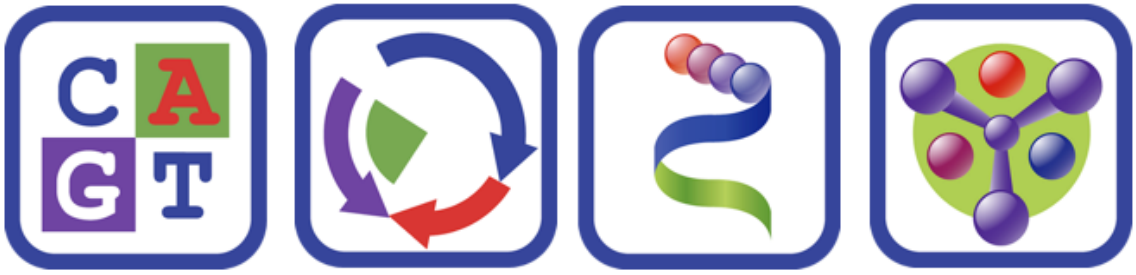
Book 2

Bioinformatics with DNASTAR

Lasergene



DNASTAR's Lasergene

Note: current lasergene version: 14

Selected webinars[1]

Introducing Lasergene 14

Published on Oct 5, 2016

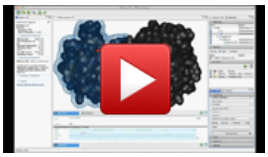
44:10

Techniques in biological research are constantly changing. These technological innovations fundamentally affect the depth and breadth of research scientists are able to accomplish. In this webinar, Tom Schwei will present an overview of Lasergene 14, including the introduction of three completely new applications (a genome browser, protein docking, and antibody structure prediction) as well as visualizing the power of combined data analysis from independent but linked genome assembly analysis approaches (e.g. RNA-Seq and ChIP-Seq) and other improvements in core molecular biology applications.

https://youtu.be/cm-qhvM3RL4

Older webinar: Introducing Lasergene 11: http://youtu.be/U_IexXTgdLE

Protein Sequence and Structure Analysis in Protean 3D



52:07

Presented by: Amanda Mitchell Originally Aired: November 14, 2012

Hear Amanda discuss the proteomics tools available in Lasergene 10.1 as she demonstrates some of the most powerful features in Protean 3D, including: epitope prediction, structural alignment, and macromolecular motion visualization.

http://youtu.be/t6vdZwktRXk

[1] http://www.dnastar.com/t-support-webinars.aspx
or
YouTube channel: http://www.youtube.com/user/DNASTARInc

Connect to Biochem. Network with VPN [1]

All Biochemistry users can access the DNASTAR LASERGENE software by VPN connection when located outside of the Biochemistry buildings.

The Biochemistry Department Intranet lists all requirements and methods to access the network remotely:

`https://biochem.wisc.edu/intranet/it/remote-access`

Below are the steps necessary to connect from an iMac.

1. Requiements

1) The CISCO ANYCONNECT software[2] installed on the DMC computers.
2) Biochemistry **username** and **password**: obtain from Biochem IT or Class instructor

2. Connect

✔ **TASK** Connect to the Biochemistry network with the Cisco VPN software.

Locate and **double-click** on the **hard drive**
Then **click on Applications** and then **Cisco**

Launch (double click) on the "**Cisco AnyConnect Secure Mobility Client.app**" which looks like this:

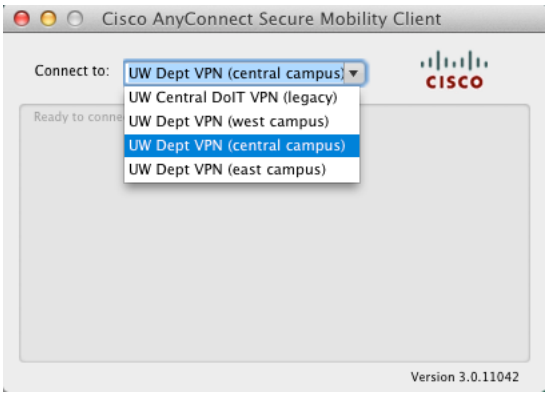
Cisco AnyConnect Secure Mobility Client.app

[1] VPN = Virtual Private Network – I allows you to "pretend" you are locally connected.
See `https://kb.wisc.edu/page.php?id=27448`
[2] Mac: `https://kb.wisc.edu/page.php?id=27573`
Win: `https://kb.wisc.edu/page.php?id=27559`

If the campus regions appear select the options that reads:

UW Dept VPN (central campus)

If something else appears, then continue to the next page.

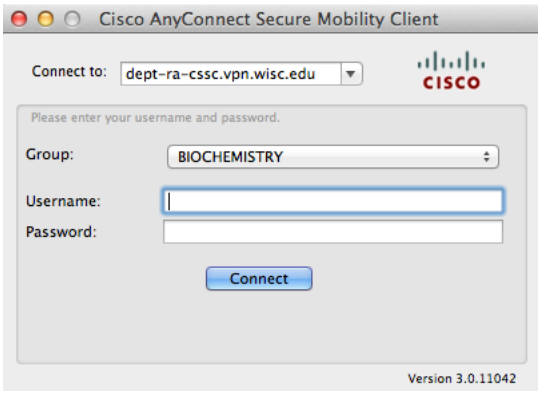


If the campus regions do not appear by default, you will need to manually enter the server name for Central Campus:

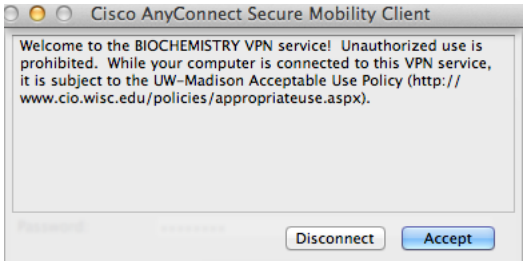
`dept-ra-cssc.vpn.wisc.edu`

This choice will make **BIOCHEMISTRY** appear within a pull-down menu.
Enter your **Biochemistry username** and **password**

Click Connect



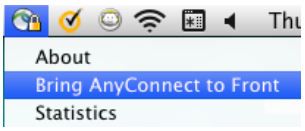
Click **Accept**



A top menu icon will appear on the Mac menu showing a locked (secure) connection.

3. Bring to Front

The top menu can be used to bring the running VPN window to the front.

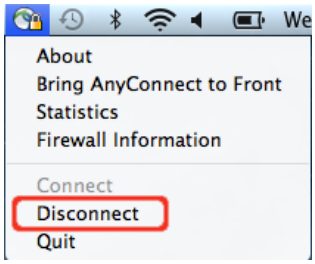


This will bring a "status" window showing the connection with information such as the IP number and how long the connection has been established.

This window can be used to disconnect from the network but it is even easier with the top menu option as shown below.

4. Disconnect

To disconnect from Departmental VPN, click on the AnyConnect VPN client icon in the top menu bar and select **Disconnect**.



IMPORTANT: YOU ARE LOGGED IN WITH YOUR NetID REMEMBER TO DISCONNECT AT THE END OF CLASS!

5. FYI: Windows VPN

Instructions for Windows are located here:
`https://kb.wisc.edu/page.php?id=27559`

6. FYI: Other Campus locations

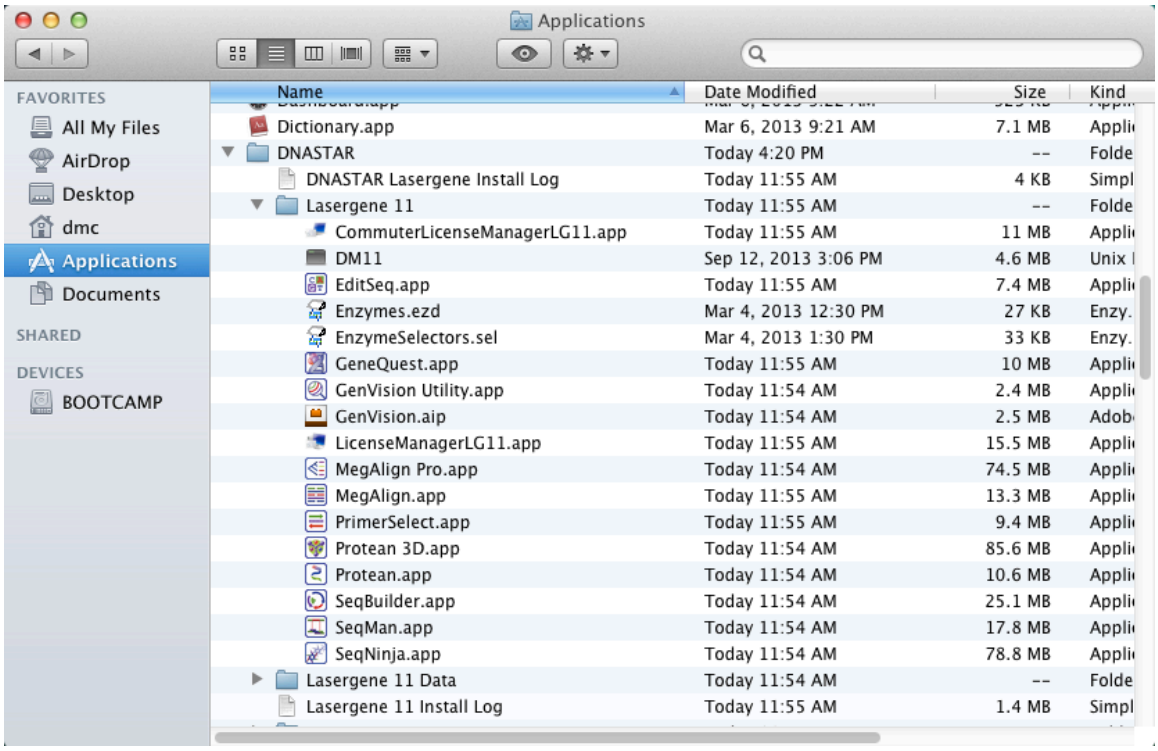
Generic instructions for the Campus are located here:
`https://kb.wisc.edu/page.php?id=27589`

The server names for all the Departmental VPN campus regions are as follows:

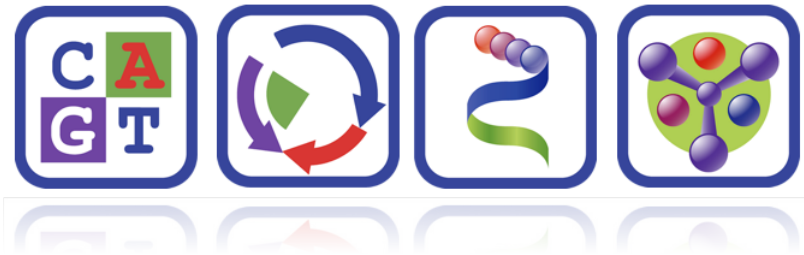
East Campus: `dept-ra-432nm.vpn.wisc.edu`
West Campus: `dept-ra-animal.vpn.wisc.edu`
Central Campus: `dept-ra-cssc.vpn.wisc.edu`

Location of the Lasergene software on iMac computers

All Lasergene software is located in the directory `/Applications/DNASTAR`



Book 2: Bioinformatics with DNASTAR Lasergene



Madison-based D N A S T A R (www.dnastar.com) is an international company developing software since 1984. The company now defines itself as "… *a pioneer in the development and sale of software used to increase life scientists' productivity using their desktop computer or on the Amazon Cloud. DNASTAR's comprehensive software suite,* Lasergene*, supports molecular biologists, geneticists, and structural biologists in meeting virtually all of their DNA, RNA, and protein sequence needs, including Sanger and next- generation sequence assembly and analysis, protein sequence and structure analysis, and protein structure prediction with easy to use, affordable, flexible computer software.*"

The following computer hands-on lab modules are self-paced exercises demonstrating some *Lasergene* applications on sequence editing (**EditSeq**), plasmid mapping (**SeqBuilder**) and protein structure (**Protean** and **Protean 3D**.)

We are grateful to D N A S T A R for providing academic teaching licenses when necessary, and to the CALS computer lab personnel for their help in installing the software onto the classroom computers.

UW Biochemistry students can access D N A S T A R software on the teaching computers (licensed via the Biochemistry Department) in room 301 on the 3rd floor of the Biochemistry Laboratories building. **Biochemistry Students** can install the *Lasergene* package following instructions on the Biochemistry Intranet (*NetID login required*):
`https://biochem.wisc.edu/intranet/it/lasergene`

Other users interested in installing the trial version of *Lasergene* can request a free trial. Simply click on the blue button 

on their web home page (`www.dnastar.com`).

All materials contained herein are intended for educational purposes only.

The standard genetic code.

The table shows the 64 codons and the amino acid for each. The mRNA is 5' to 3'

nonpolar polar basic acidic (stop codon)

1st base	2nd base U		2nd base C		2nd base A		2nd base G		3rd base
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

[A] The codon AUG both codes for methionine and serves as an initiation site: the first AUG in an mRNA's coding region is where translation into protein begins. [Nakamoto T (March 2009). "Evolution and the universality of the mechanism of initiation of protein synthesis". *Gene*. **432** (1–2): 1–6. doi:10.1016/j.gene.2008.11.001.PMID 19056476.]

Inverse Table

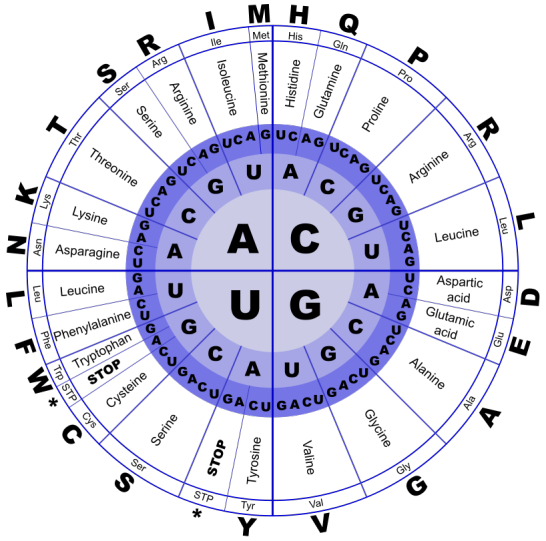
Ala/A	GCU, GCC, GCA, GCG	**Leu/L**	UUA, UUG, CUU, CUC, CUA, CUG
Arg/R	CGU, CGC, CGA, CGG, AGA, AGG	**Lys/K**	AAA, AAG
Asn/N	AAU, AAC	**Met/M**	AUG
Asp/D	GAU, GAC	**Phe/F**	UUU, UUC
Cys/C	UGU, UGC	**Pro/P**	CCU, CCC, CCA, CCG
Gln/Q	CAA, CAG	**Ser/S**	UCU, UCC, UCA, UCG, AGU, AGC
Glu/E	GAA, GAG	**Thr/T**	ACU, ACC, ACA, ACG
Gly/G	GGU, GGC, GGA, GGG	**Trp/W**	UGG
His/H	CAU, CAC	**Tyr/Y**	UAU, UAC
Ile/I	AUU, AUC, AUA	**Val/V**	GUU, GUC, GUA, GUG
START	AUG	**STOP**	UAG, UGA, UAA

Source: http://en.wikipedia.org/wiki/Genetic_code

L02: Lasergene part 1

Editseq / SeqBuilder / Protean / Protean3D

1. Table of Contents

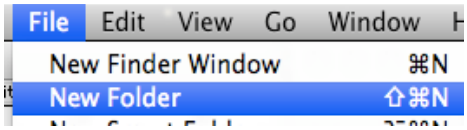


http://all-free-download.com/free-vector/vector-clip-art/genetic_code_rna_54619.html

1. Setup:

TASK

Create a folder named L02 on the desktop. This will be used to save files created during this session.

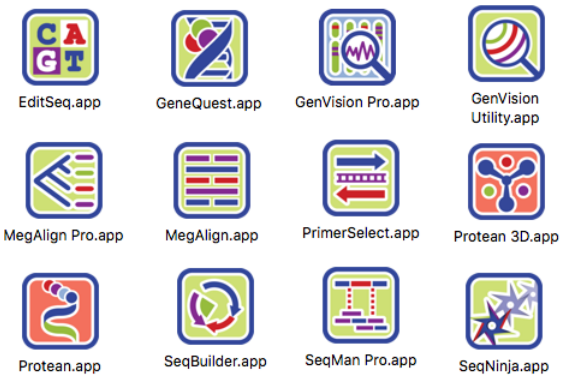


2. Lasergene modules

Lasergene is composed of modules. Most modules can share sequence files simultaneously. The suffix .app may or may not appear on your system depending on system preferences.

The modules are located within the folder:

Macintosh HD > Applications > DNASTAR > Lasergene 14



Lasergene modules

SeqMan Pro	*Sequence Assembly and Contig Management*
GeneQuest	*Gene Discovery, Sequence Annotation, Publication*
Protean	*Protein Structure Discovery, Annotation, Publication*
Protean3D	*New module for exploring 3D PDB files*
MegAlign	*Multiple and Pairwise Sequence Alignment*
PrimerSelect	*Oligonucleotide Design and Analysis*
SeqBuilder	*Edit and format. Maps and Plasmid maps.*
EditSeq	*Importing and Editing Primary Sequence Data*
GeneVision Pro	*genomic visualization application*

Note: some Lasergene modules only run in Windows (*e.g.* ArrayStar)

3. EditSeq and NCBI Sequences

The National Center for Biotechnology Information (NCBI) is the main repository for sequences. Lasergene modules can retrieve sequences directly from the NCBI server called "Entrez" (which is French for "Come in!")

EditSeq as well as SeqBuilder, PrimerSelect, and GeneQuest can import NCBI /Entrez sequences directly.

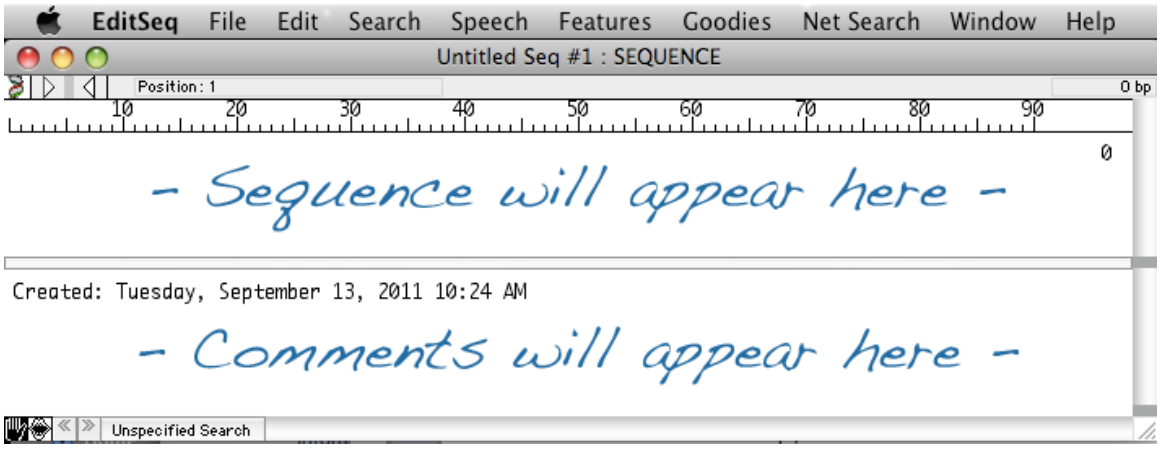
3.1 Launching EditSeq

TASK

Locate and open EditSeq:

Macintosh D > Applications > DNASTAR > Lasergene 12 > EditSeq

This will open an "Untitled Seq #1: SEQUENCE" blank window with 2 panes, the bottom pane showing the time of creation and the top pane, now blank, will contain the sequence.

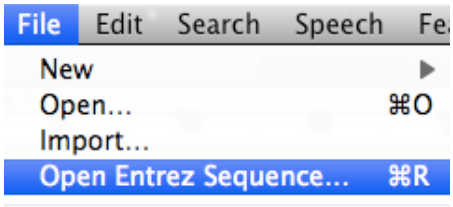


PRACTICAL: To create a new sequence file from your own sequence data you could simply paste the relevant sequence within the top panel and save the file for later use.

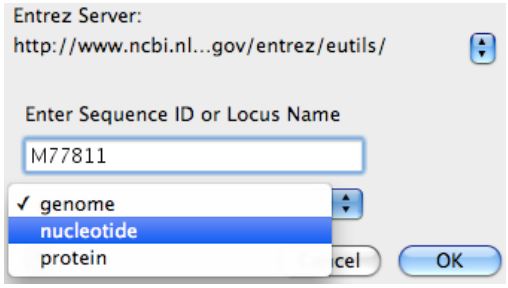
3.2 Importing sequence from NCBI

TASK

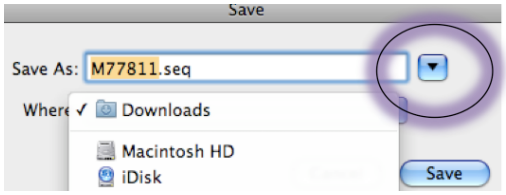
- Within the **File** menu select the option **Open Entrez Sequence…** or use the keyboard menu shortcut **⌘R**



- Within the new window **enter** the sequence accession code **M77811**
- **Change** the default look-up area from genome to **nucleotide**
- **Click OK**



- To change the default download area click on the "Where" pull-down menu and select L02 on the Desktop.



Note: The default location for downloads on a modern Macintosh is the "downloads" directory (located within the DMC username.)

Above we created a directory called L02 to contain files for these exercises.

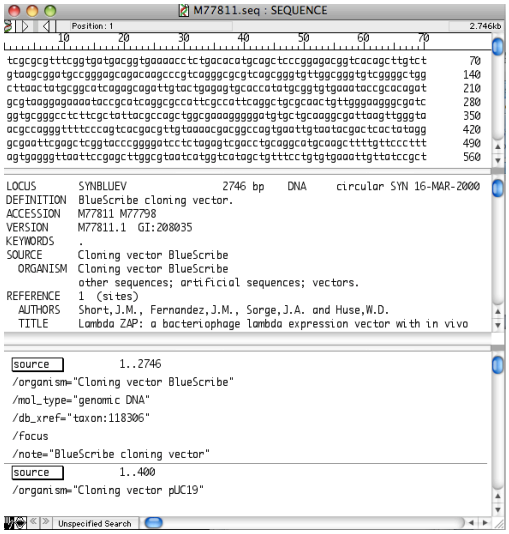
If L02 is not visible on the pull-down menu, click on the top right triangle within the blue square to expand the view (circled in above image.)

This will always allow you to navigate to the correct location.

