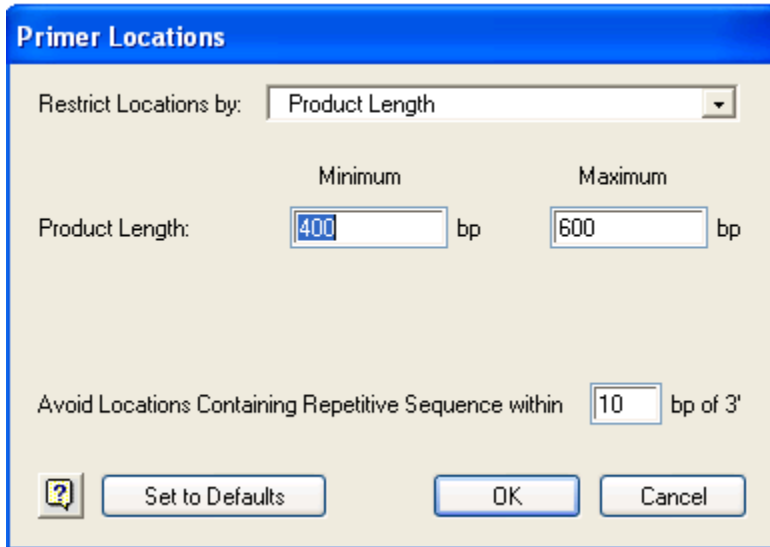
8) From the **Restrict Locations By** list, select **Upper and Lower Primer Ranges**. Text boxes appear to permit you to define these numbers.

This is the best choice when designing PCR primers for a particular region of a sequence.

9) In Upper Primer Locations, enter **1700** and **1950** respectively.

10) In Lower Primer Locations enter **2250** and **2500** respectively.

11) Click **OK** to return to the Document Window, which now appears as follows. Paired green and red triangles mark the allowable ranges for the upper and lower primers, respectively.
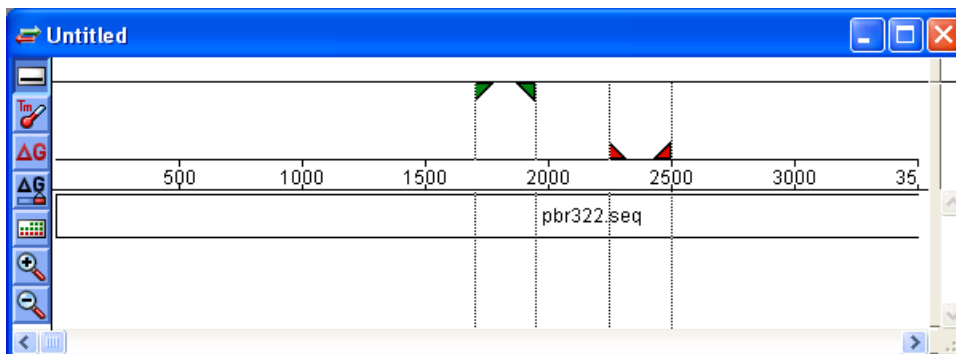
Proceed to the next tutorial.

Searching for primer pairs

Objective: To search automatically for the best primer pairs.

1) Select **Locate > PCR Primer Pairs**. The Located Primer Pairs window appears, displaying a list of primer pair candidates, with the best choices shown at the top of the window.
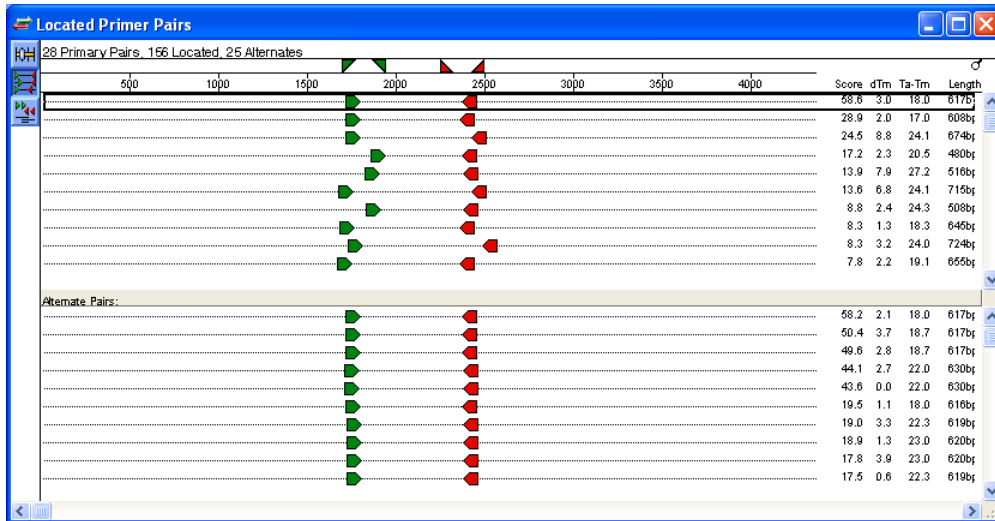
The upper pane of the Located Primer Pairs window shows the principle list of primer pairs.

Because (the **Alternate Pairs** palette tool) is active by default, the lower pane currently shows a list of alternatives for the pair currently selected in the top pane.

2) Open the advice curtain on the right side of the window by dragging ♂ (in the upper right corner) towards the left. PrimerSelect's advice about each of the primer pairs now appears in the right column of the window, as shown below.

PrimerSelect ranked these primer pairs based on criteria established in the Initial Conditions, Primer Characteristics and Primer Locations dialogs located under the **Conditions** menu. Note that the top pair is annotated as the "Best choice."

3) Double-click the primer pair at the top of the window to make it active.

4) Click (the Show Alternate Products palette tool) to change the lower pane into a display showing intended and false primer sites and potential products for the highlighted primer pair.
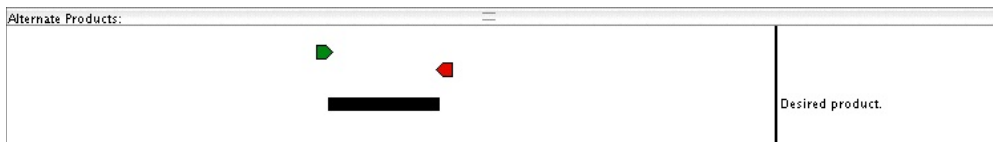
5) Click each successive primer pair in the top pane and observe how the product bar in the lower pane becomes thinner, denoting a decrease in the amount of product expected. Note that some of the primer pairs have false primer sites, indicated by pink or light green arrows, and possible secondary products, denoted by dotted lines:
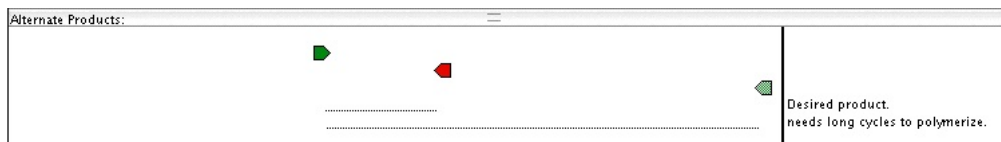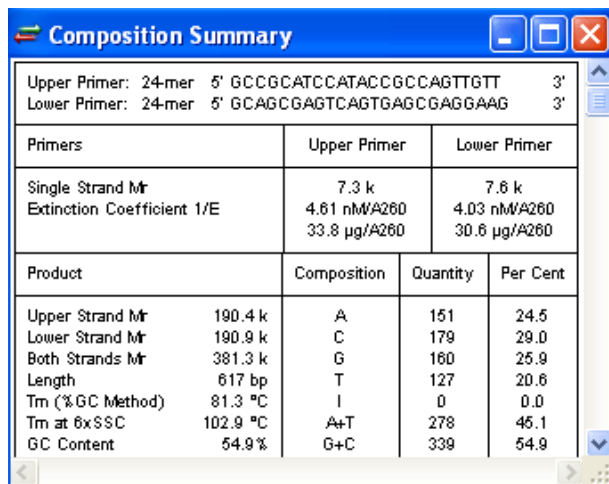
Proceed to the next tutorial.

