Latest Update: July 2014
Edited By: Talia McKay

ProMOL User Guide for Windows, Mac and Linux

Development Team:

Rochester Institute of Technology:

Eno Akpovwa, Majied Albalwi, Nicole Arroyo, Cameron Baker, Abdul Bangura, Daniel Bobo, Paule M. Boli, Sean Bourne, Chantelle Brown, Michael Bryan, Cyprian W. Corwin, Dr. Paul A. Craig, Greg Dodge, Luticha Doucette, Chanelle Francis, Alexander Grier, Brett Hanson, Kaitlin Hart, Amanda Hartung, Katrina Henry, Alison Horn, Carly Keenan, MaryEd Kenney, Haeja Kessler, Venkata Aditya Kovuri, Madolyn MacDonald, Shariq Madha, Desiree Mattews, Talia McKay, Dr. Jeffrey Mills, Scott Mottarella, Mikhail Osipovitch, Tananda Richards, Mario Rosa, Weinishet Tedla-Boyd, Charles Westin

Dowling College:

Mogjan Asadi, Isaac Awuah Asimah, Kethi Bardhi, Kostandina Bardhi, Dr. Herbert J. Bernstein, Darina Boycheva, Ricky Chachra, Clarice Chigbo, Georgi Darakev, Nikolay Darakev*, Damian Glinojecki, Armela Hasa, Jonathan Ihm, John Jemilawon, Nan Jia, Kedian Jimenez, Mia Jurjivec, Petko Kamburov, Maria Karakasheva, Barry LaPierre, Ming Li, Stavros Louris, Thomas Matchett, Gregory McQuillan, Kostadin Mitec, Daniel O'Brien, Stojan Regodic, Limone Rosa, Matt Rousseau, Paul Sussman, Georgi Todorov, Rohit Tripathi, Peter Zhivkov, Elena Zlateva

Funding Sources:

NSF-DUE 0402408	NIGMS 1R15GM078077
NIGMS 2R15GM078077-01	NIGMS 2R15GM078077-02
NIGMS 3R15GM078077-01S	NIGMS 3R15GM078077-02
Rochester Institute of Technology	Dowling College

*Brett Hanson and Charles Westin are credited as the original designers of ProMOL. They created it in 2006 as undergraduate students studying Biochemistry and Molecular Biology at RIT.

*The creation of ProMOL distribution kits was originated by Nikolay Darakev, who created the PyMOL MINGW windows build process in 2008 (see http://sourceforge.net/apps/mediawiki/sbevsl/index.php?title-PyMOL_MINGW_Build_Instructions).

Table of Contents

Introduction:

ProMOL is a plugin for the visual molecular graphics environment PyMOL. Together, these programs are used to identify potential enzyme motifs. ProMOL includes a library of active site residues for nearly 1000 enzymes. The program compares these active site residues to the sequences of proteins that are inputted by the user.

ProMOL includes the abilities to:

- make and store motif templates in a growing library of existing motif templates
- view optimal alignment between a motif template and a query structure
- calculate the RMSD and Levenshtein distance between the query protein and its motif hits
- request alignment with subsets of the motifs by template source and Enzyme Commission number

ProMOL is written in Python and was developed for several different operating systems, including Windows (XP, 7, 8), Macintosh OS X (including Lion), and Unix/Linux. There are only minor differences in the appearance or operation of ProMOL among the three operating systems.

Download and Installation:

Download and Install Python:

Python must be installed before installing PyMOL and ProMOL. The best available versions of Python that are compatible with ProMOL include versions 2.5, 2.6, and 2.7. All updates of Python after and including version 3.0 are not recommended; these versions are incompatible with the current ProMOL code.

You can download Python at:
https://www.python.org/downloads/

Download and Install PyMOL*:

Before ProMOL can be installed, PyMOL must first be downloaded. This program is compatible with Windows, Macintosh, and Unix/Linux systems. PyMOL is available as an open source program available as a source kit** from:
http://sourceforge.net/projects/pymol/

A proprietary version of PyMOL can also be downloaded directly from:
http://pymol.org

Source and binary kits of PyMOL, including instructions for installation, can be found in the following Wiki pages:
Windows: http://pymolwiki.org/index.php/Windows_Install
Linux: http://pymolwiki.org/index.php/Linux_Install
Macintosh*:** http://pymolwiki.org/index.php/MAC_Install

***Note:** "PyMOL releases 1.3 through 1.7 have been tested with ProMOL. PyMOL 1.7 is recommended.
****Note:** If you are using a computer with Windows 8 and are planning to install PyMOL from sourceforge.net, you may experience some difficulties. When you run the program, a small window will pop up warning you that the program is "Unknown." Select "More Information" and then "Run Anyway."
*****Note:** The PyMOL Wiki page outlines the process for installing PyMOL using a package installation manager (Fink, MacPorts, Homebrew).

Download and Install ProMOL:

ProMOL is available as a compressed format from: http://sourceforge.net/projects/sbevsl/files/ProMOL, or from www.promol.org

You can choose between downloading a ".zip," ".tar.gz," or ".exe" file, depending on your preferences and operating system. Furthermore, you should download the latest version of ProMOL available on the site.

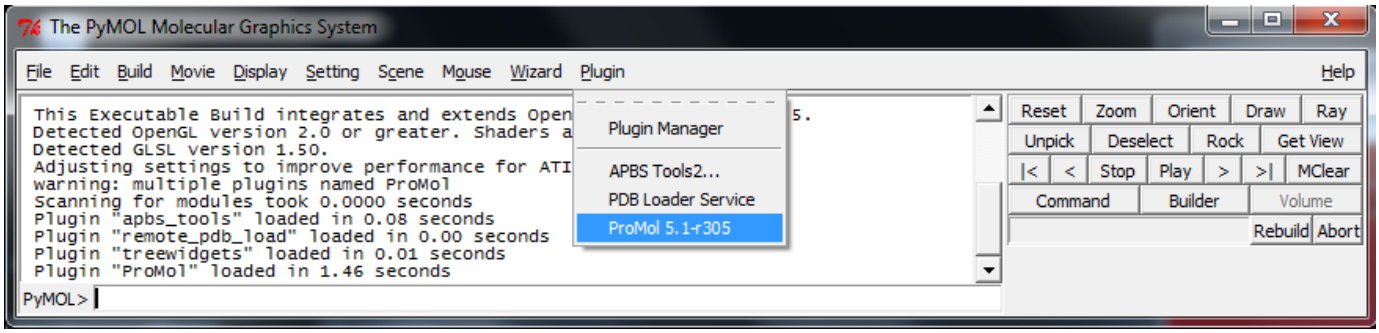

Figure 1[1]**.** PyMOL Molecular Graphics System Window

Once you have completed the installation of ProMOL, launch PyMOL and look for ProMOL in the Plugin menu (Figure 1).

[1] All images were captured on a Windows 7 operating system. Please note that the PyMOL and ProMOL interfaces are visually very similar on Windows, Linux, and MAC systems, and that general operation procedures outlined in this guide are the same for all operating systems, unless otherwise noted.

- **Windows*:**

a.