- **Click Save** to save the file within the L02 directory. It will be named M77811.seq by default.

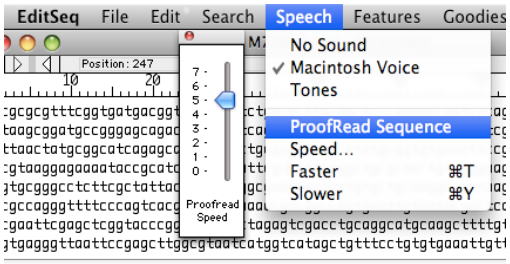
✔ READ

- The original file will now be parsed into 3 separate panels:
  - Top pane: Actual sequence in the same case as on NCBI (lower case.)
  - Middle pane: definitions from the header.
  - Bottom pane: special notes and references that can be edited.
- Each panel has a vertical and a horizontal scroll bar.



✔ INFO

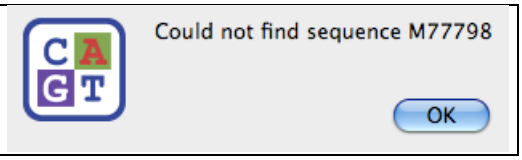
- The computer can read the sequence aloud for proofreading. Of course this is most useful when verifying short sequences.



The first 3 lines of the middle pane read:

```
LOCUS       SYNBLUEV     2746 bp    DNA     circular SYN 16-MAR-2000
DEFINITION  BlueScribe cloning vector.
ACCESSION   M77811 M77798
VERSION     M77811.1  GI:208035
```

This sequence is that of a cloning vector, last revised in the year 2000.

• Both accession number M77811 & M77798 are listed, but M77798 is obsolete and replaced by M77811. NCBI keeps version tracks.	

This cloning vector was engineered from different sequences, listed in the "source" section within the middle and bottom panes.

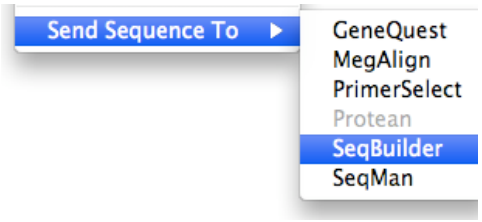
DO NOT CLOSE the window, we will use this sequence in the next exercise!

4. Sending sequence to SeqBuilder

Modules can send sequences to each other without the need to save the file and reopening it within the other module.

4.1 Sending a sequence from EditSeq to SeqBuilder

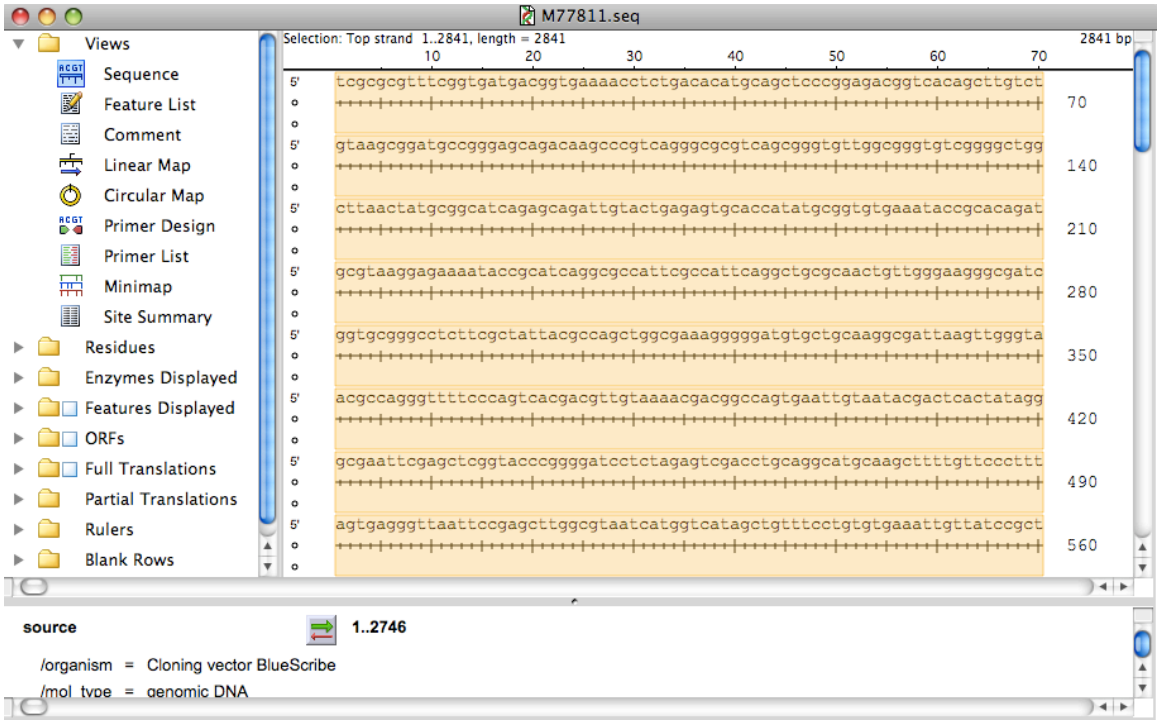
TASK



- Within the **File** menu select the last item:
  Send Sequence To > SeqBuilder

If SeqBuilder is not yet running, it will open automatically and load the new sequence.

When SeqBuilder opens, the default view is that of the sequence within a window with a side window showing viewing options.



SeqBuilder can be used to edit and display sequences. Many optional views can be seen simultaneously: Sequence view, Feature List view, Comment view, Linear map, Circular map, Primer Design view, Primer List view, Minimap, and Site Summary.

We will use the display capabilities to create an enzyme restriction map.

4.2 Create and display an enzyme restriction map

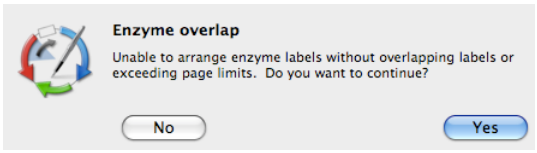
✔ **TASK**

- Within the left panel **click on** the small triangle to expand both directories: **Enzymes Displayed** and **Filter by Selectors**
- **Click on 5'Overhang**
- If necessary scroll down the left panel.
- When the 5' Overhang is clicked, scroll down on the right panel if the window is too small. The area between bases 400-500 is the polylinker where many enzymes are unique cutters.
- The polylinker enzymes are listed within the header of the NCBI file:

```
POLYLINKER  HindIII-SphI-PstI-SalI-XbaI-BamHI-SmaI-KpnI-SacI-EcoRI
```

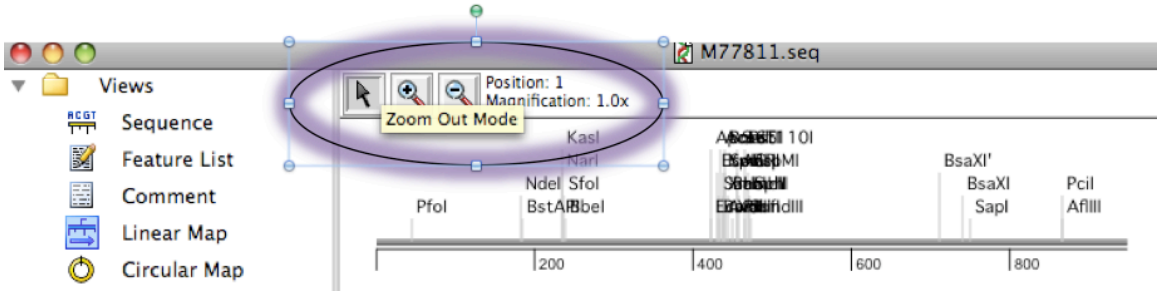
4.3 Restriction enzymes: linear map

✔ **TASK**

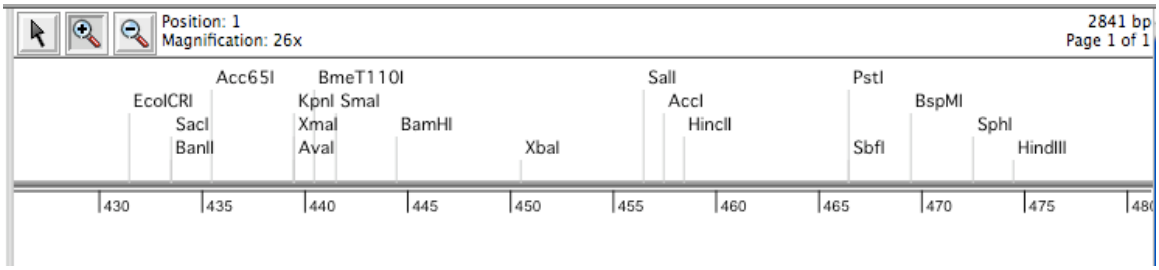


- On the left panel click on **Linear Map**
- **Click Yes** after the warning if any.

- **Click** on the magnifying glass to zoom in the desired area around 400.



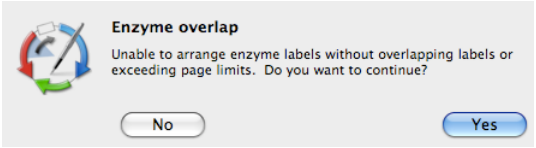
A magnification of around 24 to 36 should provide a better view:



4.4 Restriction enzymes: circular map

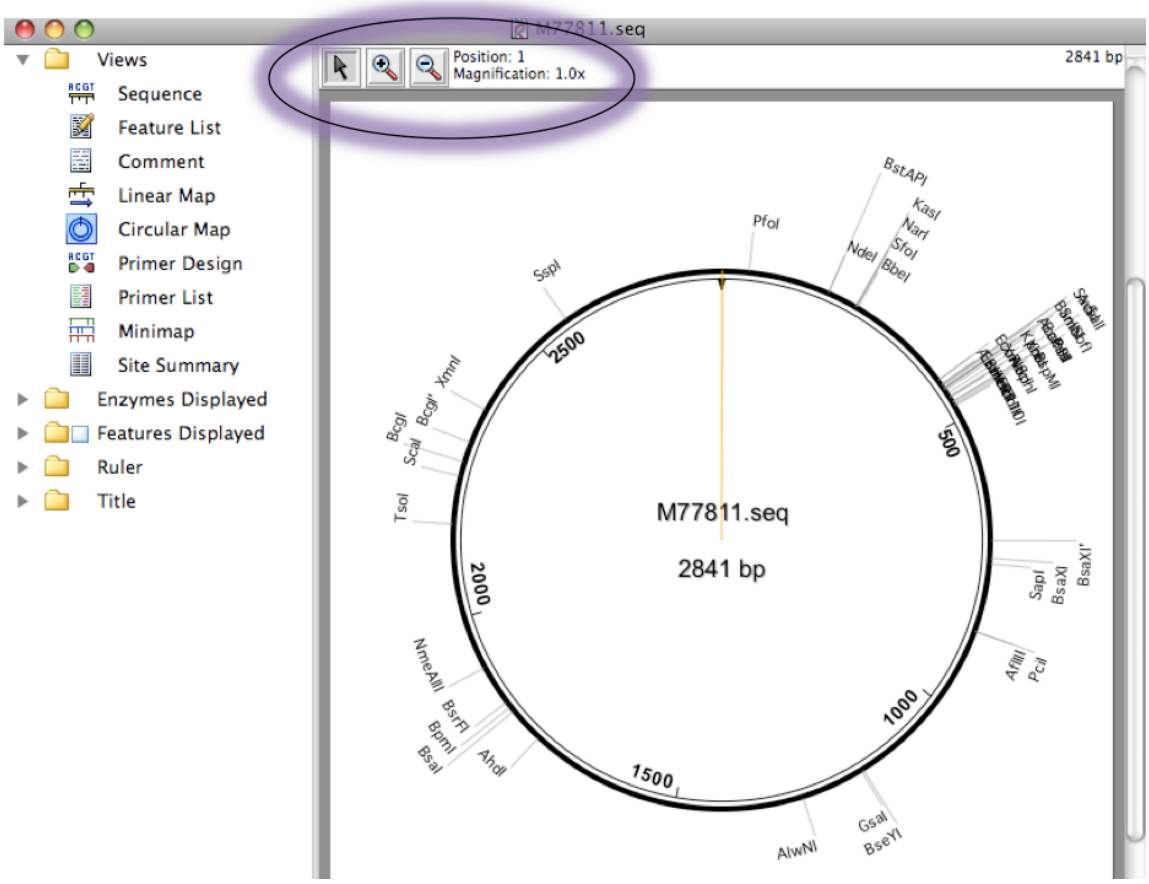
TASK

- On the left panel click on **Circular Map**
- **Click Yes** if warning appears

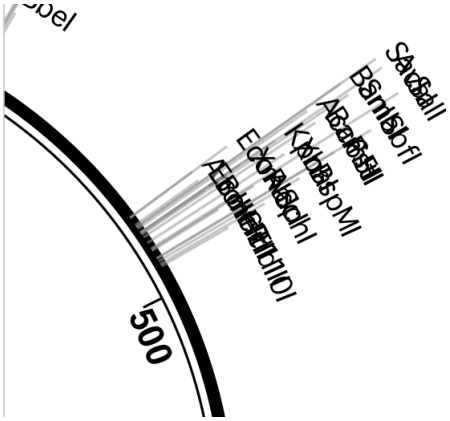


On the left panel, under Enzymes Displayed / Filter by Selectors:

Click "Unique Sites"



- **Click** on the magnifying glass to zoom in the desired area around 400. However, in this representation there are too numerous labels overlapping.

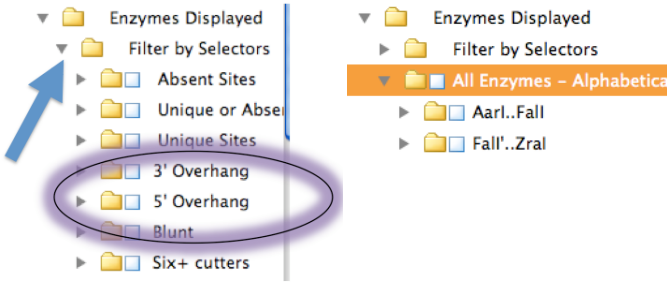


4.5 Restriction enzymes: choosing individual enzymes

Enzymes can be chosen individually from an alphabetical list.

TASK

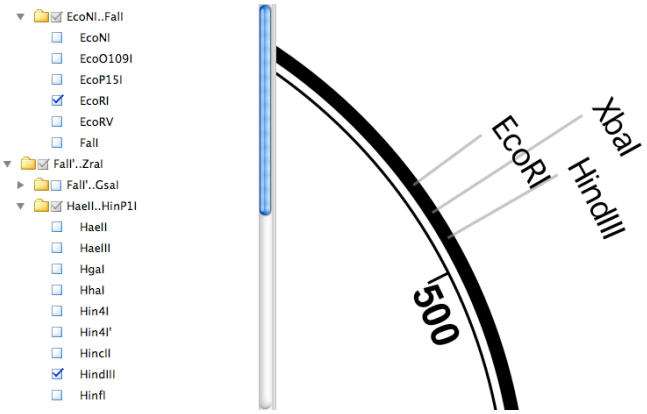
- **Unselect** the 5' Overhang and/or Unique Sites
- **Close** the **Filter by Selectors** tab
- **Open** tab **All Enzymes - Alphabetical**



The next 2 tabs are AarI..FaII and FalI'..ZraI containing all enzymes in alphabetical order.

For the sake of example we will select 3 enzymes present in the linker to be displayed: EcoRI, HindIII and XbaI

- **Click** the tabs to uncover the 3 enzymes alphabetically.
- This image shows the final result as well as the left panel with 2 of the enzymes visibly selected.

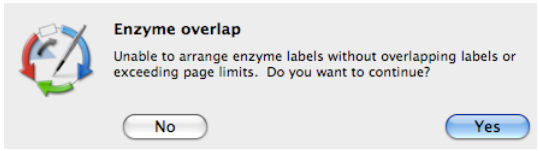


4.6 Minimap: a linear outlook of restriction sites

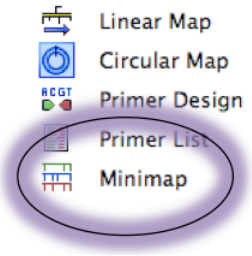
Another interesting way to visualize and find cutting enzymes is via the minimap view.

TASK

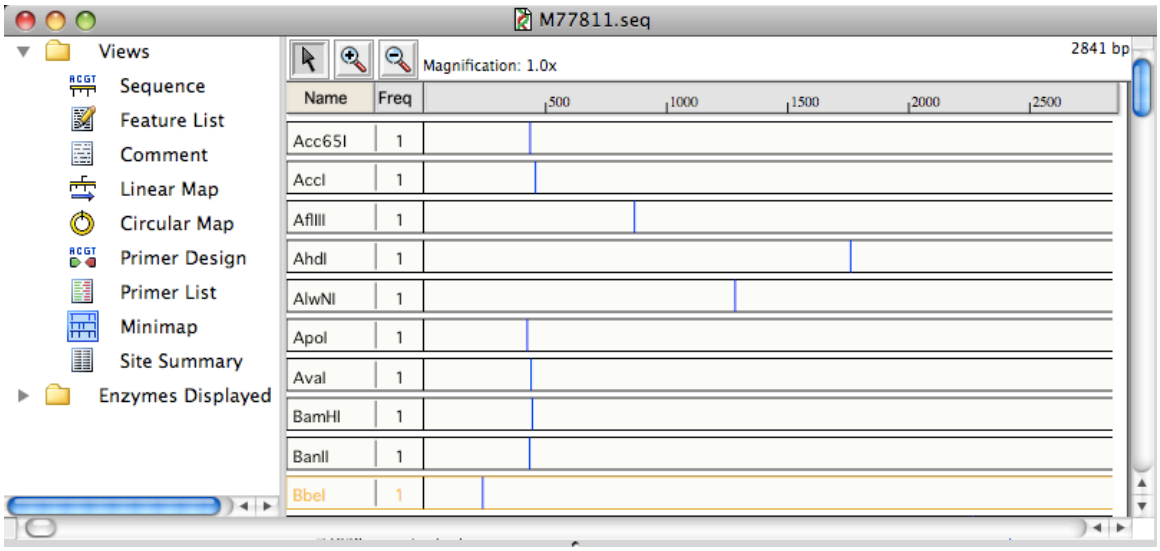
- From the **Enzymes Displayed > Filter by Selectors** button on the left panel
- **Select Unique Sites**
- **Click Yes** if warning appears.



- On the left panel **Click on Minimap**
- Regardless of the magnification within the previous view, it will be reset to a magnification of 1.0x within the Minimap view



The default Minimap view will show the list of cutting enzymes in alphabetical order, followed by the cutting frequency and an horizontal line symbolizing the sequence location with a blue vertical line locating the enzyme cutting position.

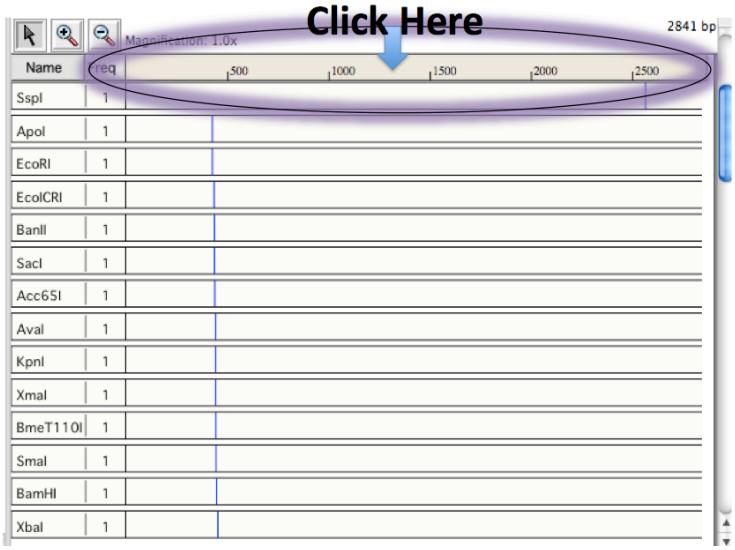


The list order can be changed by clicking within the table header. For example, to see the cutting as a function of the sequence position:

✔ TASK

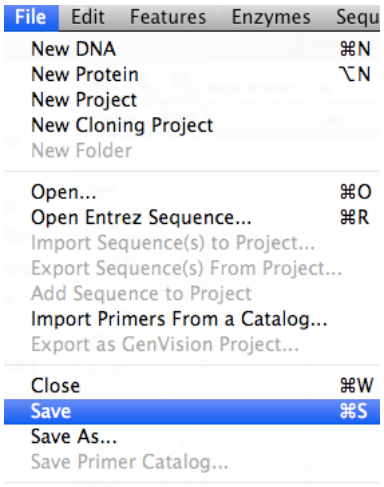
- **Click on the sequence number line at the top**
- **Scroll down** to find the polylinker area (between 4 and 500)

The view will now be be reset to the new settings.

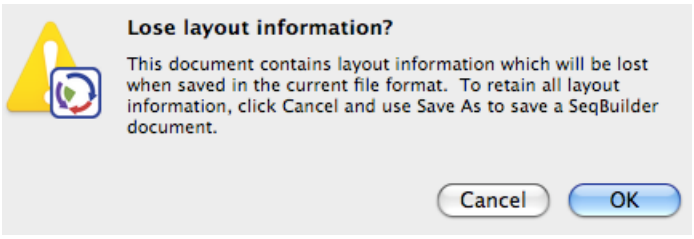


4.7 File saving from SeqBuilder

INFO

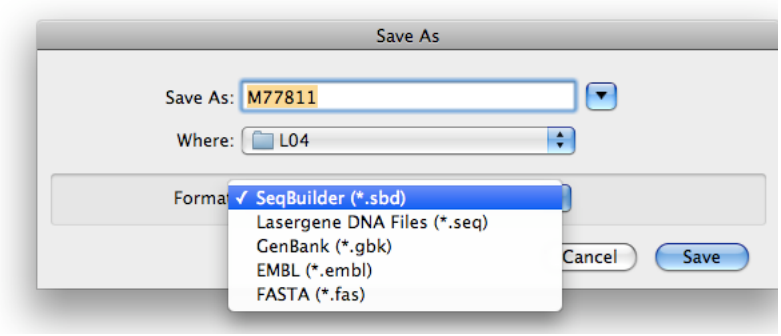


SeqBuilder files have more information than EditSeq files. Therefore trying to simply save the sequence with the **File > Save** menu cascade will bring a warning:



To preserve SeqBuilder format and layout information, use the **File > Save As…** menu cascade instead.

Note that there are other file formats available.