Viewing composition and hairpin reports

Objective: To view reports showing the composition of the primers and information about possible hairpins.

1) Double-click the primer pair at the top of the window.

2) Select **Report > Composition Summary** to view the following table describing the composition of the primer pairs.

3) Select **Report > Primer Hairpins** to view predicted primer hairpin structures in a window like the one below.
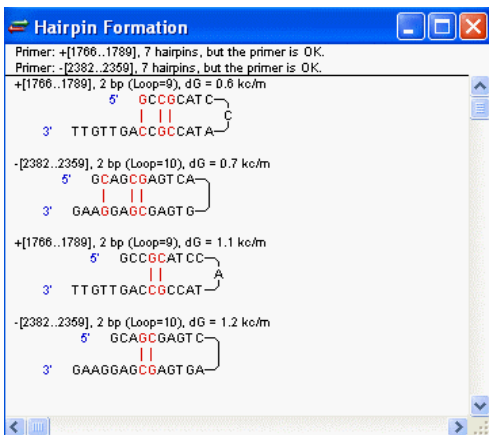
Other commands in the **Report** menu let you view self-dimers and pair dimers.

4) When you are done viewing these windows, close them before continuing with the next tutorial. Leave the Located Primer Pairs and main PrimerSelect window ("Untitled") open.

Proceed to the next tutorial.

Changing primer length

Objective: To extend the length of a primer using the Upper Primer Workbench.

1) With the top primer pair active, select **Edit > Work on Upper Primer** to open the Upper Primer Workbench shown below.

The palette tools (Show sites), (Show reading frames), (Show self-dimers), (Show primer hairpins) and (Show false priming sites) are active by default. The (Site filter tool) may also be used if desired.

About a third of the way down the Workbench, the primer sequence is shown in black. The triangular "handles" at either end allow you to shorten or lengthen the sequence.

2) Click on the primer sequence in black.

- The header in the upper left corner of the window now shows that the primer is 24 nucleotides long and has a melting temperature of 66.3ºC, while a ruler near the bottom of the window shows the single possible annealing site as a short green bar.

- Below the ruler are messages showing that there are no dimers larger than 2 bp, and that the primer is able to form only one, energetically unfavorable 2 bp hairpin.

3) Click the black triangle on the right side of the primer sequence and drag it one base to the left so that the length is shortened to 23 nucleotides. Then, 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 31 nucleotides. Check the length in the header, adjusting as necessary until your header matches the one in the figure below. The header also shows that the melting temperature for the primer has increased to 77.8ºC.

Looking at the lower part of the window, note the new, energetically favorable dimers that PrimerSelect has labeled with the warning "BAD!". Such primers may not amplify the region you want efficiently.

Proceed to the next tutorial.

Introducing a primer mutation

If desired, you can modify a primer simply by clicking on any base and typing. This tutorial, however, presents a more "structured" method for modifying a primer.

Objective: To improve the undesirable extended primer while keeping its sequence relatively unchanged. This will be accomplished by changing a Glycine residue to Serine

1) Click the **Glycine (Gly) residue** in frame 2, written in green just below the primer. A menu appears, showing all four possible Glycine codons and the word Other. The check mark shows that GGT is the triplet encoding this particular Glycine residue.

2) Select **Other** to open a menu with a list of amino acids.

3) Click **S - Serine**.

- The Gly code in the original line of green text does not change, but the duplicate sequence that appears below it in black now shows Ser where the Gly used to be. Ser appears in red to show that it is a mutation from the original primer sequence.

- Looking at the bottom of the screen, note that the two "BAD!" dimers have been replaced by one unfavorable hairpin.

4) Click the **Ser** residue and select **AGT** from the menu that appears.

- This makes a silent mutation to the Ser residue by changing the triplet that codes for Serine from TCT to AGT.

- The original Ser residue had two substitutions compared to the original extended primer (TCT vs. GGT), but now the residues differ by only a single nucleotide.

- The melting temperature has now been reduced to 63.3ºC, which is very close to the melting temperature of the original 24 bp primer (66.3ºC).

5) In the **Name** field at the bottom of the workbench, enter "Mutant Upper Primer".

6) Click **OK** to save the mutant primer.

Proceed to the next tutorial.

Creating a new primer

Objective: To create a new primer using the "mutated primer" sequence as a foundation within the Primer Catalog,

1) Select **Log > Primer Catalog** to open the primer catalog for the current project.
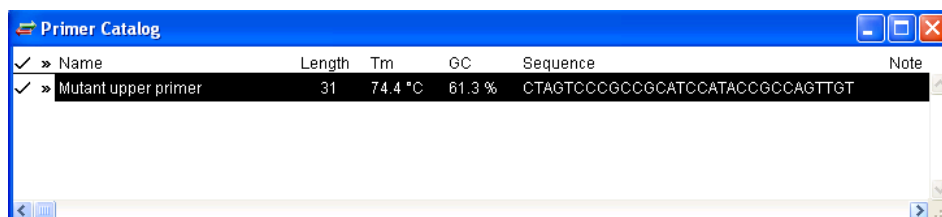
Note that the mutated upper primer created in the last tutorial is already present in the catalog. To the left of the primer is a checkmark identifying it as active, and a chevron (>> , Windows) or bullet (•, Macintosh) showing that it has passed initial secondary structure tests.

2) Double-click in the Sequence field to make it active.

3) Highlight the sequence.

4) Select **Edit > Copy**.

5) Press Return/Enter to create a line for entering a new primer. The new Sequence field will automatically be active. Select **Edit > Paste** to paste in the sequence you just copied.

6) Place the cursor at the end of the sequence and type ACGT.

7) Press the **Tab** key to enter a comment into the Note field.

8) Double-click in the Name field to activate it, and then type in a name.

9) Close the Primer Catalog window to finish. The new primer is now added to the catalog. You can limit future primer searches to only those primers in the catalog (**Locate > Only Catalogued Primers**), which is a useful way to evaluate whether primers you already have may be useful in a particular experiment.

Further Exploration

Sorting primers by characteristic

When you select **Locate** > **Primers & Probes**, PrimerSelect considers all of the desired primer characteristics before sorting primers in the Located Primer Pairs window from best to worst. To re-sort primers according to a particular characteristic, such as melting temperature, select **Locate** > **Sort Primers**. After using this command, the Located Primers & Probes window sorts the list of primers according to the specified criteria.

Selecting a primer pair

To retain information about multiple primer pairs found along the template sequence, you may place your favorite primers in a special list. To add a primer pair to the list, select it from the **Located Primer Pairs** window and select **Locate** > **Choose This Pair**. To open the list, select **Log** > **Selected Primer Pairs**. The current pair, whose amplified products are displayed beneath the template sequence in the Block View, is indicated by a check mark. To remove the check mark without discarding a pair from the list, simply click on it. Click elsewhere to make a new pair active. To discard a pair, select it and press the **Backspace** key.

Using only catalogued primers

To perform a Primers & Probes search using only primers in the Primer Catalog, select **Locate** > **Only Catalogued Primers**.

Note: Close any open PrimerSelect projects and begin afresh before using this command. Otherwise, the results will be combined with the results of any previous Primers & Probes search.

