BioLuminate 1.0

User Manual

Schrödinger Press

Revision A, September 2012

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, command input and output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

Links to other locations in the current document or to other PDF documents are colored like this: Document Conventions.

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol `|` separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

File name, path, and environment variable syntax is generally given with the UNIX conventions. To obtain the Windows conventions, replace the forward slash / with the backslash \ in path or directory names, and replace the $ at the beginning of an environment variable with a % at each end. For example, `$SCHRODINGER/maestro` becomes `%SCHRODINGER%\maestro`.

Keyboard references are given in the Windows convention by default, with Mac equivalents in parentheses, for example CTRL+H (⌘H). Where Mac equivalents are not given, COMMAND should be read in place of CTRL. The convention CTRL-H is not used.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Introduction

1.1 Overview of BioLuminate Features

BioLuminate offers a wide range of tools for protein modeling, protein engineering, protein analysis, and antibody modeling. In addition to the unique tools, BioLuminate provides access to many of the related tools in the Schrödinger software suite.

This manual documents the unique tools and capabilities of BioLuminate, and provides references to other documents for the related tools. A brief description of the tool set is given below, with links to the relevant parts of this manual or of other manuals. The descriptions are classified by function. These tools are divided between the Tools menu, where the action does not take much time and may be interactive, and the Tasks menu, where a job may need to be run that takes a larger amount of time.

1.1.1 Protein Analysis Tools

The protein analysis tools provide information on a protein and its properties. No change is made to the protein structure.

- Protein Structure Quality Viewer (Tools → Protein Structure Quality): Show reports on deviations of protein parameters from standard values, in graphical and tabular form. See Chapter 3.
- Residue Analysis (Tasks → Residue Analysis): Calculate energetic and other properties of residues. See Chapter 4.
- Consensus Visualization (Tools → Protein Consensus Viewer): Locate consensus waters, counter ions and ligands in a set of homologs to a reference protein. See Chapter 5.
- Reactive Protein Residues (Tools → Reactive Residue Identification): Identify residues that are prone to specified reactions, by matching sequence patterns and some structural information. See Chapter 6.
- Aggregation Surface (Tasks → Aggregation Surface): Predict regions on a protein surface that have a propensity for aggregation. See Chapter 7.
- Low Mode Vibrational Sampling (Tasks → Low Normal Mode Analysis): Locate and visualize large-scale vibrational motions in a protein. See Chapter 8.

- SiteMap (Tools → Binding Site Identification): Locate druggable sites on a protein. See the *SiteMap User Manual*.

1.1.2 Protein Structure Tools

The protein structure tools allow you to fix structures from the PDB or other sources that are missing information needed for modeling or are missing atoms, predict the structure of proteins by homology modeling, and predict the structure and stability of alpha helices in small peptides.

- Protein Preparation Wizard (Tools → Protein Preparation): Prepare proteins for modeling by assigning bonds, fixing structural defects, removing unwanted parts, assigning protonation and tautomeric states, and refining the structure. See the *Protein Preparation Guide*.
- Simple Homology Modeling (Tasks → Simple Homology Modeling): Predict the structure of proteins using homology modeling, where the homology is high and the alignment of the query and the template is straightforward. See Chapter 11.
- Structure Prediction (Tasks → Advanced Homology Modeling): Predict the structure of single-chain or multi-chain proteins, including multimers, by homology modeling. See Chapter 3 through Chapter 5 of the *Prime User Manual*.
- Refinement (Tasks → Loop + Sidechain Prediction): Refine protein structures by performing predictions of selected side chains or loops. Also performs minimizations. See Chapter 6 of the *Prime User Manual*.
- Peptide Helicity (Tasks → Peptide Alpha Helicity): Predict the stability of alpha helices for small peptide sequences, using molecular dynamics. See Chapter 9.

1.1.3 Alignment and Docking Tools

The alignment tools include tools that structurally superimpose two proteins (or structures), and a tool for docking one protein to another. These tools perform rigid-body translation of the structures to obtain the best alignment.

- Align Binding Sites (Tools → Binding Site Alignment): Align the sites on a set of proteins at which drug-like molecules can bind. See Chapter 7 of the *Prime User Manual*.
- Protein Structure Alignment (Tools → Protein Structure Alignment): Structurally align two or more proteins, using secondary structure information as well as coordinates. See Chapter 7 of the *Prime User Manual*.
- Superposition (Tools → Superposition): Align two or more structures by minimizing the RMSD of a selected set of atoms. See Section 10.3 of the *Maestro User Manual*.

- Protein-Protein Docking (Tasks → Protein-Protein Docking): Predict how two proteins interact, using a rigid body search algorithm. See Chapter 10.

1.1.4 Protein Mutation Tools

- Residue Scanning (Tasks → Residue Scanning): Systematically mutate protein residues to determine how energetic and other properties change, and to identify mutations that can effect desired changes. See Chapter 13.
- Cysteine Mutation (Tasks → Cysteine Mutation): Locate residue pairs that can reasonably form disulfide bonds if one or both of the residues are mutated to cysteine and perform the mutation, or locate disulfide bonds and mutate one of the residues to another type to break the disulfide bond. See Chapter 14.
- Residue and Loop Mutation (Tasks → Residue and Loop Mutation): Mutate a single residue to a standard or custom residue, invert the chirality of a residue, delete or insert multiple residues in a single loop, or swap a loop for another loop. See Chapter 12.

1.1.5 Antibody Tools

BioLuminate provides a specialized set of tools for modeling antibodies, including managing databases of antibody structures, homology modeling of antibodies, antibody humanization, and antigen-antibody docking.

- Antibody Prediction (Tasks → Antibody Modeling → Prediction): Predict the structure of the CDR region of an antibody by homology modeling, using homology and database methods. See Section 15.1 on page 91.
- Antibody Humanization (Tasks → Antibody Modeling → Humanization): Humanize an antibody by performing mutation of residues selected manually or on the basis of homology to human antibodies. See Section 15.2 on page 102.
- Antibody Database Management (Tasks → Antibody Modeling → Database Management): Select, create, and update antibody databases. See Section 15.3 on page 110.
- Protein-Protein Docking (Tasks → Protein-Protein Docking): Dock an antigen to an antibody. See Section 10.5 on page 53.

1.2 Running Schrödinger Software

Schrödinger applications can be started from a graphical interface or from the command line. The software writes input and output files to a directory (folder) which is termed the *working directory.* If you run applications from the command line, the directory from which you run the application is the working directory for the job.

Linux:

To run any Schrödinger program on a Linux platform, or start a Schrödinger job on a remote host from a Linux platform, you must first set the `SCHRODINGER` environment variable to the installation directory for your Schrödinger software. To set this variable, enter the following command at a shell prompt:

csh/tcsh:	`setenv SCHRODINGER` *installation-directory*
bash/ksh:	`export SCHRODINGER=`*installation-directory*

Once you have set the `SCHRODINGER` environment variable, you can run programs and utilities with the following commands:

`$SCHRODINGER/`*program* `&`
`$SCHRODINGER/utilities/`*utility* `&`

You can start the BioLuminate interface with the following command:

```
$SCHRODINGER/maestro -profile BioLuminate &
```

It is usually a good idea to change to the desired working directory before starting BioLuminate. This directory then becomes BioLuminate's working directory.

Windows:

The primary way of running Schrödinger applications on a Windows platform is from a graphical interface. To start the BioLuminate interface, double-click on the BioLuminate icon, on a BioLuminate project, or on a structure file; or choose Start → All Programs → Schrodinger-2012 > BioLuminate. You do not need to make any settings before starting BioLuminate or running programs. The default working directory is the Schrodinger folder in your documents folder (Documents on Windows 7/Vista, My Documents on XP).

If you want to run applications from the command line, you can do so in one of the shells that are provided with the installation and that have the Schrödinger environment set up:

- Schrödinger Command Prompt—DOS shell.
- Schrödinger Power Shell—Windows Power Shell (if available).

You can open these shells from Start → All Programs → Schrodinger-2012. You do not need to include the path to a program or utility when you type the command to run it. If you want access to Unix-style utilities (such as `awk`, `grep`, and `sed`), preface the commands with `sh`, or type `sh` in either of these shells to start a Unix-style shell.

Mac:

The primary way of running Schrödinger software on a Mac is from a graphical interface. To start the BioLuminate interface, click its icon on the dock. If there is no BioLuminate icon on the dock, you can put one there by dragging it from the SchrodingerSuite2012 folder in your Applications folder. This folder contains icons for all the available interfaces. The default working directory is the Schrodinger folder in your Documents folder (`$HOME/Documents/Schrodinger`).

Running software from the command line is similar to Linux—open a terminal window and run the program. You can also start BioLuminate from the command line in the same way as on Linux. The default working directory is then the directory from which you start BioLuminate. You do not need to set the `SCHRODINGER` environment variable, as this is set in your default environment on installation. If you need to set any other variables, use the command

`defaults write ~/.MacOSX/environment` *variable* `"`*value*`"`

1.3 Citing BioLuminate in Publications

The use of this product should be acknowledged in publications as:

BioLuminate, version 1.0, Schrödinger, LLC, New York, NY, 2012.

The BioLuminate Interface

The BioLuminate interface is a customized form of the Maestro interface that is specially designed for biologics use. It inherits most of the capabilities of the Maestro interface (though organized differently), and it has features of its own.

This chapter focuses on the features that are unique to BioLuminate. Summaries of the main Maestro features are given, with references to the *Maestro User Manual* for details. If you have never used Maestro, you should be able to gain a basic understanding of its operation from this chapter.

If you prefer to use the standard Maestro interface, you can do so. Most of the capabilities of BioLuminate are available from the BioLuminate submenu of the Applications menu, and some of them are on the Tools menu.

You can open the BioLuminate interface as follows:

- **Windows:** Double-click the BioLuminate icon on the desktop
- **Mac:** Go to Applications → SchrodingerSuite2012 and double-click the BioLuminate icon
- **Linux:** Start Maestro and choose BioLuminate in the Choose Profile dialog box, or use the command `$SCHRODINGER/maestro -profile BioLuminate`

2.1 The Main Window

The BioLuminate main window opens with the following features displayed by default:

- **Menu bar.** This is at the top of the window on Linux and Windows, and is the menu bar on the Mac.
- **Manager toolbar.** This toolbar is just below the menu bar on Linux and Windows, and at the top of the window on the Mac. Each label on this toolbar displays or hides another toolbar. By default they are all hidden, as much of their function is available in the Toggle Table. See Section 2.4 of the *Maestro User Manual* for details of the toolbars.
- **Toggle Table.** This dockable panel is displayed on the right side of the main window. You can undock it from the main window and redock it with the docking button.

  Many other panels are also dockable. You can change the docking behavior in the Preferences panel (Edit → Settings → Preferences, or CTRL+,)

- **Workspace.** This is the large black area that occupies the main part of the main window. It is where structures are displayed, along with any associated objects such as surfaces and text labels.
- **Status bar.** This bar is below the Workspace. At the left is a button that displays information on what jobs are running, which you can click to open the Monitor panel for detailed information on your jobs. When the pointer is not over an atom in the Workspace, the status bar gives information on the contents of the Workspace. When the pointer is over an atom, the status bar gives information on the identity of the atom. For more information, see Section 2.5 of the *Maestro User Manual*.

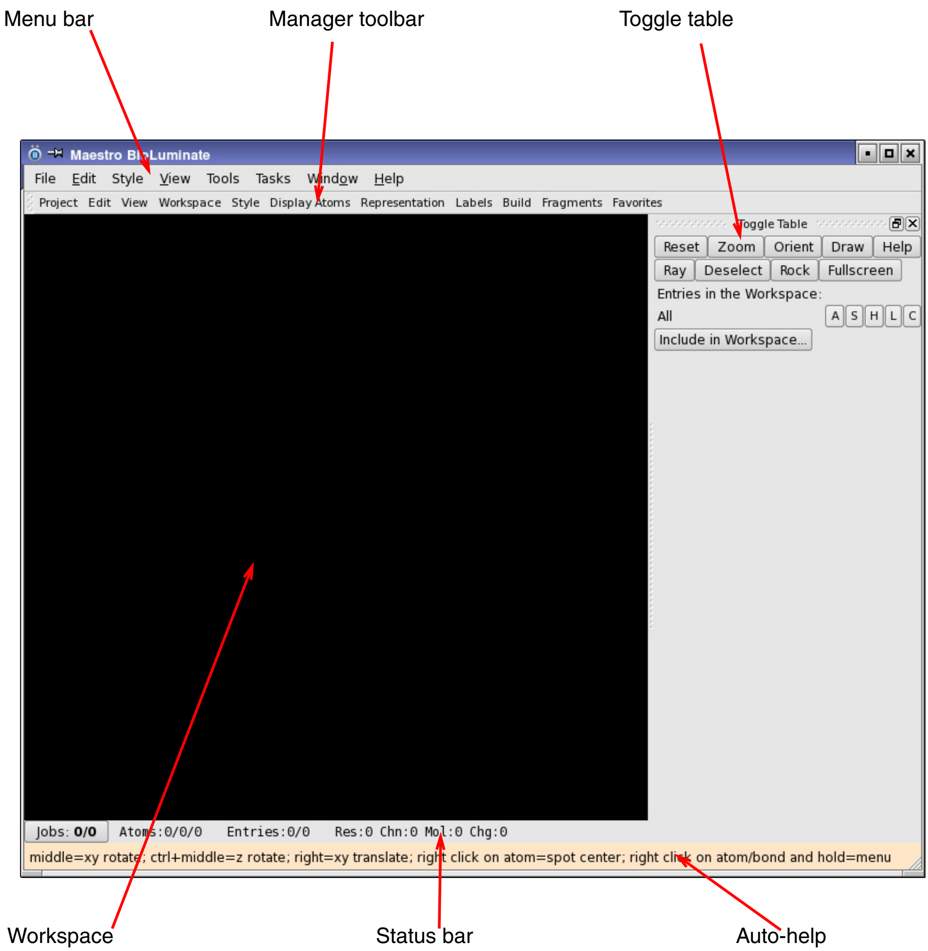


Figure 2.1. The BioLuminate main window.

- **Auto-help.** This orange-yellow bar at the bottom of the main window gives tips on the current action that can be performed in the Workspace.

There are several other components of the main window that can be displayed when needed, by choosing Edit → Settings and then choosing the component. These components include the Sequence Viewer, which displays the sequences of proteins that are in the Workspace; the Find toolbar (also opened with CTRL+F, ⌘F), which you can use to find structural components in the Workspace, like chains or residues; and the Clipping Planes window, which shows a top view of the Workspace and the planes where the structures in the Workspace are clipped for display.

2.2 Structures and Projects

When you start BioLuminate, a new, temporary project is created. Projects are the data structures that Maestro uses to store and manage molecular structures, such as proteins and ligands. Each such molecular structure in a project is stored in an *entry*. An entry can consist of multiple molecules—chains of a protein, waters, cofactors, ions, and so on. The molecular structure (coordinates, charges, and bonding information) is stored along with any properties of the structures, such as the PDB ID and crystallographic information, surfaces associated with the structure, and display information for showing the structure in the Workspace. Properties of individual atoms are stored as well as properties of the structure as a whole.

To add structures to the project, you can *import* them from an external source, such as a file or the Protein Data Bank (PDB). To import structures from a file, choose File → Import Structures. To get structures directly from the PDB (either from a local copy or from the web), choose File → Get PDB. Details of both of these methods for importing can be found in Section 3.1 of the *Maestro User Manual*.

To see a list of all the entries in the project, you can open the Project Table panel with CTRL+T (⌘T) or Window → Show → Project Table. This panel lists the entries in the project with their properties, and provides ways of doing actions on the entries and their properties, such as management tasks, sorting, grouping, plotting, import and export of entries and properties. Full details of the operation of the Project Table can be found in Chapter 9 of the *Maestro User Manual*. The menu organization in the Project Table panel in the BioLuminate interface is a little different from that in the standard Maestro interface: the Table menu in the standard interface is split between the Table menu and the Tools menu in the BioLuminate interface.

2.3 The Toggle Table

The Toggle Table panel can be used to interact with structures in the Workspace. The panel has a set of buttons at the top for quick access to some common actions. The main part of the panel consists of a row for each entry in the Workspace, with a set of buttons, or “toggles”, that can be used to perform actions. These buttons are actually cascading menus, from which you can make selections. In addition to the rows for each entry, there is a row for all entries, labeled All, and rows for selected atoms in the Workspace. The entry rows are labeled with the entry title. Below the rows is a button for including entries in the Workspace.

When the toggle table is displayed, a set of shortcut (or context) menus is also available in the Workspace, which you open by right-clicking. These menus offer the same functions as the toggles.

The interaction with the Workspace provided by the Toggle Table panel is very similar to the operation of PyMOL. If you are familiar with PyMOL, this interface should be easy to learn. If you are familiar with Maestro but not with PyMOL, you can close the Toggle Table panel and use the standard Maestro interaction with the Workspace.

The features of the Toggle Table panel are described in detail in the following subsections.

Note: The terminology used in the Toggle Table panel is the PyMOL terminology, which is somewhat different from that used in standard Maestro.

2.3.1 Quick Access Buttons

This set of buttons at the top of the Toggle Table panel provides quick access to a number of actions, some of which are also on the menus.

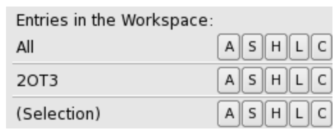
- Reset—Reset the view to the default view, in which the view axes are aligned with the coordinate axes of the structure.
- Zoom—Change the view of the Workspace so that all atoms fit inside the Workspace area.
- Orient—Orient the Workspace structures by translating and rotating them so that the center of mass is at the origin, the largest principal axis lies on the x axis, and the second-largest principal axis lies on the y axis.

  Note: This operation changes the coordinates of the structures, not just the coordinates of the view.

- Draw—Save an image of the Workspace to a file in TIFF, JPEG, or PNG format. (Same as File → Save Image.)
- Help—Open the help topic for the panel in your browser.
- Ray—Use PyMOL to draw a ray-traced image of the Workspace. This feature requires PyMOL to be installed, and either the `PYMOL4MAESTRO` or `PYMOL_PATH` environment variable must point to the directory where the PyMOL executable file resides.
- Deselect—Clear the current Workspace selection.
- Rock—Rotate the Workspace back and forth smoothly. Click once to start rotation, click again to stop.
- Fullscreen—Switch Maestro to full screen mode. To exit full screen mode, press the ESCAPE key or click this button again.

2.3.2 Table Rows

Each row in the Toggle Table has a title and a set of five action buttons, labeled A, S, H, L, and C. These buttons open cascading menus, and are described in the following sections.



There are three distinct types of rows in the Toggle Table:

- **The All row:** Actions taken in this row apply to all entries in the Toggle Table.
- **Entry rows**: These rows apply to a single project entry that is currently in the Workspace. The name of the row is the Title property of the entry. Entry rows are deleted when the entry is excluded from the Workspace, and a new row is added when an entry is added to the Workspace.
- **Selection rows:** When a group of atoms is selected, a selection row is added to the table. Actions in this row apply to the group of atoms that was selected to create this row and even apply after those atoms are no longer selected. The selection row named (Selection) always refers to the most recent group of selected atoms. If a new selection is made while a selection row is active, that row now refers to the new set of selected atoms. Only one selection can be active at a time.

  Selection rows can be renamed: renamed selection rows always refer to the same set of atoms regardless of subsequent Workspace selections. To rename a selection row, choose A → Rename Selection. Selection row names are always enclosed in parentheses.

  Selection rows are deleted when any atoms they refer to are removed from the Workspace. To delete a selection row, choose A → Delete Selection.