- click to download the .exe file
- install ProMOL 5 setup
- the destination folder should be located here:
  C:\Python27\Lib\site-packages\pmg_tk\startup

b.

- click on either the .zip or .exe file to download
- open PyMOL
- in “The PyMOL Molecular Graphics System” window, open the dropdown “Plugin” menu, and select “Plugin Manager” (see **Figure 1**)
- click the tab “Install New Plugin”
- under the header “Install from Local File” click “Choose File...”
- navigate to the following location:
  C:\Python27\Lib\site-packages\pmg_tk\startup
- click on ProMOL.py, the “Open” then “Yes” and “Ok”
- on your computer, go to the Start menu, open the Computer Folder, and navigate to:
  C:\Python27\Lib\site-packages\pmg_tk\startup
- copy the “ProMol” and “Treewidgets” folders, and “ProMol.py”
- in the same window, navigate to:
  C:\Users\”username”\AppData\Roaming\pymol\startup
- paste the “ProMol” and “Treewidgets” folders and “ProMol.py” here
- completely close, the re-open PyMOL
- To open ProMOL, click on the “Plugin” dropdown menu in “The PyMOL Molecular Graphics System” window. If you downloaded the latest version, you should now have an option for “ProMOL 5.3 r-360.” Click to open.

***Note:** New and improved versions of ProMOL are released frequently. To update ProMOL on your device, simply download and open the .zip folder of the version that you wish to install. Copy and paste “ProMOL.py” and the folders “ProMOL” and “Treewidgets” into the following destination to **replace** the old documents:

C:\Users\”username”\AppData\Roaming\pymol\startup

- **Unix or Linux**

  - expand the compressed file and unpack the tarball with:
    gunzip < sbevsl-ProMol-4-1.tar.gz | tar xvf
  - open PyMOL
  - in “The PyMOL Molecular Graphics System” window, open the dropdown “Plugin” menu, and select “Plugin Manager” (see Figure 1)
  - click the tab “Install New Plugin”
  - under the header “Install from Local File” click “Choose File...”
  - navigate to the following location:
    C:\Python27\Lib\site-packages\pmg_tk\startup
  - click on ProMOL.py, the “Open” then “Yes” and “Ok”
  - on your computer, go to the Start menu, open the Computer Folder, and navigate to:
    C:\Python27\Lib\site-packages\pmg_tk\startup
  - copy the “ProMol” and “Treewidgets” folders, and “ProMol.py”
  - in the same window, navigate to:
    C:\usr\lib\pymodules\python2.6\pmg_tk\startup
  - paste the “ProMol” and “Treewidgets” folders and “ProMol.py” here
  - completely close, then re-open PyMOL
  - To open ProMOL, click on the “Plugin” dropdown menu in “The PyMOL Molecular Graphics System” window. If you downloaded the latest version, you should now have an option for “ProMOL 5.3 r-360.” Click to open.

- **Mac OS X**:

  - refer to the note in “Download and Install PyMOL” (above) for installation of PyMOL without using a package installation manager
  - refer to the Wiki page for proper PyMOL installation of Mac OS X (see the section “Download and Install PyMOL,” above) using Fink, MacPorts, or Homebrew

a. Installation with MacPyMOLX11Hybrid.app

- “MacPyMOLX11Hybrid.app" from PyMOL 1.6 and earlier will work with ProMOL under older versions of Mac OS X, but is not a recommended combination under OS X 10.9 or when when using PyMOL 1.7. Use the package manager installation below, instead.
- expand the .tar.gz or the .zip file
- Plugins that consist of a single Python file (e.g. promol.py) can be installed by using the Plugin\Manage Plugins\Install... menu in “The PyMOL Molecular Graphics System” window (see Figure 1)
- open Finder\Applications
  - If you have a three button mouse, right click on MacPyMOLX11Hybrid.app and select “Show Package Contents.” This will take you to the directory tree for PyMOL.
  - If you do not have a three button mouse, the keyboard shortcut Command-Click will bring up the “Show Package Contents” option.
- traverse the tree to:
  pymol/modules/pmg_tk/startup
- copy “promol.py” and the “ProMOL” folder
- open Finder\Applications and navigate to the “startup” folder in PyMOL
- paste “promol.py” and “ProMOL” to the “startup” folder
- completely close, the re-open PyMOL
- To open ProMOL, click on the “Plugin” dropdown menu in “The PyMOL Molecular Graphics System” window. If you downloaded the latest version, you should now have an option for “ProMOL 5.3 r-360.” Click to open.

b. Installation Using a Package Installation Manager

- expand the .tar.gz or the .zip file
- Plugins that consist of a single Python file (e.g. promol.py) can be installed by using the Plugin\Manage Plugins\Install... menu in “The PyMOL Molecular Graphics System” window (see Figure 1)
-within the expanded file, copy “promol.py” and the “ProMOL” folder
- find the portion of the PyMOL installation tree that contains:
    /sw/lib/pymol-py##/modules/pmg_tk/startup
-paste “promol.py” and the “ProMOL” folder into this startup folder
-completely close, then re-open PyMOL
-To open ProMOL, click on the “Plugin” dropdown menu in “The PyMOL Molecular Graphics System” window. If you downloaded the latest version, you should now have an option for “ProMOL 5.3 r-360.” Click to open.

Components:

The ProMOL interface has 4 buttons (Open PDB, Fetch PDB, Random PDB, Clear) along the base of the GUI, in addition to 5 tabs (Welcome, EZ-VIZ, Motif Finder, Motif Maker, View Options) along the top of the window (Figure 2).