5. SeqBuilder Features Plasmid Maps

5.1 Creating a new Protein Translation Feature

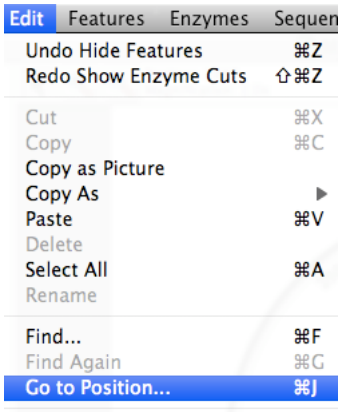
This exercise section uses the same Bluescribe vector file as in the previous exercise: M77811.seq.

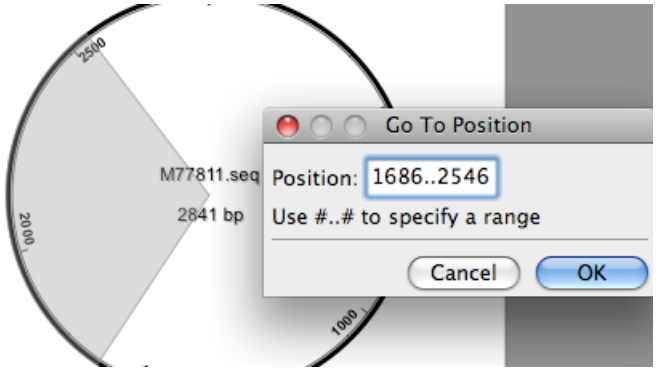
The Bluescribe vector contains an open reading frame on the negative strand, coding for ß-Galactosidase (LacZ) which is used for selecting clones.

In the following exercise we will tell SeqBuilder where the sequence is and create a graphical feature to highlight the protein on the circular plasmid map.

✔ TASK

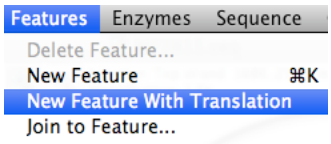
- **Open SeqBuilder**
- **File > Open…** (or **⌘O**) and open file M77811.seq saved previously (*Alternatively use the File > Open Entrez Sequence… as we have seen in a previous exercise.*)
- **Click on Circular Map** on the left panel.
- Select **Edit > Go To Position** and enter the range **1686..2546** (the corresponding region will become highlighted within the circular map. Note the 2 dots between the numbers.)

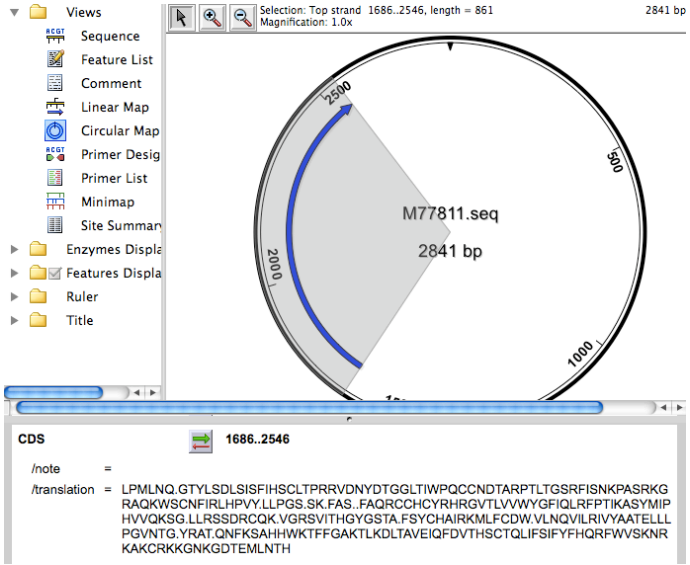




- Select **Features > New Feature With Translation** and note how a new, blue arrow will appear at the same time as a new feature entry within the bottom panel together with a new translation. However, the translation is currently wrong because the coding sequence is in fact on the *opposite strand* and therefore we will need to make a few changes.

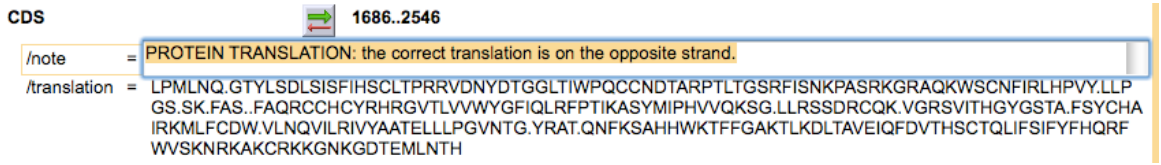
The new feature within the lower panel is labeled CDS.



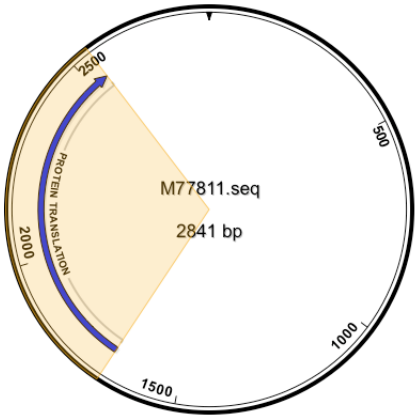


✔ TASK

- Within the lower panel click after the = sign on the /note keyword and enter a reminded that the translation is on the other strand: e.g.
  "**PROTEIN TRANSLATION: the correct translation is on the opposite strand.**"

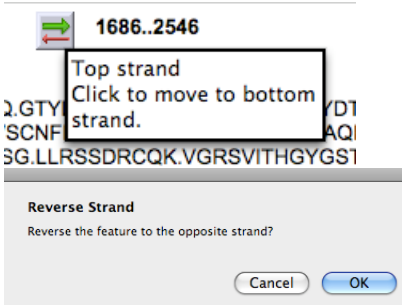


- Note that the words before the colon (here "PROTEIN TRANSLATION") will appear as the decoration title of the blue arrow within the circle.

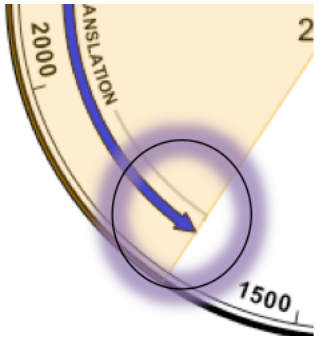


- **In the bottom panel: Hover the mouse** on the grey square with the green arrow next to CDS and <u>wait long enough</u> for the instructions to appear.

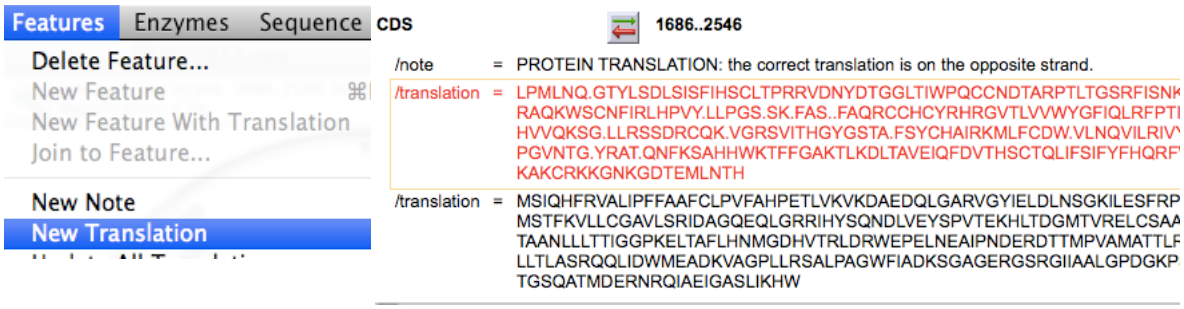
- **Click on the square** to change to the opposite strand.
- The red arrow will become larger and the translation text will change from black to red.
- **Click OK** in the warning box



- <u>Note</u>: at the same time the blue arrow within the circle will point in the opposite direction



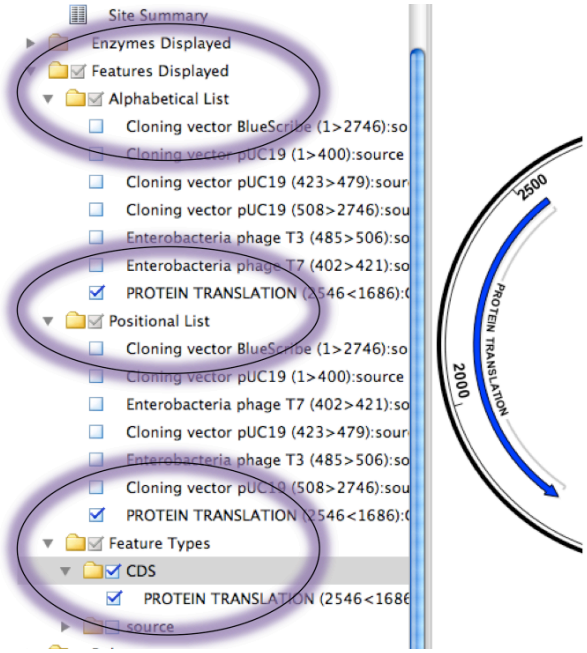
- Select menu **Features > New Translation** and a new translation, in black color will appear. This new translation is from the opposite strand and is now the correct protein sequence!



INFO - Note that the protein translation feature (the blue arrow and its named decoration) appears as a highlighted gray or yellow arc within the circle while it is selected. Similarly, it is possible to graphically select this feature by clicking on either the blue arrow or the decoration's name "PROTEIN TRANSLATION" within the circle.

In addition the new feature "CDS" was added within the "Features Displayed" list on the left side panel in all 3 categories: Alphabetical List, Positional List, and Feature Types.

Since this vector was created from various sources, these list indicate the origin of the sequences and can be toggled on and off.

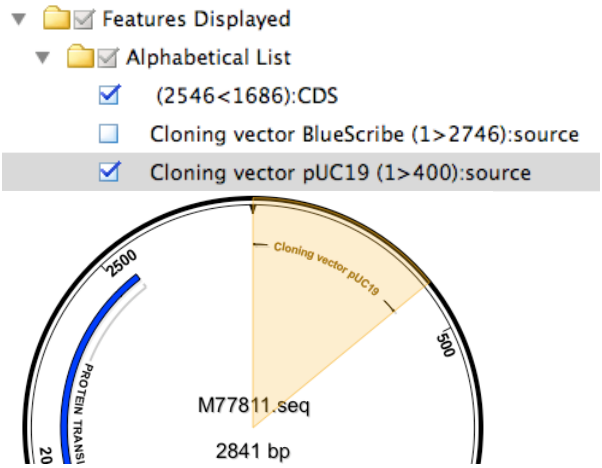


5.2 Creating a new Graphical Feature

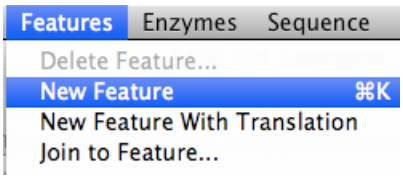
New graphical features can be added for features listed within the Genbank file. As an example we will add an arc-circle for the sequence segment corresponding to the pUC19 cloning vector:

✔ **TASK**

- Within either section of the "Features Displayed" (*e.g* under "Alphabetical List") **click on** "**Cloning vector pUC19 (1>400):source**"
- **Click on** the new description within the circle to select this area. The arc corresponding to this area will become highlighted.

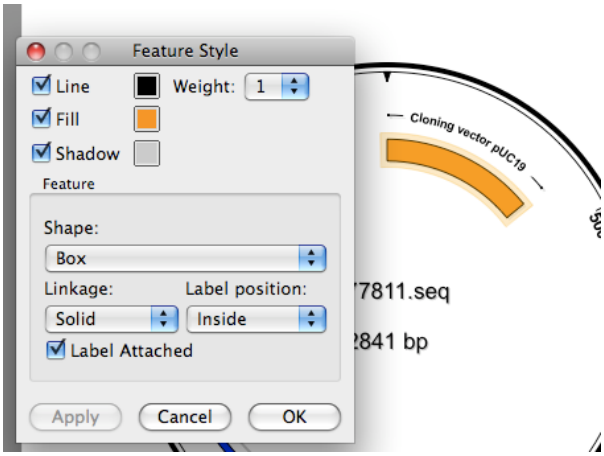


- Within the menu bar select **Features > New Feature** or use the **⌘K** shortcut.

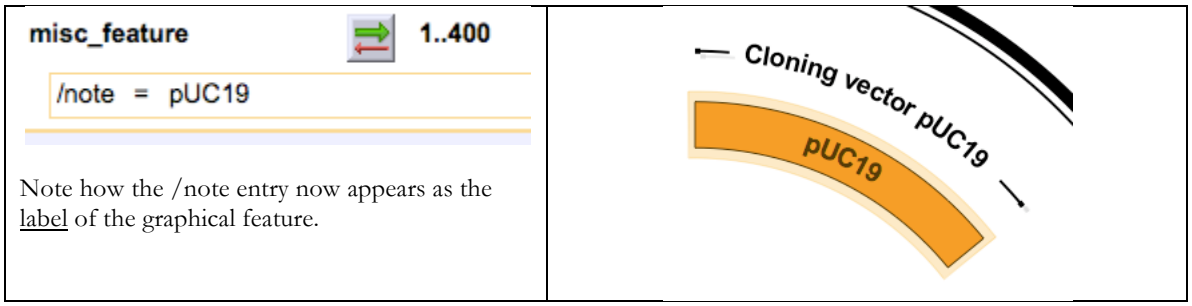


- A new thick arc will appear within the circle as the graphical representation of this area in the default position: below the text, as an orange arc.

- Select the menu cascade: **Features > Edit Feature Style**
- **Unselect Shadow**
- **Click Apply**
- **Click OK**



- Select menu cascade **Features > Edit Feature Description**
- This will bring the bottom panel section into focus with a default /misc_feature keyword and a /note keyword entry as well.
- Within the /note section **type** "**pUC19**"



5.3 Editing the Graphical Features of a demo file

SeqBuilder files end with the .sbd filename extention. We will look briefly at 2 files located within the demo data of the Lasergene 8 package.

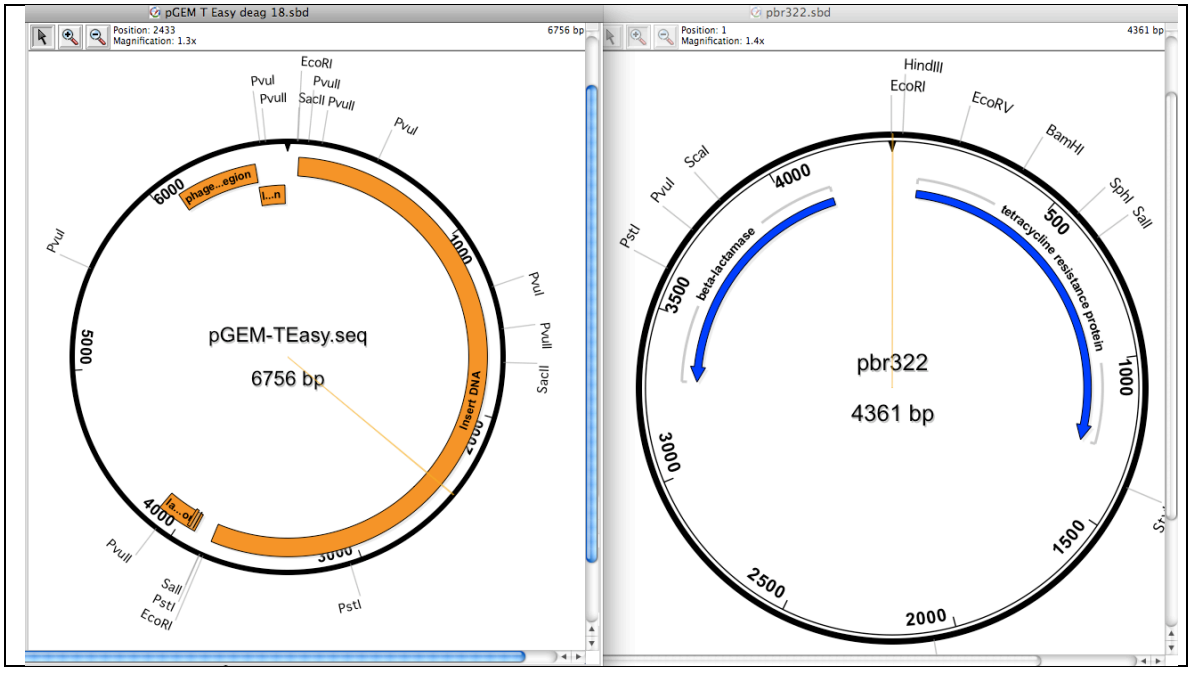
SeqBuilder should be running from the previous exercise segment. If not, launch SeqBuilder.

✔ **TASK** Using the **File > Open** menu cascade open successively the following 2 files:

pGEM T Easy deag 18.sbd

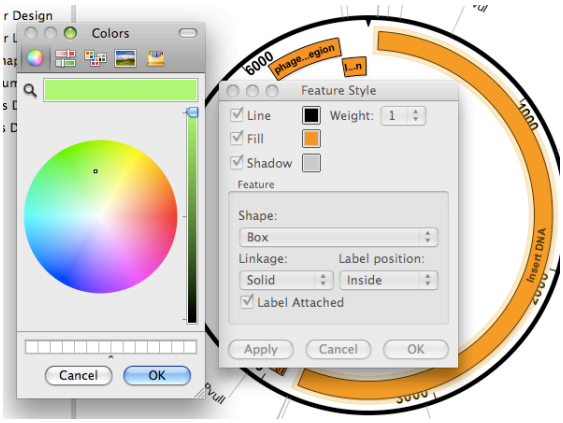
pbr322.sbd

both located within **/Applications/ DNASTAR/Lasergene 12 Data/Demo Data/**

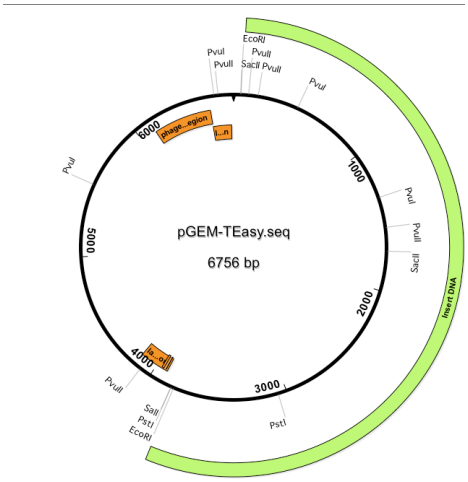


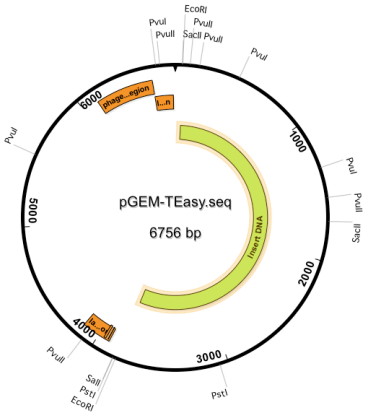
As an exercise we will change the color of the "insert DNA" feature as well as its position within the circle:

- **Double-click** on the "**insert DNA**" feature within the circle
- Select menu **Features > Edit Feature Style**
- **Click on** the **Fill** color square
- **Select a new color** e.g. light green (as shown here)
- **Click OK** (within the "colors" window)



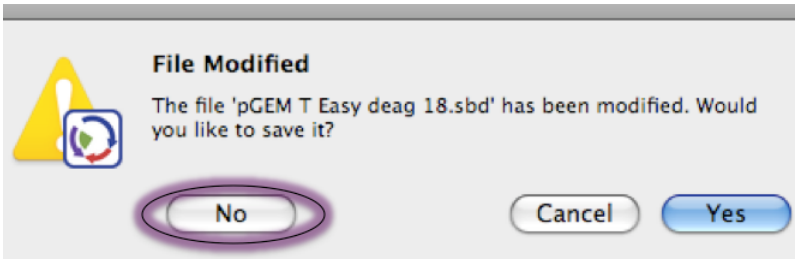
- **Click Apply** and then **Click OK** (within the "Feature Style" window)
- **Click and hold** the selected **graphical feature segment** and while holding the mouse button down **slide the feature either** to the outside or the inside of the circle to obtain *either* of these representations:





5.4 Quit SeqBuilder

✔ **TASK** Quit the SeqBuilder program: **SeqBuilder > Quit SeqBuilder** or use the **⌘Q** shortcut. Do not save the changes made to the demo files (Answer NO when asked if you want to save changes)



6. Protean – Protein Analysis

✔ **TASK**
Open the Protean module located within
/Applications/Classes/DNASTAR/Lasergene 12/

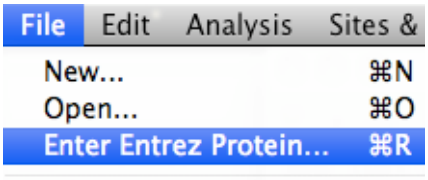
A logo will flash and seemingly not much else. However, note that the top menu has switched from the finder menu to the Protean menu:

6.1 Opening an Entrez protein sequence

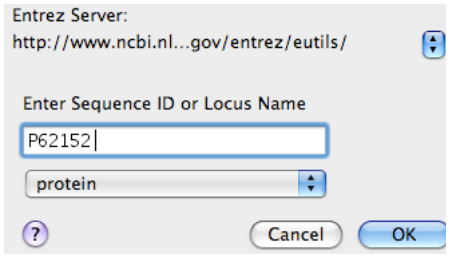
Protein sequence can be either imported from Entrez, or opened from a file. Proteins sequences files end with .pro in Lasergene.

✔ **TASK**

- Select the menu **File > Open Entrez Protein** or the **⌘R** shortcut

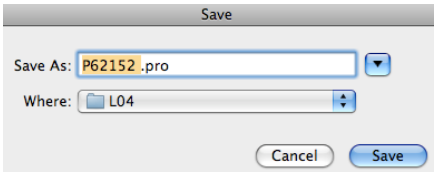


- Request protein sequence with accession code **P62152** (formerly P07181).