Using the A button submenus, selection rows can also be duplicated, copied, and extracted to define a new entity that is independent of the objects from which the selection was originally derived.

Some operations or menu items change based on whether they are being applied to the All row, an entry row, or a selection row. The descriptions below primarily describe the behavior for entry rows. When a behavior changes for the All row or a selection row, the change is noted after the description.

Clicking the name of the All row or an entry row changes the visibility of that object in the Workspace. When the object's visibility is off, the name is dimmed. This is a quick way of showing or hiding the atoms in entries. Hiding atoms does not remove them from the Workspace, so any action taken on the entry or the entire Workspace applies to the hidden atoms as well as the visible atoms.

For instance, if the Workspace contains ten entries and only one entry is visible, choosing Clean from the A menu in the All row operates on all ten structures. This may take a considerable amount of time to complete and lead to unexpected results. Similarly, if a panel imports structures from the Workspace, it imports all ten structures rather than just the visible structure.

Clicking the name of the current selection row deselects the selected atoms, but the selection remains defined. Clicking the name of any other selection row selects the atom group that the row refers to, and any currently selected atoms that are not part of this atom group are deselected.

2.3.3 The Action Menu

The A button opens the Action menu, from which you can choose a variety of actions for the structure defined by the table row. Not all of the actions are available in every table row.

2.3.3.1 Changing the View

The first four actions on the Action menu change the view (camera angle) of the structure defined by the table row. The Workspace coordinate system has the origin in the middle of the Workspace, the x axis is the horizontal axis, the y axis is the vertical axis, and the z axis is coming out of the screen.

- Zoom—Change the view of the Workspace so that all the atoms in the structure fit inside and fill the Workspace area. In the All row, this action is equivalent to clicking the Zoom button at the top of the panel.
- Orient—Orient the structure by translating and rotating it so that the center of mass is at the origin, the largest principal axis lies on the x axis, and the second-largest principal axis lies on the y axis. When you apply this operation to the selection, it is the center of

mass and principal axes of the selection that are used, but the entire structure is reoriented.

Note: This operation changes the coordinates of the structures, not just the coordinates of the view (the camera angle).

- Center—Center the structure in the Workspace, by translating the structure so that its centroid is at the center of the Workspace.
- Origin—Set the center around which rotation is performed to the centroid of the structure. The centroid need not be at the Workspace origin.

2.3.3.2 Minimizing the Structure Energy

You can minimize the energy of the Workspace or the selected atoms using the OPLS_2005 force field, by choosing Clean from the Action menu. The current selection is updated (or a new selection created) to refer to the minimized atoms. To avoid starting a lengthy calculation that is better performed in the background, Clean is limited to structures of fewer than 1000 atoms.

2.3.3.3 Changing the Appearance of Structures

The set of actions on the Preset submenu change the appearance (representation) of the structures using various preset styles. Some of these representations take a little time to set up, and a progress bar is displayed at the bottom of the table.

- Simple—Show proteins as ribbons (C alpha trace) colored by chain, ligands and bound receptor as sticks, and solvent, disulfides and ions as lines. Atom colors are not changed.
- Simple (no solvent)—Same as Simple, but no waters are shown.
- Ball and Stick—Show atoms and bonds in ball and stick, with no protein ribbons.
- B-Factor—Show the protein as tubes with residues colored by the B-factors of the residues, in a relative scheme that ranges from shades of blue, through green and yellow to red.
- Technical—Show atoms as sticks, colored with rainbow colors by residue position, and show polar contacts (hydrogen bonds) as yellow dotted lines.
- Ligand—Show proteins as ribbons colored with rainbow colors by residue position, and ligands as lines. All protein atoms within 5 Å of the ligand are shown as lines, with carbons colored with rainbow colors by residue position. Waters are shown as sticks, polar contacts are shown as yellow dotted lines, and the view is zoomed in to the atoms shown.

- Ligand Sites—This submenu shows variations on the Ligand preset that alter the way the protein or region around the ligand is shown:
    - Cartoon—Show proteins as cartoons rather than ribbons.
    - Solid surface—Show the molecular surface of the protein around the ligand as an opaque surface, colored by the nearest non-hydrogen atom.
    - Transparent surface—Show the molecular surface of the protein around the ligand as a semi-transparent surface, colored by the nearest non-hydrogen atom; show atoms and bonds as sticks.
    - Dot surface—Show the molecular surface of the protein around the ligand as a dot surface, colored by the nearest non-hydrogen atom; show atoms and bonds as sticks.
    - Mesh surface—Show the molecular surface of the protein around the ligand as a mesh surface, colored by the nearest non-hydrogen atom; show atoms and bonds as sticks.
- Pretty—Show proteins as cartoons colored with rainbow colors by residue position and ligands as sticks.
- Pretty (with solvent)—Same as Pretty but waters are shown as ball and stick.
- Publication—Same as Pretty, but protein helices are two-sided.
- Publication (with solvent)—Same as Pretty (with solvent), but protein helices are two-sided.
- Protein Interface—Color ribbons and carbons by chain, show anything not at a protein interface as cartoon ribbons, and interface residues as ball and stick. Non-carbon atoms retain their previous coloring. Interface residues are residues in a chain with more than 300 atoms that are within 4.5 Å of another chain with more than 300 atoms.
- Antibody—Show everything as cartoon ribbons colored by antibody structure. The light chain is colored in red hues, the heavy chain is colored in blue hues, and everything else is colored green. Constant regions are dark hues and the CDR regions are bright hues. The light chain L1-L3 loops are shaded orange to brown, while the heavy chain H1-H3 are shaded grey-blue to cyan.
- Default—Show everything as lines with default colors (colored by element with green carbons).

2.3.3.4 Displaying Polar Contacts

You can turn on or off the display of polar contacts (hydrogen bonds) between various groups of atoms from the Polar Contacts submenu. The display is turned on with A → Polar Contacts → Find; it is turned off with A → Polar Contacts → Remove All. The atom groupings are:

- Within Object
- Involving Side Chains
- Involving Solvent
- Excluding Solvent
- Excluding Main Chain
- Excluding Intra-Main Chain
- Just Intra-Side Chain
- Just Intra-Main Chain
- To Other Atoms In Entry
- To Other Atoms In Entry Excluding Solvent
- To Any Atoms
- To Any Atoms Excluding Solvent

Each menu choice clears any previous choice before applying the new choice.

If you want more flexibility in choosing the atom groups between which polar contacts are shown, you can use the Measurements panel. Choose Style → Measurements → H-Bonds from the menu bar at the top of the Workspace to open this panel.

2.3.3.5 Generating an Atom Selection

You can select atoms in predefined groups by choosing A → Generate → Selection and then choosing the group. The Generate item is not available for the All or Selection rows. The predefined groups are:

- All—All atoms.
- Polymer—Backbone and side-chain atoms.
- Organic—Ligand atoms.
- Solvent—Water atoms.
- Surface Residues—All residues with solvent-exposed surface area greater than 10 $Å^2$.
- Protein Interface—Residues in a chain of more than 300 atoms that are within 4.5 Å of another chain of more than 300 atoms.

Atom selections can also be generated by picking atom groups in the Workspace. To set the kind of atom group you want to pick, choose Edit → Pick Mode then the atom group name (Atoms, Residues, Chains, Molecules, or Entries), using the menu bar at the top of the Workspace. You can also set the mode by typing the first letter of the name when the pointer is in the Workspace. To pick an atom group, click on an atom in the Workspace that belongs to the group. You can see information about the atom in the Status bar when you pause the pointer over the atom.

2.3.3.6 Displaying Atoms Related By Crystallographic Symmetry

If you want to see atoms from nearest-neighbor crystal symmetry mates of your structure, you can choose A → Generate → Symmetry Mates. To show symmetry-related atoms requires crystal symmetry information to be present for the structure. You should also ensure that you have only the structure that you want to see the related information for, because this action applies to the entire Workspace.

The Symmetry Mates submenu items control the cutoff distance from the original structure for which symmetry mate atoms should be displayed. Note that no matter how far the cutoff is placed from the original structure, only the nearest-neighbor mates are created and shown. The symmetry mates are created as separate, temporary entries ("scratch entries") and are shown in the toggle table. You can remove the symmetry mates by choosing Generate → Symmetry Mates → Show None.

2.3.3.7 Modifying an Atom Selection

The Modify item is available for selection rows only, and replaces Generate on the Action menu. It allows you to alter the group of selected atoms to include or exclude other atoms. After a Modify action, the selection row applies to the new group of atoms.

The Modify actions all have a choice of atom groups to which they apply.

- Around—select all atoms or residues within a given distance from the current set of atoms, and deselect the current set of atoms. The distance is chosen from the submenu, and can encompass atoms only or be filled to entire residues that have any atoms within the chosen distance.
- Expand—Expand the current selection to include all atoms or residues within a given distance from the current set of atoms. The distance is chosen from the submenu, and can encompass atoms only or be filled to entire residues that have any atoms within the chosen distance.
- Extend—Expand the current selection to include all atoms or residues within a given number of bonds from the current set of atoms.
- Invert—Deselect the current set of atoms and select all other atoms within a given atom group. The atom groups can be chosen from the submenu:
    - Within Objects—All atoms in the entry that are not part of the selection are selected.
    - Within Chains—In each chain that contains selected atoms, the unselected atoms are selected, and the selected atoms are deselected.
    - Within Residues—In each residue that contains selected atoms, the unselected atoms are selected, and the selected atoms are deselected.

- Within Molecules—In each molecule that contains selected atoms, the unselected atoms are selected, and the selected atoms are deselected.
- Within Any—All atoms in the entire Workspace that are not part of the selection are selected.

- Complete—Add all other atoms within a given atom group to the selection. The atom groups can be chosen from the submenu.
  - Residues—In each residue that contains selected atoms, all atoms are selected.
  - Chains—In each chain that contains selected atoms, all atoms are selected.
  - Objects—In each entry that contains selected atoms, all atoms are selected.
  - Molecules—In each molecule that contains selected atoms, all atoms are selected.
  - C-alphas—All alpha carbons for residues within the selection are selected. All other atoms are deselected.
- Restrict to—Reduce the selection to only those atoms within a specific group that are currently selected. The available atom groups are:
  - Object—Restrict the selection to atoms in a specific entry, chosen from the submenu.
  - Selection—Restrict the selection to atoms in a specific selection row, chosen from the submenu.
  - Visible—Restrict the selection to atoms that are visible.
  - Polymer—Restrict the selection to backbone and side-chain atoms.
  - Organic—Restrict the selection to ligand atoms.
  - Solvent—Restrict the selection to water atoms.
  - Inorganic—Restrict the selection to atoms other than H, C, N, O, F, P, S, Cl, Br, or I.
- Include—Include additional atom groups in the current selection. The available atom groups are:
  - Object—Include atoms in a specific entry, chosen from the submenu.
  - Selection—Include atoms in a specific selection row, chosen from the submenu.
  - Visible—Include all visible Workspace atoms.
- Exclude—Exclude specific atoms from the selection. The available atom groups are:
  - Object—Exclude atoms in a specific entry, chosen from the submenu.
  - Selection—Exclude atoms in a specific selection row, chosen from the submenu.
  - Visible—Exclude atoms that are visible.
  - Polymer—Exclude backbone and side-chain atoms.
  - Organic—Exclude ligand atoms.
  - Solvent—Exclude water atoms.
  - Inorganic—Exclude atoms with element other than H, C, N, O, F, P, S, Cl, Br, or I.

2.3.3.8 Removing Selections

For the All row, Delete Selections removes any selection rows from the Toggle Table. All atoms are deselected but are otherwise unaltered.

For selection rows Delete Selection just removes the row from the Toggle Table. The atoms in this selection group are deselected but otherwise remain unaltered.

2.3.3.9 Renaming Rows

You can change the name of an entry row, and thereby change the Title of the project entry with the Rename action. A text box is displayed instead of the name, and you can type a new name in the box. If you decide you do not want to change the name after all, press ESC.

You can change the name of a selection row with the Rename Selection action. If you do this for the default selection row, the selection is preserved for future use as a named selection. You can also rename named selection.

This command not available for the All row.

2.3.3.10 Duplicating Rows

You can duplicate table rows with the Duplicate action, with the exception of the All row. The action is different for entry rows and for selection rows.

For entry rows, the action creates a new project entry below the entry that is the duplicate of the entry. Both structures remain in the Workspace and are listed in the Toggle Table.

For selection rows, this action creates a duplicate selection row in the Toggle Table. No project entry is created. This can be useful if you want to use a selection as the basis for another selection.

2.3.3.11 Removing Entries from the Workspace

To remove (exclude) an entry from the Workspace, choose A → Remove from Workspace for that row. Removing an entry from the Workspace just means that it is no longer in your working area. The entry remains in the project, and is listed in the Project Table. You can add it back to the Workspace by using the Include in Workspace button or the Project Table.

To clear the Workspace entirely, choose Remove Everything from Workspace in the All row.

2.3.3.12 Deleting Entries from the Project

To remove an entry from the project, choose Delete from Project in the entry row. The structure is removed from the Workspace and from the project. It no longer appear in the Project Table.

2.3.3.13 Creating Project Entries

You can create new project entries either from a selection or from an entry row, by choosing Copy to New Project Entry. The new project entry is created by copying the selected atoms or entries, and it is added to the Workspace. This command is the same as Duplicate if you choose it for an entry row.

To create a project entry by removing atoms from entries and placing them into a new entry, you can choose Extract to New Project Entry in a selection row. The new project entry contains the atoms in the selection, and the atoms are deleted from their current structure.

2.3.3.14 Adding and Removing Hydrogens

You can add or remove hydrogen atoms from a row by choosing Hydrogens → Add or Hydrogens → Remove. If you perform this action in the All row or an entry row, the atoms are removed from the structure. If you perform this action in a selection row, the hydrogens in the selection are removed, or they are added to complete the valences of the atoms in the selection.

You can also add hydrogens from the main menu bar with Edit → Add Hydrogens or from the Edit toolbar.

2.3.3.15 Removing Waters

You can remove waters from structures with the Remove Waters action. If you perform this action in the All row or an entry row, the atoms are removed from the structure. If you perform this action in a selection row, the waters in the selection are removed from the structure.

2.3.3.16 Computing Properties

To compute simple properties of the object (selection or entry), you can use the Compute action. The properties you can compute are:

- Atom Count—The number of atoms.
- Formal Charge Sum—Sum of atom formal charges.
- Partial Charge Sum—Sum of atom partial charges.
- Surface Areas—Compute surface areas. Solvent-Accessible surface area (SASA) uses a solvent radius of 1.4 Å while Molecular surface area uses the same algorithm but with no solvent to expand the atomic radii. The surface area is computed for the object in the context of the visible atoms in the Workspace.

The computed properties are displayed in a window that opens, and you can copy and paste the text. They are not stored in the Project Table.

2.3.4 The Show Menu

The S button opens the Show menu, which contains commands that alter the display of structures in the Workspace. There are three different types of display for structures:

- Atomic representations such as lines, sticks, ball and sticks or spheres
- Ribbon representations such as ribbons and cartoons
- Surface representations such as solid or mesh surfaces

Each atom may have one of each representation type active at once, but cannot have multiple representations of the same type active. For instance, an atom can be shown simultaneously with lines, cartoon ribbons and a mesh surface. However, an atom cannot be shown simultaneously by both ball and stick and sphere representations because they are both atomic representations. If an action is chosen to show an atom with a representation in the same category as an existing representation for that atom, the existing representation is removed and the atom is shown with the new representation. For instance, if an atom is shown as lines, and the Ball and Stick menu item is chosen, the lines representation is removed and the atom is shown with ball and stick representation.

The common representations that can be applied are:

- Lines—Set the atomic representation to lines (wire frame).
- Sticks—Set the atomic representation to sticks (tube).
- Ball and Stick—Set the atomic representation to ball and stick.
- Ribbon—Set the ribbon representation to ribbon (CA trace tube).
- Cartoon—Set the ribbon representation to cartoons.
- Label—Show any labels defined for this structure.
- Nonbonded—Set the atomic representation of any atom with no attachments (bonds) to Lines (wire frame). These atoms appear as small stars. No change is made to atoms that have attachments (bonds).
- Spheres—Set the atomic representation to spheres (CPK). The sphere radii are the van der Waals radii of the atoms.
- Nonbonded Spheres—Set the atomic representation of any atom with no attachments (bonds) to spheres (CPK). No change is made to atoms that have attachments (bonds).

These items appear on both the Show menu itself and on the As submenu. When you choose from the Show menu, the representation is added to the display. When you choose from the As submenu, the previous representation is replaced by the new choice.

For instance, a residue shown with lines and cartoon ribbons is shown as only ball and stick if S → As → Ball and Stick is chosen, but is shown as cartoon and ball and stick if S → Ball and Stick is chosen.

Two commands on the menu for entry rows and the All row create and display molecular surfaces. The surface is created if it does not exist, otherwise the color and representation of the surface is changed.

- Mesh—Display a mesh molecular surface colored by current atom color.
- Surface—Display a solid molecular surface colored by current atom color.

The remaining four items have submenus from which you can set the representation to lines, sticks or spheres:

- Organic—Set the atomic representation of backbone and side-chain atoms.
- Main Chain—Set the atomic representation of backbone atoms.
- Side Chain—Set the atomic representation of side-chain atoms.
- Disulfides—Set the atomic representation of disulfide atoms.

2.3.5 The Hide Menu

The H button opens the Hide menu, which contains commands that hide features in the Workspace. To redisplay these features, use the Show menu. Each item hides only the particular feature for the row, and leaves any other features as they are. Hiding atoms also hides any representation of the bonds to those atoms.

- Everything—Hide all features: atomic, ribbon, and surface representations and labels.
- Atoms—Hide atoms and bonds.
- Ribbon—Hide all ribbon representations.
- Cartoon—Hide all ribbon representations.
- Label—Hide labels. The labels remain defined and can be redisplayed.
- Nonbonded Atoms—Hide atoms with no attachments, such as Cl^- ions.
- Mesh—Hide surfaces.
- Surface—Hide surfaces.
- Main Chain—Hide backbone atoms.
- Side Chain—Hide side-chain atoms.
- Waters—Hide water atoms.
- Hydrogens—Hide nonpolar or all hydrogens, as chosen from the submenu.

- Symmetry Mates—Removes all crystal symmetry mates from the Workspace. This is a Workspace setting, so affects all symmetry mates in the Workspace.
- Polar Contacts—Remove polar contact (hydrogen bond) markers.
- All Others—Hide everything for all atoms in the Workspace other than the atoms defined in the row.

2.3.6 The Label Menu

The L button opens the Label menu, which contains commands that label features in the Workspace. Unless otherwise specified, labels are created for every atom the row applies to. Atom labels created by these commands are not cumulative. Any existing atom labels are removed when new ones are created.

Labels can be cleared from the Workspace with L → Clear. They can be hidden with H → Label. If they are hidden, they can be redisplayed with S → Label, while if they are cleared, they need to be created (usually by other Label menu commands) before they can be shown again.

Residues—Label the first carbon atom in each residue with the three-letter PDB code and residue number.

Chains—Label the first and last residue in each chain with the chain name.

The next set of commands offers a choice of the label content, including identifiers and numeric properties.

- Atom Name—Label each atom with its PDB atom name.
- Element Symbol—Label each atom with its element symbol.
- Residue Name—Label each atom with the three letter PDB code of the residue it is in.
- Residue Identifier—Label each atom with the residue number of the residue it is in.
- Chain Identifier—Label each atom with the name of the chain it is part of.
- B-factor—Label each atom with its PDB B-factor value, if it exists.
- Occupancy—Label each atom with its partial occupancy data if it exists.
- VDW Radius—Label each atom with its van der Waals radius.
- Other Properties—Submenu with other properties that can be used for labels:
  - Formal Charge—Label each atom with its formal charge.
  - Partial Charge (0.00)—Label each atom with its partial charge to two decimal places.
  - Partial Charge (0.0000)—Label each atom with its partial charge to four decimal places.
  - MacroModel Text Type—Label each atom with its MacroModel atom type.