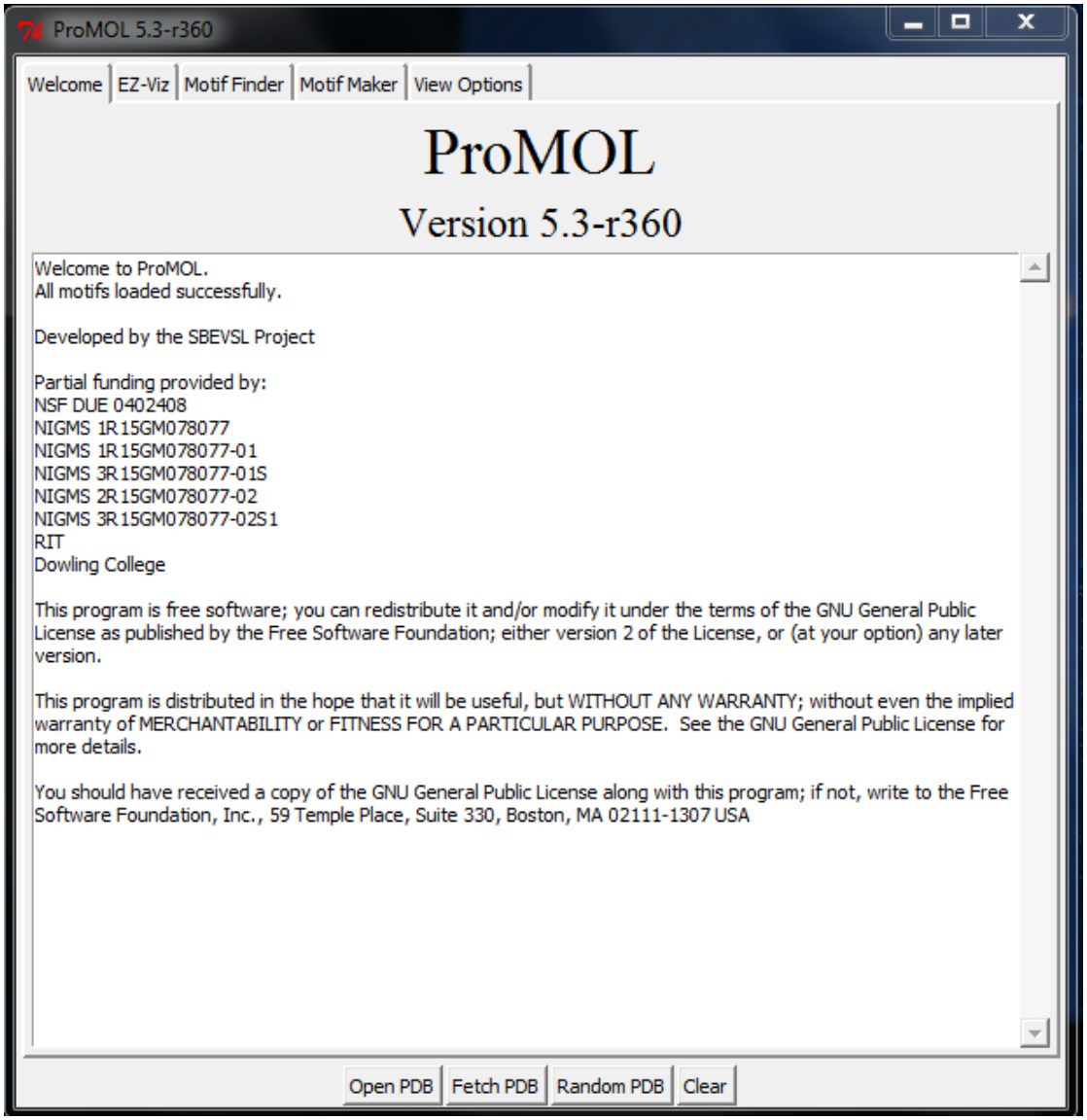

Figure 2. ProMOL GUI

Buttons on the ProMOL GUI

The options on the bottom of the ProMOL window are always available when ProMOL is open. The four buttons allow the user to locate and open Protein Data Bank files that are saved on the user's computer or from the Internet.

Figure 3. Four buttons that appear at the base of the ProMOL GUI.

- **Open PDB**
  Selecting this option opens a dialog box that will let you search for PDB files in storage devices on your computer

- **Fetch PDB**
  This button opens a dialog box that allows you to open a PDB file from the Internet. A plugin window titled "PDB Loader Service" will open, prompting for a 4-digit PDB protein identification code.

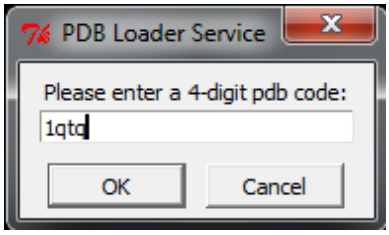

Figure 4. PDB Loader Service Window

  When a valid 4-digit PDB protein identification code is entered, the corresponding 3-D structure will be available for visual analysis in the "PyMOL Viewer" window (see Figure 7, page 14).

- **Random PDB**
  Selecting this option will load a PDB file into ProMOL from the Internet at random. The randomly selected protein will be visible in the "PyMOL Viewer" window (Figure 7, page 14). This is a useful tool to use when designing and testing motifs; it can assist in the identification of true and false positives and negatives.

- **Clear**
  This button will remove all structures that are visible in the PyMOL molecular viewing area.

Tabs on the ProMOL GUI

The options at the top of the ProMOL GUI are always available when ProMOL is open. These tabs allow the user to manipulate and interact with and compare complete protein structures and/or active site residues.

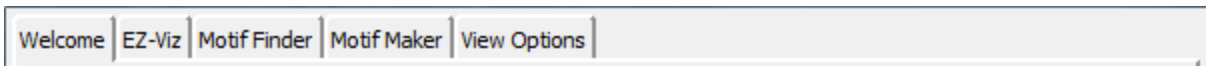

Figure 5. Tabs on the ProMOL GUI

- **Welcome Tab**
  The contents under this tab are visible by default when the ProMOL GUI is opened. It summarizes funding and licensing information (see Figure 2, page 9).

- **EZ-Viz Tab**
  After a protein is loaded into the program, this tab enables users to visually manipulate the protein structure. The user simply uses the available drop-down menus to select the desired viewing options that will automatically change in the "PyMOL Viewer" window.