Note: The entry name of this calcium-binding calmodulin is CALM_DROME and details of the protein can also be found at
`http://us.expasy.org/uniprot/P62152`
or
`http://www.uniprot.org/uniprot/P62152`

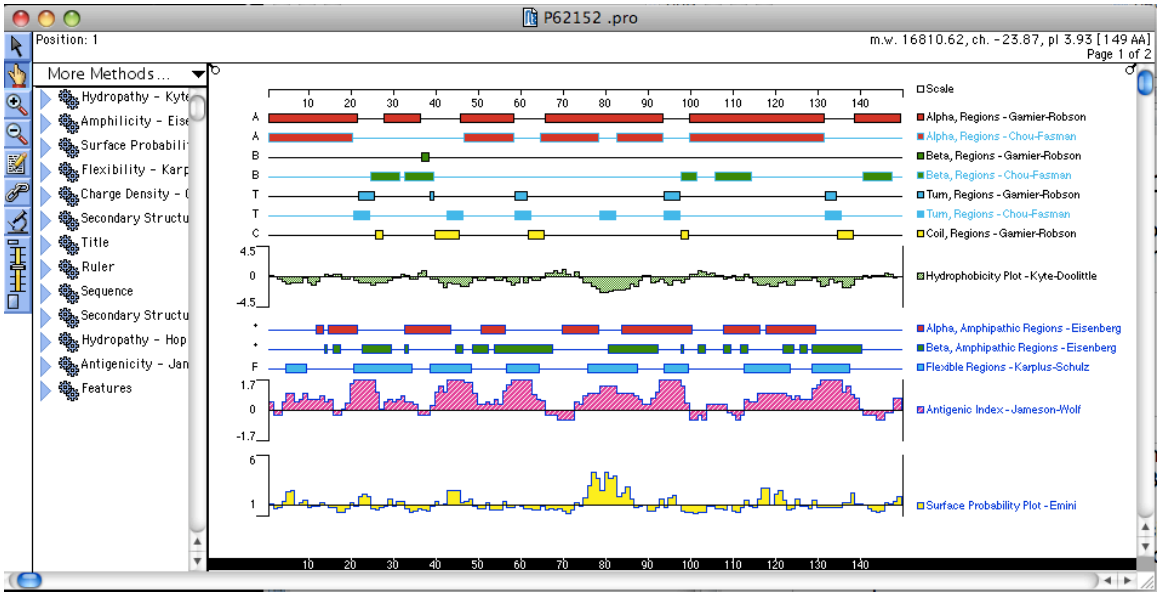
- **Save file within L02 directory** created on the desktop earlier.
- **Accept** the proposed **.pro** extension
- **Click Save**



Note: in the future when you have your own sequences, if these are not available on Entrez, simply use EditSeq to create the .pro file containing the sequence you want to analyze.

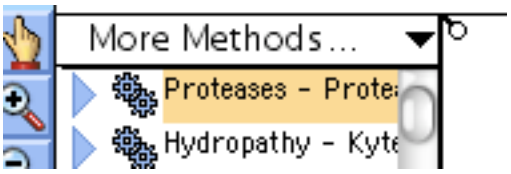
6.2 Protean Default View and Methods

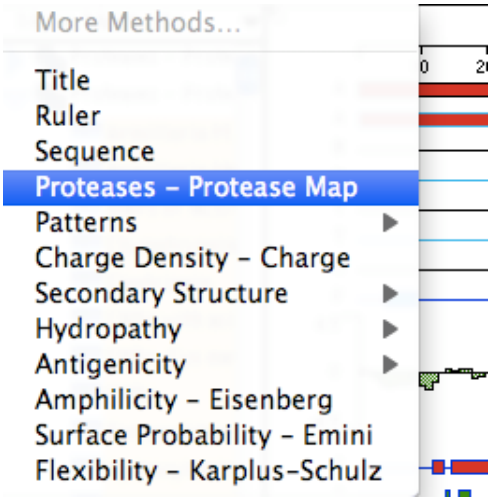
The default view is that of standard secondary structure prediction with different methods that will be reviewed later.



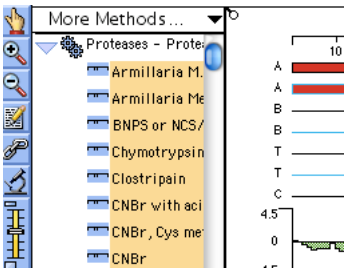
TASK

- At the top left **click on** "**More Metods**" and select **Proteases – Protease Map**
- The new analysis method will appear at the top left of the list of method currently displayed:





- **Click on the small triangle** to see the list of available enzymes and chemical treatments.



INFO

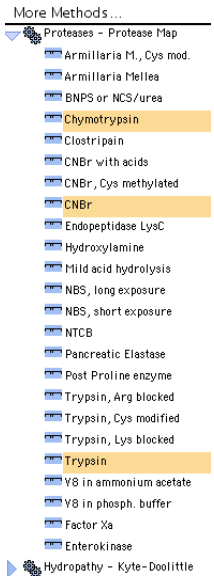
The same information is provided within a list by choosing **Sites & Features > Show Protease List** as shown in the next image:

Proteases

Name	Sites	Exceptions	Remarks
Armillaria M., Cys mod.	X^C,X^K		Cys is reversibly aminoethylated.
Armillaria Mellea	X^K		See the variant of this method with reversible modification of Cys residues.
BNPS or NCS/urea	W^X		Trp is modified. Problem: with NCS Met can be modified into a sulfone.
Chymotrypsin	W^X,F^X,Y^X,M^X,L^X,H^X	Y^P	Not very specific, can also sometimes cut after Asn, Gln, Lys and Thr.
Clostripain	R^X		Same specificity can be obtained with Trypsin after blockage of Lys residues.
CNBr with acids	M^X,W^X		Met is modified to Homoserine Lactone and Trp is also modified and blocked on its C-Terminal.
CNBr, Cys methylated	M^X,X^C		Cys is modified to Serine. Met is modified to Homoserine Lactone.
CNBr	M^X		Met is modified to Homoserine Lactone. Problem: Cys could be oxydized (but it is reversible).
Endopeptidase LysC	K^X		Same specificity can be obtained with Trypsin after blockage of Arg residues.
Hydroxylamine	N^G,N^L,N^A,N^M		Cuts rarely: Asn-Leu, Asn-Ala and Asn-Met.
Mild acid hydrolysis	D^P		Not a very efficient method.
NBS, long exposure	W^X,Y^X,H^X		Trp, Tyr and His are oxydized. A shorter exposure will not cleave after His (see corresponding method).
NBS, short exposure	W^X,Y^X		These conditions are unfavorable for histidine cleavage. Trp, Tyr and His could be irreversibly oxydized.
NTCB	X^C		Cys is 5-cyanylated and can be reduced to Ala with Ni Raney. Problem: Cys can cyclise irreversibly.
Pancreatic Elastase	A^X,G^X,S^X,V^X		Not very specific.
Post Proline enzyme	P^X	P^P	Cleave slowly Ala-Xaa. Work only with proteins or fragments less then 50 residues long.
Trypsin, Arg blocked	K^X	K^P	One of the variants of the classical Trypsin method. Same specificity can be obtained with Endopeptidase LysC.
Trypsin, Cys modified	R^X,K^X,C^X	R^P,K^P,C^P	One of the variants of the classical Trypsin method.
Trypsin, Lys blocked	R^X	R^P	One of the variants of the classical Trypsin method. Same specificity can be obtained with Clostripain.
Trypsin	R^X,K^X	R^P,K^P	Classical method: see the variants of this method using reversible modifications of either Arg, Cys and Lys1.
Y8 in ammonium acetate	E^X		See also the method with Y8 in phosphate buffer which allows to cut Asp-Xaa.
Y8 in phosph. buffer	D^X,E^X		Slow to cut when Xaa is Phe, Val or Leu. See also the method with Y8 in ammonium buffer which will not cut after Asp.
Factor Xa	IEGR^X		Restriction Protease
Enterokinase	DDDDK^X		Restriction Protease

(Note: after viewing close this window)

6.3 Enzyme and chemical digestion and SDS PAGE Gel simulations

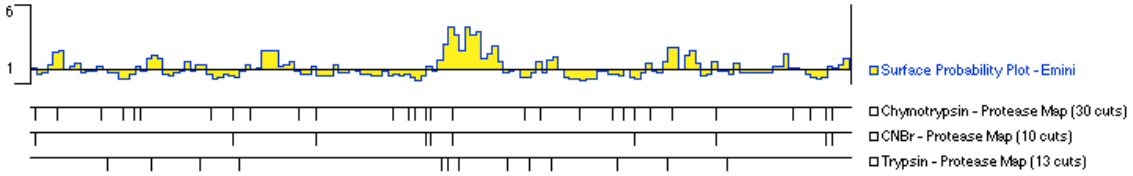


TASK

- Under **Proteases – Protease Map** within the Methods list select the following three items **Chymotrypsin**, **CNBr**, and **Trypsin**.

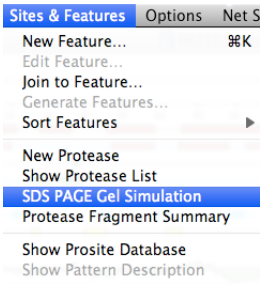
- (Note: **hold** the command key (⌘) on a Macintosh or the Control key on Windows.)

- **Drag** the methods over to the main, central panel and release the mouse.
- The new methods will be shown a the bottom of the pane below the last displayed method (here the surface probability).
- **Click in a white are to deselect** them for better viewing.
- Recognition sites are shown as vertical bars

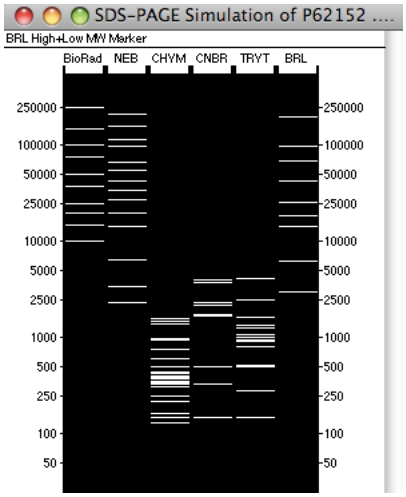


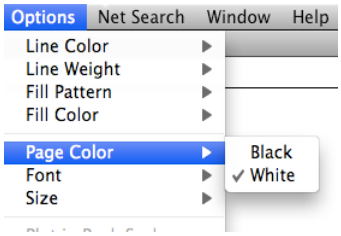
- **Holding** the **Shift key, Click** on each of these 3 lines to select them again

- Select menu **Sites & Features > SDS PAGE Gel Simulation.**
- An image of the simulated separation will appear within a new window.

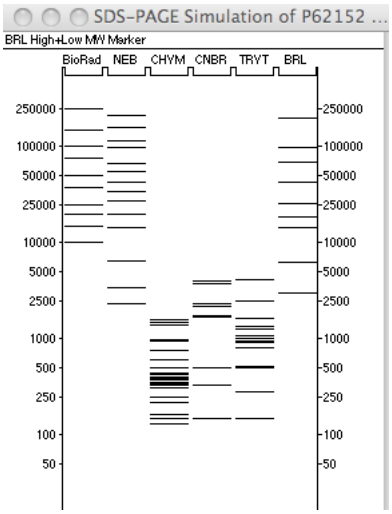


- The background color default can be changed from black to white with the Options menu items.



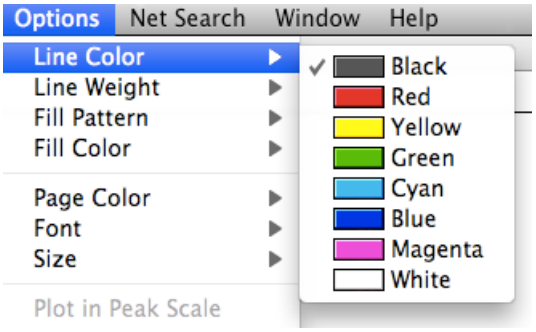


Options > Page Color > White



✔ INFO

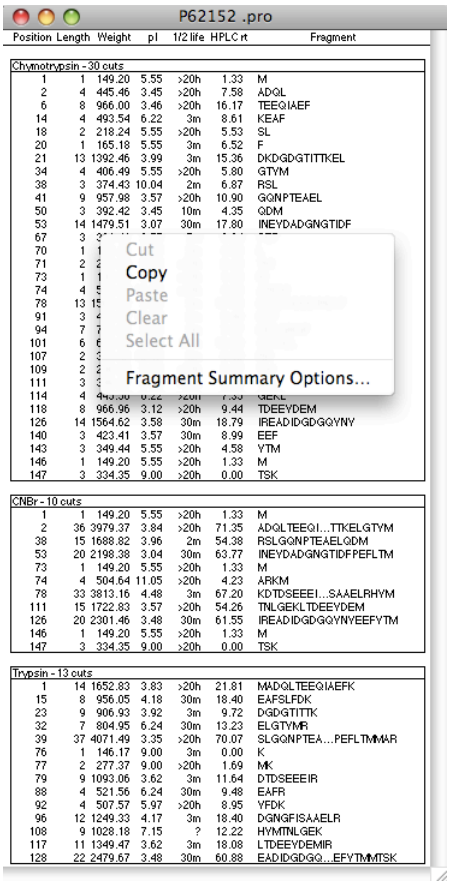
- Changing the page color from black to white automatically changes the line colors to the inverse color as well, here black.
- There are other colors to choose from as well.



Note: these colors apply ONLY to the lines selected within the Main View and NOT the SDS gel.

TASK

To see a summary of the simulated peptides select the following menu cascade: **Sites & Features > Protease Fragment Summary**. This opens a new window with the summary of the proteolytic fragments.



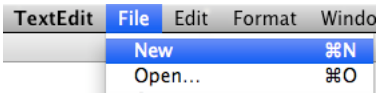
INFO

A **right-click** of the mouse would bring an option to **Copy** the contents and paste it elsewhere. Depending on the nature of the software this is pasted into, the result could be either an image or a plain text list.

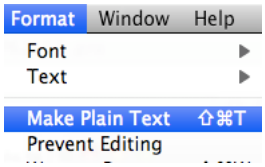
TASK

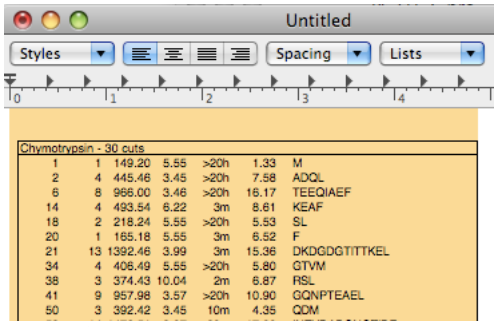
- **Right-click** (or Ctrl-Click) and select **Copy**
- **Open TextEdit** Application
- **Paste**
- **Try to move: it's a picture**

- TextEdit: **File > New**
  (or use **⌘N** shortcut)

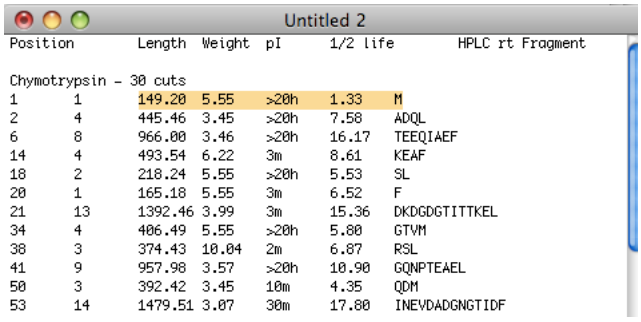


- TextEdit: **Format > Make Plain Text**
- **Paste**
- Now it is selectable plain text





Within TextEdit results in image

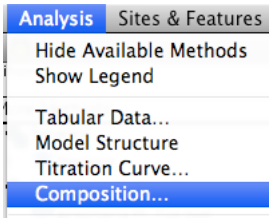


Plain text pasting results in editable text

Note: Sequence is one of the Methods and can be added for clarity within the main panel.

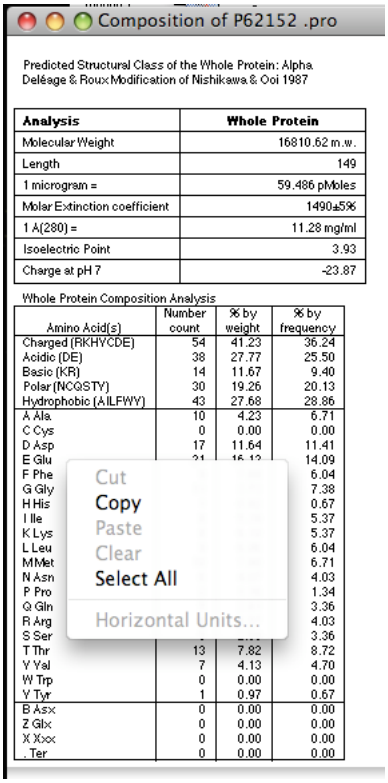
6.4 Composition

The protein composition can be obtained from the Analysis menu.



TASK

- Select the menu cascade: **Analysis > Composition**

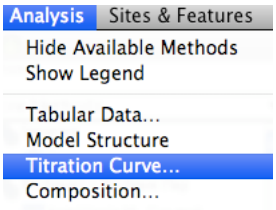


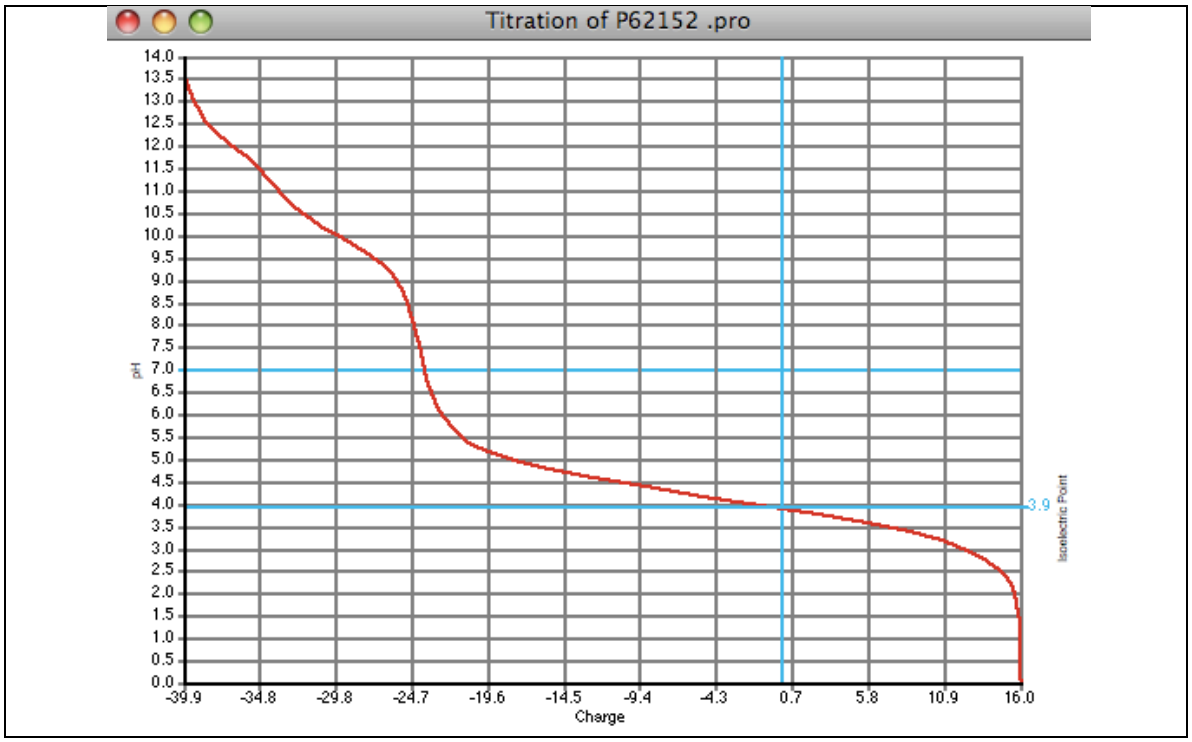
- The new window will contain the protein composition.
- In the same way as shown in the previous section, the contents can be copied.
- The resulting pasted material can be either plain text or an image depending on the software receiving the pasting.
- See above example with TextEdit

6.5 Titration Simulation

TASK

- Select the menu cascade: **Analysis > Titration Curve**



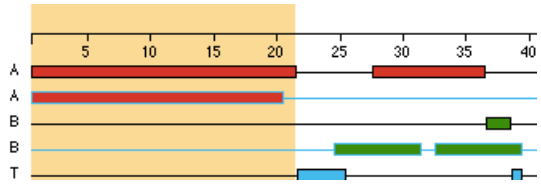


6.6 Model Structure Simulations: Helical Wheel

The prediction of secondary structure will be highlighted in the next section. However, the default view of Protean shows the predicted alpha- and beta- secondary structure. Protean also offers a graphical view for these local folds of the proteins: helical wheel, helical net, and beta net.

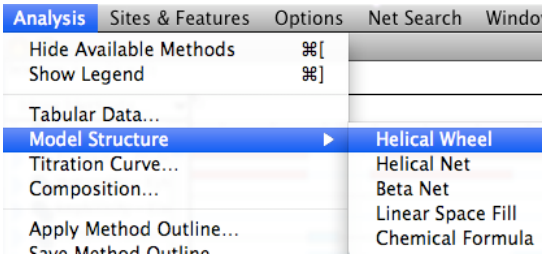
TASK

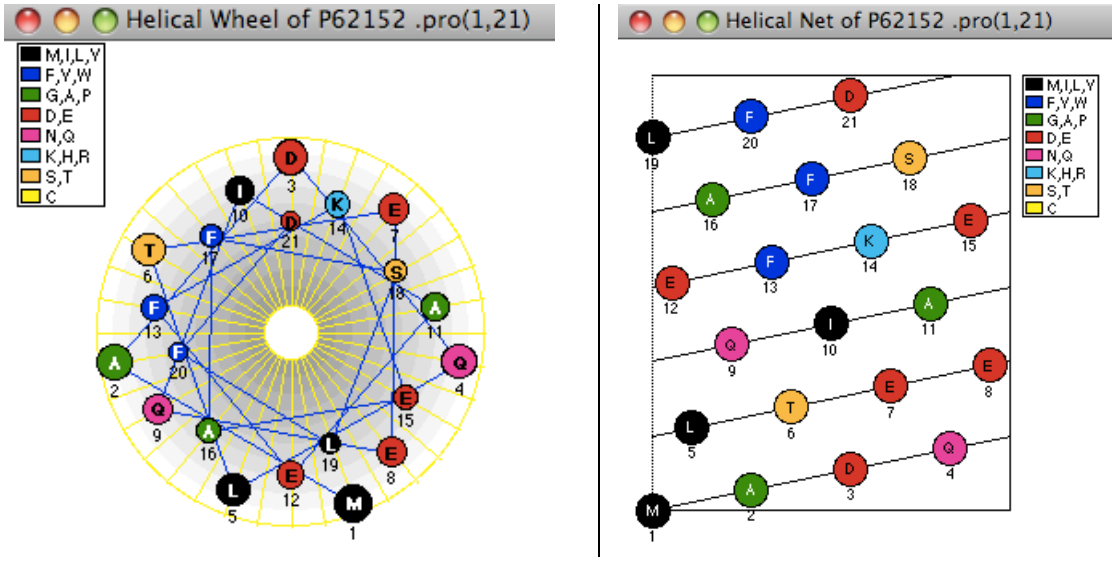
- On the main panel graphically select with the mouse the length of the first alpha helix structure on the first line: it appears as a red feature.