  - MacroModel Numeric Type—Label each atom with the numerical index for the MacroModel atom type.
  - Stereochemistry—Label each atom with E,Z and R,S stereochemistry.

- Atom Identifiers—Submenu of atom identifiers that can be used for labels.

2.3.7 The Color Menu

The C button opens the Color menu, which contains commands to color the atom representations by various coloring schemes including by element, chain, substructure, B-factor and entry. The schemes are grouped into classes, each of which is represented by a menu item.

- Color by Element—Color H, C, N, O and S atoms. Default colors are white for H, green for C, blue for N, red for O and yellow for S. There are several choices for modifying the color scheme on this submenu.
  - Reset HNOS—Set H, N, O and S atoms to their default color. Carbon atoms remain their current color.
  - Custom Color {C}HNOS—Pick the color for carbon atoms and set H, N, O, and S atoms to their default color. Clicking on the menu item opens a palette of colors to choose from for carbon atoms, while selecting Recent Color Choices lists the most recent colors chosen by this command.
  - Custom Color {H}CNOS—Pick the color for hydrogen atoms and set C, N, O, and S atoms to their default color. Clicking on the menu item opens a palette of colors to choose from for hydrogen atoms, while selecting Recent Color Choices lists the most recent colors chosen by this command.
- Color by Chain—Color atoms by chain:
  - by Chain (Carbons)—Change the color of carbon atoms only.
  - by Chain (Calpha)—Change the color of alpha carbon atoms only.
  - by Chain—Change the color of all atoms.
  - Chainbows—Each chain is colored with rainbow colors.
- Color by Substructure—Helices, sheets and loops are colored by the chosen color scheme. All other atoms retain their current color.
- Color by Spectrum—Color all residues by a spectrum of colors.
  - Rainbow (Carbons)—Color carbon atoms only with rainbow colors by residue position. The chain is divided into segments, each of which has residues of the same color.
  - Rainbow (Calpha)—Color alpha carbon atoms only with rainbow colors by residue position.

- Rainbow—Color all atoms with rainbow colors by residue position.
- B Factors—Color atoms by the residue B factor.
- B Factors (Calpha)—Color alpha carbon atoms only by the residue B factor.

- Auto—Color groups of atoms via a cycle of colors:
    - Carbons—color carbon atoms by the next color in the cycle.
    - All Atoms—color all atoms by the next color in the cycle.
    - Carbons by Object—color carbon atoms a different color for each entry.
    - All Atoms by Object—color all atoms a different color for each entry.
- Custom Color All Atoms—Change the color of all atoms to a single custom color. Clicking on the menu item opens a palette of colors to choose from, while selecting Recent Color Choices lists the most recent colors chosen by this command.

2.4 Shortcut Menus

The Workspace has two shortcut (context) menus when the Toggle Table panel is open. One is for the selected atoms, and one is for the entire Workspace. To show the shortcut menu for the Workspace, right-click and hold in an empty part of the Workspace. To show the shortcut menu for the selection, right-click and hold over one of the selected atoms. If you right-click and hold over an unselected atom, the selection changes: it's the same as picking that atom first, then right-clicking and holding on it.

Most of the menu items are the same as those on the Toggle Table buttons or button menus. The Selection shortcut menu has a Disable item, which turns off the selection. The Workspace shortcut menu has Enable and Disable actions, which you can use to display or undisplay any of the Workspace rows. These actions are also available from some of the submenus. This menu also has a Select action, which you can use to create a selection from the visible (displayed) atoms. It also allows you to operate on the visible atoms only or on all atoms in the Workspace.

Analyzing Protein Quality

The quality of a protein structure is often measured by deviations from values reported in the PDB. You can analyze a protein and display tabular and graphical reports on its quality in the Protein Structure Quality Viewer panel, which you open by choosing Tools → Protein Structure Quality.

If there is a protein in the Workspace, it is analyzed when you open the panel. Otherwise, you can display a protein in the Workspace and click Analyze Workspace to perform the analysis.

At the top of the panel, the protein table lists the chains in the protein that is analyzed along with various measures of the overall structure quality. You can analyze multiple proteins and they are all listed in the table, and you can select multiple chains in a single protein for reporting, but you cannot select multiple proteins.

The remainder of the panel consists of two tabs that show different data: the Ramachandran Plot tab, and the Protein Report tab.

3.1 Ramachandran Plot

The Ramachandran Plot tab displays a Ramachandran plot of the dihedral angles. Glycines are plotted as triangles, prolines are plotted as squares, all other residues are plotted as circles. The plot has three regions: favored, allowed, and disallowed, which are colored by the color scheme selected from the option menu at the bottom of the tab. There are two schemes: Green/Cyan/Red and Red/Yellow/White, for coloring the favored, allowed and disallowed regions.

Pausing the cursor over a point displays information for that residue at the top of the panel, and highlights the residue in the Workspace. Clicking on a point selects the point and zooms the Workspace image in to that residue, and highlights it with pale yellow markers. The point is displayed as an outline instead of solid black. The residue information is displayed at the top of the panel. Click again on the point to deselect it.

The plot area has a toolbar, which is described in Section 3.3.

Below the plot, you can select options to change the appearance of the residues in the Workspace structure for each of the three regions. The appearance is changed by coloring the residues and modifying the molecular representation. The coloring is applied when you analyze the Workspace or change the color scheme, so you should set these options first, and then click Analyze Workspace or change the color scheme. Deselecting any of these options does not

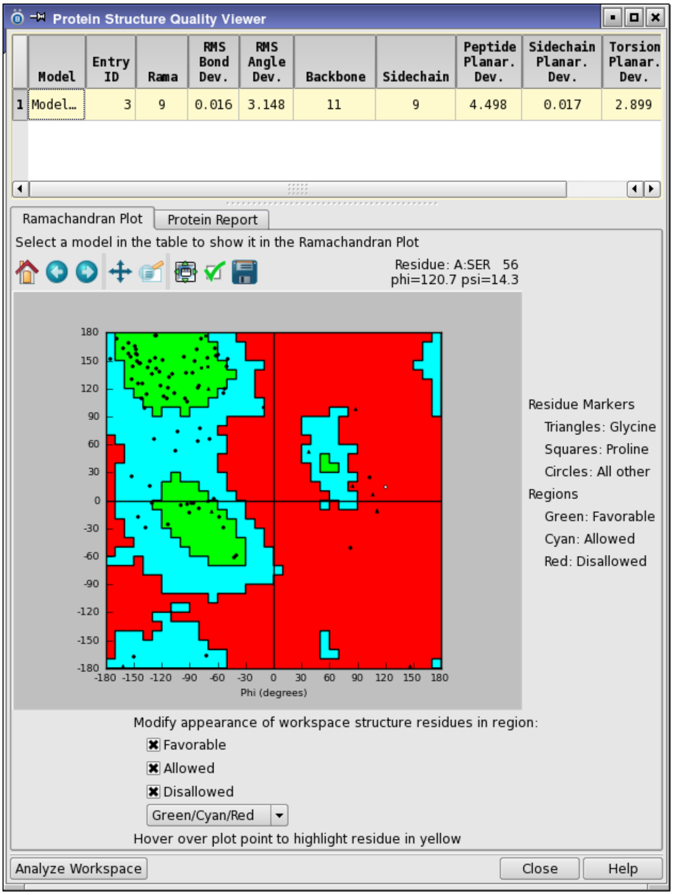


Figure 3.1. The Protein Structure Quality Viewer panel, Ramachandran Plot tab.

revert the color scheme to the original scheme, so you must change the scheme manually to revert it.

3.2 Protein Report

The property table displays a list of values of the property chosen from the Display menu. Selecting table rows zooms the Workspace view in to the structural features listed in those rows, highlights them with a change in representation, and selects them. The average of the selected property over the entire structure is displayed below the table.

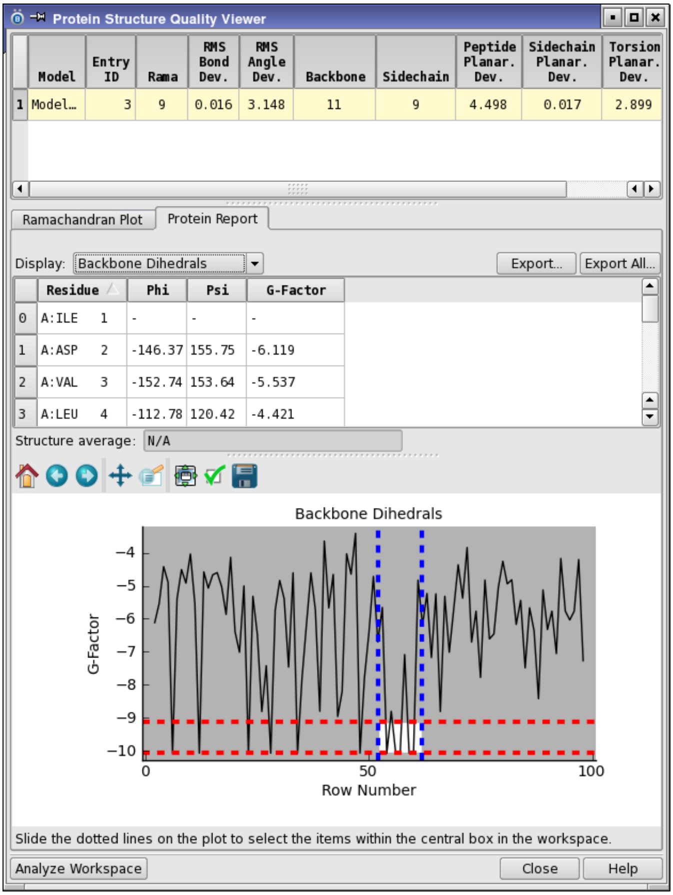


Figure 3.2. The Protein Structure Quality Viewer panel, Protein Report tab.

The property is plotted as a function of the row number in the table in the area below the table. The plot area has a toolbar, which is described in Section 3.3. The red dashed horizontal lines and blue dashed vertical lines can be dragged to highlight a portion of the plot of interest, which is given a white background, and the rest is gray. The atoms associated with the highlighted portion of the plot are interactively selected in the Workspace, and the rows for the points in the highlighted region are selected in the table.

If you want to export the values in the table, click Export or Export All. Export exports the current property table as a text file. Export All exports all property tables as a text file. Both buttons open a file selector in which you can navigate to the location and name the file.

3.3 Plot Toolbar

Both tabs have graphical displays, for which a toolbar provides some tools for manipulation of the plot and for saving an image of the plot. This is a generic toolbar, and some of the actions may not be useful in the current context. The toolbar buttons are described below.

Reset
Reset the plot to the original pan and zoom settings.

Back
Display the previous view of the plot in the view history

Next
Display the next view of the plot in the view history

Pan/zoom
Pan the plot with the left mouse button, zoom with the right mouse button.

Zoom to rectangle
Drag out a rectangle on the plot to zoom in to that rectangle.

Configure subplots
Configure the margins and spacing of each plot in the panel.

Edit axis and curve parameters
Make settings for the title, range, labeling, and scale of the axes; the color, style, and width of lines; and the color, style, and size of markers.

Save image
Save an image of the plot to file. Opens a file selector in which you can browse to a location, select the image format, and name the image.

Analyzing Residue Properties

Identifying stable or unstable residues, or residues with desirable or undesirable properties may be a useful precursor to mutation studies. The Residue Analysis panel analyzes a protein to produce properties of the residues, including hydropathy, various energies, solvent-accessible surface areas, and rotatable bonds. These properties can be used to identify residues with either desirable properties or undesirable properties, which may suggest residues that could be mutated to improve the protein properties.

To open the Residue Analysis panel choose Tasks → Residue Analysis in the main window.

Before analyzing a protein, you should prepare it with the Protein Preparation Wizard. Calculating the energetic properties requires an all-atom structure with bond assignments. To analyze the properties of a protein, first display it in the Workspace, and then click Analyze Workspace. A job is run to calculate energetic properties. This job can take several minutes, and progress is reported in a bar at the bottom of the panel.

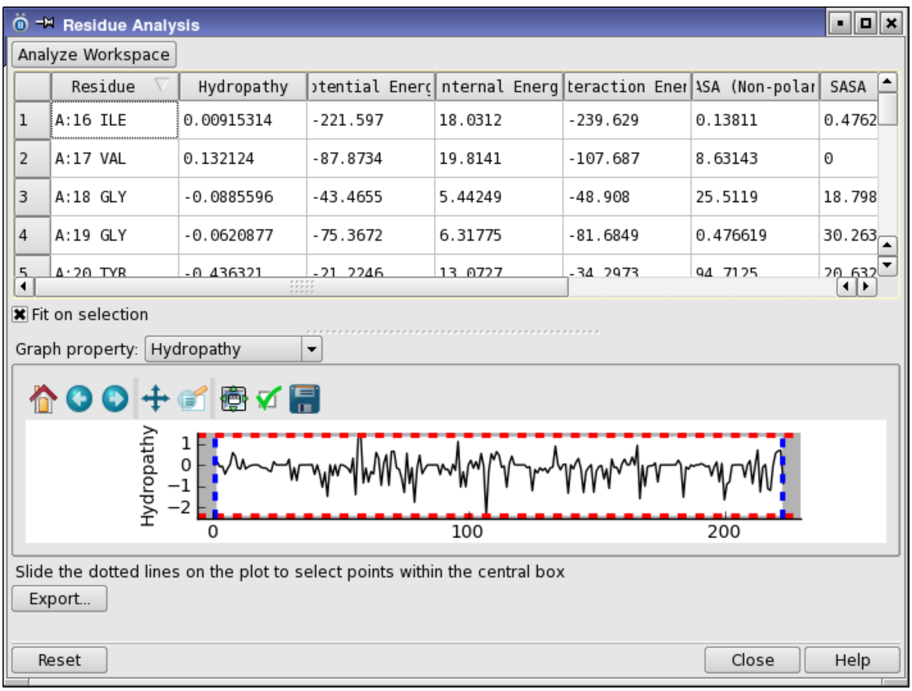


Figure 4.1. The Residue Analysis panel, with results.

When the job finishes, the table is filled in and the first property is plotted below. You can sort the table by clicking on the heading of the column you want to sort by. A second click changes the sort direction. The table columns are described in Table 4.1.

Table 4.1. Property columns in the Residue Analysis panel.

Column	Description
Residue	Residue identity: chain, residue number and insertion code, 3-letter name.
Hydropathy	Hydropathy calculated using the Kyte-Doolittle scale [6], normalized by the solvent accessible surface area.
Potential Energy	Sum of the residue-based internal energy and the non-bonded interaction energy (vdW, electrostatic) between the residue and the remainder of the system.
Internal Energy	Sum of energies arising from intra-residue bonded interactions (bonds, angles, torsions) and intra-residue non-bonded interactions (vdW, electrostatic).
Interaction Energy	Energy of interaction between this residue and all other atoms.
SASA (Non-polar)	Solvent-accessible surface area of nonpolar atoms of this residue
SASA (Polar)	Solvent-accessible surface area of polar atoms of this residue
SASA	Total solvent-accessible surface area of this residue
Rotatable bonds	Number of rotatable bonds in this residue

You can export the table data to a CSV file by clicking Export and then providing the file name in the file selector that opens.

To highlight one or more residues in the Workspace, select the table rows. If you have Fit on selection selected, the view zooms in (or out) so that these residues occupy most of the Workspace.

To examine a particular property for all residues, you can make a plot of the property as a function of residue position. Choose the property from the Graph property option menu to display the plot. You can use the plot to select residues in the table and highlight them in the Workspace, by moving the dotted lines to enclose the residues you are interested in. For example, you might want to select residues that have significantly larger or significantly smaller values of the properties than the average, by moving the upper or the lower red dotted line to enclose just those data points.

The toolbar provides tools for manipulation of the plot and for saving an image of the plot. This is a generic toolbar, and some of the actions may not be useful in the current context. The toolbar buttons are described below.

Reset
Reset the plot to the original pan and zoom settings.

Back
Display the previous view of the plot in the view history

Next
Display the next view of the plot in the view history

Pan/zoom
Pan the plot with the left mouse button, zoom with the right mouse button.

Zoom to rectangle
Drag out a rectangle on the plot to zoom in to that rectangle.

Configure subplots
Configure the margins and spacing of each plot in the panel.

Edit axis and curve parameters
Make settings for the title, range, labeling, and scale of the axes; the color, style, and width of lines; and the color, style, and size of markers.

Save image
Save an image of the plot to file. Opens a file selector in which you can browse to a location, select the image format, and name the image.

If you want to analyze another protein, click Reset to clear all the panel data.

Identifying Consensus Molecules

The Consensus Visualization panel helps you to identify conserved waters, counter ions, and ligands for a protein. To open this panel, choose Tools → Protein Consensus Viewer.

Homologs of the target protein are identified by a BLAST search. You can select a subset of these homologs to determine the consensus between them for the locations of waters, counter ions, and ligands. These homologs are aligned, both by sequence and by structure, to the target protein. The consensus between the positions of the waters, counter ions and ligands is then determined. Consensus analysis can help you to quickly identify moieties that are repeated among multiple structures, such as structurally important waters that should be included in modeling studies.

For a tutorial introduction to consensus visualization, see Chapter 5 of the *BioLuminate Quick Start Guide*.

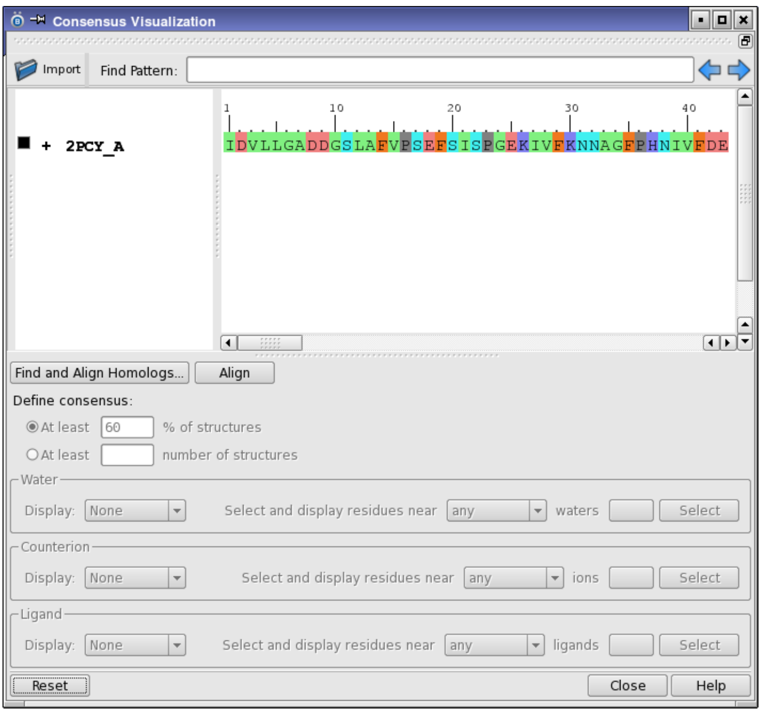


Figure 5.1. The Consensus Visualization panel after import.