EditSeq

Overview

EditSeq is provided with every Lasergene system to enable you to work on nucleic acid and protein sequences of all sizes from a variety of formats, including GenBank, FASTA, MacVector, GCG®, Text, ABI®, and data from the clipboard. You can even access NCBI's databases by accession number or utilizing the integrated Internet interface to search BLAST and Entrez text databases.

With EditSeq you can manually edit sequence, comments, and annotations, conveniently separated into three adjustable and editable panes. Dynamic links between sequence and annotations exhibit automatic updating of feature coordinates with sequence editing and inclusion of features with copying/pasting of sequences. You can trim a long sequence by specifying a portion of the sequence by coordinates.

EditSeq's search functions let you find positions and open reading frames instantly. You can complement, reverse complement, or view statistics about any sequence or portion of a sequence. You may also translate a DNA sequence or back-translate a protein sequence using customized, expression-specific genetic codes.

A sequence saved using EditSeq in Lasergene DNA or protein sequence format may be opened by any other Lasergene application.

10 Quick-Start Steps for Using EditSeq

Here are ten basic steps for using EditSeq:

1) Create a new EditSeq document (**File > New**) or open/import an existing sequence file (**File > Open** or **File > Import**).

2) Enter or edit data into the Sequence Pane and Comments Pane by typing, or by using options from the **Edit** menu.

3) Format the document as desired by using the **Edit** menu.

4) Use the **Search** menu to locate ORFs, specific sequence strings, or to move the cursor to a numerical position or line.

5) Proofread the sequence using tones or a speech synthesizer by selecting options from the **Speech** menu.

6) Use the **Goodies** menu to open a new sequence window containing the complement or inverse of a highlighted portion of sequence. You may also translate a DNA sequence or reverse translate a protein sequence.

7) Use the **Goodies** menu to view DNA statistics or protein statistics for any highlighted portion of a sequence.

8) Create or edit feature annotations using the **Features** menu.

9) Search online for additional sequences using the **Net Search** menu.

10) Save the sequence document (**File > Save**) or export the sequence as a GenBank Flat File, FastA file, or GCG file (**File > Export**).

EditSeq Tutorials

The tutorials in this chapter should be followed in order, as they are designed to walk you through an example EditSeq project from start to finish. The Further Exploration section at the end of the chapter highlights optional features you may wish to use as you gain experience with the software.

Opening a subset of a sequence

All Lasergene applications allow you to utilize a subset of a sequence if so desired.

Objective: To open a portion of an existing sequence file.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch EditSeq.

2) Select **File > Open**.

3) Locate *tethis21.seq* from the **Demo Data** folder and single-click on the file to highlight it. Single-clicking a sequence file lets you use the Set Ends feature and also causes sequence information to be displayed in the lower half of the dialog. Note that the sequence length is 906 bp.

4) From the Open dialog, click **Set Ends**.

5) Type 50 and 850 into the 5' and 3' end boxes, as shown below.
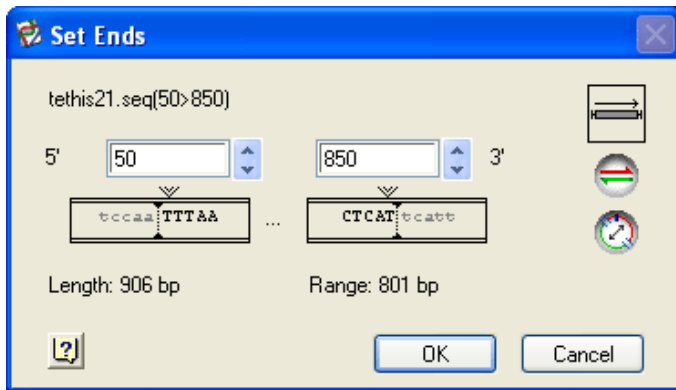
6) Click **OK** to return to the Open dialog. Note that the sequence range is now 801 bp.

7) Click **Open** to open the sequence in EditSeq.
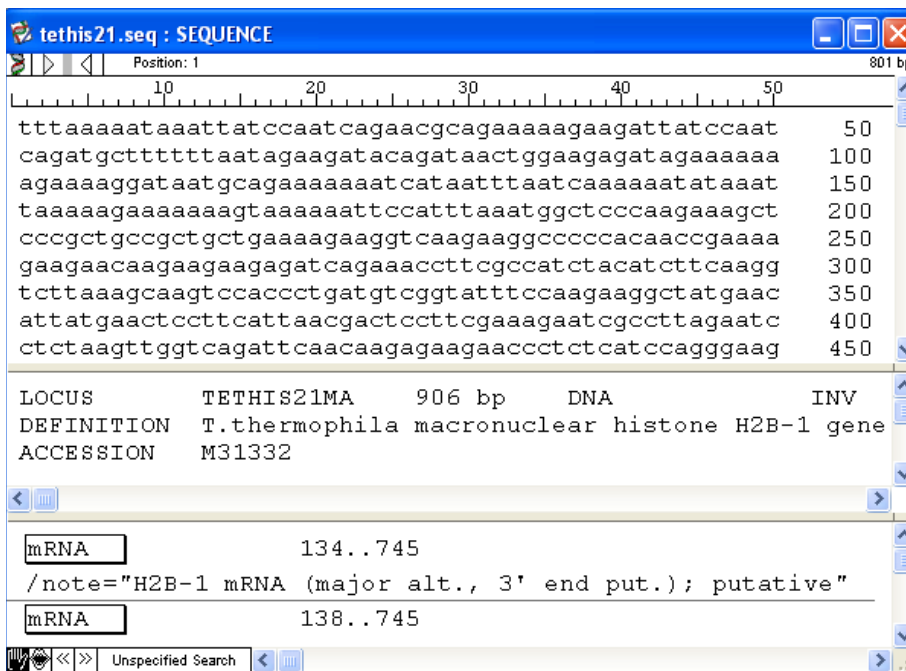
The upper right corner of the header shows that 801 bp of the original sequence has been opened in the upper pane of the window.

The top pane contains the sequence, the middle pane contains the sequence comments, and the bottom pane lists the features (annotations) included in the 50-850 bp range of the original sequence.

Proceed to the next tutorial.

Locating open reading frames

Objectives: To locate potential ORFs.

1) Select **Search > Find ORF**. The Search DNA dialog appears.
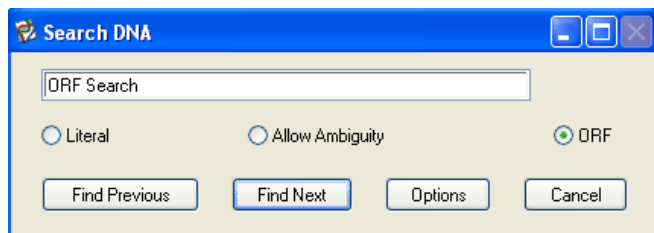
2) Click **Find Next** to close the dialog and locate the first ORF. EditSeq locates and highlights the first ORF in the Sequence Pane. The coordinates for the highlighted ORF appear in the EditSeq header.

3) Click ⧸»⧸ (**Find Next**) at the bottom left of the EditSeq window until you have located and highlighted the ORF at position 345-455.

Proceed to the next tutorial.

Adding a feature with a translation

Features are annotations that document points of interest in the sequence. The current sequence already contains features, which can be viewed in the Features Pane at the bottom of the EditSeq window. EditSeq also lets you add your own features to a sequence.

Objectives: Create a new feature and an associated translation for a segment of the sequence, then update an obsolete translation.

1) With bases 345-455 still highlighted, select **Features > New Feature with Translation**. The new feature and translation appear at the bottom of the Features Pane (the bottom pane in the EditSeq window).