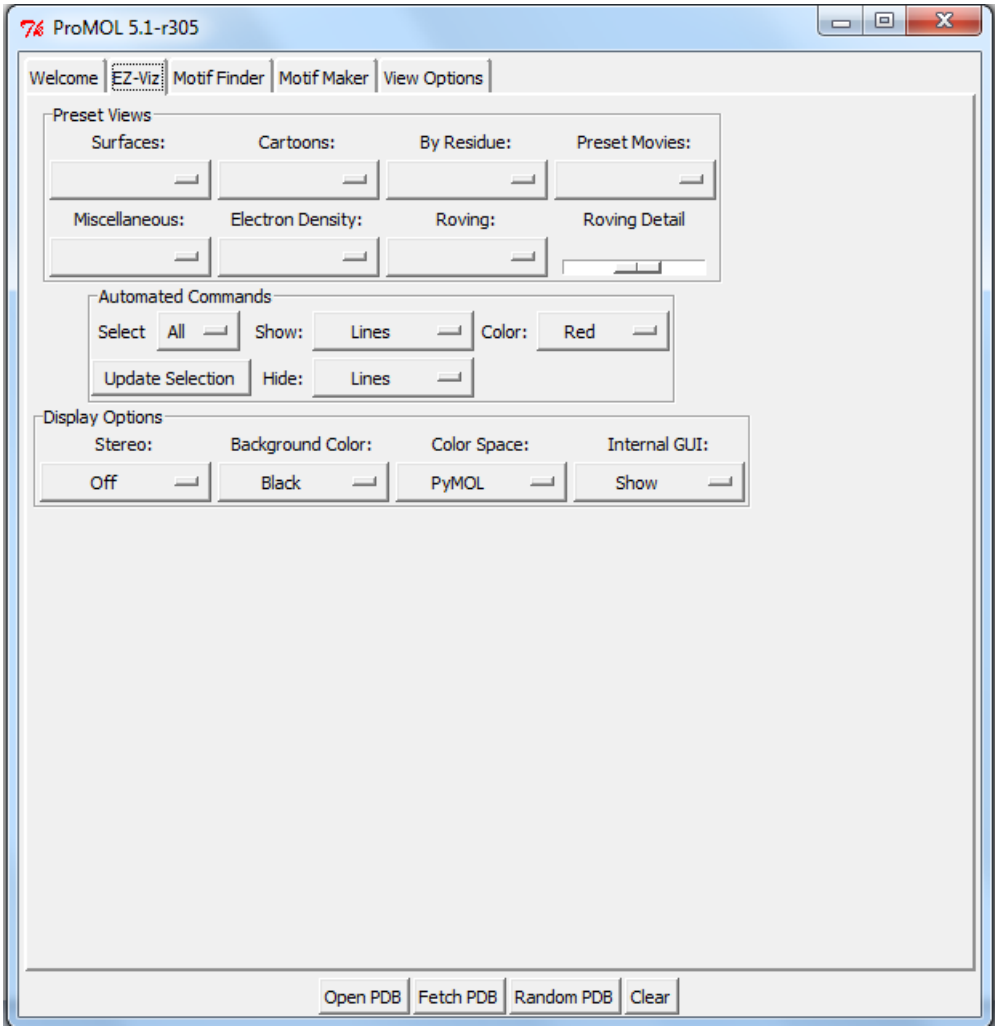

Figure 6. ProMOL's EZ-Viz Tab

To use EZ-Viz, simply open a PDB file in PyMOL (using the "Open PDB," "Fetch PDB," or "Random PDB" buttons at the bottom of the GUI). In this example, the "Fetch PDB" option was used. The PDB protein identification code "1qtq" (glutaminyl tRNA synthetase, EC 6.1.1.18) was entered (see Figure 4, page 11).

The structure that was entered will appear in the "PyMOL Viewer" window as a simple stick figure, in which each stick represents a bond between two atoms in the structure (Figure 7, page 14).

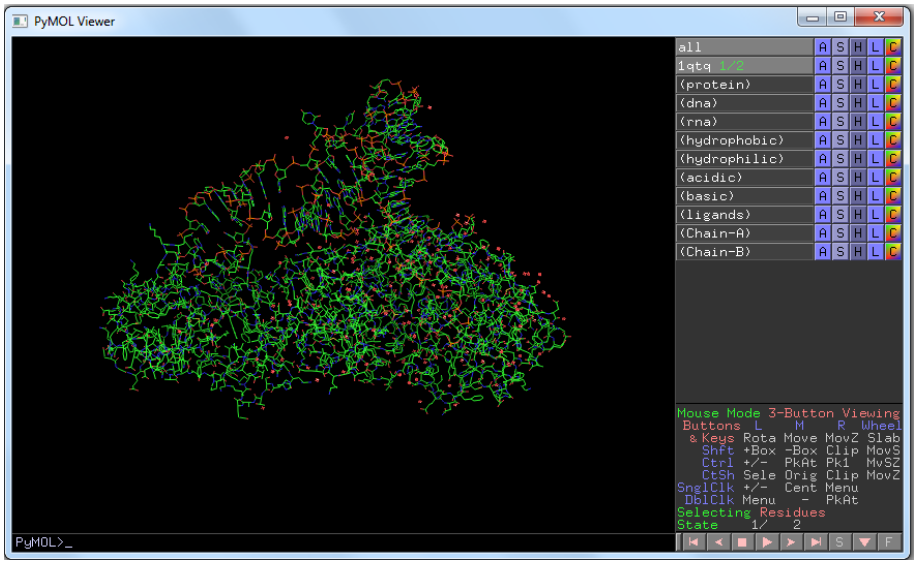

Figure 7. Protein structure of 1qtq in the “PyMOL Viewer” window

The appearance of the structure can be changed using the drop-down menus in the “Present Views” area under the EZ-Viz tab. In Figure 7, the model of 1qtq was generated by selecting “Cartoons” in the “Cartoons:” menu. This view shows the protein as a cartoon, any present DNA as a stick structure, and any present water molecules as spheres. The “Automated Commands” area allows the user to “Select,” “Show,” “Color,” and “Hide” specific aspects of the visual representation.

As an example of how the visual representation of a protein structure can be modified, Figure 8 on page 15 summarizes the different components that were selected to produce the model that was produced of 1qtq, seen in Figure 9 on page 15. In the “Preset Views” area, “Cartoon” was selected in the “Cartoons:” menu, and “Aromatics” was selected in the “By Residue:” menu. In the “Automated Commands” area, “Sticks” was selected in the “Show:” menu, “CPK” was selected in the “Color:” menu, and “Water” was selected in the “Hide” menu.

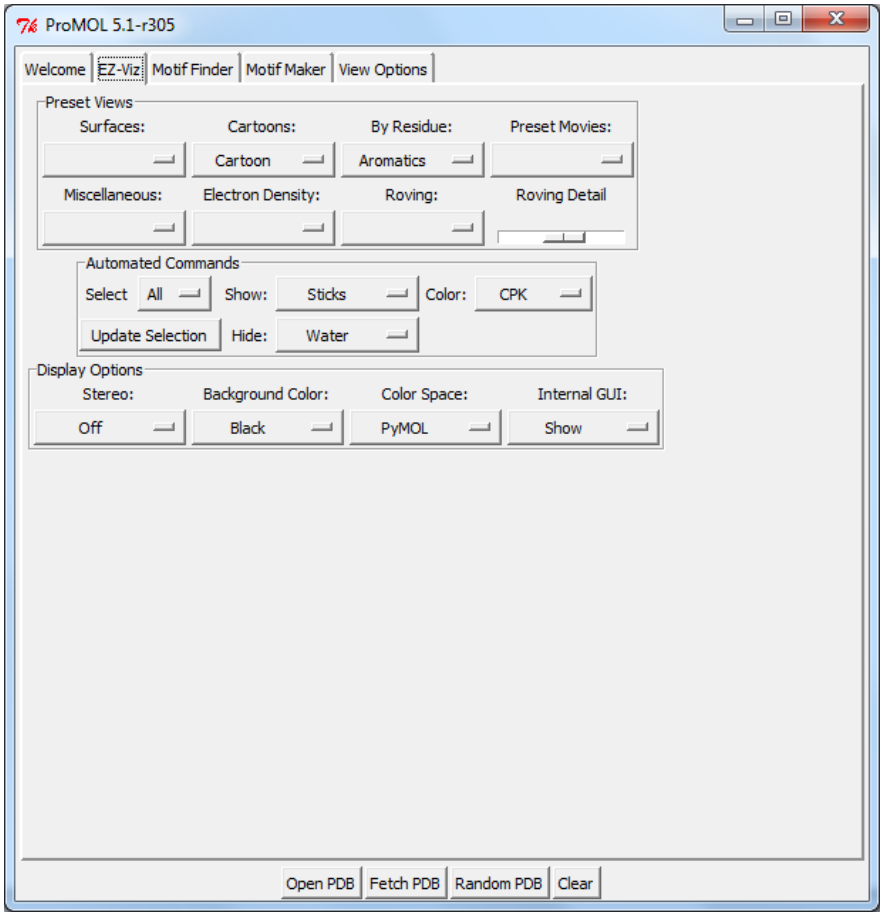

Figure 8. Modifications specified in the EZ-Viz tab to manipulate the visual representation of the protein 1qtq.