5.1 Choosing the Target Protein

First, you must import your target protein. You can do this by clicking Import and choosing a source. The choices are:

- Browse for File—Open a file browser in which you can navigate to the desired location and select the file that contains the structure. The allowed file types are Maestro and PDB.
- From PDB ID—Import the structure from the specified PDB ID. Opens the Enter PDB ID dialog box, in which you can enter the PDB ID of the structure. The structure is retrieved from a local copy of the PDB if it is available, or from the RCSB web site, depending on the preference set for PDB retrieval.
- From Workspace—Import the structure that is displayed in the Workspace.

If you choose to prepare the protein beforehand in the Protein Preparation Wizard panel, you should ensure that you do not delete any of the molecules for which you are seeking a consensus. In particular, you might want to deselect Delete waters beyond *N* Å from het groups, or make the distance large enough to ensure that you have the relevant waters.

5.2 Finding and Aligning Homologs

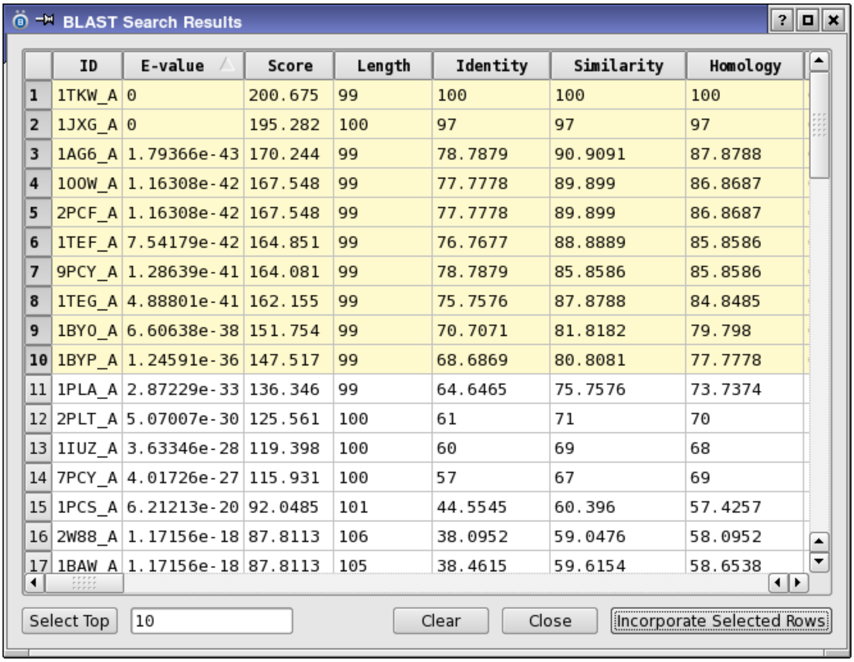
The homologs that are used to identify consensus molecules can be imported, or you can run a BLAST search to locate homologs.

If you already have a set of homologs, you can simply import them with the Import button. Once they are imported you can align them structurally by clicking Align.

To run a BLAST search for homologs, click Find and Align Homologs. First, the Blast Search Settings panel opens. You do not usually need to change the BLAST search settings. When you are satisfied with the settings, click Start Job. A Job Progress dialog box replaces the Blast Search Settings dialog box, and displays the log file from the BLAST search.

After a few minutes, the job finishes and the BLAST Search Results dialog box opens, with the results of the search. The top ten results are selected in the table by default. You can change the number of top rows to select by entering the number of rows in the text box below the table, and clicking Select Top. You can also manually select rows in the table.

When you have finished selecting rows, click Incorporate Selected Rows. If you do not have a local installation of the BLAST or PDB databases, the search is done on the web, and a warning is displayed: “Multiple Sequence Viewer is attempting to access a remote server. Would you like to continue?” You can select Do not ask this question again, to prevent it from opening each time a structure is downloaded, then click OK.



	ID	E-value	Score	Length	Identity	Similarity	Homology
1	1TKW_A	0	200.675	99	100	100	100
2	1JXG_A	0	195.282	100	97	97	97
3	1AG6_A	1.79366e-43	170.244	99	78.7879	90.9091	87.8788
4	1OOW_A	1.16308e-42	167.548	99	77.7778	89.899	86.8687
5	2PCF_A	1.16308e-42	167.548	99	77.7778	89.899	86.8687
6	1TEF_A	7.54179e-42	164.851	99	76.7677	88.8889	85.8586
7	9PCY_A	1.28639e-41	164.081	99	78.7879	85.8586	85.8586
8	1TEG_A	4.88801e-41	162.155	99	75.7576	87.8788	84.8485
9	1BYO_A	6.60638e-38	151.754	99	70.7071	81.8182	79.798
10	1BYP_A	1.24591e-36	147.517	99	68.6869	80.8081	77.7778
11	1PLA_A	2.87229e-33	136.346	99	64.6465	75.7576	73.7374
12	2PLT_A	5.07007e-30	125.561	100	61	71	70
13	1IUZ_A	3.63346e-28	119.398	100	60	69	68
14	7PCY_A	4.01726e-27	115.931	100	57	67	69
15	1PCS_A	6.21213e-20	92.0485	101	44.5545	60.396	57.4257
16	2W88_A	1.17156e-18	87.8113	106	38.0952	59.0476	58.0952
17	1BAW_A	1.17156e-18	87.8113	105	38.4615	59.6154	58.6538

Figure 5.2. The BLAST Search Results dialog box.

If an information box opens stating that problems were found when importing a structure, you can select Do not show this dialog again to prevent it from opening for each structure that has problems, and click OK. The structures are imported without any preprocessing, so they might have structural defects. For the purposes of this panel, it is generally acceptable to use structures from the PDB that have structural issues.

The homologs you selected are aligned, added to the sequence viewer, and displayed in the Workspace. All atoms are marked in all of the homologs.

5.3 Viewing Consensus Molecules

After the results are available, the next task is to decide how to define a consensus between the set of structures. There are two choices, available under Define consensus: a minimum percentage or a minimum number of structures. A match between the parent and a homolog for a particular molecule (such as a water) is obtained when any atom in the molecule in a homolog is within 2 Å of any atom in the same type of molecule in the reference (parent) structure. Consensus occurs when a match is found for the specified number or percentage of homologs.

For each of the three types of molecules, you can perform the following actions:

- Choose whether to display all, only the consensus, or none of the molecule of the given type, from the Display option menu. For example, viewing all the molecules gives an indication of whether a consensus exists, whereas viewing the consensus shows whether there is a strong enough consensus to consider the molecule as conserved.
- Select and display residues near the molecules of the given type. You can choose to display residues near any of the molecules or only the consensus molecules. The action is not performed until you click Select.
- Change the color of the residues that are near the molecules of the given type, by clicking on the color button, and choosing a color in the color selector that opens.

When displaying the molecules, the identity of any consensus molecule can be ascertained by moving the cursor over the structure in the Workspace and viewing the text in the status bar, below the Workspace.

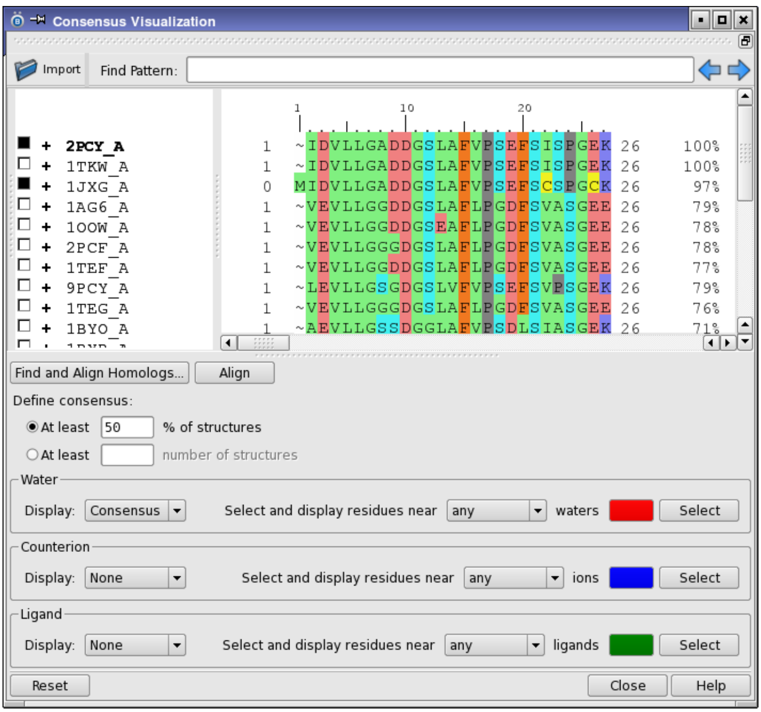


Figure 5.3. The Consensus Visualization panel with results.

Identifying Reactive Residues

You may need to identify reactive residues in a protein, so that you can mutate them to improve the protein properties. This can be done in the Reactive Protein Residues panel, which you open by choosing Tools → Reactive Residue Identification.

Reactive residues are identified by matching residue patterns in the sequence. Four patterns are provided by default, for the common reactions: deamidation, oxidation, glycosylation, and proteolysis. You can use these patterns or you can set up and use your own patterns.

To identify reactive residues, include the protein you want to analyze in the Workspace, and click Analyze Workspace. The structure in the Workspace is analyzed to identify residues that match the patterns.

The results are listed in the table, which shows the reaction type, the reactive residues identified, the solvent-accessible surface area of the reactive residues, their percentage exposure to

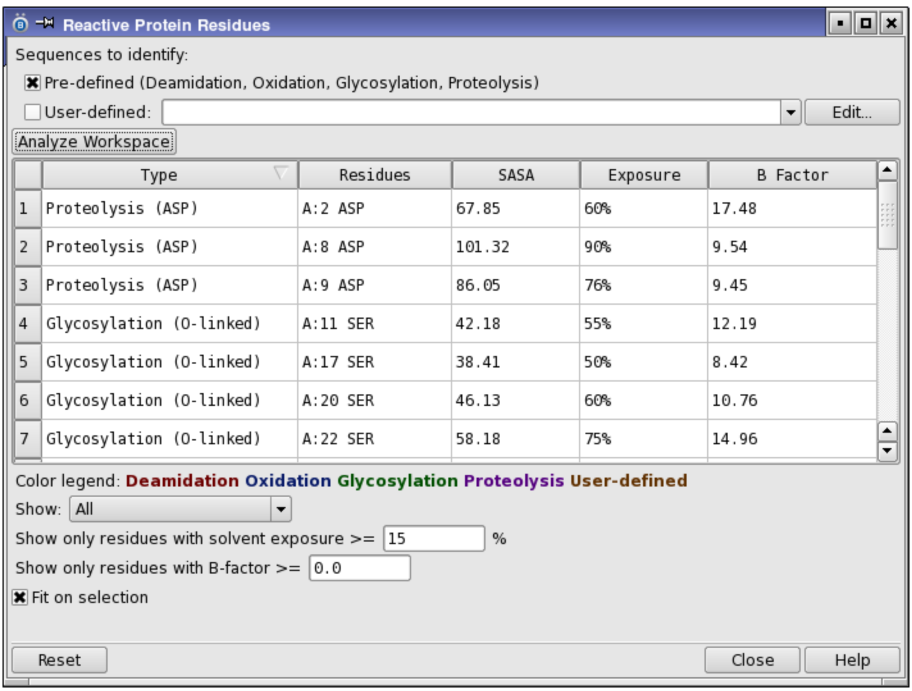


Figure 6.1. The Reactive Protein Residues panel.

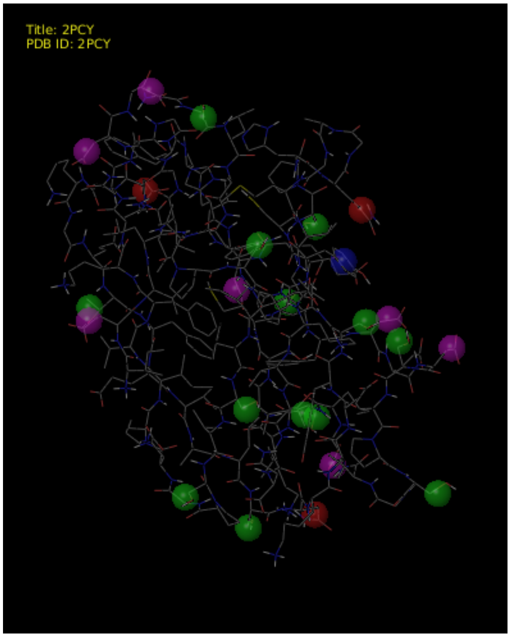


Figure 6.2. Reactive residue sites in 2PCY.

solvent, and their B-factors (if available). The B-factor shown is the average over all atoms in the residue.

You can use the Show option menu to show only the results for a particular reaction type. You can also apply filters on the percentage solvent exposure and the B-factor.

To sort the table by the values in a column, click on the column heading. Click again to change the direction of the sort. The column by which the table is sorted is indicated by an arrow on the right side of the heading, which also indicates the sort direction.

The reactive sites are marked with spheres in the Workspace. The spheres are colored according to the reaction type, and the color legend for the spheres is given below the table. When you select a table row, the residue is highlighted in the Workspace. If you have Fit on selection selected, the view zooms in to that residue

If you want to define your own reactive groups, click Edit, to open the Edit Patterns dialog box. This dialog box lists the patterns in a table, giving the pattern name, the definition, and a hotspot index. To edit an existing pattern, double-click the table cell that you want to edit, and enter the changes. The lower part of the panel explains the syntax for the patterns in the Definition column. The syntax is an extended PROSITE syntax, which allows you to specify secondary structure and some properties:

- Standard IUPAC one-letter (upper case) codes are used for all amino acids.
- Lower case x is used for any amino acid.
- Each element of a pattern is separated with a - symbol.
- Residues that are permitted at a given position are listed between square brackets, e.g. [ACT] means one of Ala, Cys, or Thr, or in other words, only Ala, Cys, or Thr can appear at this position.
- Residues that are not permitted at a given position are listed between curly brackets, e.g. {GP} means not Gly and not Pro, or in other words, any residue but Gly or Pro can appear at this position.
- Repetition is indicated using parentheses, e.g. A(3) means Ala-Ala-Ala, G(2,4) means between 2 to 4 Gly residues.
- The following lower case characters can be used for residue types:
  - a—acidic residue: [DE]
  - b—basic residue: [KR]
  - o—hydrophobic residue: [ACFILPWVY]
  - p—aromatic residue: [WYF]
- The following lower case characters can be used to restrict residue types by property:
  - s—solvent-exposed residue
  - h—residue in helical region
  - e—residue in extended (beta strand) region
  - f—flexible residue, defined as having a B-factor above the average over all residues

  These four characters can be appended to a residue type to restrict the type of residue, e.g. Ah means Ala in a helical region.

Some examples of valid and invalid patterns are given below, with comments.

N-{P}-[ST]	Asn-X-Ser or Asn-X-Thr, X is not Pro
N[fs]-{P}[fs]-[ST][fs]	as above, but all residues flexible or solvent exposed
Nfs-{P}fs-[ST]fs	as above, but all residues flexible and solvent exposed
Ns{f}	Asn, solvent exposed and not in flexible region
N[s{f}]	Asn, solvent exposed or not in flexible region
[ab]{K}f{s}	acidic OR basic, except for flexible and non-solvent-exposed Lys
Ahe	Ala, helical and extended - no match is possible
A[he]	Ala, helical or extended

A{he}	Ala, not helical or extended
[ST]	Ser or Thr
ST	Ser and Thr - no match possible

The hotspot index is the index of the reactive residue in the pattern.

Predicting Aggregation Regions

Protein aggregation often occurs via hydrophobic regions. You can locate these regions using the Aggregation Surface panel, which you open by choosing Tasks → Aggregation Surface. Aggregation regions are defined by identifying clusters of exposed hydrophobic residues, and creating a molecular surface that is colored red in the regions near these residues.

Once you have located potential aggregation regions on a protein, you might want to mutate residues in these regions to reduce the tendency to aggregate.

7.1 Creating an Aggregation Surface

Follow the steps below to create a surface that displays the likely aggregation regions.

1. (Optional) Select atoms to define the structure for the surface.

   If you want to create a surface for the entire protein and you have only one entry in the Workspace, you can skip this step. Otherwise, you can select atoms to create a surface for an entire entry, a chain, or the selected atoms with their hydrogens.

2. Choose an option for the part of the Workspace structure that you want to create the surface for.

   There are several options. If no atoms are selected in the Workspace, only the first is available.

   - Workspace—Compute the surface for the structure in the Workspace. You must have only a single entry in the Workspace.

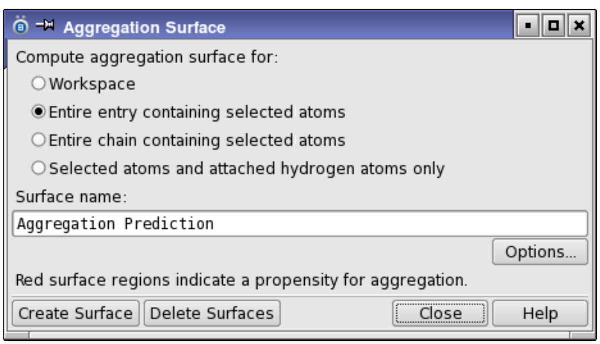


Figure 7.1. The Aggregation Surface panel.

- Entire entry containing selected atoms—Compute the surface for the entry that contains the selected atoms. This is useful if you want to view how the aggregation surface on one protein matches residues in another protein.
- Entire chain containing selected atoms—Compute the surface for the entire chains that contain the selected atoms. This is useful for computing a surface for one chain of a multi-chain protein, for example.
- Selected atoms and attached hydrogen atoms only—Compute the surface for the selected atoms with their attached hydrogens only. This option allows you to calculate the surface for only part of a protein chain, for example.

3. Enter a name in the Surface name text box, if you want to change the default name.

   This name is used in the Manage Surfaces panel (Style → Create and Manage Surfaces → Surface Manager) to identify the surface.

4. Click Create Surface.

   A progress bar is displayed above the buttons while the surface is being created. Surface creation should take less than a minute.

When the surface is created, it is displayed in the Workspace and other surfaces are hidden. To display more than one surface, use the Manage Surfaces panel (Window → Show → Surface Manager).

The aggregation surface is a molecular surface that is created for a large probe molecule, representing part of a protein. It is colored red in the regions near the hydrophobic residue clusters. The side chains of these residues and the alpha carbons are also colored red, and the remaining backbone atoms are colored gray. The side chains are displayed in ball-and-stick representation, and the backbone is displayed as lines. Residues that are in contact with these residues are displayed as lines in gray. All other residues are hidden, so you see only the residues that contribute to the aggregation regions and very near neighbors. The surface is semi-transparent, so you can see the atoms inside the surface. The residues involved in the clusters are colored red in the Workspace sequence viewer.

Because the surface color scheme depends on the atom coloring, the surface is removed if you change the atom color scheme. For a similar reason, the surfaces are removed when you close the panel. However, if you don't make any changes, the atom visibility and the color scheme for the residues are still in effect when you close the panel even though the surface is removed.

If you want to remove the surfaces and redisplay all atoms, click Delete Surfaces.