2) Assume that this portion of the sequence is an exon of interest. Click on the box marked **CDS** at the beginning of the feature description to view the feature name options.

3) Select **Gene > Exon**.

4) Click directly on the word **/note** to display a menu of alternative qualifiers and choose **/function** from the list.

5) Insert the cursor between the quotation marks following the "/function" field. Type a note about the feature if you wish. (In the example below, we've typed "histone binding region"). Click anywhere outside the note area to complete the procedure.
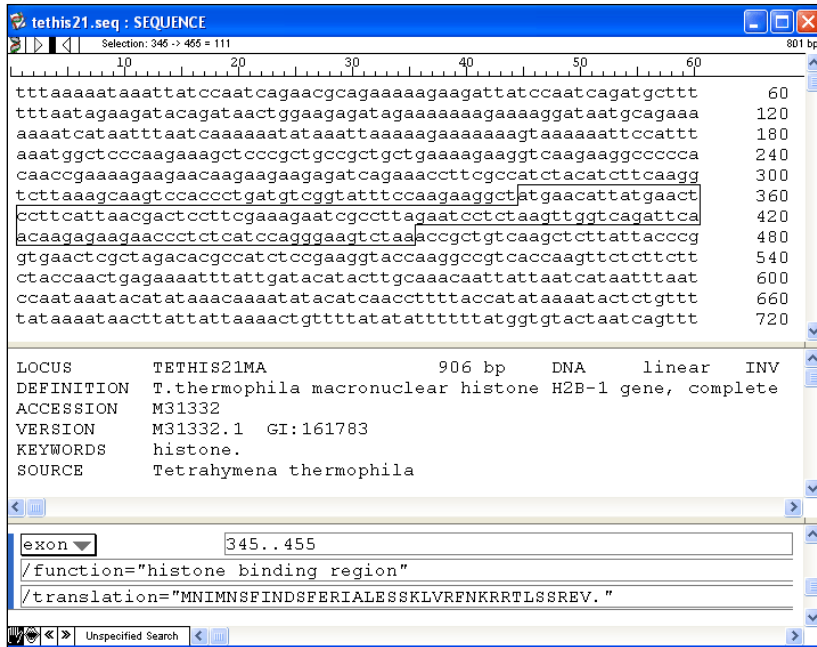
6) Double-click anywhere in the Sequence Pane between bases 345-455 to insert your cursor and then type three random bases. The translation in the new feature turns red to notify you that changes have been made to the sequence. The translation is now obsolete.

7) Select **Features > Update Translation** or **Features > Update all Translations** to update the feature translation. After updating, the translation will again appear black.

Proceed to the next tutorial.

Copying and translating sequences

Objective: To copy the ORF and its associated feature into a new EditSeq document, then to translate the ORF.

1) Highlight the modified ORF, which now runs from 345-458, by double-clicking anywhere on that feature in the Feature Pane.

2) Copy the ORF using **Edit > Copy**.

3) Create a blank EditSeq document by selecting **File > New > New DNA**.

4) Paste the ORF into the blank document using **Edit > Paste**.

   Note that the feature you created earlier has also been pasted into the document. When you copy and paste sequence, features wholly included in the copied portion are also included. Features only partially included in the copied sequence are excluded.

5) Select the ORF using **Edit > Select All**.

6) Select **Goodies > Translate DNA** to translate the sequence in a new protein sequence window.
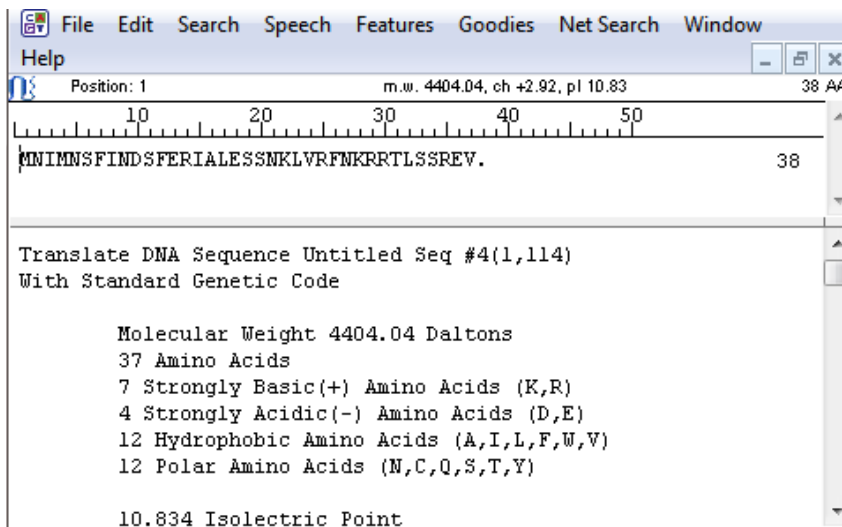
The lower pane of the translation window displays statistics about the translated sequence. Note that in the graphic above, the number of amino acids listed by category do not add up to the total number of amino acids (7+4+12+11≠36). That's because the four categories are not comprehensive and do not include such items as non-polar amino acids.

7) Close all open EditSeq windows by selecting **File > Close**. When prompted to save the document, click **No** (Win) or **Don't Save** (Mac).

Further Exploration

Importing and exporting sequences

- To open a sequence file created with an outside software program, use drag and drop or select **File** > **Import**.

- To save a sequence as a GenBank Flat File, FastA file or GCG file, use **File** > **Export**.

- To save all open windows of a particular type (DNA or protein) into a single file having a GenBank Flat File, FastA file or DNA Multi-seq file format, select **File** > **Export All As One**.

Changing layout options

- To optimize fonts, sizes and cases select **Edit** > **Font**, **Size**, **To Uppercase**, **To Lowercase**, or **Reverse Case**.

- To make a sequence easier to read, group it into blocks using **Edit** > **Layout**.

- To display amino acids using their three-letter code designations, select **Edit** > **Use 3 Letter Codes**.

Proofreading a sequence

- To listen to a sequence as you type, or to playback a sequence that has already been entered, select **Speech** > **Proofread Sequence**.

- To hear the sequence read via a speech synthesizer, select **Speech** > **Windows Voice/Macintosh Voice**. To hear a tone, instead, select **Speech** > **Tones**.

- To optimize playback speed, select **Speech** > **Speed**, **Faster**, or **Slower**.

Complementing, reversing and translating sequence

- To create a new window containing the complement of a highlighted portion of sequence, select **Goodies** > **Reverse Complement**.

- To translate a highlighted portion of DNA sequence or reverse translate a highlighted portion of protein sequence, select **Goodies** > **Translate DNA** or **Goodies** > **Reverse Translate Protein**.

- To translate or reverse translate using a genetic code other than the Standard Genetic Code, select **Goodies** > **Genetic Codes**. The genetic code file can be edited using **Goodies** > **Edit Genetic Code**.

- To reverse the order of a selected sequence, select **Goodies > Reverse Sequence**.

Integration

Overview

SeqBuilder, GeneQuest, PrimerSelect, Protean, Protean 3D, and MegAlign are able to share much of the same data by means of synchronous updating. Synchronous updating allows you to work on the same file in any or all of the six Lasergene applications at the same time. Changes made in one application will automatically be applied to the same data in the other shared applications.

Synchronous updating

Synchronous updating allows you to work on the same file in any or all of six Lasergene applications at the same time. Changes made in one application will automatically be applied to the same data in the other shared applications. Use synchronous updating with SeqBuilder, GeneQuest, PrimerSelect, Protean, Protean 3D, and MegAlign.

Note: Synchronous updating is not supported for *.pdb files.

To use synchronous updating, simply open the exact same file in more than one application. Edits that would affect the presentation of that sequence in the other application will appear when you bring the focus to the other application. The following examples demonstrate some of the functionality that you can access using this integration feature.