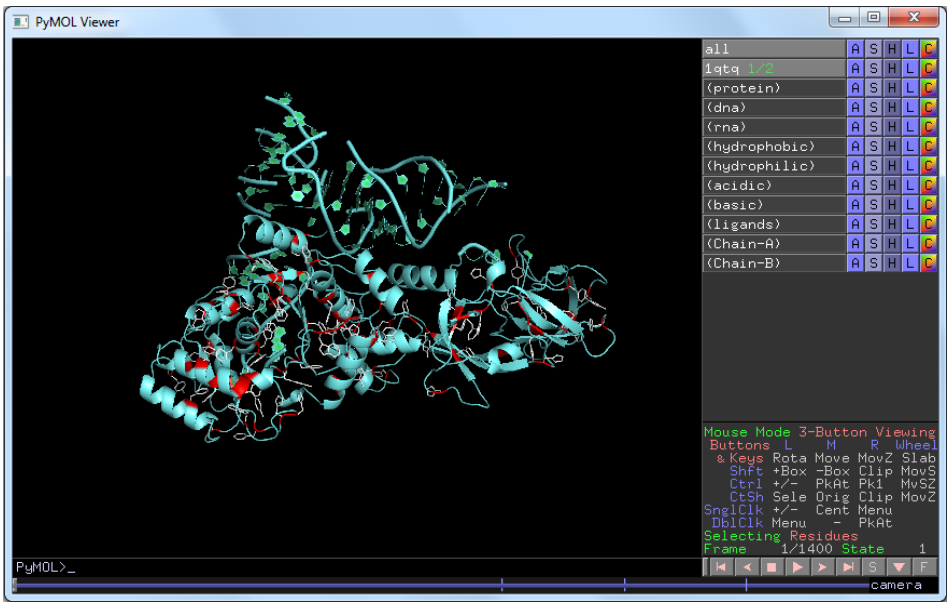

Figure 9. The image of 1qtq that was produced in the “PyMOL Viewer” window after making modifications using EZ-Viz.

- **Motif Finder Tab**
  This tab allows the user to compare the sequence of a queried protein to the active site residues in a database of 934 enzymatic motifs.

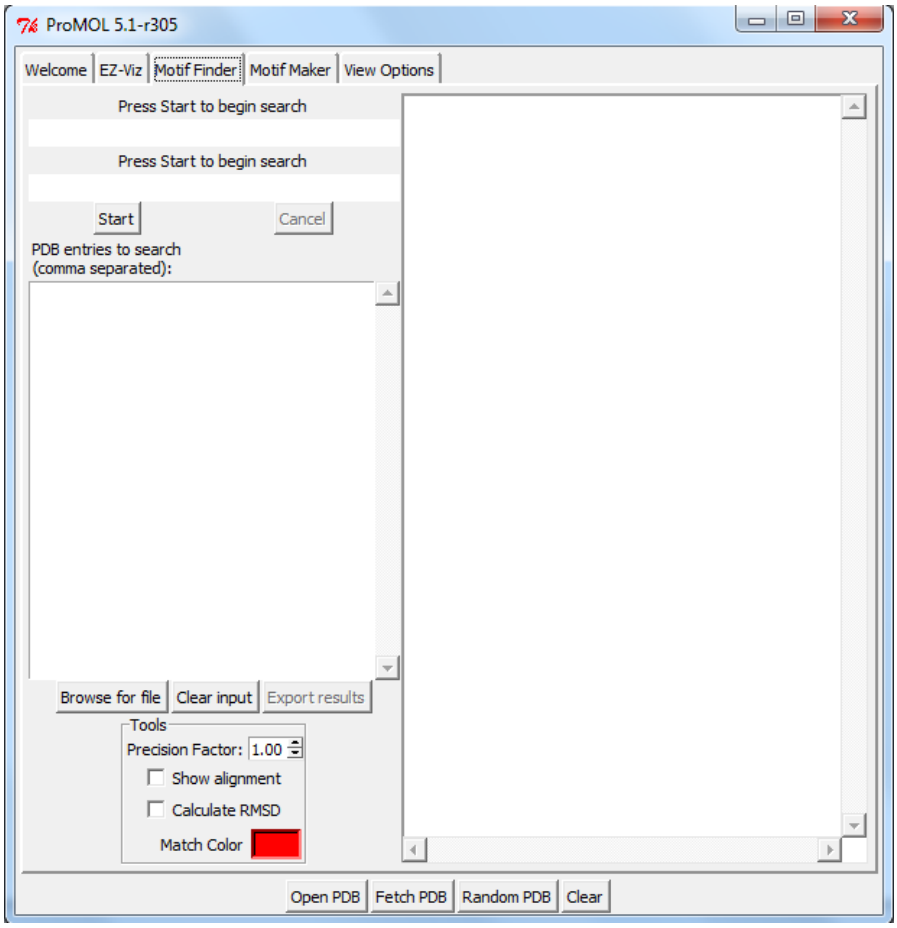

Figure 10. Motif Finer in the ProMOL GUI

Enter the 4-digit PDB protein identification code(s) you wish to compare to the motif database in the “PDB entries to search” area. If more than one protein code is listed, simply separate the entries with a comma (no space). ProMOL will automatically search the Protein Data Bank for the protein sequence file.

The button “Browse for File” located below the text area can be used to upload files saved on the user’s computer. The file must be in a .pdb format.

The “Tools” box (Figure 11, page 17) options allow the user to refine a search. The “Precision Factor” is set to 1.00 by default. Increasing the numerical value will decrease the specificity that ProMOL uses to locate query and motif residues and

how they are oriented in space. “Show alignment” allows the user to view the residue matches once the program has completed analysis of queried structures. The “Calculate RMSD” option will calculate and display RMSD values in the “PyMOL Molecular Graphics System” window when a motif structure match is selected (Figure 12). RMSD stands for “Root Mean Square Distance,” and it is a measure of “closeness of fit” when two structures are aligned. It is measured in Angstroms. This additional data is useful, but note that it will add significant time to the motif search. “Motif color” and “Query Color” will change the color of the motif and query residues in the visual results. Default colors are white and red, respectively.

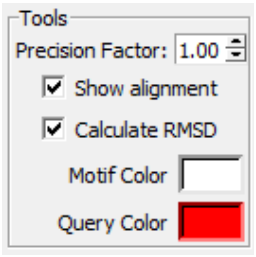

Figure 11. “Tools” box in Motif Finder

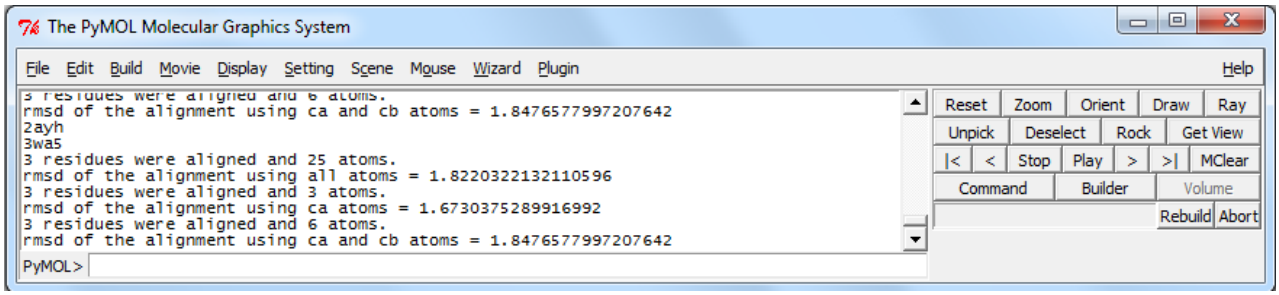

Figure 12. RMSD values in the “PyMOL Molecular Graphics System” window

When the “Search” button is selected, a pop-up window will appear (Figure 13, page 18). It will prompt the user to select which set(s) of motifs he or she would like to search against. ProMOL contains five sets of motifs based on different approaches for catalytic site alignment.