7.2 Setting Options for Aggregation Surfaces

You can control some of the parameters of the quality and appearance of the aggregation surface and the parameters that are used for locating aggregation regions in the Aggregation Surface - Options dialog box. To open this dialog box, click Options in the Aggregation Surface panel. The parameters you can set are:

- Surface grid spacing—Set the grid spacing in angstroms for generation of the surface. A smaller number results in a smoother surface, but takes longer to generate the surface.
- Probe radius—Set the radius of the probe for defining the surface. This is the radius of the sphere that is rolled over the van der Waals surface to create a Connolly surface. The large default radius is intended to model a protein probe. Hydrogens are included when creating the surface. See Section 12.1.2 of the *Maestro User Manual* for details on the construction of the surface.
- Transparency—Set the default transparency of the surface. The transparency can be changed in the Surface Display Options dialog box—see Section 12.4.2 of the *Maestro User Manual*.
- Radius to find neighbors—Specify the distance cutoff for finding hydrophobic neighbors to identify aggregation regions. The distance between any side-chain heavy atom in one hydrophobic residue and any side-chain heavy atom in another hydrophobic residue must be less than this cutoff for the residues to be counted as neighbors.
- Hydrophobic neighbors required for site—Specify the minimum number of hydrophobic neighbors required to include a residue (site) in an aggregation region.
- Buried residue SASA—Specify the maximum solvent-accessible surface area (SASA) for a residue to be regarded as buried. Buried residues are not included in aggregation regions. The SASA is measured for the heavy atoms only, and does not include hydrogens.

7.3 Using the Results for Mutation Studies

To reduce the likelihood that the protein will aggregate, you might want to mutate residues in the aggregation regions.

One way of doing this is to perform a set of mutations to reduce the size of the aggregation regions. You can create a set of structures that have single mutations at selected sites, which you choose from the residues in the aggregation regions, as follows:

1. Right-click in the Workspace (not on an atom) and choose Visible → Select.

   The visible atoms, which include all the aggregation residues, are selected, along with the neighbors. If you want to hide the innkeepers first, you can do so, but this is not really necessary as you will be able to choose the residues that you mutate later.

2. Choose Tasks → Residue Scanning → Perform Calculation to open the Residue Scanning panel.

3. Select Analyze only selected Workspace residues.

4. Click Analyze Workspace. You may be prompted to choose a chain, as only one chain at a time is mutated.

The residues that were found to contribute to the aggregation region are listed in the table in the Residues tab. You can then select any of them for mutation, define the mutations, and run the job. See Chapter 13 for details of setting up a residue scanning job.

Another option is to mutate a single residue or a loop. Mutating a loop (“loop swap”) is useful if you want to mutate more than one residue and the residues are adjacent. For this purpose, you can use the Residue and Loop Mutation panel. The loop to change is defined by selecting the residues in the Workspace. You can take advantage of the fact that you only have the aggregation residues and their neighbors visible in the Workspace to choose the residues for the loop to swap. See Chapter 12 for details of using this panel.

Locating Large-Scale Motions

Finding large-scale motions in proteins can provide information on the flexibility or rigidity of parts of the protein, on domain movements, or give some clues to biochemical processes. Trajectories from molecular dynamics simulations can show the large-scale motions of proteins, but these simulations are not resolved into individual modes, and they require a large amount of computational resources. The lowest vibrational modes of a protein can be determined and visualized using the Low Mode Vibrational Sampling panel. You can open this panel by choosing Tasks → Low Normal Mode Analysis → Calculate or Visualize.

Before running the calculation, you should ensure that the protein is properly prepared, using the Protein Preparation Wizard (Tools → Protein Preparation). You should remove waters and solvent molecules so that you are analyzing just the protein (and its ligands, if any).

First, the input structure is minimized (using the PRCG method for a maximum of 10000 iterations to a gradient convergence threshold of 0.05 kJ $mol^{-1}Å^{-1}$). This ensures that the structure is at its minimum, which is important for generating the vibrational modes. The vibrational modes are generated as a set of structures sampled at regular intervals along a full cycle of the vibrational mode. The rotational and translational modes (the trivial modes) are discarded. The vibrational modes are then visualized by displaying the structures in sequence as a "movie" in the Workspace.

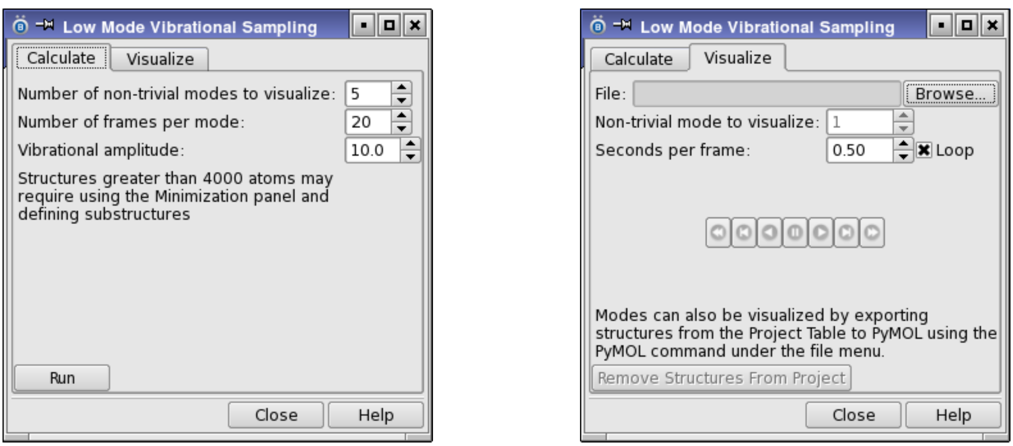


Figure 8.1. The Low Mode Vibrational Sampling panel.

To set up the calculation:

1. Select the number of vibrational modes you want to view in the Number of non-trivial modes to view text box.
2. Set the number of frames to generate per vibrational cycle in the Number of frames per mode text box.

   Each frame is a snapshot of the structure at a particular point in the vibration between the classical limits. The full cycle of a vibration is divided evenly to define the coordinates used for the snapshots, so this number should be divisible by 4.
3. Set the maximum amplitude of vibration in the Vibrational amplitude text box.

   This is the maximum displacement of the fastest-moving atom. For proteins, the fastest moving atom could potentially move a large distance if the motion involves a long loop, for example. This choice is somewhat arbitrary: you might for example want to exaggerate the motions to make them easier to see.
4. Click Run.

A job is started that generates the series of structures for each vibration. The limit on the number of atoms that can be processed depends on the memory available on the machine, but with 4GB of memory, it is about 5000. For example, 1ETT, with about 4800 atoms, runs successfully. The job can take several hours to run, depending on the size of the protein.

To visualize the vibrational modes:

1. Import the results of the job by clicking Browse and selecting the `.com` file for the job in the file selector that opens.
2. Specify the mode you want to view in the Non-trivial mode to visualize box.
3. Specify the duration of each frame, in seconds.

   The speed at which the structures can be displayed depends on the size of the structure. For structures of 5000 atoms, an interval of 0.1 sec is possible and produces a reasonable animation.
4. Select Loop to loop continuously through the frames, so that the vibration goes through multiple cycles.
5. Use the play controls to start and stop the animation.

   You can rotate the Workspace during the animation to get a better view of the moving parts of the structure.

When you have finished viewing the modes, you should clean up the structures in the project, by clicking Remove Structures from Project.

Predicting Alpha Helix Stability

Determining the stability of alpha-helical peptides is important for their use as therapeutic agents. You can obtain an estimate of the stability by using the Peptide Helicity panel, which you can open by choosing Tasks → Peptide Alpha Helicity. The primary purpose of this panel is to provide information on the relative stabilities of a series of small peptides, although you can also obtain information on a single peptide.

The alpha helical tendency of one or more peptides is determined from a molecular dynamics simulation by tracking i to i+4 hydrogen bond formation, and other indications of helical structure. The prediction is made entirely from the sequences, by building them as idealized alpha-helices, and then performing a molecular dynamics simulation in water using simulated annealing, to simulate experimental melting experiments. At the end of the simulation, averages are taken to determine the values of properties that can be used to determine the helicity of the sequences.

The simulations are set up in the Start Simulations tab, and the results are presented in the Results tab.

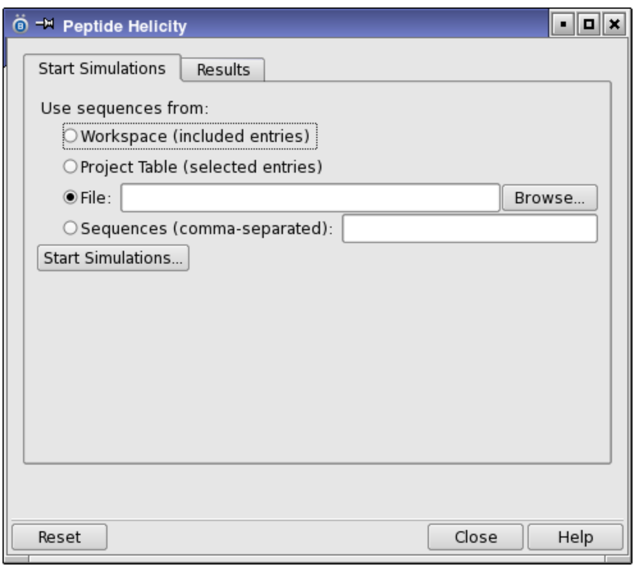


Figure 9.1. The Residue Scanning Viewer panel.

To run a simulation:

1. Select an option for the sequences to simulate under Use sequences from.

   The options are Workspace (the included entries), Project Table (the selected entries), File, or Sequences. If you choose File, enter the file name in the text box or click Browse to navigate to and select the file. If you choose Sequences, type or paste the sequences into the text box, separated by commas. Each sequence that you provide is simulated separately in a single job.

2. Click Start Simulations.

   A Start dialog box opens, in which you can set the job name, select a host, and specify the number of processors. Click Start in this dialog box to submit the job to a host for execution.

   Since the simulations are likely to take many hours, it is a good idea to run the job on a multiprocessor host. The simulation uses a 3D domain decomposition of the simulation box, so you can specify the number of processors for each dimension in the decomposition (labeled x, y, and z). See Section 3.10 of the *Desmond User Manual* for more information. A good choice is 2 processors for each, on an 8-core CPU.

When the simulation finishes, you can review the results of the simulations in the Results tab. Click Load Output File read the results, which are in a CSV file. A file selector opens, in which you can navigate to and select the output CSV file. The results are presented in a table, whose columns are described in Table 9.1. The averages are taken over the length of the simulation.

If you want to run another calculation, click Reset to clear all panel data.

Table 9.1. Columns of the Results table in the Peptide Helicity panel.

Column	Description
Title	Structure title
Structure No	Structure number
α-Propensity	Average helical propensity, defined as the fraction of (i, i+4) backbone hydrogen bonds.
H-bonds	Average number of alpha helical H-bonds
Residues Count	Average number of residues involved in two H-bonds
Residues Fraction	Average fraction of residues satisfying the phi/psi criteria for helicity individually
Dihedrals	Average fraction of alpha-helical dihedrals, as a fraction of the maximum number present in the idealized helix.

Protein-Protein Docking

The question of whether one protein binds to another, and where, can be addressed by protein-protein docking. Protein-protein docking in BioLuminate is performed using the Piper program [5], under license from Boston University. The job can be set up in the Protein-Protein Docking panel. In this panel you can set up jobs to dock two arbitrary proteins, dock an antigen to an antibody, or dock one protein to itself to form a dimer or a trimer.

One protein is treated as the "receptor" and the other as the "ligand". In the general case, it does not matter which protein is treated as the receptor and which protein is treated as the ligand. For antibody-antigen docking, the receptor is the antibody and the ligand is the antigen. The algorithm samples all possible orientations of the two proteins, subject to whatever constraints are applied. It uses a grid to locate the best poses of the two proteins, with a maximum resolution in the poses of about 5°. The docking is performed as a rigid-body optimization: there is no subsequent minimization of the interfacial region.

To open the Protein-Protein Docking panel, choose Tasks → Protein-Protein Docking.

For a tutorial introduction to protein-protein docking, see Chapter 10 of the *BioLuminate Quick Start Guide*.

10.1 Preparation of Proteins for Docking

Although it is not necessary to prepare proteins for docking, it may be advisable to do so. Proteins from the PDB often do not have coordinates for side chains on the surface of the protein, or even for whole chains that are solvent-exposed, as these are usually more mobile and it can be difficult to determine their coordinates from the X-ray data. If these missing parts of the structure are not in the favored binding region, the docking should produce good results. If the poorly defined parts of the structure are in the binding domain, pre-docking preparation of the structure can appreciably improve results. An example of the effect of missing side chains in the binding region is given in Chapter 10 of the *BioLuminate Quick Start Guide*.

To prepare your protein for docking, you can use the Protein Preparation Wizard (see the *Protein Preparation Guide*). When you prepare your structure, you should select Fill in missing side chains using Prime to predict the side chains, and Fill in missing loops using Prime to predict the loops. The loop prediction used in the protein preparation is the faster look-up method, rather than the more extensive ab initio loop building. Both of these predictions can take several minutes.

If you are concerned about the accuracy of the surface side chains or loops, you can do a more extensive prediction in the Refinement panel (Tasks → Loop and Sidechain Prediction). If you have access to the X-ray data, you could consider performing some refinement of the structure with PrimeX, the X-ray refinement program. Choose Tasks → Advanced Tasks → Protein X-Ray Refinement → Display Toolbar, and use the toolbar to perform the refinement tasks. See the *PrimeX User Manual* for more information.

In the docking experiment, however, the accuracy of the methods may not be sufficient to distinguish between conformations, so an extensive prediction of the surface side chains or loops is probably not necessary. The presence of the side chain or loop in the right region is likely to be more important than its exact conformation. If you want to test the effects of loop or side chain conformations, you can generate multiple conformations and dock each of them.

10.2 Docking a Protein to Another Protein

To dock a protein to another protein, without any special conditions on the type of protein, choose Standard in the Mode section of the Protein-Protein Docking panel.

Next, choose the receptor protein and the ligand protein. It does not matter which of your two proteins you choose as the receptor and which you choose as the ligand. To select the structures, click Receptor or Ligand in the Protein Structures section, and choose a source from the menu that is displayed. The menu items are the same for each menu:

- Browse for File—Opens a file selector so that you can browse to the location of the file and select it. The structure is added to the project and to the Workspace.
- From the Workspace—Use the structure that is in the Workspace. You should ensure that only the desired structure is included in the Workspace. If you have prepared the protein with the Protein Preparation Wizard, the structure should already be in the Workspace.

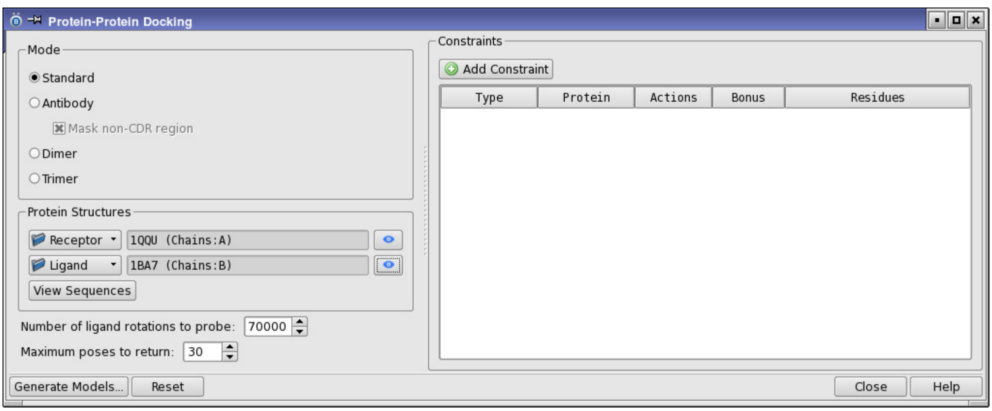


Figure 10.1. The Protein-Protein Docking panel.

- From PDB ID—Opens a dialog box in which you can specify a PDB ID. The protein you specify is imported from the PDB, either from a local copy or from the web site. It is added to the project and to the Workspace. Proteins from the PDB are likely to have atoms missing.

When you import the structures, if hydrogens are missing, you are prompted to add them. If the protein has multiple chains, you are prompted to choose the chains. You can choose more than one chain for the docking.

When the protein is imported, the structure is added to the project and included in the Workspace. If the structure is removed from the Workspace you can add it with the Include and zoom button (eye icon).

To view the sequences of the proteins you imported in a sequence viewer, click View Sequences.

When you have selected the proteins to dock, you can set two parameters to control the docking. The first is the number of ligand orientations to the receptor that are sampled. You can set the value in the Number of ligand rotations to probe box The default of 70,000 corresponds approximately to sampling every 5° in the space of Euler angles, and is the maximum value allowed. Decreasing the number of rotations generally degrades the results, but decreases the run time.

The second parameter is the number of poses to return, which you can set in the Maximum poses to return box. Each pose is the center of a cluster that results from clustering the top 1000 results of rigid docking of the ligand. If more clusters are found than the number of poses to return, the clusters are ranked by size and poses are chosen from the largest clusters. If fewer clusters are found than the maximum number of poses to return, one pose per cluster is returned.

After setting the parameters, click Generate Models to run the docking job. A Start dialog box opens, in which you can choose whether to append the results to the project (Append new entries) or leave them on disk (Do not incorporate), choose a host to run the job on, name the job, and start it. Docking jobs are run on a single processor. A typical docking job with the default parameters takes several hours.

10.3 Applying Constraints

You can add constraints in the docking process, by providing an additional attractive term to the potential, remove the attractive potential, or declare residues to be buried, for residues that you select. Constraints are implemented as a bias during the docking process. They increase or decrease the likelihood of the specified interaction, but do not guarantee that the specified restraint will be met by all top ranking poses.

Constraints can be added in the Constraint section of the panel. Each constraint is represented by a row in the table in this section. To add a new constraint row to the table, click Add Constraint. All of the settings needed to define the constraint can be made in the cells of the table.

Table 10.1. Columns in the Constraints table.

Column	Description
Type	Choose the type of constraint to apply. The types are: Attraction—Increase the attractive potential for participation in binding. The value in the Bonus column is added to the default value of 1 to define the scaling factor for the attractive potential. Buried—Increase the van der Waals radius and the repulsive potential for buried residues, and set the attractive potential for ligand or antigen residues to zero. Repulsion—Set the attractive potential to zero, leaving only the repulsive van der Waals potential, which is not modified.
Protein	Choose the protein that the constraint applies to. The choices on the menu are the same as the labels on the button menus in the Protein structures section.
Actions	This column has two action buttons, one for deleting the constraint, and the other for zooming in on the atoms that define the constraint.
Bonus	Define the increase in the attractive potential for Attraction constraints. The value must be in the range 0.11 to 0.99. This value is added to 1 to define a scaling factor for the attractive potential.
Residues	Lists the residues affected by the constraint, and provides tools to select the residues. If you click in a table cell in this column, you can choose from two sources of the selected residues: From Workspace selection—Use the current Workspace selection to define the residues for the constraint. Choose Workspace atoms—Open the Atom Selection dialog box, so you can select the residues for the constraint. The selection you make is filled out to complete any residues that are only partly selected.

To define the constraint, you must choose the constraint type in the Type column, choose the protein to apply it to in the Protein column, set a value in the Bonus column if it is an attractive constraint, and then select the residues to apply the constraint to by using the Residues column. You can select the residues in the Workspace, then choose From Workspace selection, or you can choose Choose Workspace atoms, then use the Atom Selection dialog box to select the residues for the constraint. See Section 6.5 of the *Maestro User Manual* for information on using the Atom Selection dialog box.

To delete a constraint, click the red minus icon in the Actions column.

10.4 Creating Dimers and Trimers

You can dock a protein to itself to create a dimer or a trimer. To do so, choose Dimer or Trimer in the Mode section of the Protein-Protein Docking panel. When you do, only one of the menus in the Protein structures section is available, and it is labeled Monomer. You can select the monomer protein in the same way as for the general case, and set up and run the job in the same way, as described in Section 10.2.

The dimer and the trimer are subject to symmetry constraints: the dimer must have a twofold axis of rotation (C_2 axis), and the trimer must have a threefold axis of rotation (C_3 axis).