Note: On the left panel, with blue squares, you may need to switch from the "object selector" in the shape of a hand ( ) to the "range selector" in the shape of an arrow ( ).

- From the menu select: **Analysis > Model Structure > Helical Wheel**
- Then select: **Analysis > Model Structure > Helical Net**



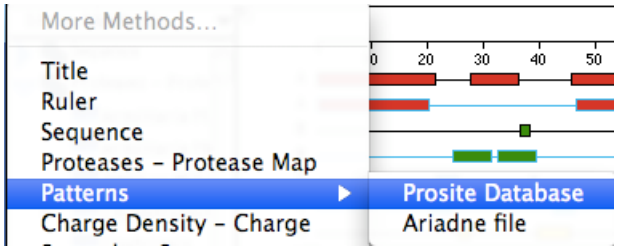


Both methods are useful for locating hydrophobic or hydrophilic patches and particularly useful for hydropathic helices.

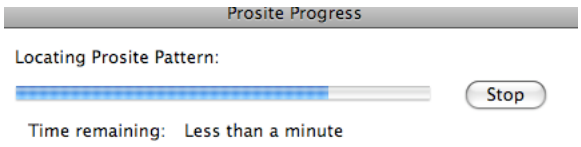
6.7 Patterns Search: Prosite

✔ **TASK**

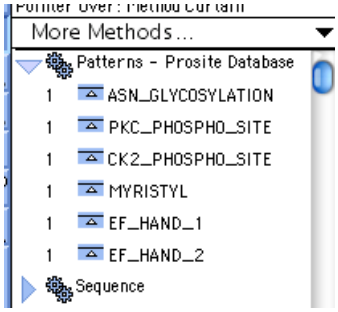
- Within "More Methods" choose **Patterns > Prosite Database**



- **Drag** the listed item **Patterns – Prosite Database** onto the main panel. A progress window will appear.



- The results will be displayed both within the graphical panel and under the Methods list

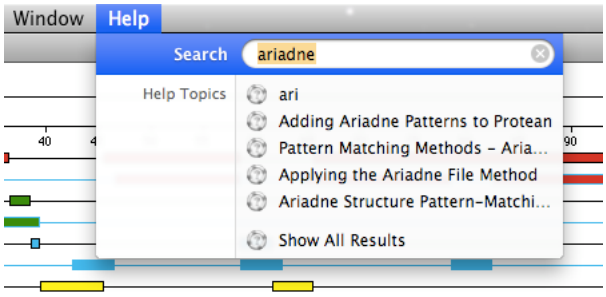


Note: The EF-hand pattern is typical of calmodulin and typically exhibit a *helix-turn-helix* motif with acidic residues chelating a calcium ion.

6.8 Help - Patterns Search: user-specified descriptor (Ariadne File)

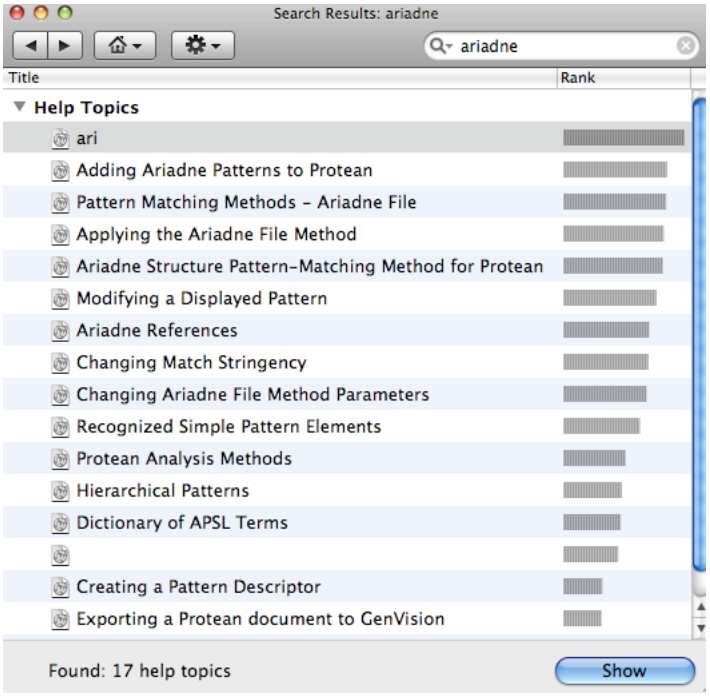
Help can be found with the last menu item on the menu bar from all Lasergene modules.

Here is an example of Help search for the Pattern Search method Ariadne.



Select "**Show All Results**" to display the list if entries in the Help search.

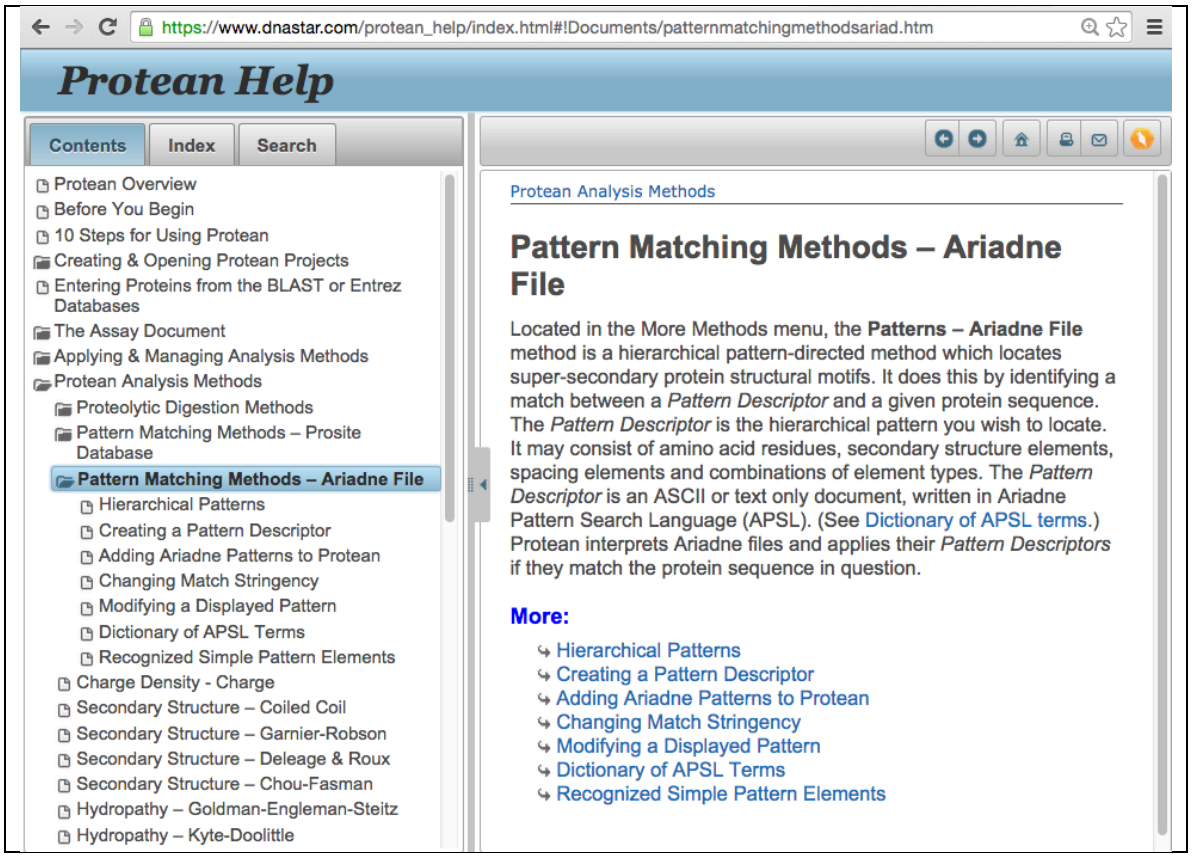
The results are shown in a new window and are sorted by relevance.



The Ariadne method is different from the Prosite (http://prosite.expasy.org/) pattern search and is based on a vocabulary of keywords and variables rather then ambiguous sequence codes.

To see an example for creating a search pattern file, see the help entry titled: "**Creating a Pattern Descriptor**" towards the end of the list and "**Recognized Simple Pattern Elements**."

To understand the meaning of descriptors with an example, see "**Hierachical Patterns**." See `https://www.dnastar.com/protean_help/index.html#!Documents/patternmatchingmethodsariad.htm` or the short URL version: `http://bit.ly/1FVLLV7`



6.9 Quit Protean

We will reopen Protean in the next exercise.

7. Protean: Secondary Structure Predictions

INFO Basics of sequence-based protein structure prediction:

Very often the only information at hand is the primary sequence of a protein, often deducted from DNA sequence. Based on the protein sequence and amino acids biochemical properties, some software predict the conformation of protein secondary structure elements: *alpha helices, beta sheets, turns and random coils*.

Most of the secondary structure predictions are based on tabulations of observed amino acids conformations in solved 3D protein structures, and anticipate that similar residues will occur in analogous configurations. In the algorithms of Chou and Fasman (1978), Levitt (1978) and Garnier, Osguthorpe and Robson (1978), amino acids are classified as formers, neutral or breakers of the secondary structure elements.

Chou and Fasman used 19 protein structures, with 2,473 amino acids and Levitt used 11,569 residues in globular proteins to deduct predictive rules, generalized here:

1- find any cluster of 4 consecutive helix-forming residues within any length of 6 amino acids
2- propagate helix in both directions from this nucleus until:
3- at least 4 helix-breakers are found (tetra-peptide)
4- beta sheet rules are similar except that 3 out of 5 beta-formers are required to nucleate a sheet
5- in the case of a tie between alpha and beta, the helix usually wins
6- turns require 4 out of 4 residues that prefer turn configurations.

The algorithm described by Garnier *et al.* is similar to that of Chou and Fasman and is also based on observed amino acids conformations. However, this method considers a window of 17 residues (8 residues on each side of the current amino acid on the sequence) and asks whether the flanking neighbors favor helix, sheet or turn configurations. Therefore the method is not only residue specific but also sequence specific. The program will also give different predictions if the user already knows that a protein contains a particular alpha/beta ratio.

The algorithms and tables have been refined considerably but the principles remain the same today and are good probably to 55% on average. But secondary structure predictions are not very reliable, especially for proteins that are not soluble or globular. The Chou and Fasman algorithm may be better at predicting structures with large helical content.

Chou P.Y. and Fasman G.D. 1978 Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol Relat. Areas Mol. Biol.* **47**:45-148
Levitt, M. 1978 Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**: 4277-4285
Garnier, D.G. Osguthorpe, D.J. And Robson, B. 1978 Analysis of the accuracy and implications of simple methods for predicting the secondary structure of proteins. *J. Mol. Biol.* **88**:873-894

7.1 Open the Protean software

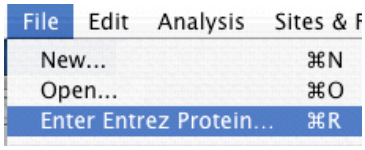
✔ **TASK**

Double-click Protean from the following location:
/Applications/DNASTAR/Lasergene 12/

7.2 Open a protein sequence and display predictions

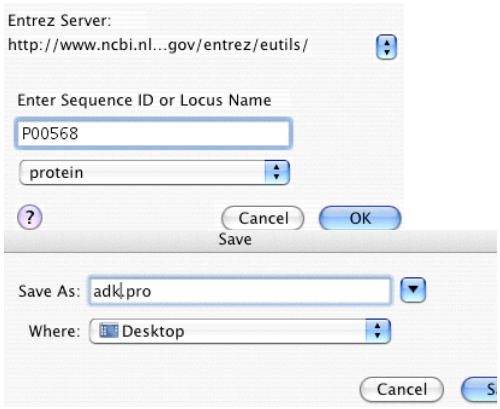
✔ **TASK**

Open a protein sequence from the Entrez database with the following Protean menu cascade:
File > Enter Entrez Protein…



On the next form **type P00568** as the accession number of a human adenylate kinase.

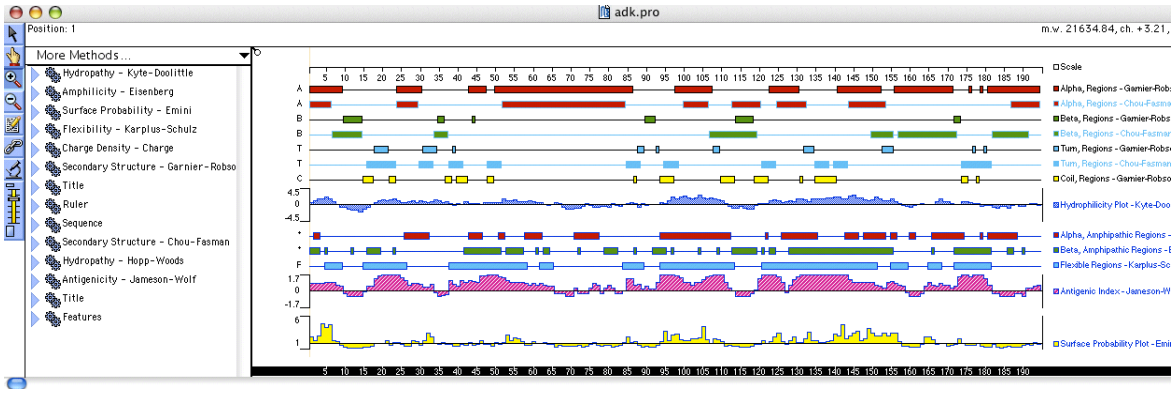
Click OK



Within the next window **change** the **name** to **adk.pro** and save the file in L02 on the Desktop.

Click Save

As soon as you click save, the file is saved in the L02 folder in a DNASTAR format, and the structure analysis is immediately displayed in a complex, default prediction panel:



The far left column contains icons to zoom or reduce the size of the display (magnifying glass with a + or – sign within).

The next column is a list of methods used to calculate the current display. The column can be made wider with the "curtain pull" icon: 

Pull this to the right to make the left panel wider.

The middle panel is that of the graphical display. Each method used for calculating the display is listed on the right hand panel.

7.3 Recalculating graphics (changing defaults)

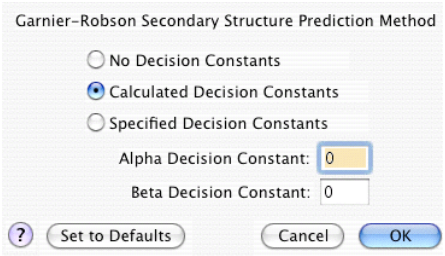
INFO

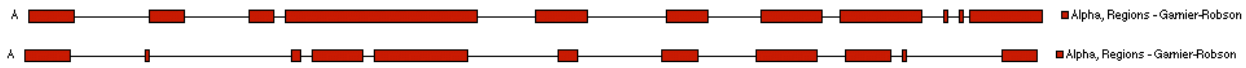
Each graphical plot is an object. Double clicking on one such object calls a dialog panel that displays the defaults used to calculate this plot; at this point parameters can changed and the plot recalculate.

For example the first plot depicts the prediction of alpha elements by the method of Garnier-Robson. The prediction is based on a "decision constant" based on the percentage of alpha and beta structures that are either known by other methods, or that could be inferred. The algorithm calculates a value for this decision constant, which can affect the results.

TASK

Double click on the first panel, titled **Alpha, Regions – Garnier-Robson**
This opens the calculation panel.
Change to "**No Decision Constants**" and **Press OK**
Note the change that occurs in the display of the predicted alpha helices as shown below (top is original default, bottom is recalculated)



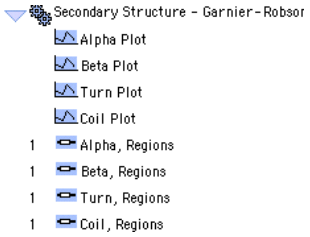


Note: Constants are based on three protein classes: with less than 20%, 20-50% and >50% alpha or beta sheet. Circular dichroïsm data is an experimental method that can help estimate the alpha-helical content of a structure. If such data is know, use the "Specified Decision Constants" option and enter values for both alpha and beta.

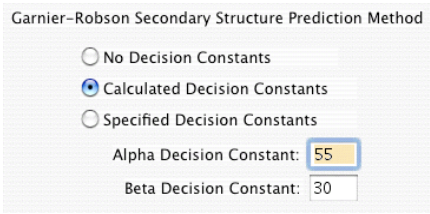
7.4 Recalculating options from the methods panel (left)

✔ TASK

On the left panel, **click** on the **triangle** next to "Secondary Structure - Garnier-Robson." This will display subset items.
Double-click on **Alpha Plot**. This will open the same recalculation panel, this time with the calculated decision constants filled in.

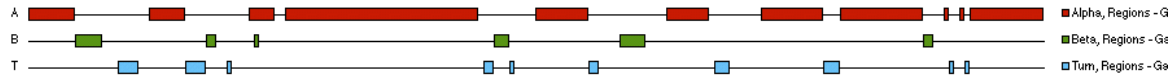


Click on "Specified Decision Constants"
Enter 55 and **30** for Alpha and Beta respectively
Click OK

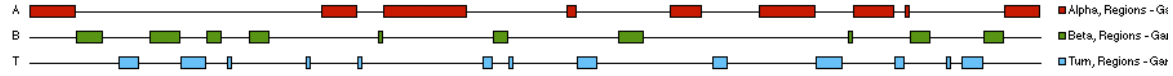


Note that changes occur for all the Garnier-Robson objects, as depicted below

Original default:

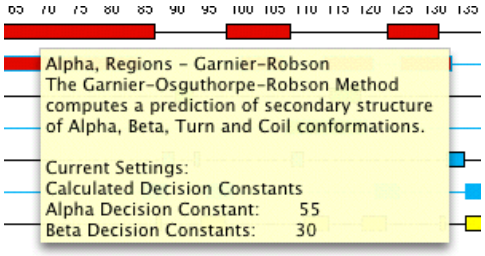


Recalculated



✔ TASK

Place the mouse cursor above the Alpha, Regions – Garnier-Robson and **wait a few seconds**. A yellow window will reveal the nature of the plot and some of its parameters, such as the Decision Constants, here 55 and 30 respectively for alpha and beta.



7.5 Reorganizing the panel plots

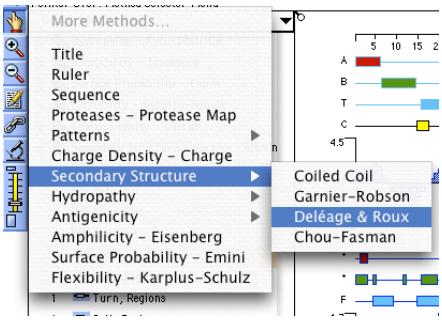
As each plot is an independent object, it can be dragged and moved within the graphics panel.

✔ TASK

As an exercise **move all** the Garnier-Robson panels and place them in the same order as in the image above.

7.6 More methods

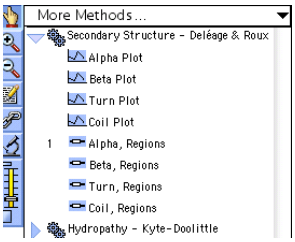
For clarity and simplicity, only a limited number of plots (based on specific methods such as Garnier-Robson) are shown at first.



TASK

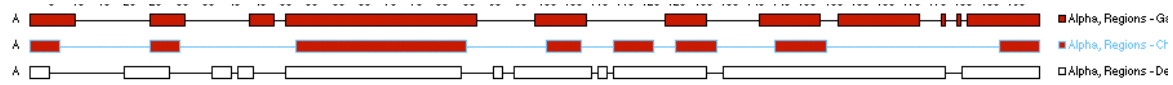
Other methods or objects can be added from the More Methods… pull down menu. For example, **select** the **Secondary Structure > Deléage & Roux** method.

This will create a new entry on the left panel called "Secondary Structure – Deléage & Roux"



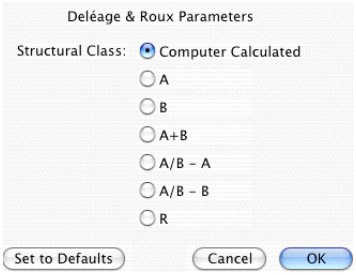
Click on the triangle at left to expand the visible sub-contents.

Click on the **Alpha, Regions** button and drag it onto the graphics panel. A new alpha helical prediction plot is displayed. You can rearrange the panels order and place all three alpha helix structure prediction objects in a row as shown below:



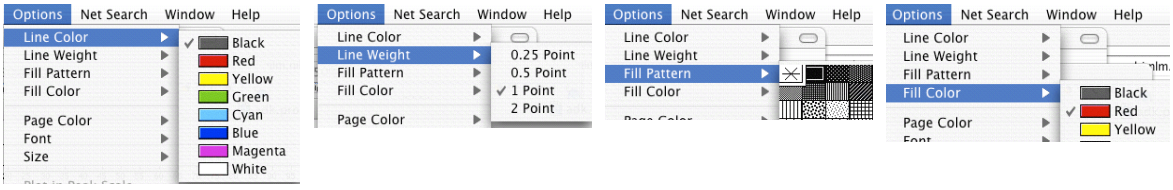
Note: as before the calculation methods can be altered by double clicking on Alpha, Regions icon.

Here the calculation method is based on a structure class of protein, such as alternate alpha/beta topology.



7.7 Changing colors and patterns of plots

The other 2 helical plots have a red filled look. These cosmetic changes can be changed from the Options menus and submenus:

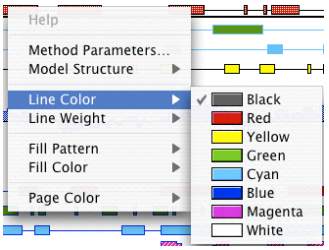


The default pattern (*) is empty and should be switched to completely filled (black square) or patterned designs.

To make this new panel look like the other two, follow the following menu cascade sequences:

✔ **TASK** **Click on** the new Deléage & Roux Alpha, Regions panel object
Options > Fill Pattern > black-square
Options > Fill Color > Red

Note: Graphical interfaces are usually very redundant. All the options menus are also available by Ctrl-Clicking on the object, to open a contextual menu.