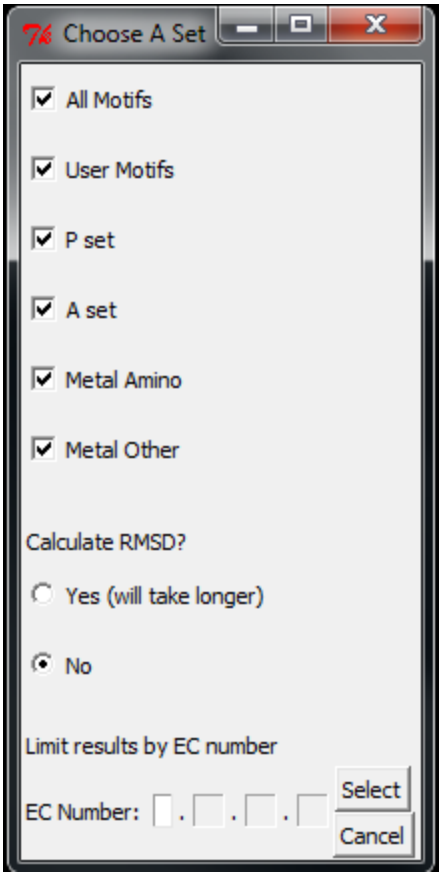

Figure 13. "Choose A Set" window

a. All Motifs

Selecting "All" will compare queried structure(s) to motifs in all sets.

b. User Motifs

Meaning "User" set, this collection represents the motifs that the user creates in ProMOL via the "Motif Maker" tab. The saved files can be found in the following locations:

-Windows: C:\Users**username**\AppData\Roaming\SBEVSL\ProMOL\UserMotifs

-Linux: $Home\.sbevsl\ProMOL\UserMotifs

-MAC: C:\Users**username**\Library\ApplicationSupport\SBEVSL\ProMOL\UserMotifs

c. P Set
These motifs were built using the “ProMOL Motif Maker,” utilizing the same catalytic residues identified in the Catalytic Site Atlas. These motifs were selected because they have been identified as representative members of a variety of EC classes.

d. A Set
This “Automated” set was generated by a script that prompted ProMOL to automatically generate motifs by acquiring data from the Catalytic Site Atlas.

e. Metal Amino Set
The motifs in this set are comprised of the amino acid active site residues and the metal ions and/or prosthetic groups on the protein structure which are associated with enzyme activity.

f. Metal Other Set
This set contains the metal ions and/or prosthetic groups associated with enzyme active sites. There are no amino acid residues incorporated into this set.

Selecting the “Calculate RMSD” option in this window will display the RMSD values within the results list (see Figure 15, page 20).

The motif set with which the query structures are compared can also be limited by specifying the EC class of the motifs that ProMOL considers. It is possible to enter one, two, three, or all four numerical values of the EC class of interest. This section is optional.

Once a motif set has been selected and the search has been started, two progress bars will appear above the “PDB Entries to Search” area on the ProMOL GUI. The top bar indicates the search progress for an individual structure, while the bottom bar indicates the overall progress for the full list of structures that are being evaluated.

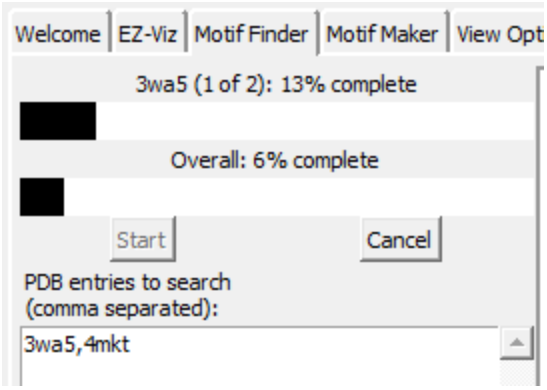

Figure 14. Progress of ProMOL search

When the search is complete, the results will appear in the box in the right side of the ProMOL GUI. Results are reported by motif EC class in a scrollable window.

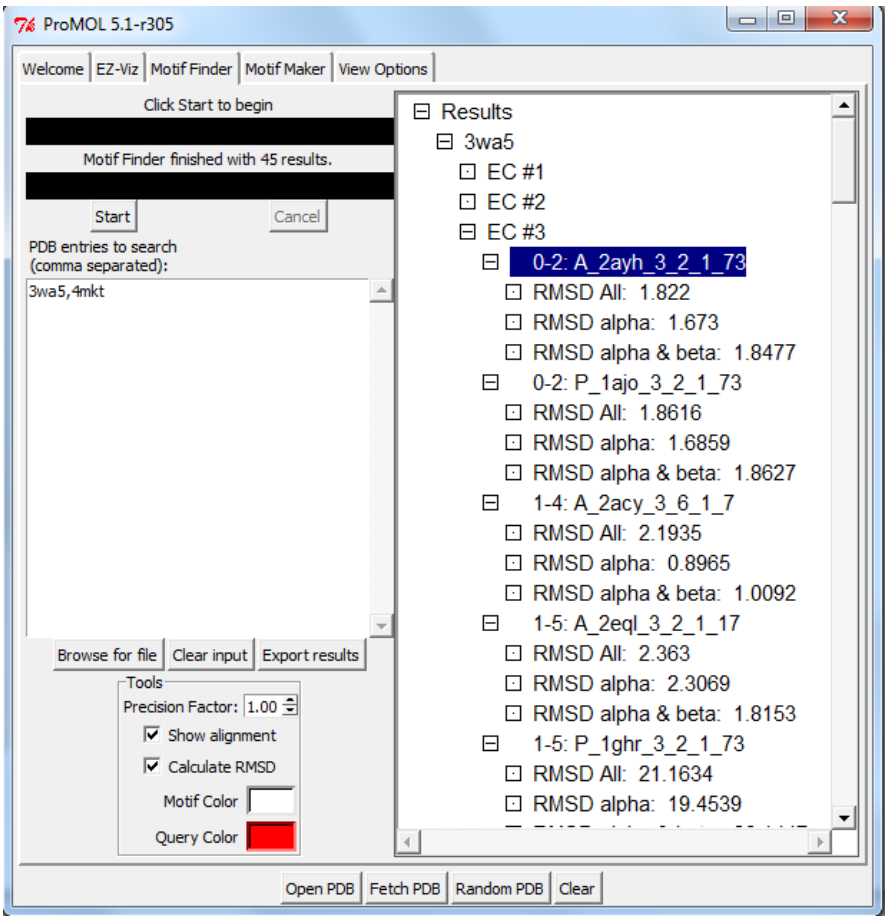

Figure 15. Results of a “Motif Finder” search

Because the EC class of the motifs that ProMOL considered was limited to EC 3 in this example, this is the only class represented in the results. In Figure 15, the

highlighted box represents one motif match. In order from left to right, the data in the highlighted box indicate: [Levenshtein distance: Motif Set_PDB Code of Motif_EC Class of Motif].