Opening and modifying a file in two applications

In this example, you will see how you can open the same file in two applications, in this case, MegAlign and SeqBuilder. However, you may use any two of SeqBuilder, MegAlign, PrimerSelect, GeneQuest, and Protean in a manner similar to the following example.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\Lasergene 11 Data\Demo MegAlign

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo MegAlign

1) Launch MegAlign.

2) Select **File > Open**. The Open dialog appears.

3) Select *Calmodulin Alignment.meg* from the **Demo MegAlign** folder.

If, while working on this file, you decided that you needed to edit one of the sequences, you could do so using SeqBuilder, without closing the file. The remaining steps in this example will edit 2 residues in the Black Mold Calmodulin sequence.

4) Highlight the **Black Mold Calmodulin** sequence in the list of Sequence Names.

5) Select **File > Send Sequence To > SeqBuilder**. SeqBuilder opens and displays the Black Mold Calmodulin sequence.
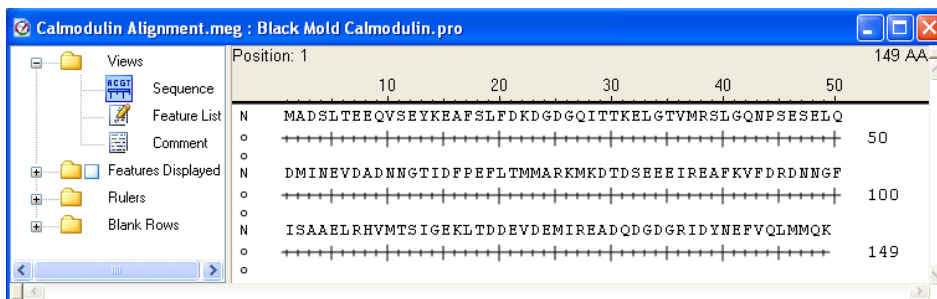
6) Modify the sequence by changing the 4th residue from an **S** to a **Q** and the 10th residue from a **V** to an **I**. Here is how this appears in SeqBuilder:
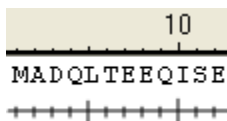
7) Move the focus to MegAlign. Note that the change has been made in the file in MegAlign, as shown below:

Note: Only the sequence has been updated. This may affect the alignment. The alignment is *not* automatically updated. Therefore, after editing, running the alignment again is recommended.

- If you save the changes made in either application using **File** > **Save**, you will save the changes in the original MegAlign document.

- If you save the changes made in either application using **File** > **Save As** and you rename the file or change the file format, the link between the two applications will be broken and synchronous updating will no longer occur.

Adding a feature to a file in two applications

In this example, you will see how adding a feature to a sequence in one application will result in adding the same feature to the sequence in another application. This example uses GeneQuest and PrimerSelect. However, you may use any two of SeqBuilder, MegAlign, PrimerSelect, GeneQuest, Protean, and Protean 3D for this example.

The data for this tutorial can be found in the following location:

- **Windows 7 & Windows 8:** C:\Users\Public\Public Documents\DNASTAR\Lasergene 11 Data\Demo Data

- **Windows XP:** C:\Documents and Settings\All Users\Shared Documents\DNASTAR\ Lasergene 11 Data\Demo Data

- **Macintosh OSX:** Hard Drive:Applications:DNASTAR:Lasergene 11 Data:Demo Data

1) Launch GeneQuest.

2) Select **File > Open** to open the project *Nematode R01H10 Assay.dad* From the **Demo Data** folder.

Note: You may need to click to view the image as it appears below.

---
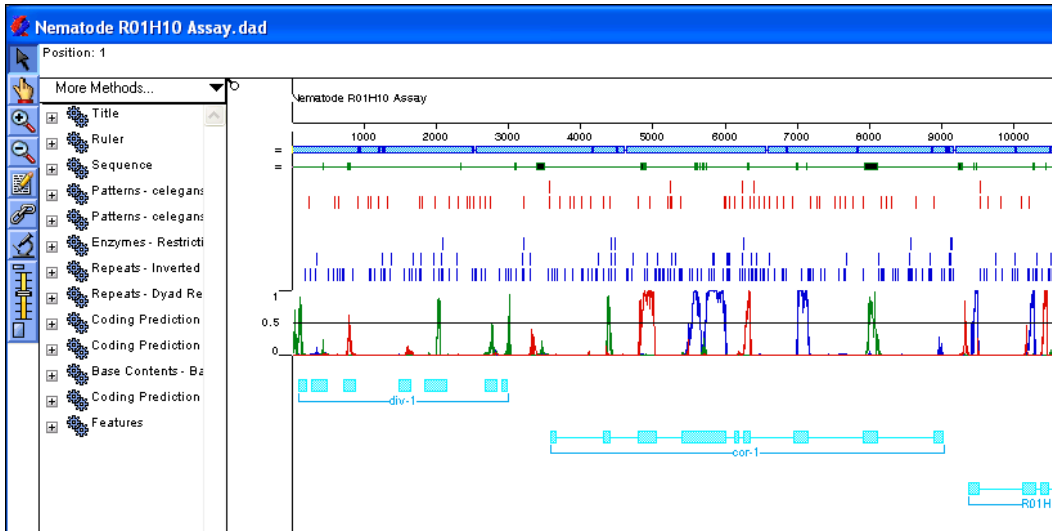
3) Select **File > Send Sequence to > PrimerSelect** to open the file in PrimerSelect. Click on to display the Tm plot view.
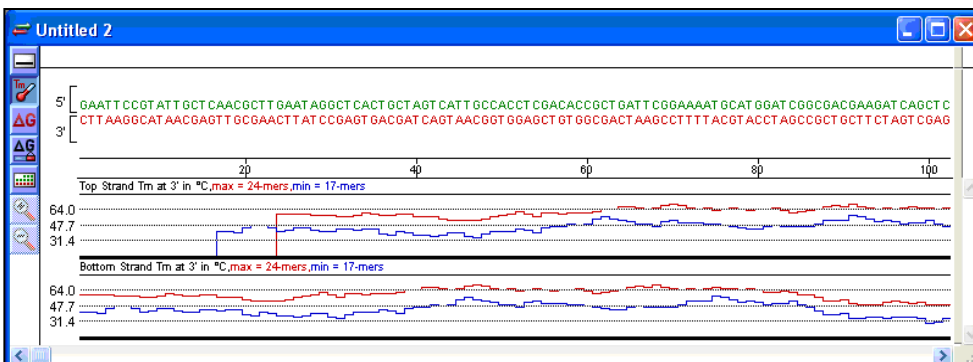
4) Return to GeneQuest and create a new feature:

   a) Click and then select a range of bases.

   b) Select **Sites & Features > New Feature**. The Feature Editor appears.

   c) In the Title field, name the new feature **Test feature**.

   d) Click **OK**.

5) Return the focus to PrimerSelect.

6) Select **Conditions > Sequence Positions and Limits**. The Sequence Positions and Limits dialog appears.

7) Click **From Feature Table**. Note that the new feature appears at the bottom of the list.
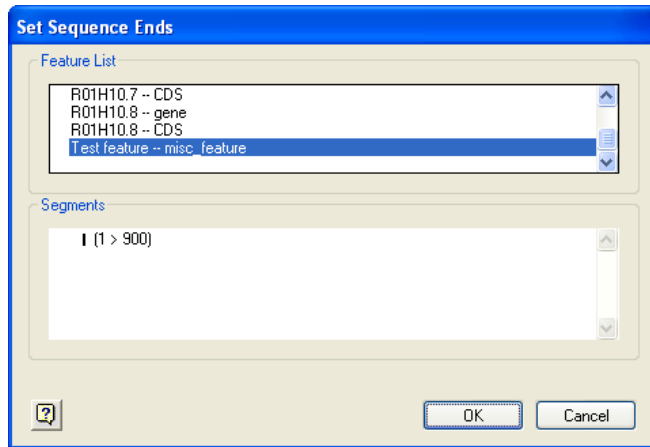
Note: If you save this sequence file from either application as anything but the original *.dad file, the synchronous updating between GeneQuest and PrimerSelect will be lost.