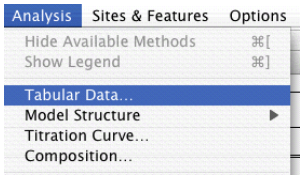


Other menus are also present in addition to the Options

7.8 Tabular Data

The graphical panel presents an overview of the computed data. However the pertinent information calculated for each residue can be displayed in a table. Only the objects present on the graphics panel are shown on the tabular panel.

✔ **TASK**
To display the data in table form, select the following menu cascade:
Analysis > Tabular Data…



Tabular Data for adk.pro

Colors														
Min		.	.	.	.	.	.	.	.	-1.93	.	.	.	-0.60
Max		A	A	A	B	B	T	T	C	2.68	*	*	F	3.10
Mean		0.619	0.443	0.768	0.0928	0.294	0.113	0.247	0.18	0.47	0.459	0.479	0.608	0.83
Res	Pos	Gami... Alpha	Chou-... Alpha	Delea... Alpha	Gami... Beta	Chou-... Beta	Gami... Turn	Chou-... Turn	Gami... Coil	Kyte-... Hydro...	Eisen... Alpha	Eisen... Beta	Karpl... Flexi...	James... Antig...
Met	1	A	A	A	.	.	.	.	.	0.58	.	*	.	0.75
Glu	2	A	A	A	.	.	.	.	.	1.01	*	*	.	0.75
Glu	3	A	A	A	.	.	.	.	.	1.44	*	*	.	0.75
Lys	4	A	A	A	.	.	.	.	.	1.52	.	.	.	0.75
Leu	5	A	A	.	.	.	.	.	.	1.96	.	*	F	0.90
Lys	6	A	A	.	.	.	.	.	.	1.67	.	.	F	0.90
Lys	7	A	.	.	.	B	.	.	.	0.78	.	.	F	0.90
Thr	8	A	.	.	.	B	.	.	.	0.08	.	.	F	0.60
Lys	9	A	.	.	.	B	.	.	.	-0.82	.	.	F	0.45
Ile	10	.	.	.	B	B	.	.	.	-0.87	.	.	.	-0.30
Ile	11	.	.	.	B	B	.	.	.	-1.26	.	.	.	-0.60
Phe	12	.	.	.	B	B	.	.	.	-1.64	.	*	.	-0.60
Val	13	.	.	.	B	B	.	.	.	-1.54	.	.	.	-0.60
Val	14	.	.	.	B	B	.	.	.	-1.93	.	.	.	-0.60
Gly	15	.	.	.	.	.	.	.	C	-1.34	.	.	F	-0.05
Gly	16	.	.	.	.	.	.	T	C	-0.80	.	*	F	0.45
Pro	17	.	.	.	.	.	.	T	C	-0.06	.	*	F	0.45
Gly	18	.	.	.	.	.	T	T	.	0.46	.	*	F	1.25
Ser	19	.	.	A	.	.	T	T	.	1.00	.	*	F	1.55
Gly	20	.	.	A	.	.	T	T	.	1.34	.	.	F	1.85
Lys	21	.	.	A	.	.	T	T	.	1.02	.	.	F	2.30
Gly	22	.	.	A	.	.	.	T	C	1.23	.	.	F	2.25
Thr	23	.	.	A	.	.	.	T	C	1.62	.	*	F	3.00
Gln	24	A	A	A	.	.	.	.	.	1.03	.	.	F	2.10
Cys	25	A	A	A	.	.	.	.	.	0.52	.	.	F	1.35

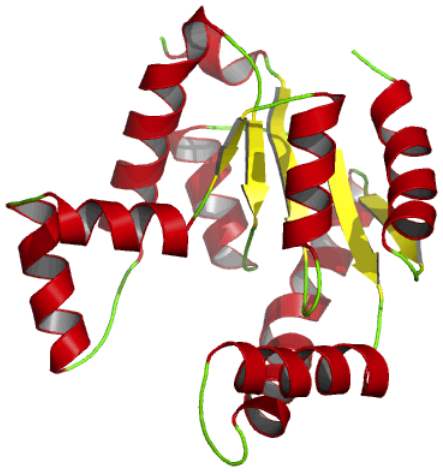
INFO

Note: One or more column can be selected and then pasted within a spreadsheet such as Microsoft Excel for further analysis. When the columns are copied and pasted within Excel, the first 2 columns (A and B) of the new spreadsheet are automatically populated with the residue name (A) and sequence position number (B) for convenience. On this example the antogenic and surface index (last 2 columns of the Tabular data) are shown when pasted into Excel.

	A	B	C	D
1	Minimum		-0.6	0.05
2	Maximum		3.1	4.54
3	Mean		0.83	1.18
4	Residue	Position	Antigenic Ind	Surface Probabilit
5	Met	1	0.75	2.22
6	Glu	2	0.75	1.43
7	Glu	3	0.75	2.24
8	Lys	4	0.75	4.54
9	Leu	5	0.9	3.78
10	Lys	6	0.9	4.36
11	Lys	7	0.9	1.53
12	Thr	8	0.6	1.3
13	Lys	9	0.45	0.56
14	Ile	10	-0.3	0.21
15	Ile	11	-0.6	0.11
16	Phe	12	-0.6	0.05

7.9 Reality check

TASK (optional)



It happens that this protein has been crystallized and is therefore a known 3D structure with PDB ID code `3ADK`.

If you have time, you can use the program PyMOL to "fetch" entry 3ADK and represent it as a cartoon structure:

Names Panel: **3adk > S > Cartoon**

colored by secondary structure:

Names **Panel: 3adk > C > ss > choose** option with red helix

You can then display the sequence (Top menu **Display > Sequence On**) to more easily see the amino acids involved in alpha helices.

From the tabular data of Protean (top menu **Analysis > Tabular Data…**), you can also see which individual amino acids are predicted as alpha helices by various methods. It is therefore possible to see how well some helices are predicted and even count them.

8. Protean 3D

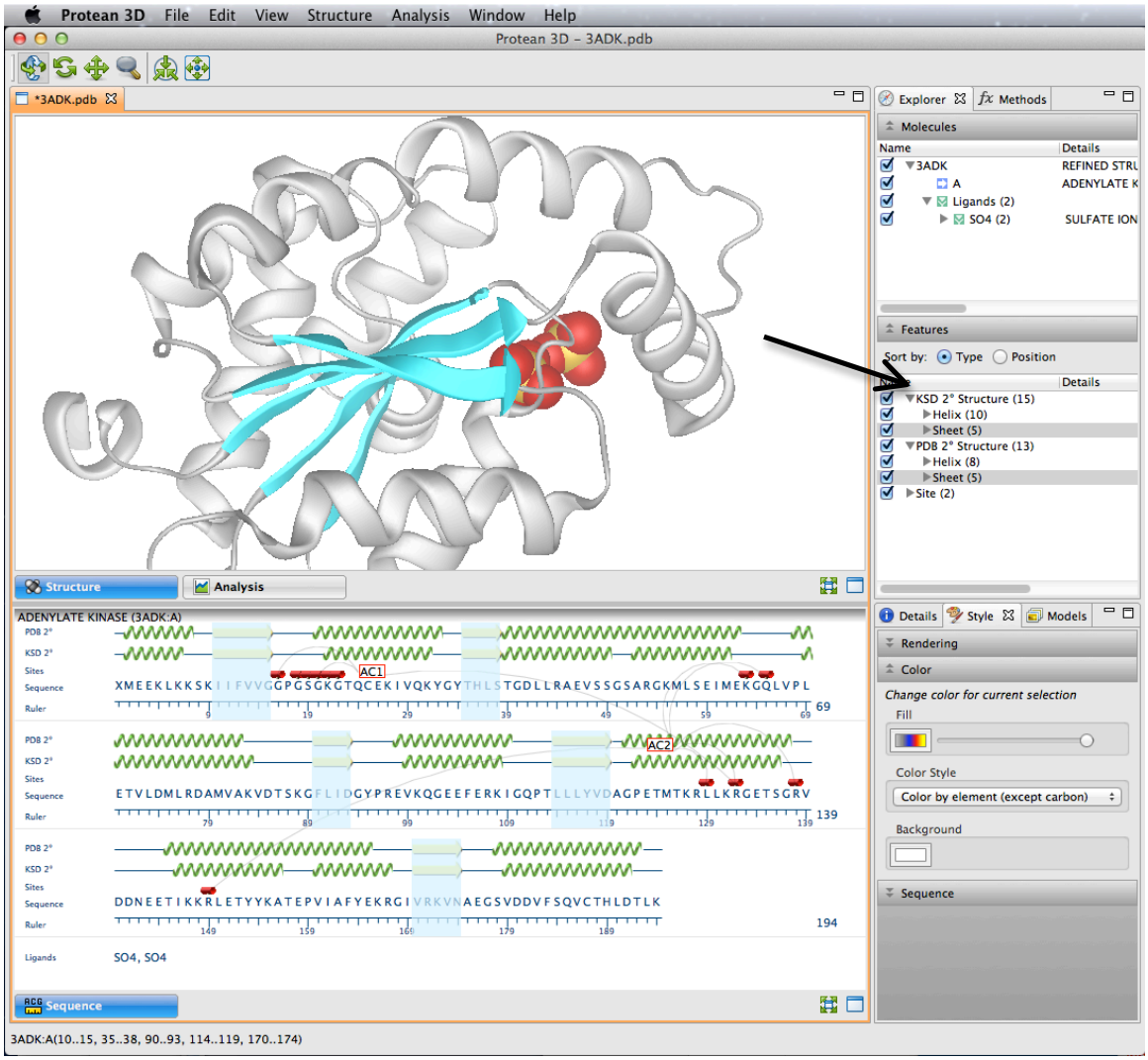
Protean3D is a new addition to Lasergene for protein analysis that includes 3D (*i.e.* PDB files) and variations of the various "methods" explored within Protean.

8.1 Try Protean3D

Optionally you can explore the 3ADK structure with Protean3D that you will find with the other Lasergene modules in **/Applications/DNASTAR/Lasergene 12/**

With this version it is possible to open the file directly from the Protein Data Bank web site (**www.rcsb.org)** with the menu cascade **File > Open From PDB…**

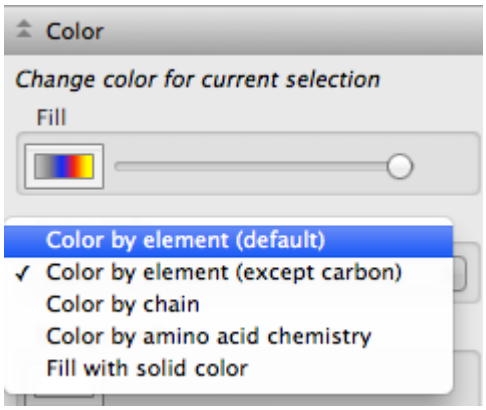
Protean 3D will open the file and automatically show the protein as ribbons and any solvent as spheres.

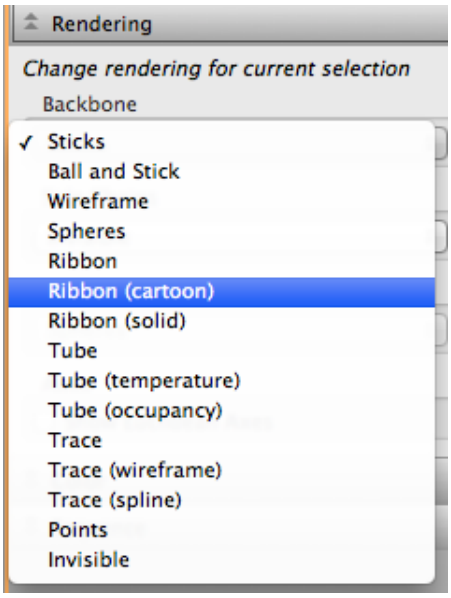


In the middle panel to the right the "Features" section will list the secondary structure found within the PDB file as well as in the KSD[1] secondary structure database (see arrow in image.)

On the image above it was chosen to select the Beta sheets from both offered methods. This has the effect to color the sheets in cyan-blue within the 3D viewer as well as the bottom panel that illustrates in cartoon form the secondary structure.

If you have time you can explore other aspects of Protean3D such as the "Rendering" and "Color" panels.





Here the background was changed from the default black to white for printing purposes.

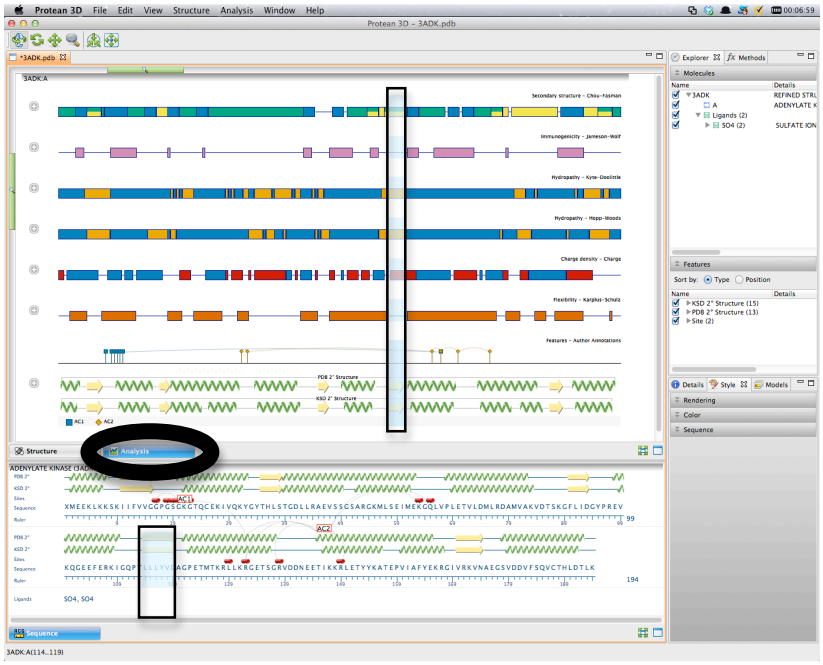
8.2 Analysis panel

All or most of the methods available in Protean are also present in Protean3D and visible when clicking on the Analysis button (circled in image below) next to the Structure button in the middle pane

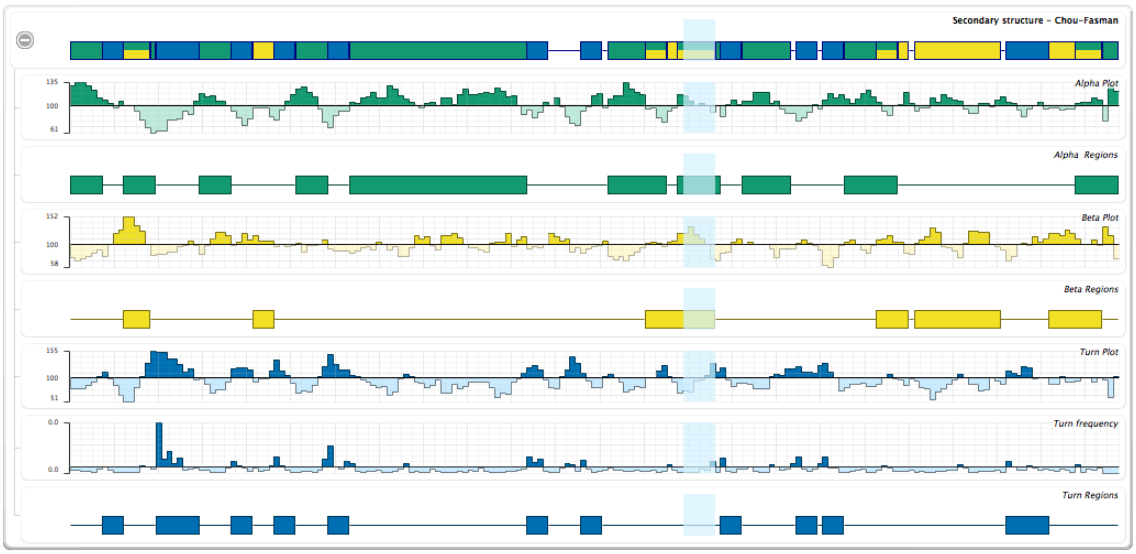
[1] **Kinase Sequence Database** is a collection of protein kinase sequences grouped into families by homology of their catalytic domains. http://sequoia.ucsf.edu/ksd/

Example: click on one beta-sheet arrow within the Sequence panel (the bottom panel) for example around amino acid residue 114-119.

Once clicked the same area will be highlighted on all methods (enhanced with black outline here.)



Clicking on the ⊕ sign at the left will open each method to full display, for example with Chou and Fasman here, where the blue highlight is also shown.



To close click on the ⊖ sign.

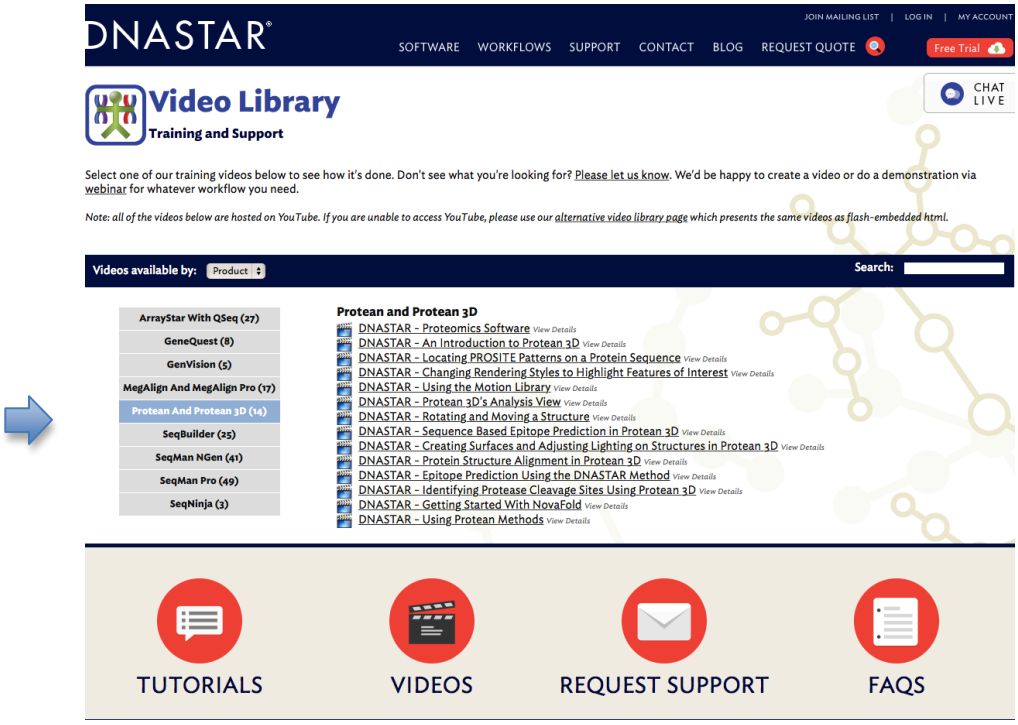
8.3 Learn more from DNASTAR tutorials and videos

To learn more about Protean3D DNASTAR provides tutorials online including video tutorials! The web Protean3D page is at:
`http://www.dnastar.com/t-protean-3d.aspx`

LASERGENE®
Protean 3D™

From the web page:
"Protean 3D is Lasergene's application for exploring macromolecular structure, motion, and function. Rich, synchronized graphical views allow you to see the 3D molecular structure, the annotated sequence, and the analysis of applied prediction methods simultaneously, enabling easy identification and analysis of secondary structure elements. Protean 3D also provides access to the Motion Library, where you can browse and search over 300 animated and annotated macromolecular conformational changes, and to NovaFold, which allows you to predict three-dimensional structures for protein sequences."

Video tutorials are at: `http://www.dnastar.com/t-protean-videos.aspx`



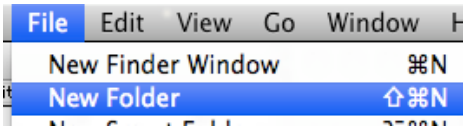
9. Quit All Programs

1) Quit Protean and/or PyMOL
2) Close all windows.

L02 – Supplemental: Primer Design

1. L02 - Setup:

✔ **TASK** If it does not exist yet, create a folder named L02



BACKGROUND

Primers are small, oligonucletotides complementary to a specific location of a nucleic acid strand. Primers are used widely in PCR (Polymerase Chain Reaction) to amplify specific regions of nucleic acids, for example genomic DNA fragments or mRNA molecules. In this case primers are designed in pairs. For the generation and subsequent cloning cDNA fragments primers can be designed as a single sequence.

Primer design is one of the most critical parameter in designing PCR experiments as most of the outcome of the experiment depends on many of the primer's properties. Here are some critical or important parameters to consider in the design of primers:

Primer length: oligonucleotides between 18 and 24 bases in length are the most sequence specific (at the optimal annealing temperature). If the primer is too long the annealing is less efficient.

Annealing temperature: It has been determined empirically that a good annealing temperature is approximately 5°C lower than the melting temperature. (Tm).

Melting Temperature (Tm) : a PCR reaction uses 2 primers. Both primers should have a similar melting temperature. If the Tm are too dissimilar the high Tm primer will mis-prime at lower temperatures while the low Tm primer may not be efficient at the higher temperatures. Nearest neighbor thermodynamic calculations is the most accurate method to calculate the melting temperatures of primers. However in the 18 to 24 length range the formula Tm = 2(AT) + 4(GC) can be used as an approximation. A melting temperature of 55°C - 72°C gives the best results.

Specificity: depends in part on the primer length. Repetitive elements in the DNA sequence may be responsible for random annealing resulting in a smear on the gel. Short 3' complementary stretches may cause non-specific annealing at lower temperatures.

Complementarity within primer sequences: Primers should not contain stretches of more than 3 complementary base pairs in a row, as this would result in "snap back" configurations: partially double-stranded structures of the primer folding back onto itself. Similarly, sequence similarity should be limited between the 2 primers used in a reaction to avoid primer dimers .

G/C content, homopolymeric, (T, C) and (A, G) stretches: Ideally the chosen primers will have a near random sequence with between 45 to 55% in GC content. PolyC and polyG promote non-specific annealing while polyA and polyT stretches can facilitate "breathing" after annealing. Similarly polypyrimidine (T, C) and polypurine (A, G) stretches should be avoided.

3'-end sequence: a G or C residue should be included at the 3' end of primers since a "GC Clamp" helps correct binding and minimizes breathing.