The “0-2” in the highlighted box (Figure 15, page 20) represents the Levenshtein distance, which reflects the number of differences between the query structure and the template motifs that are listed there. In simple terms, the Levenshtein distance from the word “horse” to the word “house” is 1 since you need to change the 3rd letter from an “r” in horse to a “u” in “house.” A Levenshtein distance of 0 means that there are no changes; all of the amino acids in the active site of the template structure have been found in the active site of the query structure. "In some cases, a range of Levenshtein distances (e.g. "0-2") is given. In that case there have been multiple matches to the motif, usually in different chains.

The RMSD calculation (measured in Angstroms) does three different comparisons between the two active sites: all atoms, alpha carbons only, and alpha/beta carbons only. Smaller RMSD values indicate a better match. The distances between each atom from each of the residues in a catalytic site are measured using PyMOL’s integrated measurement function. These distances are used to create the 3-D motif definition for that catalytic site; the motif definition are then used to predict the presence of a catalytic site on the structure being tested.

If a user wished to export the result as a .csv file, simply click on the “Export Results” button located in the above the “Tools” box in the ProMOL GUI. This file is easily viewed by importing it into an Excel spreadsheet.

When a motif is selected, the active site residues of the motif and the corresponding residues to which ProMOL aligned residues in the query structure appear in the “PyMOL Viewer” window, like in Figure 16 on page 22. The visualization is 3-D and interactive. To change display options, you can right click in the viewer window (a drop-down menu will appear), click on the boxes next to the protein descriptions in the top right corner of the viewer, and/or utilize the drop-down menus in the “PyMOL Molecular Graphics System” window.

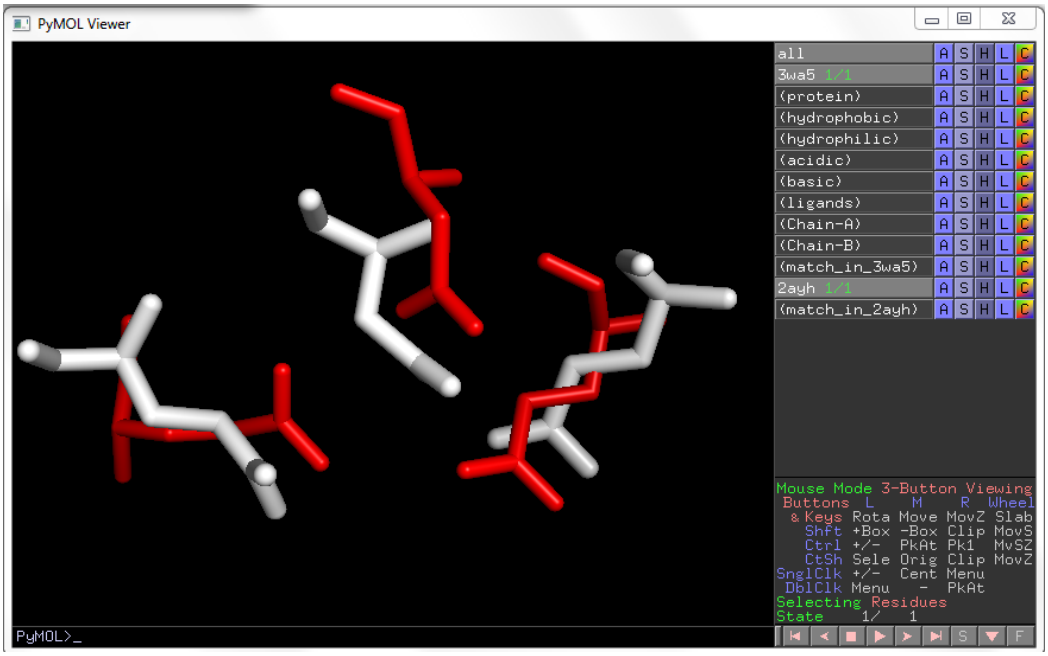

Figure 16. Alignment of active site residues on 2ayh (motif/white) with residues on 3wa5 (query/red) in the “PyMOL Viewer” window

To identify the individual residues, click on the residue(s) of interest. The “PyMOL Molecular Graphics System” window will indicate the protein, chain, residue (3-digit code), the sequence position of that residue, and the carbon atoms that were aligned, in that order.

- **Motif Maker Tab**
  The “Motif Maker” tab (Figure 17) enables the user to build enzyme motifs based on definitions in the Protein Data Bank (PDB) and Catalytic Site Atlas (CSA).