Online Database Searches

Overview

SeqBuilder, GeneQuest, PrimerSelect, Protean, MegAlign, EditSeq and SeqMan allow you to search online databases, such as the National Center for Biotechnology Information's (NCBI's) BLAST and Entrez servers, for sequences or proteins that match all or part of the current file. Perform all online searches from the **Net Search** menu of these Lasergene applications. Some options vary from application to application, as noted in each section below. Additionally, applications may be able to use different types of results, or use similar results in different manners.

Setting Up Preferences

Before performing an online database search, you must setup Lasergene's online search preferences. This may be done in any Lasergene application. Once preferences are setup in one Lasergene application, they are shared by all other Lasergene applications.

Note: If multiple Lasergene applications are open while you are setting up the preferences in one, the other applications must be closed and re-launched before the changes will be shared between all of the applications.

To access online preferences, choose an appropriate option from the list below.

Windows

- For SeqMan, select **Project > Parameters > Servers**.

- For all other Lasergene application, select **Edit > Preferences**.

Macintosh

- For SeqMan, select **SeqMan > Preferences > Servers**.

- For all other Lasergene applications, select **[Application Name] > Preferences**. For example, in SeqBuilder, select **SeqBuilder > Preferences**.

Specifying online server addresses

To enter or change your BLAST and Entrez (text) server addresses:

1) Open the Preferences dialog using any suitable option from the preceding section, *Setting Up Preferences*.

2) Select the **Servers** tab. The upper portion of the ensuing dialog allows you to store BLAST Server addresses, while the lower portion lets you store Entrez Server addresses.

- To add a new address to either box, click the associated **Add** button and type in the address.

- To delete an address, click the **address** and click **Delete**.

3) To set the default server that Lasergene will use when performing a BLAST or text search, click the name of the server in either box and then click **Choose**.

Note: To change the default server later, repeat this process but choose a different server in step 2.

4) Click **OK** to accept the changes or click **Restore** to reset the default parameters. Click **Cancel** to leave settings as they were prior to making changes.

Server URLs

- For NCBI's BLAST 2.0, use http://www.dnastar.com/blast/ncbi-blast.html

- For NCBI's Entrez server, use http://www.ncbi.nlm.nih.gov/entrez/eutils

Note: For other servers, please consult your system administrator.

Specifying a download folder

To select a default folder for storing downloaded sequences:

1) Open the Preferences dialog using any suitable option from the section, *Setting Up Preferences*.

2) Select the **Internet** tab.

3) Click **Set Folder**. A dialog will appear where you may specify the desired folder. When you are satisfied with your choice, click **OK** (Win) or **Choose** (Mac).

Performing a BLAST Search

You can search NCBI's BLAST databases for a sequence that contains a section that matches the current selected portion of the sequence. You must select a portion of the current sequence before beginning a BLAST search.

1) Select a portion of the current DNA or protein sequence by highlighting it.

2) Select **Net Search** > **BLAST Selection**. (In GeneQuest, **Net Search > BLAST ORF** is also available). The Blast Query dialog appears. This dialog contains the following options:

- **Server**: BLAST Servers that have been setup using **Preferences** appear here. You may change to a different server by choosing one from the dropdown list, or select **New Server** from the list and then type a new URL into the **Server** field.

- **Programs**: The following programs are available to select, although some servers may not offer all of these programs:

  ✓ **blastn** (searches a nucleotide sequence database with a nucleotide query sequence)

  ✓ **blastp** (searches a protein sequence database with an amino acid query sequence)

  ✓ **blastx** (searches a protein sequence database with the six-frame translation of a nucleotide query sequence)

  ✓ **tblastn** (searches the six-frame translation of a nucleotide sequence database with a protein query sequence).

  ✓ **tblastx** (searches the six-frame translation of a nucleotide sequence database with the six-frame translation of a nucleotide query sequence).

  Note: When using **BLAST ORF** in GeneQuest, the program is automatically selected for you as **blastn**.

- **Databases**: Select from the databases available on the server by selecting one from the dropdown list.

3) Click **OK**.

After initiating a BLAST search, the BLAST search results window appears. When the server returns results, this window lists any sequences that match the current sequence. You may continue to work with other Lasergene application activities while waiting for BLAST search results, but you cannot launch another BLAST search until you close the current BLAST window.

BLAST search results window

Upon completing a BLAST search, the BLAST results window appears.
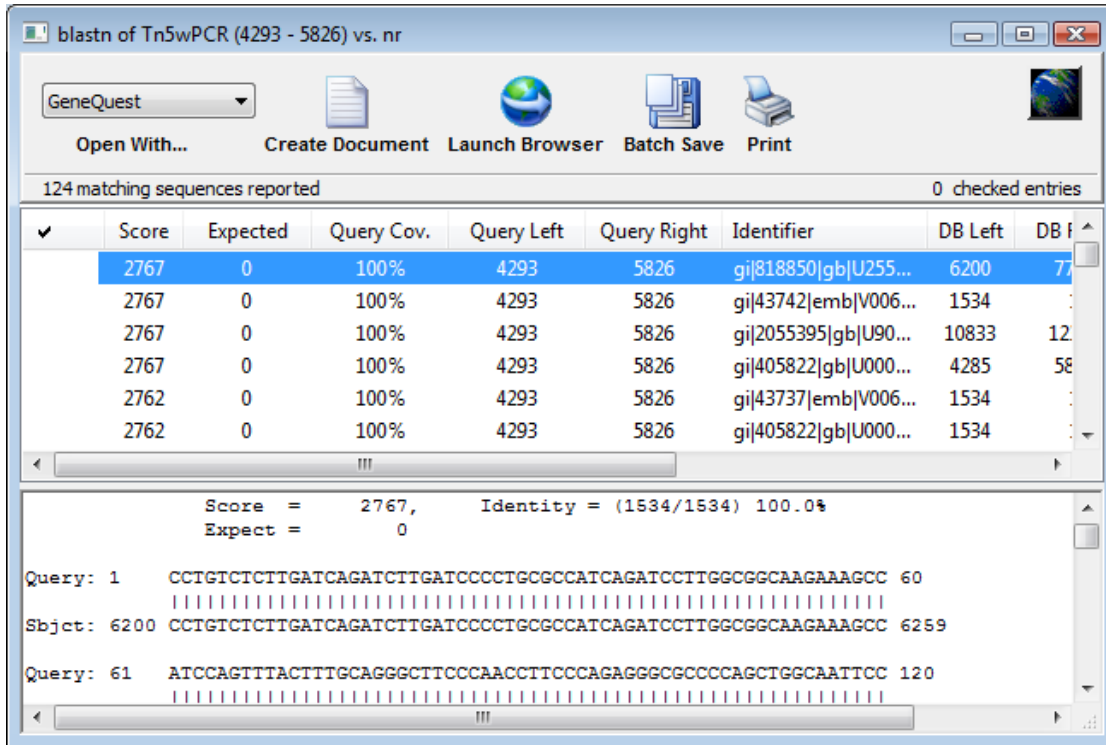
The upper pane of the BLAST search results window contains the name of possible matches in order of probability, while the lower pane contains the alignment of the query (upper sequence) to the highlighted database entry (lower sequence).