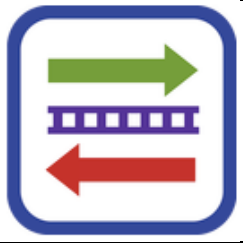
- Summary Design Rules -

1. Primers should be 17-28 bases in length;
2. Base composition should be 50-60% (G+C);
3. Primers should end (3') in a G or C, or CG or GC: this prevents "breathing" of ends and increases efficiency of priming;
4. Tms between 55-80 ºC are preferred;
5. 3'-ends of primers should not be complementary (ie. base pair), as otherwise primer dimers will be synthesised preferentially to any other product;
6. Primer self-complementarity (ability to form 2o structures such as hairpins) should be avoided;
7. Runs of three or more Cs or Gs at the 3'-ends of primers may promote mispriming at G or C-rich sequences (because of stability of annealing), and should be avoided.

Adapted from: Innis, M.A. and D.H. Gelfand. 1990. Optimization of PCRs. In PCR protocols: A guide to methods and applications (ed. M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White), pp. 3–12. Academic Press, San Diego, CA.-See
http://bioweb.uwlax.edu/GenWeb/Molecular/Seq_Anal/Primer_Design/primer_design.htm#designrules
or short URL: http://bit.ly/1jN2U9v

2. DNASTAR PrimerSelect

DNA* Lasergene 12 modules are located within
Macintosh HD > Applications > DNASTAR > Lasergene 12

The DNA* Lasergene module for primer design is PrimerSelect The program can be used for PCR primer selection, sequencing projects (primer walking) or to introduce mutations.	

2.1 Launch PrimerSelect

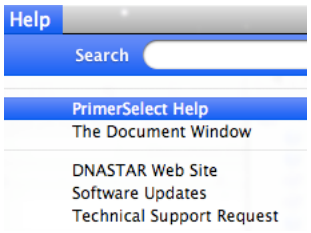
✔ **TASK**

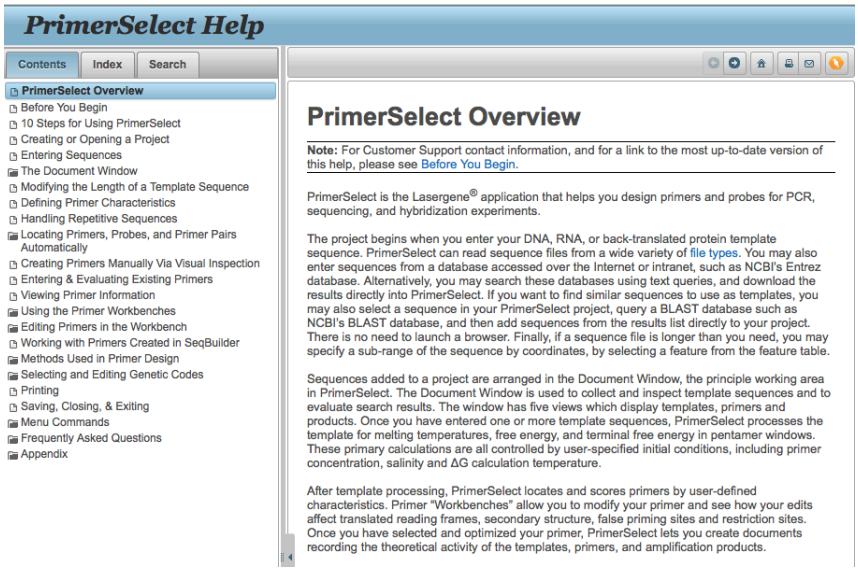
- **Double-click** the PrimerSelect icon

2.1.1 PrimerSelect Help

✔ **INFO**

Help is available from the Help menu bar:





Or online at `https://www.dnastar.com/primerselect_help/`

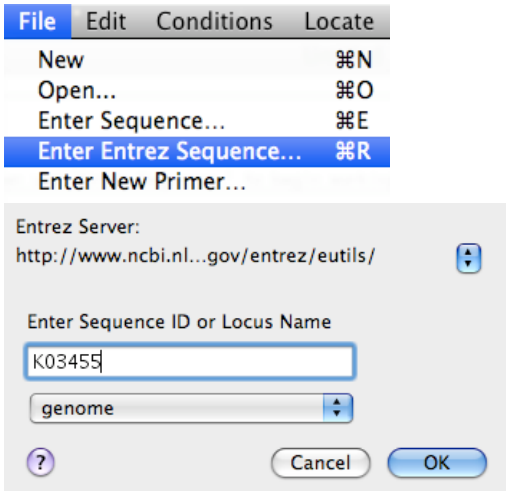
2.2 Importing a sequence

We will select primers flanking the region of the HIV-1 envelope poly-protein in order to amplify it. The same primer pairs could then be used systematically on various isolates to collect and compare sequences.

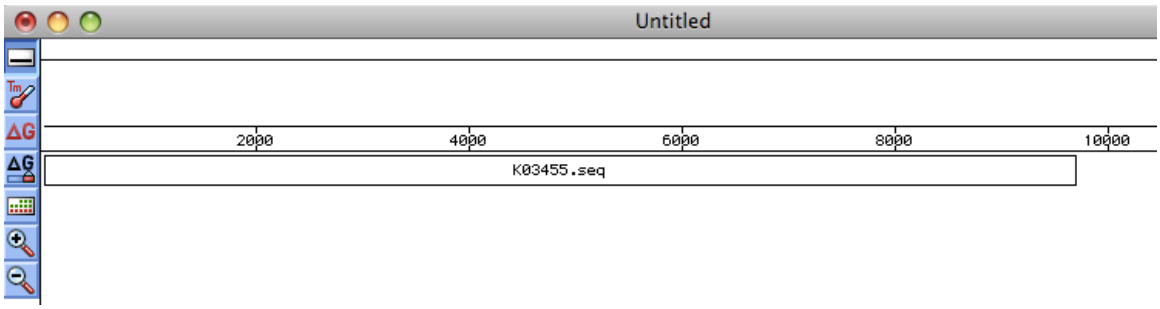
Within this sequence the coding region for the envelope poly-protein is from 6225 to 8795.

✔ **TASK**

- Within the **File** menu select the option **Open Entrez Sequence…** or use the key-board menu shortcut **⌘R**
- Within the new window **enter** the sequence accession code **K03455**
- This is a complete **genome**.
- **Save** the file within the **L02** directory creat-ed above
- **Click OK** and save as file **K03455.seq**



The new sequence will be shown as a simple long rectangle, shorter than 10,000 bases long within a new, untitled window:



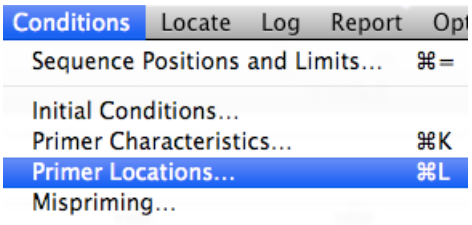
Various icons at the left of the PrimerSelect window are switches to display various properties of the template sequence, such as theTm over a window range ( ) or free energy plots (e.g.). The default is the schematic presentation, toggled with the button at the top.

2.3 Limiting the Annealing Region

Allow approximately 200 bases as a priming area:

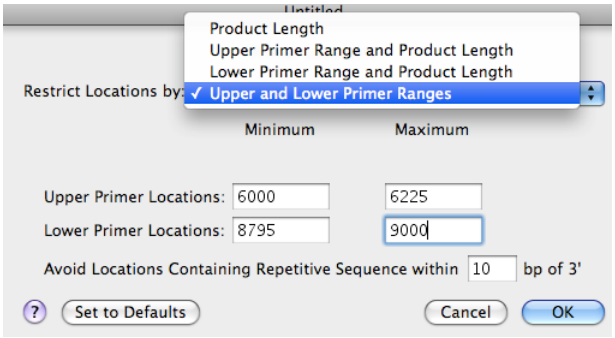
TASK

• Select **Conditions > Primer Locations...** or choose the shortcut **⌘L**

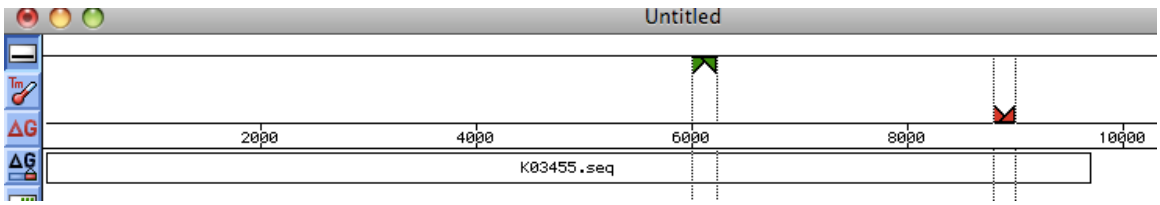


• From the pull-down menu at the top of the dialog window select the "**Upper and Lower Primer Ranges**" and enter 6000 to 6225 for the Upper Primer Locations and 8795 to 9000 for the Lower Primer Location.

• **Click OK**



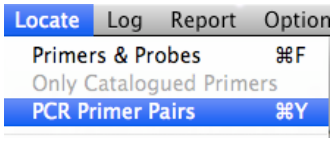
Note that triangular marks are now displayed within the schematics presentation

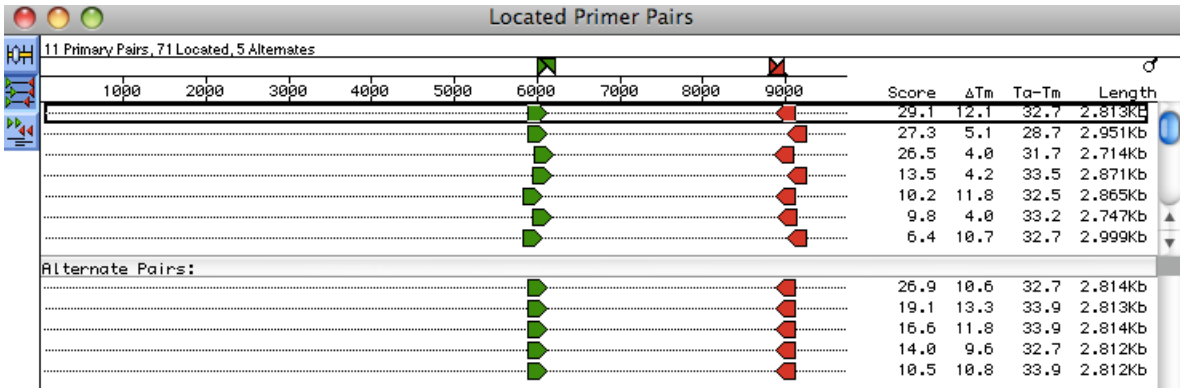


2.4 Calculating primer pairs:

TASK

• Select Menu Bar: Locate > PCR Primer Pairs
The results appear in a separate window.





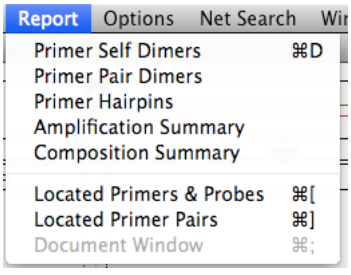
• On the <u>original "Untitled" window</u> **Click on** the **sequence view** icon  on the left side

• This will present the sequence, in split windows

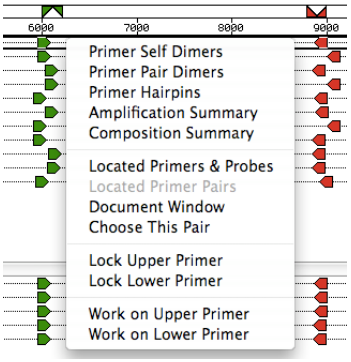
2.5 Evaluating primer pairs: menu "Report".

INFO

A report on primer pairs can be called on either from the menu bar "**Report**" or using the keyboard shortcut **⌘D**.



Another method is to **right-click** on the primer to be studied.

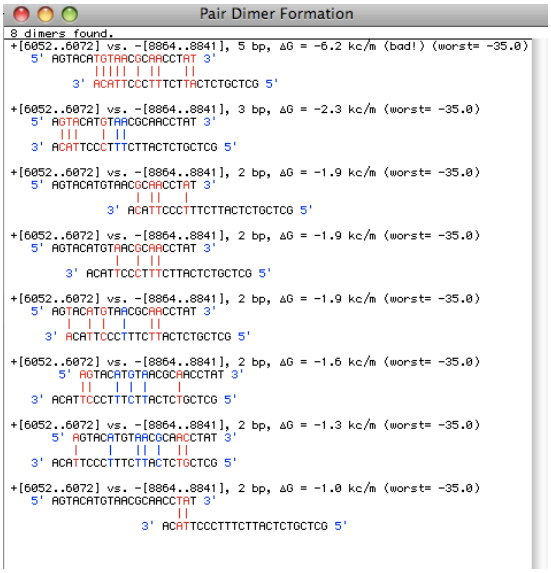


We will use the first pair at the top of the list as the print-out examples.

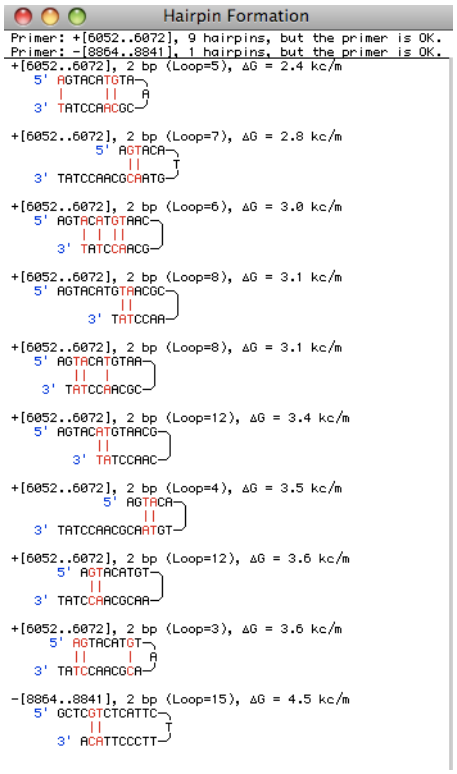
Using either method to report (Report menu or right-click) **explore the 5 menu options**:

Primer Self Dimers

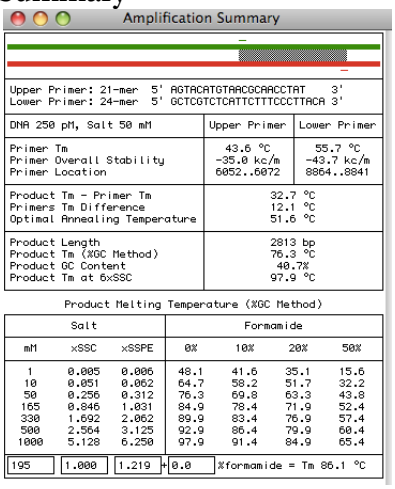
Report > Primer Pair Dimers



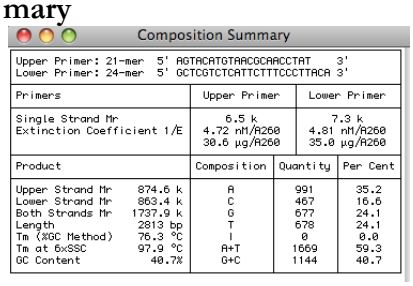
• **Menu Bar: Report Primer Hairpins**



• **Menu Bar: Report > Amplification Summary**



Menu Bar: Report > Composition Summary

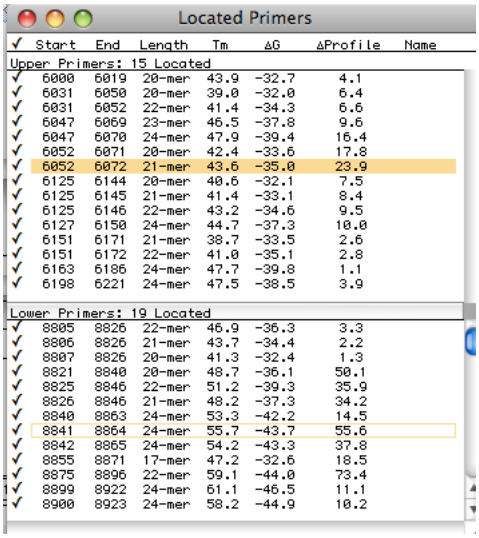


2.6 Locating Primers

✔ **TASK**

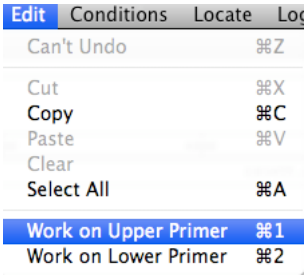
Primers location along the sequence can be reviewed with the either the menu cascade **Report > Located Primer Pairs** or with a **right-click > Located Primers & Probes** on top of a primer within the "Located Primer Pairs" graphical window as illustrated above.

The primer pairs selected within the "Located Primer Pairs" graphical window will appear highlighted within the resulting report window.



2.7 Upper and lower primers

It is possible to specify whether you want to work specifically with one of the primers within a primer pair with the **Edit** menu bar options: **Work on Upper Primer** and **Work on Lower Primer.** These 2 options can also be engaged from the right-click menu shown above.



✔ **TASK**

Engage working with the upper primer of the first line with e.g. the menu cascade

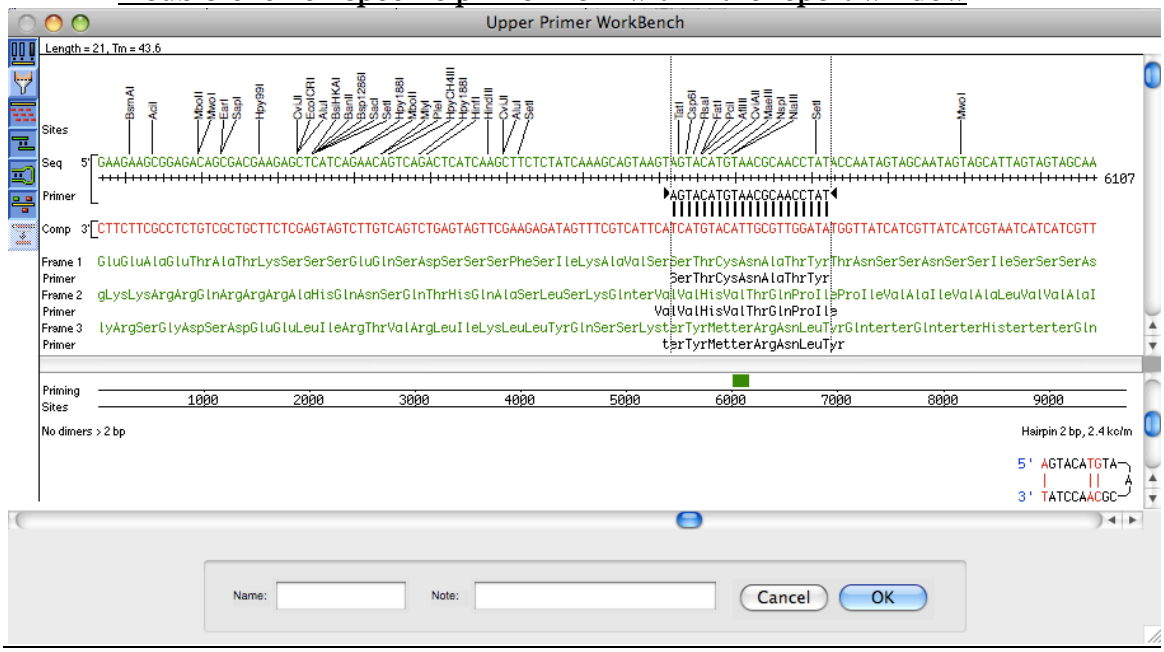
Edit > Work on Upper Primer

Or the keyboard shortcut **⌘1**

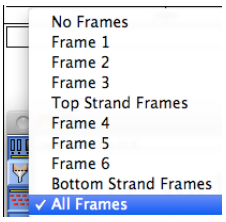
This will pop up yet another graphical window labeled "Upper Primer Workbench" showing the primer bound to the sequence.

NOTE: To work on UPPER primer only:

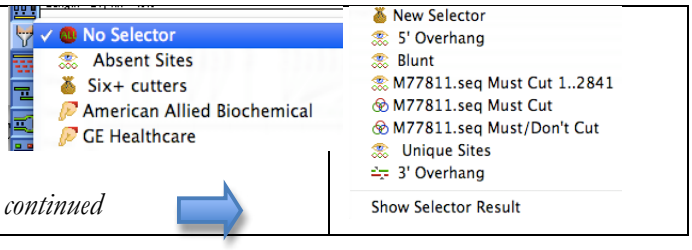
- **Double-click on specific primer from within the report window**



The buttons at left are used to show or hide the translation frames all of them shown by default.



The enzyme restriction sites are also shown by default and can be altered with the selector button  with commercial enzyme options at the top.



3. Designing Primers in SeqBuilder

Acknowledgment: this section is adapted directly from the guide "Getting Started with DNASTAR® Lasergene® - For Macintosh® and Windows® - Version 8.0"

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:

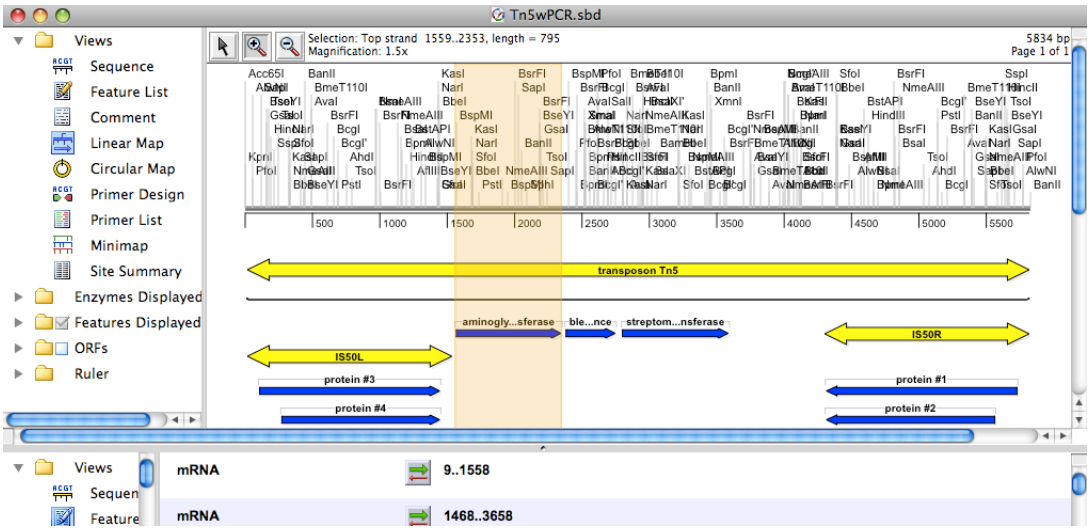
Hard Drive > Applications > DNASTAR > Lasergene 12 > Demo Data

3.1 Creating and Modifying Primer Pairs for a Region of Interest

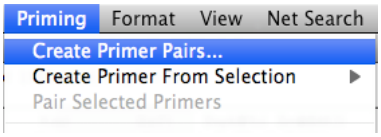
Objective: To create a primer pair for a region of interest on your sequence.