10.5 Docking an Antigen to an Antibody

The docking of an antigen to an antibody can be performed, with constraints for the target region of the antibody. To dock an antigen to an antibody, choose Antibody in the Mode section. If you want to prevent docking to the non-CDR region, ensure that Mask non-CDR region is selected. The receptor is the antibody and the ligand is the antigen.

As for the other types of docking, you can choose the antibody and antigen in the Protein structures section. After prompting you to add missing hydrogens, the antibody is analyzed to locate the CDR regions and determine which are the light and heavy chains. A progress bar is displayed while the analysis is done. When the analysis finishes, another dialog box opens, prompting you to select the chains to use for the antibody (receptor). You must select two chains: a light chain and a heavy chain. When selecting the antigen, an alert box opens, warning you about the chains. You can dismiss this box.

Apart from these changes, docking an antigen to an antibody is set up and run in the same way as a general protein-protein docking job.

Homology Modeling of Proteins

If you have a protein sequence and want to build a homology model of the protein, there are three ways that you can proceed in the BioLuminate interface.

Proteins for which you expect a high homology with the template and that require only a straightforward alignment to the template can be modeled in the Homology Model panel. In the default mode, any missing loops are predicted using a curated database of known loops in the PDB. This approach is very fast and a full homology model can typically be generated using this panel in 2-5 minutes. To open this panel, choose Tasks → Homology Modeling → Simple Homology Modeling. The use of this panel is described below.

Proteins where the homology is not as high or where alignment of the template and the query is required can be modeled either in the Multiple Sequence Viewer panel (opened with Tools → Multiple Sequence Viewer), or with the Structure Prediction panel (opened with Tasks → Homology Modeling → Advanced Homology Modeling). For information on using these panels, see the *Multiple Sequence Viewer* document for the Multiple Sequence Viewer and the *Prime User Manual* for the Structure Prediction panel.

If you are interested in homology modeling of antibodies, see Chapter 15.

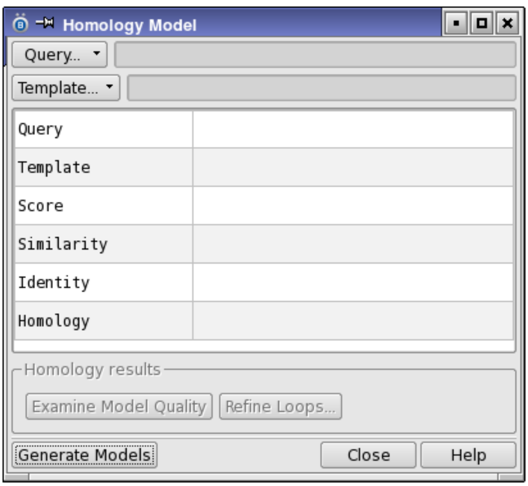


Figure 11.1. The Homology Model panel, initial view.

For a tutorial introduction to homology modeling using the Homology Modeling panel, see Chapter 6 of the *BioLuminate Quick Start Guide*. For a tutorial introduction to advanced homology modeling, see the *Prime Quick Start Guide*.

To build a homology model:

1. Choose a source for the query (or reference) sequence using the Query button. There are two choices:
   - From Workspace—Use the sequence of the structure in the Workspace as the query.
   - Browse for File—Opens a file selector so you can locate and import the sequence file for the query.

   When you have chosen a query, the box to the right of the button displays the text (Query structure) and the title of the query, if it has a title, or Query, if no title is available.

2. Choose a source for the template structure on which to build the model, using the Template button.

   You can either read in a template structure from a Maestro file or a PDB file (Browse for File), or you can run a BLAST search, as follows:

   a. Choose Template → BLAST Homology Search.

   b. Change settings if desired in the BLAST Search Settings dialog box, and click Start Job.

      When the job finishes, the Job Progress dialog box is replaced by the BLAST Search Results dialog box. This dialog box lists the homologs found with various measures of the match: E-value, Score, Identity, Similarity, and Homology.

   c. Choose a template from the list of homologs, and click Select Template.

      Blast search results are listed in order of decreasing homology score. By default, the first structure in the list is chosen, and in most cases this will be the most appropriate selection.

      If you do not have a local installation of the BLAST or PDB databases, a warning is displayed by default: “Multiple Sequence Viewer is attempting to access a remote server. Would you like to continue?” If you have not already turned this warning off, Select Do not ask this question again, to prevent it from being displayed each time a structure is downloaded, and click OK.

      The template is selected and the BLAST Search Results dialog box closes.

   When you have chosen a template, the box to the right of the button displays the text (Template structure) and the title of the template. The table rows in the Homology Model panel are filled in with information on the template.

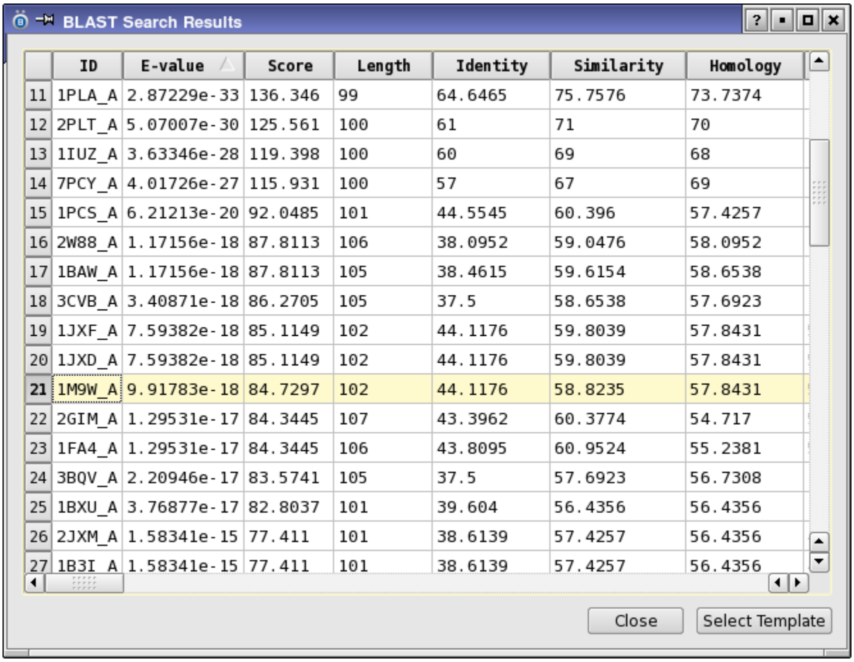


Figure 11.3. The BLAST Search Results dialog box.

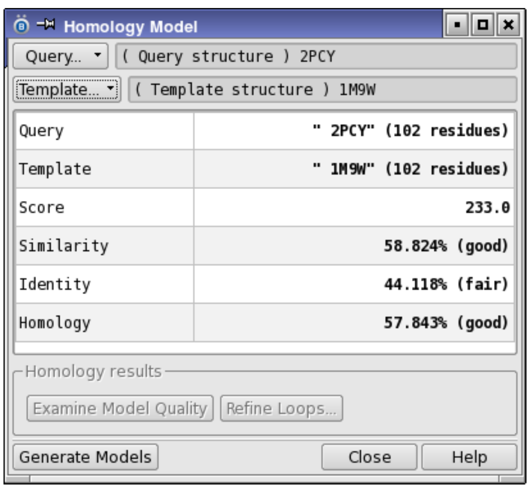


Figure 11.2. The Homology Model panel after selecting a template.

3. Click Generate Models.

   The Homology Model - Start dialog box opens. You can name the job and select a host to run it on. The job includes alignment of the template and the query using ClustalW, and the structure is built on the basis of the template and an analysis of structural elements in the PDB for non-templated regions (a "knowledge-based" selection of the coordinates).

When the job finishes, the model is added to the Workspace, in cartoon representation. The cartoon is colored by how the template was used: dark blue for residues for which all coordinates were taken from the template, cyan for residues for which the backbone was taken from the template, and red for residues that were entirely modeled, not using the template. The title given to the model includes information on the query and the template.

If the job fails, it is likely that there is insufficient homology or poor alignment between the reference and the template to build a model. In this case you should use the Advanced Homology Modeling panel (Tasks → Homology Modeling → Advanced Homology Modeling) or the Multiple Sequence Viewer (Tools → Multiple Sequence Viewer).

If you want to examine the quality of the model structure, click Examine Model Quality, to open the Protein Structure Quality Viewer panel, where you can view reports on the protein structure, a Ramachandran plot, and plots of protein properties

If you want to refine loops in the model, click Refine Loops. The Refinement panel opens with the Refine loops task selected. You should only need to refine the loops if they were not predicted from the template. Click Non-Template in the Refinement panel to list only the loops that did not come from the template. See Chapter 6 of the *Prime User Manual* for information on this panel.

Residue and Loop Mutation

At some point in a workflow, you might want to mutate a single residue, or replace a single loop with another loop. You can do this in the Residue and Loop Mutation panel, which you open from the Tasks menu.

BioLuminate provides other tools for mutations of more than one residue. The Residue Scanning panel allows you to mutate a protein at multiple sites to generate a set of proteins, each with a single mutation—see Chapter 13 for information. If you want to mutate residues to or from cysteine and break or form disulfide bonds, you should use the Cysteine Mutation panel—see Chapter 14 for information.

The workflows in this panel are described in the following sections.

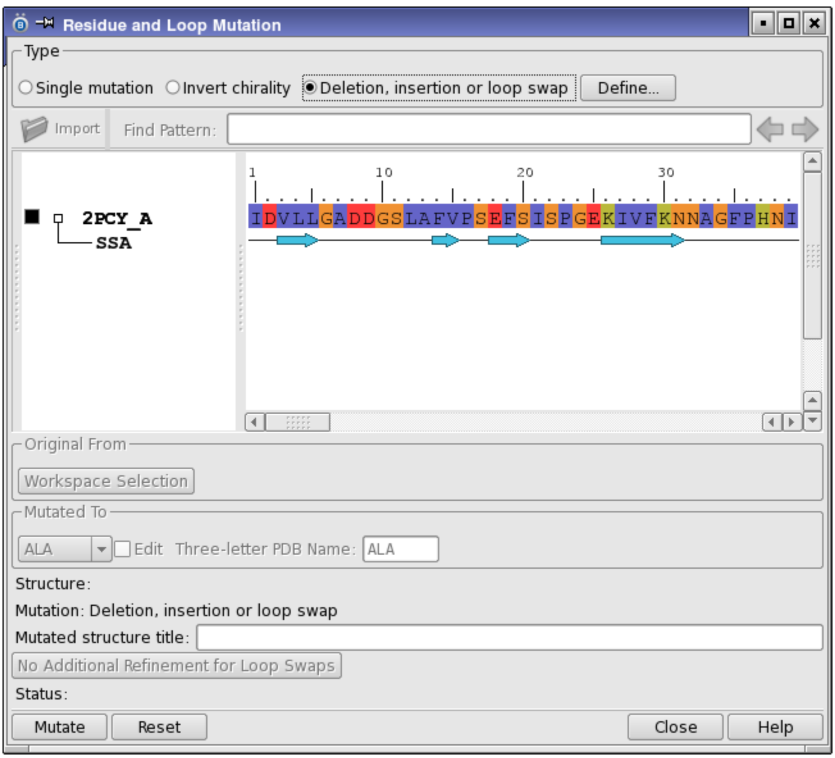


Figure 12.1. The Residue and Loop Mutation panel.

12.1 Residue Mutations

There are three types of single-point mutation that you can make in this panel: mutation to a standard amino acid, mutation to a nonstandard amino acid, and chirality inversion. You can choose which type to do with the options in the Type section. Select Single mutation for both standard and nonstandard amino acids; select Invert chirality to invert the chirality of the amino acid.

The workflows for each of these mutations is summarized here and detailed in the following sections.

To mutate a single residue to a standard amino acid:

1. Select Single mutation in the Type section.
2. Select the residue to be altered either in the Workspace or in the panel's sequence viewer.
3. Click Workspace Selection in the Original from section.
4. Choose the amino acid from the option menu in the Mutated to section.
5. Click Mutate.

To mutate a single residue to a nonstandard amino acid:

1. Select Single mutation in the Type section.
2. Select the residue to be altered either in the Workspace or in the panel's sequence viewer.
3. Click Workspace Selection in the Original from section.
4. Choose a standard amino acid from the option menu in the Mutated to section.
5. Select Edit.

   The Workspace structure is replaced by the chosen standard amino acid with the side-chain atoms shown in line (wire frame) representation.
6. Edit the side-chain atoms in the Workspace to produce the desired nonstandard amino acid.
7. Deselect Edit.
8. Provide a three-letter PDB residue name in the dialog box that opens.
9. Click Mutate.

To invert the chirality of a single residue:

1. Select Invert chirality in the Type section.
2. Select the residue to be altered either in the Workspace or in the panel's sequence viewer.
3. Click Workspace Selection in the Original from section.
4. Click Mutate.

12.1.1 Selecting the Residue to Mutate

You can select the residue to mutate by picking it in the Workspace structure or in the sequence viewer, then clicking Workspace Selection in the Original From section to register your choice.

To find a particular residue in the Workspace, you can use the Find tool. Type CTRL+F (⌘F) in the Workspace to display the Find toolbar below the Workspace. You can then choose Residue number or Residue type from the Find menu to choose residues by number or type, then enter the number in the text box or choose the type from the menu, and click the N or P button.

To find a residue in the sequence viewer in the Residue and Loop Mutation panel, enter the residue letter in the Find Pattern text box, and click the arrow keys to step through the occurrences. You can also enter multiple residues to find a pattern, e.g. `D-A-P`. The pattern uses PROSITE syntax, which is explained in the tool tip. When you have found the residue you want, clear the Find Pattern text box, then click on the residue to select it, or simply select it in the Workspace. The residues that are found are selected in the Workspace, so clicking on one of them changes the selection to the desired residue.

If the sequence viewer doesn’t have a sequence in it, you can click Import and choose From Workspace to load the sequence of the Workspace structure.

12.1.2 Defining the Mutation

When you have selected the residue to mutate, you can define the mutated residue. The controls to do so are only available if you selected Single mutation in the Type section. If you selected Invert chirality, the mutation is already defined. You can mutate to a standard amino acid or a custom amino acid. In both cases, you start by choosing a standard amino acid from the option menu.

To begin editing a standard amino acid in the Workspace to produce a custom amino acid, check Edit. The amino acid replaces the protein in the Workspace, with the backbone represented as ball and stick and the side chain as lines.

To add groups to the structure or to change elements, you can use the Build and the Fragments toolbars. Click Build or Fragments on the Manager toolbar at the top of the main window to display these toolbars. With the Fragments toolbar, you can select a fragment, then click on an atom to replace that atom with the fragment. With the Build toolbar, you can sketch a structure with the Draw tool, change the element with the Set Element tool. You might also want to add hydrogens after sketching a structure, which you can do from the Edit toolbar with the Add H tool. You should also use the Clean Up tool after sketching the structure and adding hydrogens, to ensure that the structure is not distorted.

For more information on building structures, see Chapter 5 of the *Maestro User Manual*.

When you have finished editing, clear the Edit check box. A dialog box prompts you to provide a 3-letter PDB name for the custom amino acid. The default is USR. You can edit the name in the Three-letter PDB Name text box after creating a custom amino acid.

The text on the amino acid option menu is set to Custom when you finish editing.

After the mutation is defined, the mutation is reported in the panel, and a title for the mutated structure is entered in the Mutated structure title text box. You can edit this title if you wish.

12.1.3 Refining the Mutated Structure

If you want to refine the mutated structure, click Advanced Refinement Options and make selections in the Refinement Options dialog box. It is usually a good idea to do some sort of refinement, to allow the protein to adjust to the new residue. This is particularly so if you created a custom residue, which might not be in the most favorable conformation.

There are two types of refinement available: minimization and molecular dynamics simulation. You can choose one or the other or both. The minimization is run first.

Minimization is the fastest option, and is a good choice if the mutated residue is in approximately the right conformation. If you choose to do a minimization, you can run it in the gas

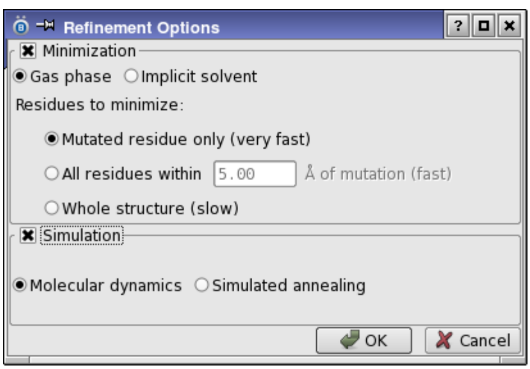


Figure 12.2. The Refinement Options dialog box.

phase or in implicit solvent. The residues minimized can be limited to just the altered residues, all residues within a given distance of the altered residues, or the entire structure.

Molecular dynamics is much more time-consuming, but it can sample other parts of conformational space, particularly if you run simulated annealing. Once the structure is mutated and any minimization is done, the protein is prepared for the simulation by adding explicit water molecules. The Molecular Dynamics panel or the Simulated Annealing panel opens, and you can make settings and run the simulation, which uses the Desmond molecular dynamics program. For more information on these panels, see Chapter 3 of the *Desmond User Manual*.

If you decide you do not want to run the simulation, which can take many hours, you should delete the entry group in the Project Table that was created for the simulation, as follows:

1. Select the entry group (click the row with the number in square brackets).
2. Choose Entry → Unlock.
3. Choose Entry → Delete.

For information on using the Project Table, see Chapter 9 of the *Maestro User Manual*.

12.2 Insertions, Deletions, and Loop Swaps

You may want to perform larger structural changes than a single residue mutation, such as deleting or inserting multiple residues, or replacing a loop with another loop. To do this, select Deletion, insertion, or loop swap in the Type section, then click Define to define the changes to be made in the Insertions, Deletions, and Loop Swaps panel.

The basic procedure is summarized below, and details are given in the following sections.

To insert or delete residues or modify a loop:

1. Select Deletion, insertion or loop swap in the Type section.
2. Click Define.

   The Insertions, Deletions, and Loop Swaps panel opens.
3. Choose the loop to be modified.

   For best results select at least two residues on either side of the residues to be modified. The selection can be made either by using the sequence viewer or by selecting residues in the Workspace.
4. Click the Workspace Selection button to load the original loop into the table.

5. If desired, load a new structure and select residues to specify a starting set of residues for the replacement loop.

   By default, the replacement loop starts out identical to the original loop.

6. If desired, edit the replacement loop by clicking on residues in the table and specifying an insertion, deletion or mutation at that point.

7. Choose the number of models to generate and the prediction method.

8. Click Accept.

   The Insertions, Deletions, and Loop Swaps panel closes.

9. Click Mutate.

No refinement of the loop modification is offered, so you must perform any refinement independently.