The following columns are available in upper pane of the BLAST search results window:

- **Score** - The similarity score for the result.

- **Expected -** The expectation value. This corresponds to the **Expect** value displayed in the lower pane of the window.

    Note: Detailed information about how "score" and "expectation" are derived is available at http://www.ncbi.nlm.nih.gov/BLAST. In general, a higher score and lower expectation connote a better match.

- **Query Cov. -** The percentage of the queried sequence that aligns to a match in the database. For example, if you BLAST a range of 596 bases, and 565 of those bases align to a sequence in the database, then the **Query Cov.** value will be 565/596, or 95%. Note that in SeqMan, if a BLAST search is performed from the Project Summary window, then the **Query Cov.** column will also list the contig that contains the matching result.

> **Note: Query Cov.** differs from the **Identity** value shown in the lower portion of the BLAST window. Identity is the percentage of bases within the alignment that match. For example, if 565 bases align to a sequence in the database, and there are 3 mismatches within the alignment, then the **Identity** value would be 562/565, or 99.5%.

- **Query Left -** The left end coordinate of the queried sequence that corresponds to the left end coordinate of the matching result (DB Left).

- **Query Right -** The right end coordinate of the queried sequence that corresponds to the right end coordinate of the matching result (DB Right).

- **Identifier -** The sequence identifier for the result.

- **DB Left -** The left end coordinate of the result sequence that corresponds to the left end coordinate of queried sequence (Query Left).

- **DB Right** - The right end coordinate of the result sequence that corresponds to the right end coordinate of queried sequence (Query Right).

- **Description -** The description, from NCBI, of the result sequence.

To sort results in the BLAST search results window, click on a column header to sort the rows by that column.

For additional information on using the BLAST results window, see *Working with search results windows*.

Performing a Text Search

Lasergene provides a graphical interface for constructing Boolean queries of text databases like NCBI's Entrez database. Search results are returned as a list from which you may select any or all sequences, and then save them or open them as documents in the current application.

1) In any Lasergene application, select **Net Search** > **New Text Search**. The Entrez Query dialog appears.
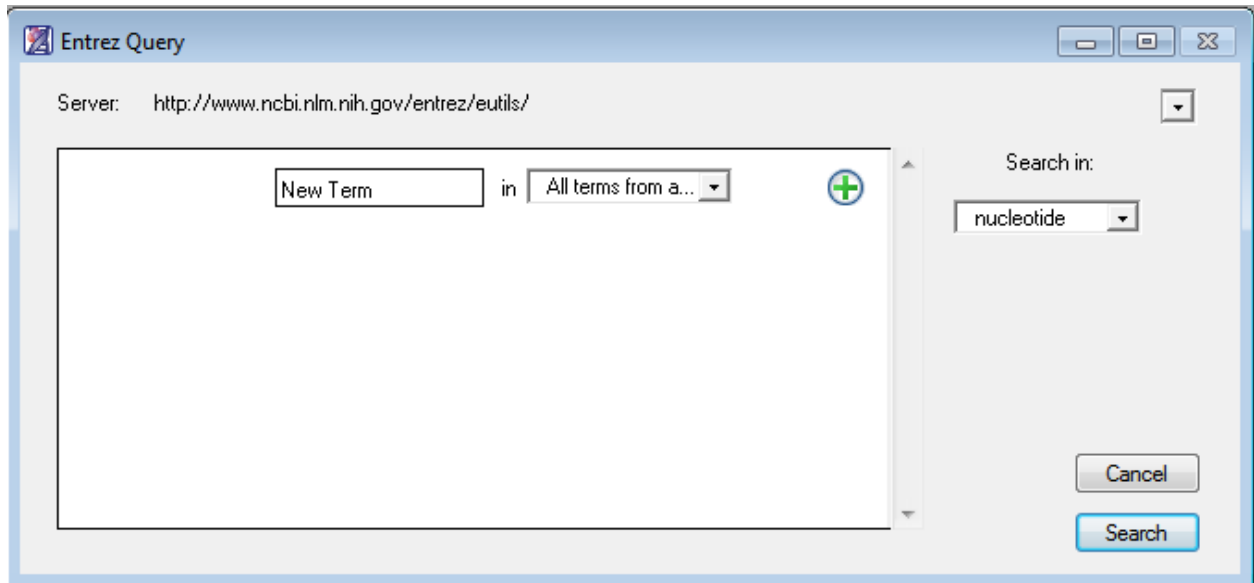
- The **Server** field lists the currently selected Entrez server. You may change to a different server by clicking the arrow button at the top right of the dialog. Choose a server from the list, or select **A New Server** and then type a new URL.

- Use the dropdown menu in the lower right of the window to choose the database to search. This menu will reflect the available databases on the selected server.

2) Create the search criteria using the fields provided:
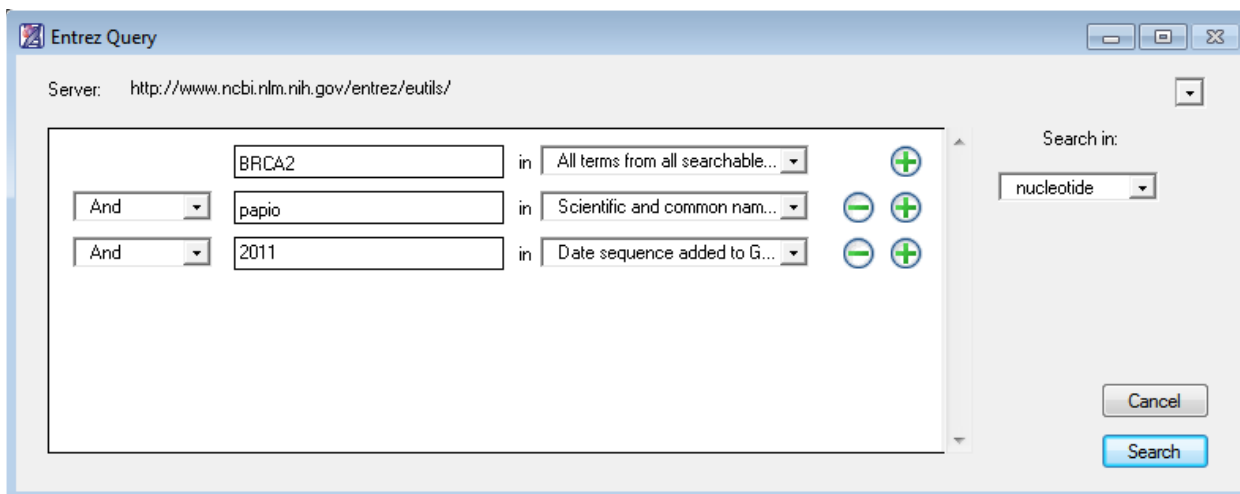
- Enter the first (or only) criterion into the text box on the left.

- You may change the search field heading from the dropdown list to the right of that text box. The default is **All terms from all searchable fields**, which will search every word of every listing for the query term in the database. The field headings vary depending upon the online database that you are using.

- To add a criterion to the query, click the ⊕ button.

    ✓ Choose **And**, **Or**, or **And Not** from the list box on the left of the new row.

    ✓ **And** requires both the first and the second criterion to be present.

    ✓ **Or** will look for matches containing either criterion.

    ✓ **And Not** will search for matches containing the first criterion but not the second.

- To remove a criterion, click the ⊖ button in the row that you want to remove.

3) Once you have defined all of your search criteria, click **Search** to send the query to the database. When the search is complete, the text search results window appears, showing a list of sequences that match your query.