✔ **TASK**

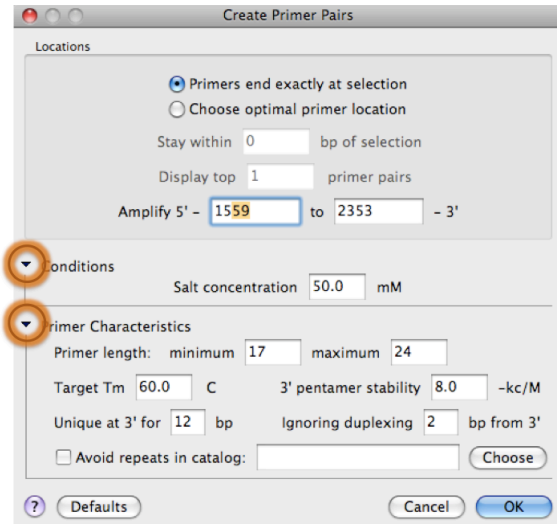
- **Launch SeqBuilder**.
- Select **File>Open** to navigate to the Demo Data folder and choose `Tn5wPCR.sbd`. The sequence opens in the Linear Map view.
- Choose the neomycin/kanamycin resistance feature labeled aminoglycoside-3'-O-phosphotransferase as the region of interest by single clicking on that feature. The sequence range 1559-2353 will be selected.



- Select menu **Priming>Create Primer Pairs**. The Create Primer Pairs dialog appears.



- Click the triangles next to **Conditions** and **Primer Characteristics** to expand those sections so that the dialog appears as follows:

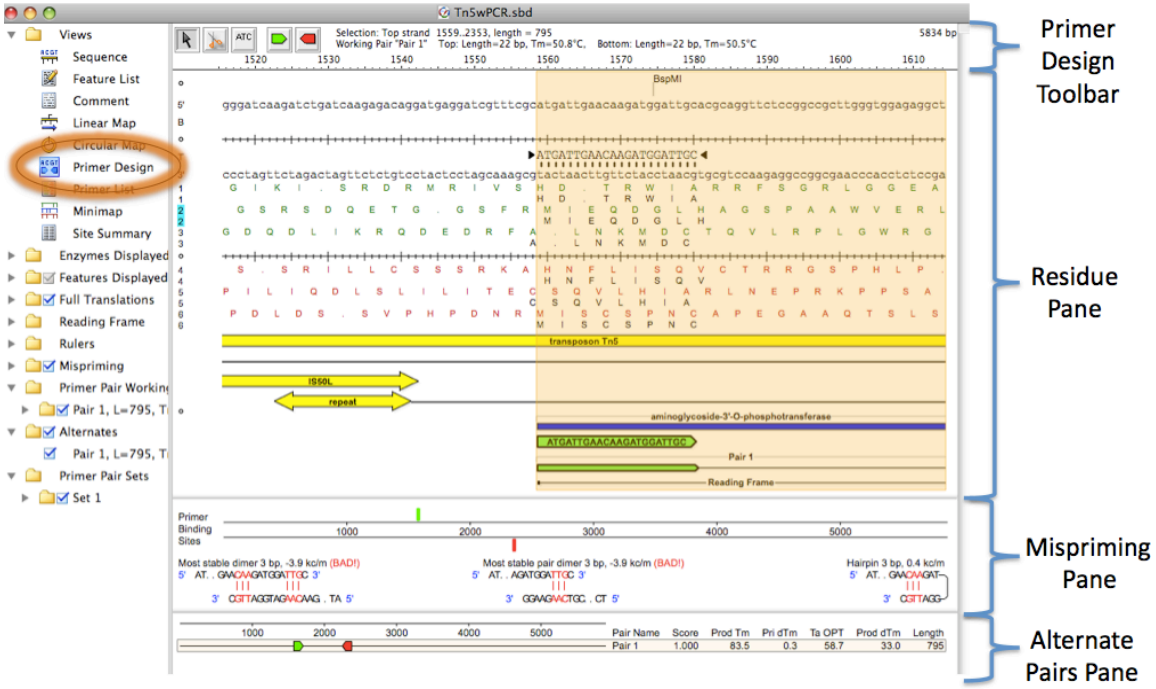


✔ **INFO**: The **Locations** section of the Create Primer Pairs dialog allows you to specify parameters that define where SeqBuilder will search for primers on your sequence. The **Conditions** section allows you to specify the initial salt concentration, which affects the calculation of predicted melting 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.

- **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 speci-

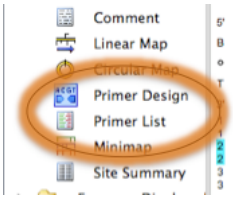
fied that lies exactly within the selected region and display it in the Primer Design View. By default, the view focuses on the top strand primer first.

- 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:

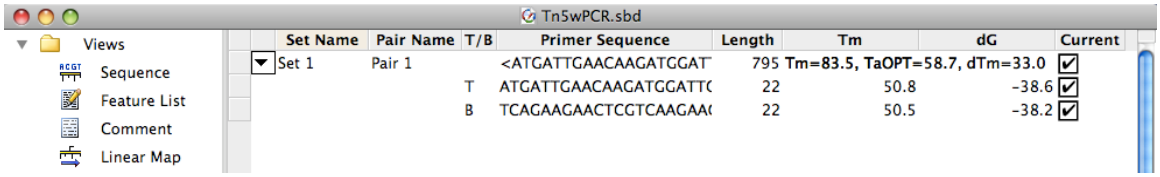


- Notice in the Mispriming Pane, the **Most stable 261-mer** and **Most stable pair dimer** are <u>both</u> labeled "BAD!". This indicates that the final pentamer value of the dimer exceeds the 3' pentamer stability threshold defined in the Primer Characteristics section of the Create Primer Pairs dialog. 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.

- **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.

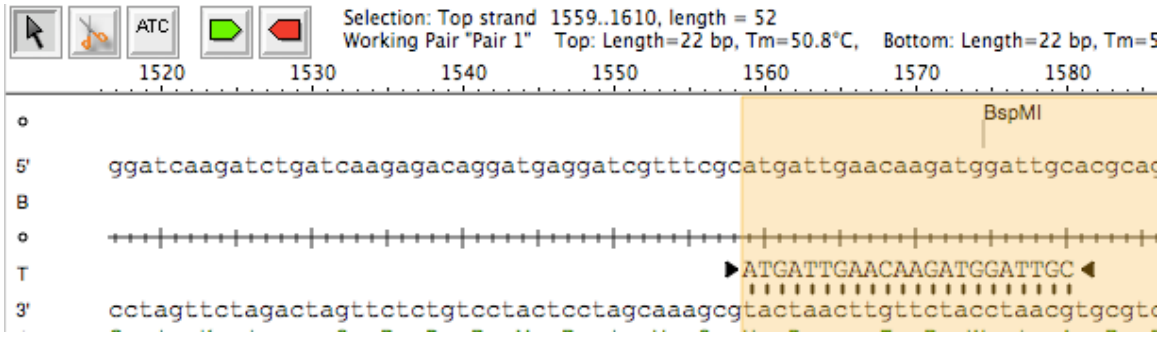


- Click the triangle next to Set 1 to expand the primer pair and view each individual primer within the pair:

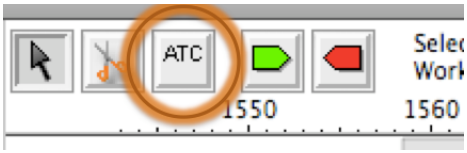


- Notice that the melting temperatures for the individual primers are both close to 50° C, which is lower than our Target Tm of 55°C.

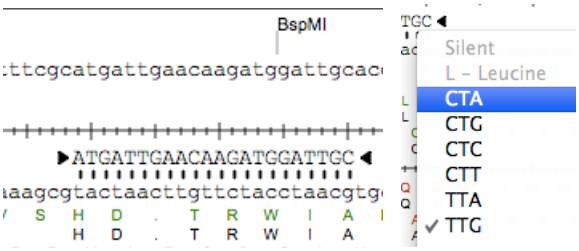
- Switch back to the Primer Design view to view the top strand primer:



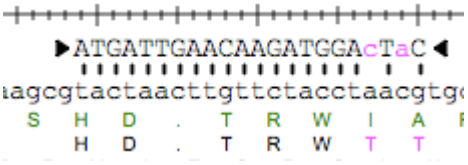
- Select the codon change mode by clicking on the ATC button in the Primer Design Toolbar. Notice that your cursor changes and that each codon is highlighted as you hover over the sequence.



- 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**.

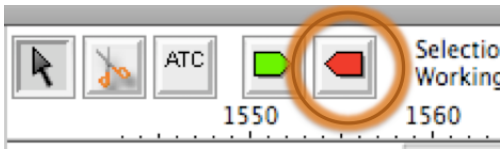


- Notice that changes to the primer sequence and translation are shown in magenta.

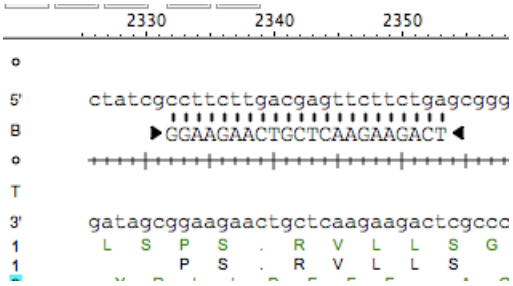


- Notice also that the Mispriming Pane has now updated to show that there is no longer a stable dimer conformation.

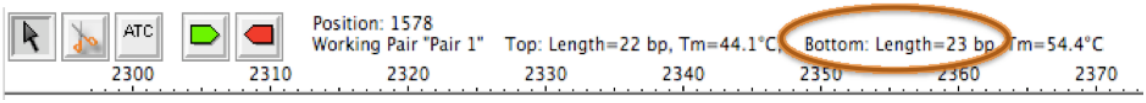
- Display the bottom strand primer by **clicking the**  **button** from the Primer Design Toolbar.



- The Primer Design View scrolls to show the bottom strand primer in the center of the view:

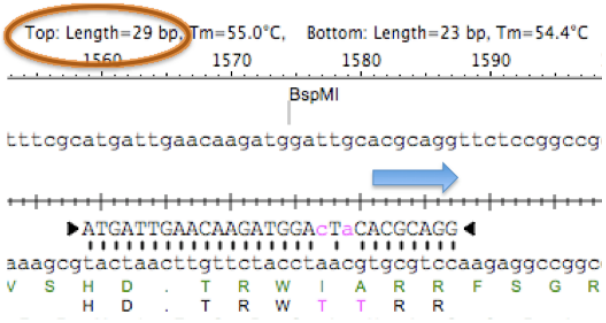


- **Increase the length** of the bottom strand primer from 22 bp **to 23 bp** by clicking and dragging the triangle shown on the **left (3') end** of the primer sequence. The length of the primer is displayed and updated automatically in the Primer Design Toolbar shown at the top of the view:



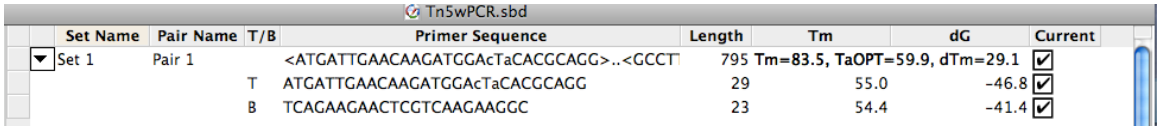
- Notice that once the primer length is increased to 23 bp, the melting temperature for the bottom strand primer increased to 54.4°C, which is much closer to our Target Tm. Also notice that the **Most stable pair dimer** has been updated in the Misprimming Pane and no longer has the indication "BAD!"

- **Switch back to the top strand primer** by clicking on the  button. Increase the length of the top strand primer from 22 bp to 29 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 55.0°C.

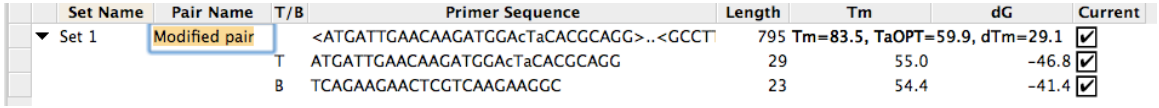


- Now that the primers have been modified to our satisfaction, change back to the Primer List view by selecting it from the curtain on the left.

- 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.



- Name the modified pair by clicking on Pair 1 in the Pair Name column and typing Modified pair.



4. Primer3, free web-based application

Acknowledgment: Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.
`http://jura.wi.mit.edu/rozen/papers/rozen-and-skaletsky-2000-primer3.pdf`
`Archive link: http://bit.ly/1j5Nj48`

Note: Primer3 is open source: `http://primer3.sourceforge.net/`

Objective: create a primer based on a free web resource

4.1 Launch web browser

✔ **TASK**: Open ***one*** the following URL:

`http://frodo.wi.mit.edu/`
Above URL will switch to: `http://bioinfo.ut.ee/primer3-0.4.0/`
`http://fokker.wi.mit.edu/primer3/input.htm`
Above URL will switch to: `http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm`
`http://biotools.umassmed.edu/bioapps/primer3_www.cgi`

4.2 Enter data

✔ **READ** As input sequence you can use for the HIV-1 reference genome you retrieved in a previous exercises. Simply open EditSeq to retrieve the version saved earlier with the accession number: K03455, and use the Copy/Paste method.

As a target sequence we will this time focus on the viral polymerase (*pol*) which spans nucleotides 2358 to 5096.

✔ TASK

- **Open K03455.seq** with EditSeq
- Select and **Copy** sequence in the clipboard
- Switch to the web browser
- Under **Sequence ID** enter **K03455** or another ID meaningful to you
- Under **Targets** enter **2358,200 5096,200**
  (Enter numbers on the same line, separated by blank space. The represent the surrounding regions for the target primers around the coding region of *pol*)

Note: This selection is different than the one we used in DNASTAR PrimerSelect. This selection will create two 200 length targets, one at the beginning and one at the end of the coding region of *pol*.

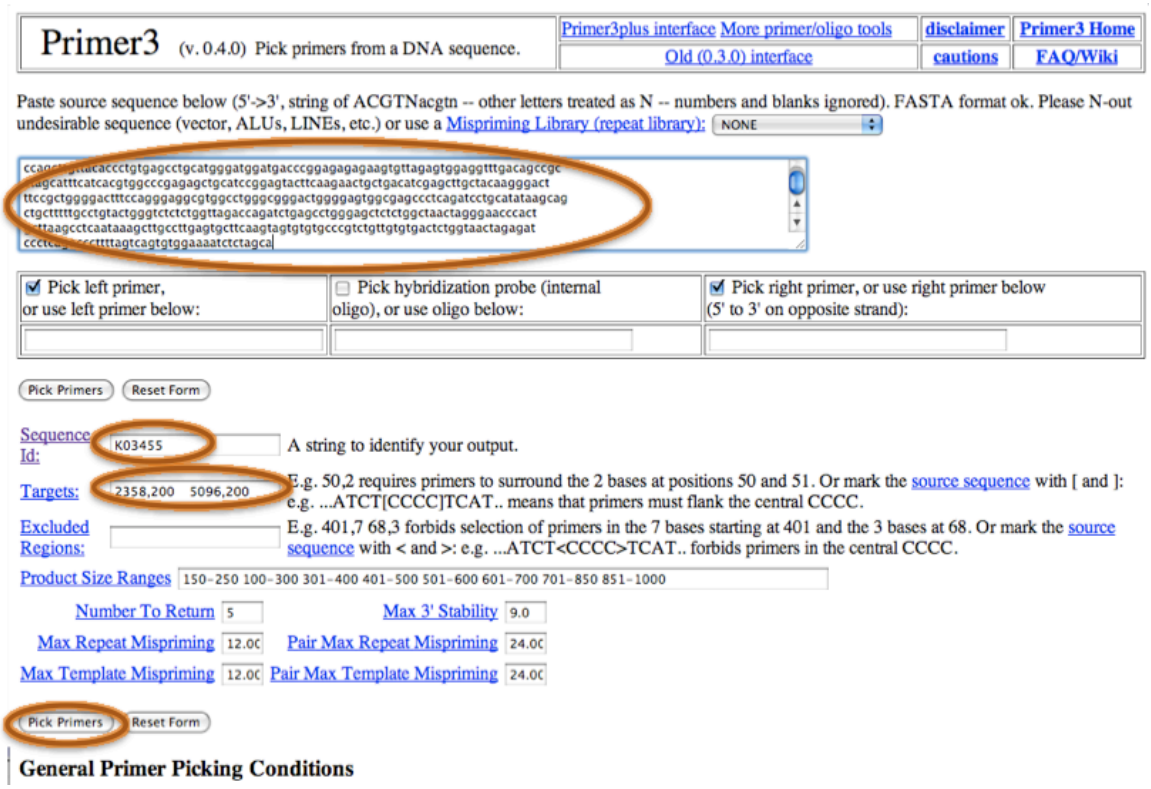
✔ INFO

We will leave the other optional areas within the page such as

General Primer Picking Conditions
Other Per-Sequence Inputs
Objective Function Penalty Weights for Primers
etc.

with their respective default values which are sometimes blanks.

The top portion of your browser should look similar to the following:



TASK

- **Click** on the "**Pick Primers**" button

A new window will appear summarizing results.

- **Explore the proposed data**.

If you get result for only one of the region, try each Target region separately: 2358,200 alone and then 5096,200 alone.

>>>>>> and <<<<<< mark the primer sequence on the result page and ******* mark the product

Example output:

Primer3 Output

```
PRIMER PICKING RESULTS FOR K03455

No mispriming library specified
Using 1-based sequence positions
OLIGO            start  len       tm      gc%   any   3' seq
LEFT PRIMER       5058   20    59.97    50.00  3.00  0.00 ggtgatgattgtgtggcaag
RIGHT PRIMER      5354   21    58.95    52.38  4.00  1.00 tggtctgctagttcagggtct
SEQUENCE SIZE: 9719
INCLUDED REGION SIZE: 9719

PRODUCT SIZE: 297, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00
TARGETS (start, len)*: 2358,200 5096,200

    1 tggaagggctaattcactcccaacgaagacaagatatccttgatctgtggatctaccaca


 2341 atacagtattagaagaaatgagtttgccaggaagatggaaaccaaaaatgataggggaa
                      ****************************************

 2401 ttggaggttttatcaaagtaagacagtatgatcagatactcatagaaatctgtggacata
      ************************************************************

 2461 aagctataggtacagtattagtaggacctacacctgtcaacataattggaagaaatctgt
      ************************************************************

 2521 tgactcagattggttgcactttaaattttcccattagccctattgagactgtaccagtaa
      **********************************

 2581 aattaaagccaggaatggatggcccaaaagttaaacaatggccattgacagaagaaaaaa


5041 atggaaaacagatggcaggtgatgattgtgtggcaagtagacaggatgaggattagaaca
                      >>>>>>>>>>>>>>>>>>>>                 *****

 5101 tggaaaagtttagtaaaacaccatatgtatgtttcagggaaagctaggggatggttttat
      ************************************************************

 5161 agacatcactatgaaagccctcatccaagaataagttcagaagtacacatcccactaggg
      ************************************************************
```

```
 5221 gatgctagattggtaataacaacatattggggtctgcatacaggagaaagagactggcat
      ************************************************************

 5281 ttgggtcagggagtctccatagaatggaggaaaaagagatatagcacacaagtagaccct
      ***************                                       <<<<<<<

 5341 gaactagcagaccaactaattcatctgtattactttgactgtttttcagactctgctata
      <<<<<<<<<<<<<

 5401 agaaaggccttattaggacacatagttagccctaggtgtgaatatcaagcaggacataac
```

```
KEYS (in order of precedence):
****** target
>>>>>> left primer
<<<<<< right primer

ADDITIONAL OLIGOS
                    start  len      tm     gc%   any   3' seq

 1 LEFT PRIMER       5058   20   59.97   50.00  3.00  0.00 ggtgatgattgtgtggcaag
   RIGHT PRIMER      5355   22   60.30   50.00  4.00  1.00 ttggtctgctagttcagggtct
   PRODUCT SIZE: 298, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00

 2 LEFT PRIMER       5058   20   59.97   50.00  3.00  0.00 ggtgatgattgtgtggcaag
   RIGHT PRIMER      5354   22   59.00   50.00  4.00  2.00 tggtctgctagttcagggtcta
   PRODUCT SIZE: 297, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

 3 LEFT PRIMER       5058   20   59.97   50.00  3.00  0.00 ggtgatgattgtgtggcaag
   RIGHT PRIMER      5354   23   59.82   52.17  4.00  2.00 tggtctgctagttcagggtctac
   PRODUCT SIZE: 297, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00

 4 LEFT PRIMER       5071   21   60.26   52.38  3.00  0.00 tggcaagtagacaggatgagg
   RIGHT PRIMER      5354   21   58.95   52.38  4.00  1.00 tggtctgctagttcagggtct
   PRODUCT SIZE: 284, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 1.00

Statistics
         con   too    in    in          no    tm    tm  high  high        high
         sid  many   tar  excl   bad    GC   too   too   any    3'  poly   end
        ered    Ns   get   reg   GC% clamp   low  high compl compl     X  stab    ok
Left   42479     0   226     0   995     0 18788 10805     0     0   297   616 10752
Right  62569     0   223     0  1723     0 29856 13976     0    12   257   770 15752
Pair Stats:
considered 180834628, no target 1335657, unacceptable product size 179498950, high end
compl 3, ok 18
primer3 release 1.1.0


(primer3_results.cgi 0.4.0 modified for WI)
```

5. DNASTAR Resources and Support

Support for DNASTAR Lasergene is at
`http://www.dnastar.com/t-support-training.aspx` :
This is a repository of tutorials on all DNASTAR software and not only Lasergene.

The "**Getting Started with DNASTAR® Lasergene® - For Macintosh® and Windows® - Version 8.0**" publication is available online, the direct link to the PDF file (*GettingStartedGuide8.0.pdf*) is on an FTP server. Quick link: `http://bit.ly/1suGSLx`

6. Ending session

- **Quit** all programs
- **Close** all Macintosh windows.
- **Move** to the trash the files you created today

END OF LABORATORY

Class notes

http://www.gutenberg /images/ornament_ix.jpg
http://www.gutenber h/images/image_38.jpg