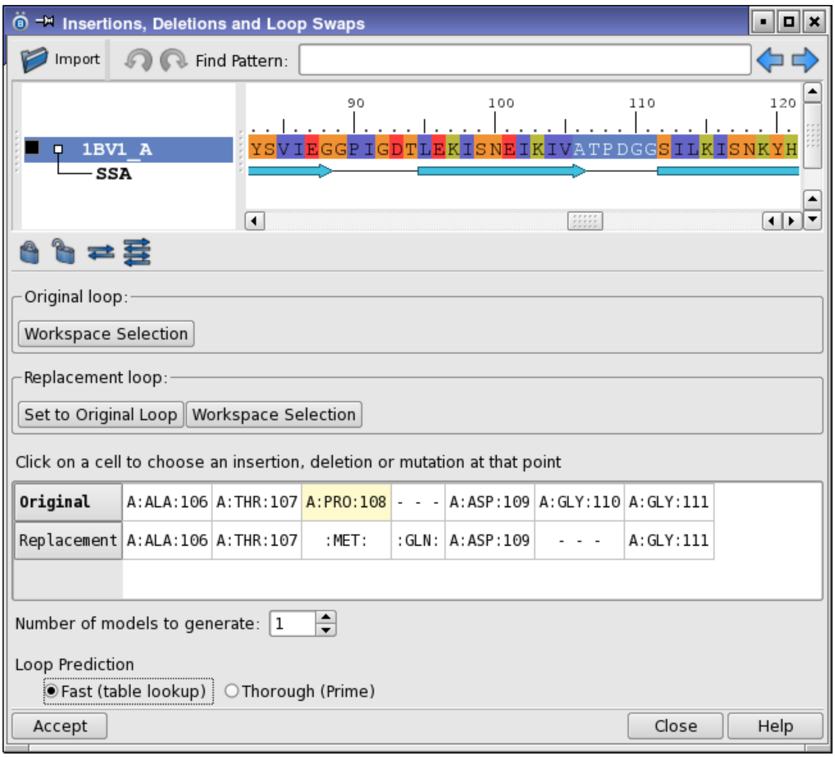


Figure 12.3. The Insertions, Deletions, and Loop Swaps panel.

12.2.1 Choosing the Original and Replacement Loop

When the Insertions, Deletions, and Loop Swaps panel opens, the sequence viewer at the top is loaded with the sequence of the structure in the Workspace. If you want to use a different s structure, you can replace the Workspace structure with another structure, or you can import a structure from a file or from the PDB. To load another structure, click Import, and choose the source from the menu that is displayed. The choices are:

- Browse for File—Open a file browser in which you can navigate to the desired location and select the file that contains the structure. The allowed file types are Maestro and PDB.
- From PDB ID—Import the structure from the specified PDB ID. Opens the Enter PDB ID dialog box, in which you can enter the PDB ID of the sequence. The sequence and structures are retrieved from a local copy of the PDB if it is available, or from the RCSB web site, depending on the preference set for PDB retrieval.
- From Workspace—Import the structure that is displayed in the Workspace.

The sequence for the imported structure is added to the sequence viewer. This allows you to import multiple structures, and choose which structure to take a loop from.

When you have the structure that you want to mutate, you can select the loop to mutate either in the Workspace or in the sequence viewer in the panel.

Selecting the loop in the Workspace allows you to visually identify the loop. To ensure that you are picking residues in the Workspace to define the loop, choose Edit → Pick Mode → Residues, or type the letter R with the pointer in the Workspace. You can then pick residues that belong to the loop you are interested in. The residues that you pick are highlighted in the Workspace (yellow dots) and in the sequence viewer (white letter on blue background).

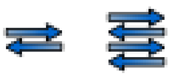
Selecting the loop in the sequence viewer allows you to easily search for patterns in the sequence, or to use the secondary structure assignment (SSA) to identify loops. The SSA has no annotation where there are loops (the absence of any other secondary structure). To search for patterns in the sequence, enter the pattern in the Find Pattern text box, with a dash separating each residue from the next, e.g. `D-A-P`. The syntax is summarized in the tool tip for the text box, and is given in more detail on page 39. You can use the arrow keys to step through the patterns. All of the matches are selected in the Workspace, so you might have to select the loop that you want in the Workspace, or by clearing the Find Pattern text box, then selecting the residues to use.

When you select the residues for the loop to mutate, you should select at least two residues on either side of the residues that you plan to modify. These extra residues are “stem” residues, which are used by the Prime software when rebuilding the loop, to properly fit the new loop onto the structure. For example, if a single residue is being deleted, the original loop should consist of five residues - the residue to delete and the two residues on either side of it.

After you have selected the residues, click Workspace Selection in the Original loop section, to register the selection and copy the structure for the loop mutation job. The original loop is used by default for the replacement loop, which you can modify as described in the next section.

If you want to replace the loop with a loop from another structure, you can place another structure in the Workspace, and select a loop from this structure. If the structure is not already in the sequence viewer, you can import it as described above. You can also select the loop in the same way as for the original loop.

You might want to align the sequence for the replacement to the sequence for the original structure, so that you can select the loop by its alignment. To do the alignment, make sure that the original sequence is at the top of the sequence viewer (use the shortcut menu for the sequence name to move it if necessary), and add the replacement sequence to the selected sequences. You can then use either the pairwise alignment tool or the multiple alignment tool to align the sequences (using ClustalW).



You can also do manual alignment, with locking and unlocking of gaps. See the *Multiple Sequence Viewer* document for more information on doing manual alignment.

When you have selected the residues in the desired structure for the replacement loop, click Workspace Selection in the Replacement loop selection. If you decide not to use this loop, you can click Set to Original Loop to change the replacement loop back to the original loop structure.

12.2.2 Editing the Replacement Loop

If you want to modify the replacement loop residues to create the new loop that will replace the original loop, you can do so in the table in the lower part of the panel. Clicking on a table cell opens a menu that allows you to delete or mutate the current residue or insert a new residue before or after the current residue. The change is applied to the Replacement row, regardless of whether you define the change for the original row or the replacement row. The Insert Before, Insert After, and Mutate items display a list of 20 standard amino acid residues that you can select.

When you choose to insert a residue, a new column is added to the table before or after the current position. The cell in the Original row has three dashes, to indicate that there is no residue in the sequence at this position. The cell in the Replacement row has the residue name between two colons, to indicate that the chain and residue number are not yet assigned. Insertion codes are added for the inserted residues when the job is run.

If you delete a residue, the cell in the Replacement row is set to three dashes, to indicate the deletion.

Likewise, if you mutate a residue, the cell in the Replacement row has the new residue name between two colons.

12.2.3 Choosing the Output and Method

The replacement loop is built with the Prime structure building software. By default, only the best structure is returned, but if you want to examine more than one structure, you can specify the number of new structures to be returned in the Number of models to generate box.

You can also choose whether to create the loops via a fast table-lookup method or via a full Prime loop prediction, which builds loops by sampling multiple conformations and scoring them, to produce the best loop structures. The fast method takes only a minute or so, where as the thorough loop sampling can take hours. To make the choice, select Fast or Thorough under Loop prediction. No additional refinement beyond these methods is done for new loop predictions.

When you have finished selecting options, click the Accept button to return to the main Residue and Loop Mutation panel. Click the Mutate button on that panel to begin the loop swap job.

Scanning for Residue Mutations

This chapter describes how to scan a protein for potential residue mutations, generate mutated structures, and compare the properties of the mutated structures. The mutation sites and the mutations can be selected manually or selected automatically based on homology modeling or 3D structural criteria.

Some examples of the use of residue scanning are:

- improvement of protein-ligand affinity
- identifying protein-protein interface hotspots
- identifying residue mutations that can improve stability.
- mutating unpaired, solvent-exposed Cys residues to reduce undesired reactivity

A residue scanning job can be set up in the Residue Scanning panel, and the results examined in the Residue Scanning Viewer panel. To open the Residue Scanning panel, choose Tasks → Residue Scanning → Perform Calculations in the main window.

13.1 Selecting and Analyzing the Protein

The first step is to select the protein and to display it in the Workspace. The protein must be one that has been prepared for use in modeling. If it has not been prepared, we recommend that you prepare it with the Protein Preparation Wizard (on the Tools menu and the Tasks menu). Details of preparing a protein can be found in the *Protein Preparation Guide*.

If you have not already opened the Residue Scanning panel, open it by choosing Tasks → Residue Scanning → Perform Calculations in the main window. You can display the protein before or after opening the panel.

The first part of the procedure is to analyze the protein, which you do by clicking Analyze Workspace. In this process, you may be asked some questions:

- If your structure has not been prepared for modeling, you are asked if you want to use the Protein Preparation Wizard to prepare the structure. You cannot proceed if you do not have a protein that is an all-atom, 3D structure with bonding information. After preparing the protein, display it in the Workspace again and click Analyze Workspace.

- If your structure has multiple chains, a dialog box opens, prompting you to choose a chain. Mutations can only be performed on a single chain in the residue scanning run. If you want to mutate multiple chains, you must mutate them in separate runs.

When the structure has been analyzed, the residues table is filled in with all the residues in the chosen chain.

Instead of analyzing an entire protein or an entire chain, you can select residues and analyze only the selected residues. To do this, select the residues in the Workspace structure, select Analyze only selected Workspace residues in the Residue Scanning panel, then click Analyze Workspace.

13.2 Setting Up and Running the Mutation Job

The tasks involved in setting up the mutation job are done from the Residues tab: selecting residues, selecting mutations, setting parameters for refinement. The Homology Suggestions tab provides a way of automatically selecting residues—see Section 13.3 on page 74—but the remaining tasks must be done in the Residues tab.

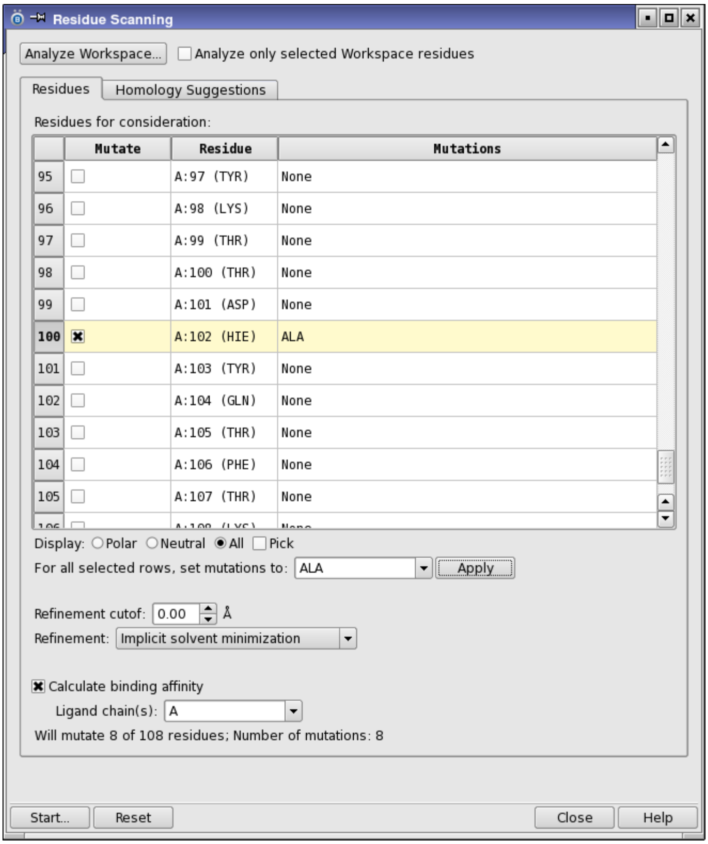


Figure 13.1. The Residue Scanning panel, Residues tab.

13.2.1 Choosing Residues to Mutate

The residues that are available for mutation are listed in the Residues in the Workspace table in the Residues tab. The residues are identified by the chain name, the residue number (and insertion code), and the 3-letter residue name. If you want to show only the polar or the nonpolar residues, select Polar or Neutral underneath the table. To redisplay all residues, select All.

There are two main ways of selecting residues to mutate: manually, or based on analysis of homologs to the structure of interest. Manual selection is described in this section, and using homology is described in Section 13.3 on page 74.

- To select a residue for mutation, check the box in the Mutate column for that residue.
- You can also select residues for mutation and select the mutations at the same time—see Section 13.2.2 for details.

When you select rows in the table, the Workspace view zooms in to the residues you have selected. The selected residues are highlighted with green carbons, and the remaining residues are dimmed. (You can adjust how far the view zooms in by making a setting in the Preferences panel. Choose Edit → Settings → Preferences, select Fitting under Workspace in the tree on the left, then enter a value in the Fit margin text box.)

13.2.2 Choosing the Mutations

When you have chosen the residues to mutate, the next step is to choose the mutations for those residues. You can choose mutations for individual residues, or you can apply a set of mutations to selected residues.

To apply a set of mutations to selected residues:

1. Select the residues in the table that you want to apply the same mutations to.

   You can select Pick and pick residues in the Workspace to select them in the table.

2. Choose the residue types that you want to use for the mutations from the For all selected rows, set mutations to menu.

   The residue types include groups such as Neutral, Polar, Hydrophobic, Positive, and Negative. You can choose more than one residue or group from the list, and the new residues are added to the list. Each residue type that you select is checked on the list.

3. Click Apply.

The Mutate check box is checked for any of the selected rows that are not already marked for mutation, and the mutation is set to the residues that you chose from the option menu. The mutations are listed in the Mutations column of the table.

A text message is displayed at the bottom of the tab that reads Will mutate M of N residues; Number of mutations: K. The number of mutations is the total number of structures generated. Only one site is mutated at a time, so the number of structures is linear in both the number of sites mutated and the number of mutations at each site.

To define the mutations for a single residue:

1. Click in the Mutations column for the residue.

   An option menu is displayed in the table cell.

2. Choose the residue types from the option menu.

   The mutations are displayed in the box at the top of the menu.

3. Press ESC to dismiss the option menu (or click in the Residue column).

   The mutations are now listed in the table cell.

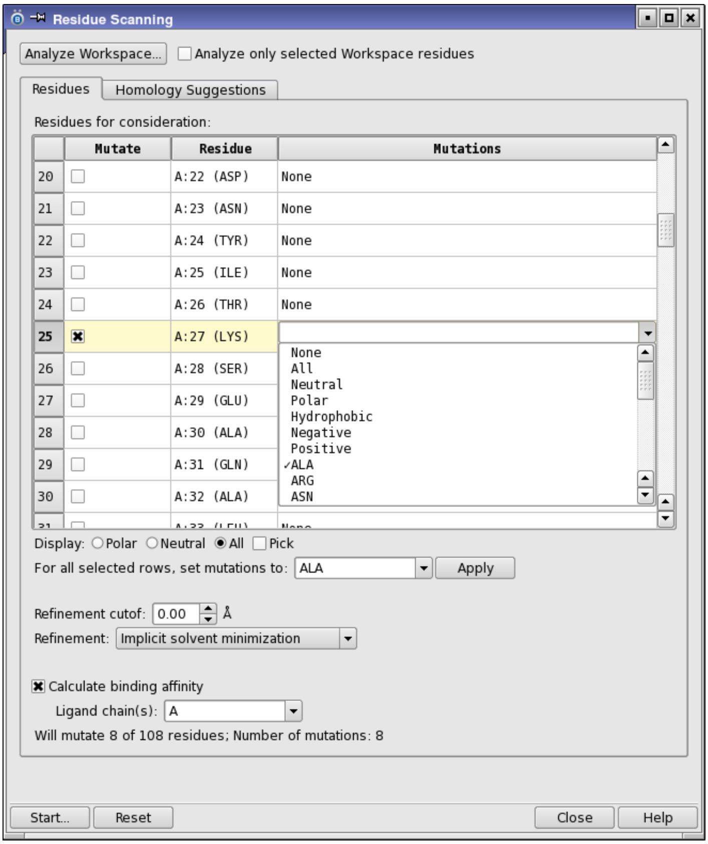


Figure 13.2. Option menu for defining mutations for a single residue.

13.2.3 Choosing Refinement Options

When the residues are mutated, the protein structure around the mutation site should be allowed to relax in response to the mutation. Relaxation is carried out based on the refinement method selected, which is described below.

The region around the mutation site that is relaxed during the calculations is defined by a cutoff distance. Any residue that has an atom within the specified distance of a hypothetical Arg residue at the mutation site is included in the refinement. A hypothetical Arg residue is used to ensure that the set of residues refined is identical, regardless of the initial or mutated residue identities. This ensures that comparisons of the properties of the mutated structures are not affected by the choice of residues that are relaxed.

The method used for the refinement of the selected residues can be chosen from the Refinement option menu. The choices are:

- Gas-phase minimization—Minimize the energy of the residues using the internal force-field minimizer. This is the fastest method, but does not include solvation effects.
- Implicit solvent minimization—Minimize the energy of the residues with an implicit (continuum) solvation model, using Prime.
- Side-chain prediction—Prior to minimization, a thorough exploration of possible side chain conformations is performed.
- Side-chain prediction (cbeta)—Prior to minimization, a thorough exploration of possible side chain conformations, including the CA-CB torsion, is performed.
- Side-chain prediction (bb)—Prior to minimization, a thorough exploration of both potential side chain and potential backbone conformations is performed.

For systems that consist of more than one entity (chain), the binding affinity changes with residue mutation can be calculated. The affinity changes are calculated between the chains that constitute a “ligand” group and the remainder of the system. The ligand group is defined as the chain for which mutations are being calculated (the chain you selected) plus any additional chains you wish to add to the ligand group. The binding affinity is calculated with the Prime MM-GBSA technology.

To include binding affinity calculations in your job:

1. Select Calculate binding affinity.
2. Choose the chains from the option menu below this option.

13.2.4 Running the Job

To run the job, you first set job parameters, and then submit the job to a host for execution. Click Start to open the Start dialog box. In this dialog box, you can choose whether to append the results to your project, or leave the results on disk in the current directory, where you can examine them later. You can also choose a job name, which is used to name the files associated with the job. When you have made your choices, click Start in the dialog box to start the job.

If you run the job on a multiprocessor host, you can divide the job into subjobs and distribute them over multiple processors. The minimum work a subjob can do is to mutate one residue to another residue, so you should not specify more subjobs than there are output structures. For optimal load balancing, the number of subjobs should be a few times the number of processors.

A status bar showing the progress of the job is displayed above the Start button. The job takes several minutes per residue mutation to run, depending on the refinement options.

13.3 Using Homologs for Identifying Mutations

In addition to manual selection of residues, you can use homology to suggest residues to mutate and residues to mutate to, on the basis of variability or conservation of residues across the set of homologs, or on structural proximity or properties. You can do this in the Homology Suggestions tab.

13.3.1 Obtaining Homologs from a BLAST Search

If you don't have homologs for your protein, you can run a BLAST search to find homologs.

1. Click Run a Blast Search.

   The BLAST Search Settings dialog box opens, so you can change settings if you want.

2. Click Start Job in this dialog box, after changing any settings.

   The job starts, and its log file is displayed in the Job Progress dialog box.

At the end of the job, the BLAST Search Results dialog box opens, so that you can choose how many of the homologs you want to use. You can do one of the following to select homologs:

- Select the homologs in the table (with shift-click and control-click).
- Enter the number of the top homologs in the text box at the bottom of the panel and click Select Top.

When you have selected the homologs, click Incorporate Selected Rows to add the structures to the sequence viewer.

13.3.2 Importing Homologs

If you already have a set of homologs for a structure that you want to use to identify potential mutation sites, or if you have other structures that you want to use as homologs, you can import them.

- If you have a set of homologs in a file, click the Import button and choose Browse for file. A file selector opens so you can locate and import the file.
- If you want to manually import structures from the PDB, click Import and choose From PDB ID, then specify the PDB ID in the dialog box that opens.
- If the structures of the homologs are in the Workspace, click Import and choose From Workspace.