Text results search window

Upon completing a text search, a text results window appears.
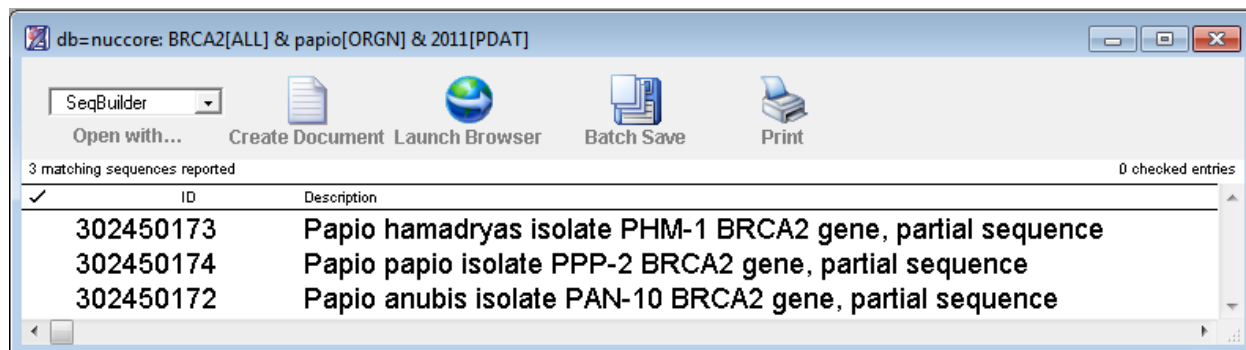
This window lists all of the results that match your search criteria. If desired, you can search the results of a text search to reduce the number of matches.

For additional information on using the text search query results window, see *Working with Search Results Windows*.

Changing a text query

You may modify a text query to obtain different results without having to completely recreate a complex query. To do this, select **Net Search > Change This Query**. The Entrez Query dialog appears containing the same search criteria that you last used, allowing you to edit as necessary before performing a new search.

Searching text search query results

You may search the results of a text query rather than re-searching the entire text database. This can help you identify the most useful results of a previously-performed query.

To do this, first make sure the text search results window is active by clicking on it. (If you have several text results windows open, select **Net Search > Current Results** to locate the desired window).

Then, select **Net Search > Search These Results** to open a text query window. Enter the desired search criteria, and then click **Search** to search the current results.

When the search is complete, the text search results window will be displayed, showing a list of sequences that match your query.

Working with Search Results Windows

The BLAST and text search results windows permit many similar functions and have common buttons along the top of their windows:

- **Open With:** Use this to open one or more results in the selected Lasergene application. See *Opening Results in a Lasergene Application* for additional information on using this option.

- **Create Document/Add to Project/Put in Document**: The name of this button varies between Lasergene applications. Use this option to open one or more results in the current application. See *Opening Results in a Lasergene Application* for additional information on using this option.

  Note: In SeqMan, the **Add to Project** button will appear as **Collect Features** if the sequence you selected to BLAST is not already associated with trace or flowgram data. For more information, open the SeqMan help and refer to the topic *Collecting Features from a BLAST Search*

- **Launch browser**: Use this option to open one or more results in your default web browser. See Viewing results in a web browser for additional information on using this option.

- **Batch Save**: Use this option to save a group of results as sequence files. See *Saving multiple results* for additional information on using this option.

- **Print**: Click this option to print one or more results. See *Printing BLAST Search Results* or *Printing Text Search Results* for additional information on using this option.

Selecting one or more results

You can select one or more results from a results window by using any of the following methods. Once selected, you may then open, view, save, or print that selection of results.

- Click a result to select only that result.

- Check each result by clicking to the left of the result in the results window. Use this method to select a group of consecutive or non-consecutive results. This method has an advantage over using **Ctrl+click** as you will not lose selected results if you forget to hold down the button when choosing a result.

- Press and hold **Ctrl/Cmd** while clicking each desired result on the results window. Use this method to select a group of consecutive or non-consecutive results.

- Click a result, and then press and hold **Shift** while clicking another result to select those two results and all of the intervening results.

Opening results in a Lasergene application

You can view results by opening one or more in a Lasergene application. To do this, first select one or more results from your search results window. Then, do one of the following:

- To open the results in a different Lasergene application than the current one, select a Lasergene application from the list box above the words **Open With**.

- To open the results in the current Lasergene application, click **Create Document/Add to Project/Put in Document**. In the resulting dialog, select one of the following options from the **Save** field:

- **Top** opens the specified number of matches, starting with the first sequence on the list. Specify a number in the text box (default is 10).

- **Next** opens the specified number of sequences, starting with the first one currently highlighted. Specify a number in the text box (default is 10).

- **Checked** opens sequences with a checkmark beside them.

- **Selected** opens only highlighted sequences.

- **All** opens all sequences.

To save the file(s) in a location other than the default directory, click the **Set Location** button. Finally, click **OK** again to save the sequence files and either add them to the current project or open them in a different Lasergene application.

Viewing results in a web browser

You can view detailed information about your search results using a web browser. To do so, first select one or more results from the search results window. Then, click the **Launch Browser** button. Alternately, you may select **Net Search > Open with Web Browser**.

Each selected match will be displayed in a separate browser window.

Note: Popup blocking software may prevent Lasergene from opening multiple web browser windows simultaneously. Check the documentation for your popup blocking software for additional information.

Saving multiple search results

You may save multiple search results as sequence files from the search results window. To do this, first select one or more results from the search results window. Then, click the **Batch Save** button. In the resulting dialog, select one of the following options from the **Save** field:

- **Top** opens the specified number of matches, starting with the first sequence on the list. Specify a number in the text box (default is 10).

- **Next** opens the specified number of sequences, starting with the first one currently highlighted. Specify a number in the text box (default is 10).

- **Checked** opens sequences with a checkmark beside them.

- **Selected** opens only highlighted sequences.

- **All** opens all sequences.

To save the file(s) in a location other than the default directory, click the **Set Location** button. Finally, click **OK** again to save the sequence files to the specified directory.

Note: When using Batch Save, the selections in the results window that have been downloaded / opened are denoted with a black diamond or arrow

Printing BLAST search results

To print BLAST search results, first select one or more results from the search results window. Then, click the **Print** button. In the resulting dialog, select one of the following options from the **Print** field:

- **Top** prints the specified number of matches, starting with the first sequence on the list. Specify a number in the text box (default is 10).

- **Next** prints the specified number of sequences, starting with the first one currently highlighted. Specify a number in the text box (default is 10).

- **Checked** prints sequences with a checkmark beside them.

- **Selected** prints only highlighted sequences.

- **All** prints all sequences.

Click **Print** to proceed with printing using the standard Print dialog for your computer's operating system.

Printing text search results

To print one or more results from a text search results window, first select one or more results. Then, click the **Print** button. The standard Print dialog for your computer's operating system will be displayed, allowing you to proceed with printing your selected results.

Copying results

Selected results may be copied to the clipboard from the BLAST or text search results window by using **Edit > Copy**.

When you copy results from the upper pane of the BLAST or text search results window, the information includes identifier information and a description for each selected result.

You may also copy text from the bottom pane of the window from a BLAST search results window. In this case, the text you select is exactly what will be copied.

Note: When pasting data into a word processing application, choose a mono-spaced font such as Courier so that the alignments display correctly.

Saving search results as text files

To save your BLAST or text search results as a text file:

1) Make the results window the active window.
2) Select **File** > **Save BLAST Report** (SeqMan only) or **File > Save As** (other Lasergene applications).
3) Find a location to save the document using the **Save In** menu.
4) Enter a new name in the **File Name** window (the default name is "Untitled").
5) Click **Save**. Lasergene will automatically add the correct text file extension (*.txt) when saving the file.