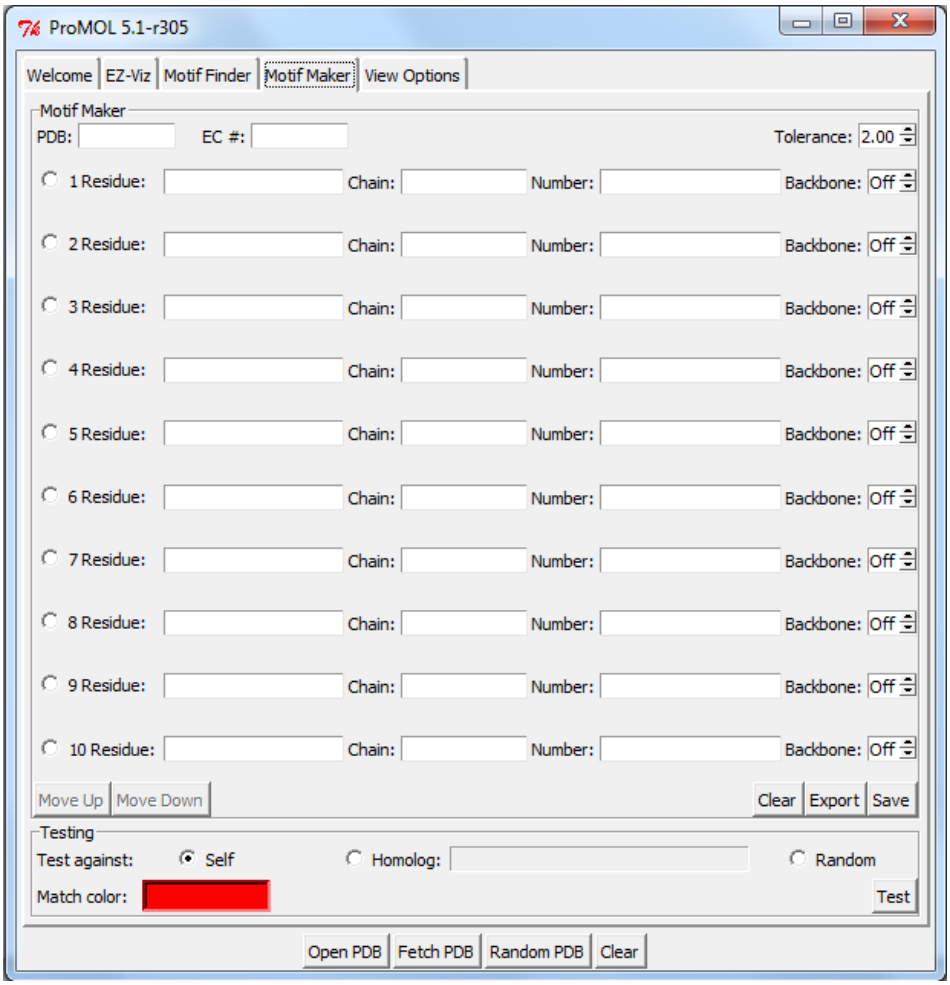

Figure 17. Motif Maker

To create a motif, search for the enzyme’s annotated entry in the Catalytic Site Atlas (CSA). Use the information found in the “Annotated by Homology” section in the bottom half of the web page to build the motif. In the ProMOL GUI, make sure to add the following data: PDB code, EC number, the three-digit identification code of a minimum of three residues (the more the better), the chain on which the residue resides, and the PDB number.

When a new motif is built, it should be tested against itself and twenty or more proteins to look for true positives (structures in which ProMOL clearly identifies the existing motif correctly) and false positives (structures in which ProMOL incorrectly identifies as containing the motif).This can be done by using the “Test” button found in the “Testing” area at the bottom of the GUI. If additional residues appear when testing the motif against “Self,” the “Tolerance” (found in the top right corner of the GUI) should be reduced. Turning the “Backbone” setting to “On” will add the backbone atoms for the residues into the motif definition, making it more precise.

Homologs to the motif, which can be found in the motif’s CSA entry, are used to look for true positives. In the case of several poor or non-existence matches, the creation of the motif should be reconsidered. Possible reasons for poor motifs matches include residue substitutions in the structure (sometimes done in x-ray crystallography via site-directed mutagenesis to create a more stable crystal structure), the absence of the catalytic site in the homolog (homology is based on entire sequences, not catalytic site residues), or an overly constrained motif maker definition file.

The “Random” option in the “Testing” area should be used to search for false positives. If false positives are found, try decreasing the tolerance of the motif.

Once all of the adjustments have been made, clicking on the “Save” button will add the motif into the ProMOL database, and will save the information to your computer.

- **View Options Tab**
  This tab allows the user to adjust the manner in which PyMOL generates images of protein structures. The “Preset Views” box of the EZ-Viz tab enables user to generate basic visualizations. The “View Options” tab can then be used to adjust the appearance of cartoons, spheres, sticks, surfaces, and ambient light.

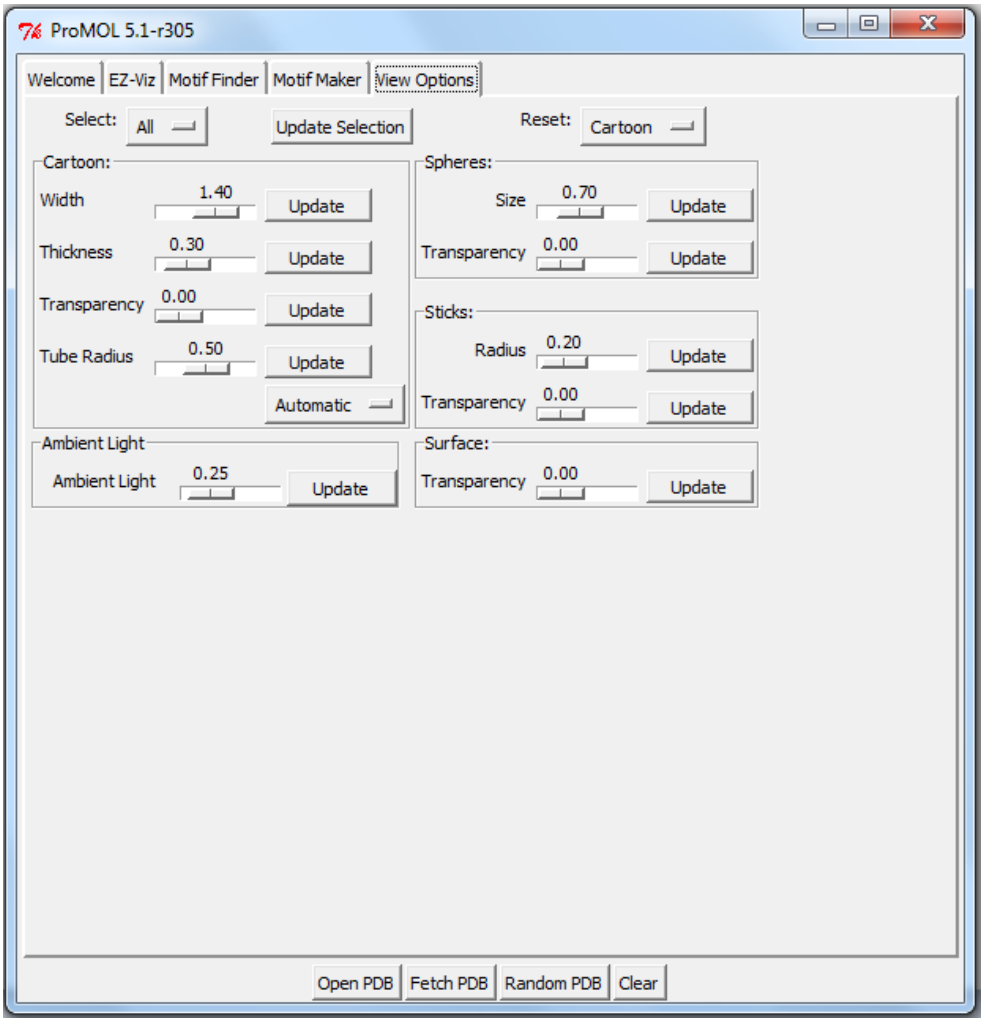

Figure 18. View Options Tab in the ProMOL GUI

Additional Information and Recommended Resources:

ProMOL is a tool that was originally created through the efforts of the SBEVSL (Structural Biology Extensible Visualization Scripting Language) project. The objective of this project is to characterize proteins of unknown function found in the Protein Data Bank via *in silico* and *in vitro*. In addition to ProMOL, the following resources have been fundamental in the development of this effort.

Protein Data Bank: http://www.rcsb.org/pdb/home/home.do
Catalytic Site Atlas: http://www.ebi.ac.uk/thornton-srv/databases/CSA/
BLAST:http://blast.ncbi.nlm.nih.gov/Blast.cgi
Pfam: http://pfam.xfam.org/
Dali: http://ekhidna.biocenter.helsinki.fi/dali_server/start
ExPASY: http://www.expasy.org/
Sauc: http://iterate.sourceforge.net/sauc/

Please visit the following website for contact information and to learn more about the ProMOL project:

ProMOL: http://www.promol.org/