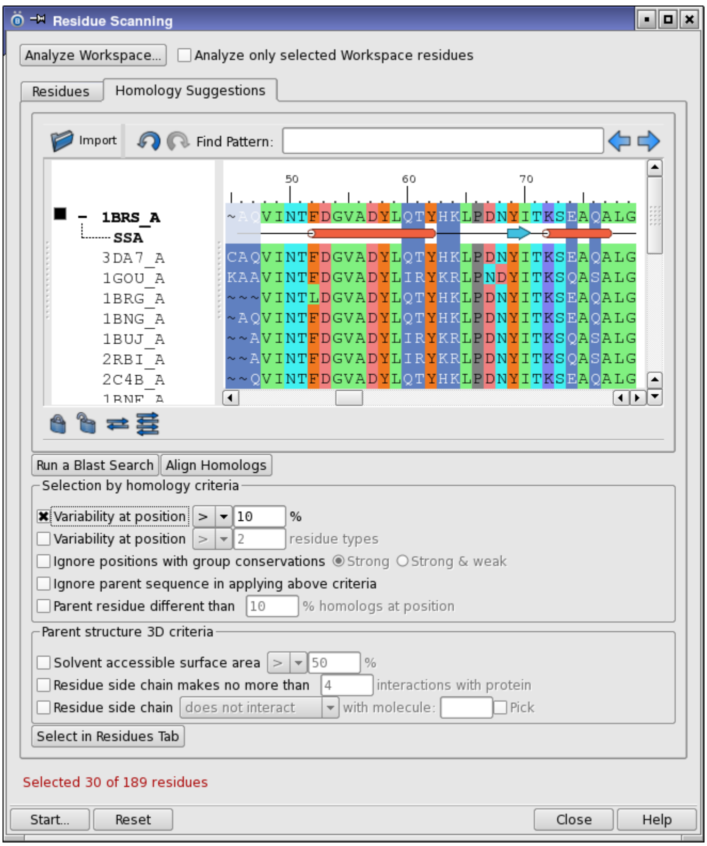


Figure 13.3. The Residue Scanning panel, Homology Suggestions tab.

13.3.3 Aligning Homologs

To make use of the homologs, they must be aligned to the parent (query), if they are not already aligned. You can run a job to align the homologs, or you can do a manual alignment, or both.

To run the alignment job, click Align Homologs, or click the Multiple Alignment toolbar button.

The alignment, which uses ClustalW, is usually very quick. A dialog box may be displayed briefly before the results are incorporated. The alignment adds gaps in the sequence viewer as necessary.

To align sequences manually, you can drag residues to the right or the left, to fill or create gaps. If you have already created gaps that you want to preserve during another alignment, you can click the Lock Gaps toolbar button. The gap symbol changes to - to indicate that the gap is locked. To unlock the gaps again, click the Unlock Gaps toolbar button.

13.3.4 Selecting Residues by Homology

When the sequences are aligned, residues can be selected on the basis of the of the alignment. To choose residues to mutate, you may want to select residues at positions where the variability among the aligned homologs is high, or residues that vary in residue type, or residues that differ from most of the homologs. The Selection by homology criteria section has a number of options that you can set to apply selection filters on the basis of the amount of variation. Each of the options that you select is applied, so the residues must meet all specified criteria.

- Variability at position >/< *N* %—Select residues based on the percentage variability at the residue position. You can apply a minimum or a maximum variability by choosing from the option menu, and specify the percentage threshold in the text box.
- Variability at position >/< *N* residue types—Select residues based on the residue type variability at the residue position. Choose whether to apply a minimum or a maximum variability from the option menu, and specify the threshold for the number of residue types that can vary in the text box.
- Ignore positions with group conservations—Ignore (do not select) residues that are strongly conserved or that are either strongly or weakly conserved. Select the appropriate option for strong conservation or both strong and weak conservation.

Strong conservation means that all residues at a particular position are in one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW. Weak conservation means that all residues at a particular position are in one of the following groups: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY. These definitions are those used by ClustalW.

- Ignore parent sequence in applying above criteria—Ignore the parent (query) sequence when applying the variability or conservation criteria. The selection is then based on the variability in the homologs only.
- Parent residue different than *N* % homologs at position—Select residues for which the parent residue is different from more than the specified percentage of homologs at the residue position.

When residues are selected on the basis of homology, a default set of mutations is also defined. The mutations for a given residue include all the variants found at that residue position.

13.3.5 Selecting Residues by Structural Attributes

Residue selection filters can also be applied on the basis of solvent-accessible surface area and contact with other parts of the protein or with a ligand or cofactor. To apply these criteria to residue selection, select one or more of the following options:

- Solvent accessible surface area—Select this option to select residues by their solvent-accessible surface area (SASA) relative to an isolated residue of the same type, and set a threshold for the maximum or minimum allowed relative SASA. This option is useful for locating surface or buried residues.
- Residue side chain makes no more than *N* interactions with protein—Select this option to filter out residues whose side chains have multiple interactions with other protein residues, and set the maximum number of residues with which the side chain has interactions.
- Residue side chain interacts/does not interact with molecule *N*—Select this option to select residues by their interaction with a particular molecule. Choose whether to allow or disallow the interaction from the option menu, and specify the molecule in the text box, or select Pick and pick the molecule in the Workspace. This is useful for mutating residues near a ligand, or for mutating residues at protein-protein interfaces, for example.

Interactions are determined by a distance cutoff: any residue that has atoms within 4 Å of the side chain is considered to interact.

13.3.6 Making the Selection

When you have set up all the criteria for residue selection, click Select in the Residues Tab. The rows are selected in the table in the Residues tab, for the residues that meet all the criteria, and the Mutate check box is checked for each of these residues. For each selected residue, a set of mutations is defined that correspond to the residues observed for that position in the homologs that were used in the Homology Suggestion tab. You can then use the other tools in the Residues tab to modify any of the selections before running the mutation job.

13.4 Viewing the Mutation Results

When the residue scanning job finishes, the Residue Scanning Viewer panel opens, displaying changes in properties for each mutation, and a graph of one of the properties against the row number.

If you want to examine results from a job that was completed earlier, you can open this panel by choosing Tasks → Residue Scanning → View Results. You can then click Import to locate the Maestro file (`.maegz`) that contains your results and import it into the current project.

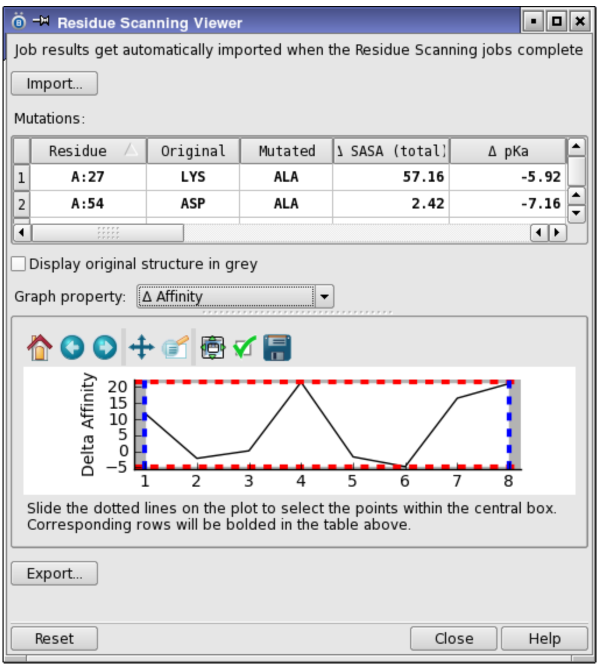


Figure 13.4. The Residue Scanning Viewer panel.

The results are listed in the Mutations table. Each mutant is identified by the residue position and the original and mutated residue names. The properties that were calculated for each mutant are listed in the table. These properties are described briefly in Table 13.1. Some of these properties are described in more detail in the sections below.

All properties are calculated after the refinement is performed, and so include relaxation of the protein after mutation. The energy and stability properties are calculated for the ligand group that you specified for the binding affinity calculations: the remainder of the protein is ignored.

You can sort the table columns by clicking on the column headings. You can plot any of these properties against the mutation (table row) by choosing the property from the Graph property option menu. If you want to export the table data as a CSV file, click Export, and navigate to a location and name the file.

Table 13.1. Mutation properties.

Property	Description
Δ SASA (total)	Change in total surface area due to the mutation.
Δ SASA (nonpolar)	Change in surface area of nonpolar atoms due to the mutation
Δ SASA (polar)	Change in surface area of polar atoms due to the mutation
Δ pKa	Change in pK_a of the mutated residue, calculated using PROPKA [1].
Δ Affinity	Change in binding affinity of the mutated protein (and any other specified chains), treated as the ligand, to the rest of the system, treated as the receptor. The calculations are carried out with Prime with an implicit solvent term.
Δ Hydropathy	Change in hydrophobic or hydrophilic nature of the mutated residue, as defined on the Kyte-Doolittle scale (see, for example, http://en.wikipedia.org/wiki/Hydrophobicity_scales). A positive values indicates a more hydrophilic residue and a negative value indicates a more hydrophobic residue.
Δ Total rotatable bonds	Change in the total number of rotatable bonds.
Δ Stability (gas)	Change in the stability of the protein due to the mutation, calculated without solvation (i.e. in the gas phase). The stability is defined as the difference in energy between the folded and unfolded states (see Section 13.4.2 on page 80).
Δ Stability (solvated)	Change in the stability of the protein due to the mutation, calculated using the Prime energy function with an implicit solvent term (unless you choose gas-phase refinement). The stability is defined as the difference in free energy between the folded state and the unfolded state (see Section 13.4.2 on page 80).

You can select a region in the graph using the horizontal and vertical dashed lines, which can be dragged to create the selection. The rows corresponding to the selected region of the graph are highlighted in the table above, and the residues are highlighted in the Workspace.

If you select a table row, the view zooms in to the mutated residue. To display the original structure, select Display original structure in grey. The parent protein is displayed and colored grey. You can then see how the mutation is positioned in relation to the original residue.

13.4.1 Binding Affinity Prediction

The change in the binding affinity of the protein due to the mutation is calculated from a thermodynamic cycle, which can be represented as follows:

$$\begin{tikzcd} \mathrm{R{+}L} \arrow[r, "\Delta G_1"] \arrow[d, "\Delta G_3"] & \mathrm{R{\cdot}L} \arrow[d, "\Delta G_4"] \\ \mathrm{R{+}L'} \arrow[r, "\Delta G_2"] & \mathrm{R{\cdot}L'} \end{tikzcd}$$

where R is the receptor, L is the ligand in the parent, and L' is the mutated ligand. R+L and R+L' represent the separated receptor and ligand. R·L and R·L' represent the receptor bound to the ligand. The change in binding affinity is

$$\Delta\Delta G(\text{bind}) = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

Experiment measures ΔG_1 and ΔG_2, but it is ΔG_3 and ΔG_4 that are calculated, to optimize the cancellation of error in the computational models. The calculations are done with Prime MM-GBSA, which uses an implicit (continuum) solvation model.

13.4.2 Stability Prediction

The stability of the protein due to the mutation is calculated from a thermodynamic cycle, which can be represented as follows:

$$\begin{tikzcd} \mathrm{L(u)} \arrow[r, "\Delta G_1"] \arrow[d, "\Delta G_3"] & \mathrm{L(f)} \arrow[d, "\Delta G_4"] \\ \mathrm{L'(u)} \arrow[r, "\Delta G_2"] & \mathrm{L'(f)} \end{tikzcd}$$

where L(u) is the unfolded parent ligand, L(f) is the folded parent ligand, L'(u) is the unfolded mutated ligand, and L'(f) is the folded mutated ligand. The change in stability is

$$\Delta\Delta G(\text{stability}) = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

Experiment measures ΔG_1 and ΔG_2, but it is ΔG_3 and ΔG_4 that are calculated, to optimize the cancellation of error in the computational models. For the purpose of the model, the unfolded ligand is represented as a tripeptide, Gly-X-Gly, where X is the residue that is mutated. The assumption is that the remaining interactions in the unfolded state are negligible. The calculations are done with Prime MM-GBSA, which uses an implicit (continuum) solvation model.

Locating Possible Mutations for Disulfide Bridges

Disulfide bridges between cysteine residues add to the stability of a protein structure. Mutating residues to form or break disulfide bridges offers a way of controlling the stability of a protein. This chapter describes how to run a cysteine mutation calculation to locate and rank possible disulfide bridges. The calculation is set up and run in the Cysteine Mutation panel, in the Run tab, and the results are presented in the Results tab.

To open the Cysteine Mutation panel, choose Tasks → Cysteine Mutation.

14.1 Selecting and Analyzing the Protein

The first step is to select the protein and to display it in the Workspace. The protein must be one that has been prepared for use in modeling. If it has not been prepared, we recommend that you prepare it with the Protein Preparation Wizard (on the Tools menu and the Tasks menu). Details of preparing a protein can be found in the *Protein Preparation Guide*.

The protein must be analyzed to locate possible residue pairs that could be mutated to cysteines, or to locate disulfide bridges that could be broken by mutation to other residues, which you do by clicking Analyze Workspace.

Instead of analyzing an entire protein, you can analyze the Workspace selection. To do this, select the desired residues in the Workspace structure, select Analyze only selected Workspace residues in the Cysteine Mutation panel, then click Analyze Workspace.

If your structure has not been prepared for modeling, you are asked if you want to use the Protein Preparation Wizard to prepare the structure. You cannot proceed if you do not have a protein that is an all-atom, 3D structure with bonding information. After preparing the protein, display it in the Workspace again and click Analyze Workspace.

When the structure has been analyzed, the Residue pairs for mutation table is filled in with all the residue pairs that meet the criteria for forming or breaking a disulfide bridge. The criterion for identifying potential cysteine pairs is a Cβ–Cβ distance between the residues that is less than the distance specified in the panel. For Gly, the distance is taken from the alpha hydrogen.

14.2 Choosing Residue Pairs to Mutate

The Residue pairs for mutation table lists the pairs of residues that have the potential to form or break disulfide bonds. The table rows that are shown are controlled by the options below the table. The table columns are described in Table 14.1. You can sort the table by the values in any column, by clicking on the column heading.

Table 14.1. Columns of the Residue pairs for mutation table.

Column	Description
(Index)	The first column contains the index of the residue pair. When the table is filtered to show only certain residue pairs, this index remains the same (i.e. it is not a table row number).
Type	Mutation type, which can be one of the following: X-X -> S-S: Mutation of two residues to Cys with formation of a disulfide bond. S-X -> S-S: Mutation of one residue to Cys with formation of a disulfide bond to a nearby cysteine S-S -> S-X: Mutation of one Cys residue of a bonded pair to break a disulfide bond.
Residue 1	Identity of the first residue in the pair, given by the chain letter, the residue number and insertion code, and the 3-letter residue name. For Cys-Cys pairs, this residue is the residue that is mutated, so the pair is listed twice, in opposite order, to allow selection of only one of the pair to mutate.
Residue 2	Identity of the second residue in the pair, given by the chain letter, the residue number and insertion code, and the 3-letter residue name.
β-carbon Distance	Distance in angstroms between the beta carbons (CB) of the two residues. In the case of Gly, the distance is taken from the alpha hydrogen (HA) that would be replaced by the beta carbon of the Cys.

There are several ways to control what is shown in the table.

a. Use the Display options:

- Cys-Cys—Show only cysteine-cysteine pairs, for S-S -> S-X mutations. The first residue listed is the one mutated, so a given pair is listed twice, in opposite order, to allow selection of either residue of the pair for mutation.
- Cys-X—Show only pairs with one cysteine, for S-X -> S-S mutations.
- X-X—Show only pairs of non-cysteine residues, for X-X -> S-S mutations.
- All—Show all pairs.

b. Use the X->Cys replacement residues option menu to filter on the residues that will be replaced with Cys.

The option menu contains individual residues, which you can select independently, an All option to select all residues (except Pro), a None option to clear the list, and a Conservative option to select conserved residues (GLY, ILE, LEU, VAL, ALA). The residues that you choose are displayed in the main part of the option menu; the complete list is shown in a tooltip if it is too long. The table is updated for each selection that you make.

c. Use the Beta carbon distance cutoff text box

You can specify the maximum allowed distance between the beta carbons of the residues in a pair. This distance is used to filter the table to show only residues with a smaller distance. For Gly, the distance is taken from the alpha hydrogen.

These display options allow you to restrict the list of residue pairs to those that are of interest, so that it is easier to select the pairs in the table that you want to mutate. The job mutates only the selected pairs in the table, so you must make a selection before you run the job.

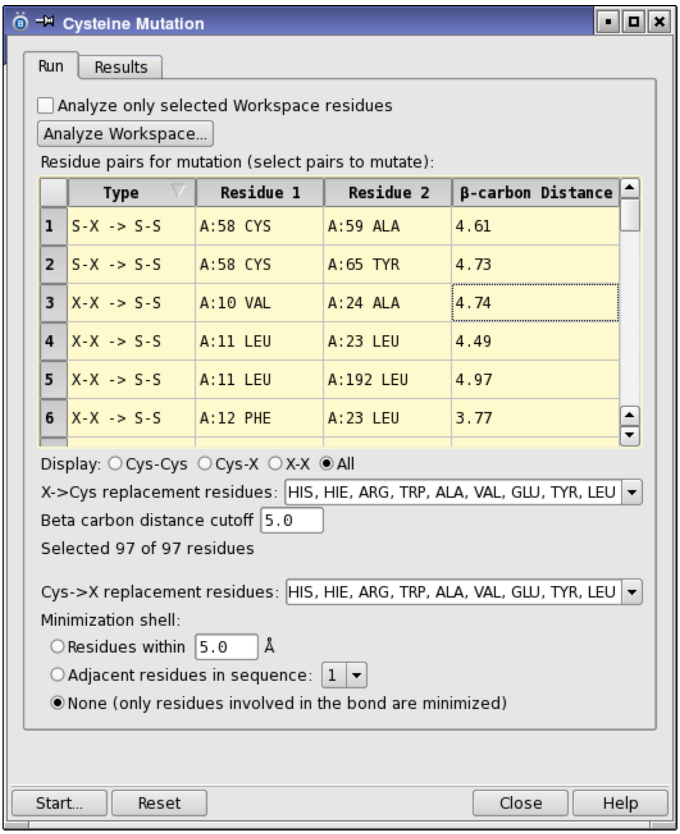


Figure 14.1. The Run tab of the Cysteine Mutation panel.

Each selected pair is mutated independently: there are no simultaneous mutations.

If you have Cys-Cys pairs whose bond you want to break by mutating one of the cysteines to another residue, you must also select the mutations, from the Cys -> X replacement residues option menu. The option menu works in the same way as the X -> Cys replacement residues option menu, described above.

14.3 Relaxing the Structure Around the Mutation Site

The two residues in any pair that is mutated are minimized after the mutation, to relieve strain. You can include additional residues around the mutation site in the minimization, to extend the relaxation of the protein. There are three options for selecting the residues, listed under Gas phase optimization shell:

- Residues within *N* Å—Optimize all residues that have atoms within the specified distance of the mutated residue pair.
- Adjacent residues in sequence—Optimize the *N* residues next in the sequence on either side of the residues in the pair, where *N* is selected from the option menu.
- None—Do not optimize any residues but the two residues in the pair.

14.4 Running the Job

When you have selected the residue pairs that you want to mutate, and set any options for relaxing the structure around the residue pairs, click Start. The Start dialog box opens, so that you can choose an output option and name the job before running it. There are two output options:

- Append new entries—Append the mutated structures to the project.
- Do not incorporate—Leave the mutated structures in the working directory.

The Maestro format file containing the structures is copied to the working directory when the job finishes, regardless of the option. If you choose not to incorporate the results into the project, you can always do so later by importing the output file.

When you have chosen the output option and named the job, click Start. The job is run locally. The progress of the job is displayed in a status bar at the bottom of the panel.

14.5 Examining the Results

The results of a cysteine mutation job are automatically displayed in the Results tab if you chose to append new entries in the Start dialog box. To view results of any other cysteine mutation job, click Load Results from Previous Run. This button opens a file selector, in which you can browse to the output Maestro file (`-out.maegz`) and load it.

The results of the mutation job are displayed in the Mutations table. The table columns are described in Table 14.2.

Selecting a row in the table zooms in on the mutated pair in the Workspace. The mutated residues are displayed in ball-and-stick representation, and the rest of the structure uses a darker color scheme. If you want to see the original residues as well, select Display original structure in gray.

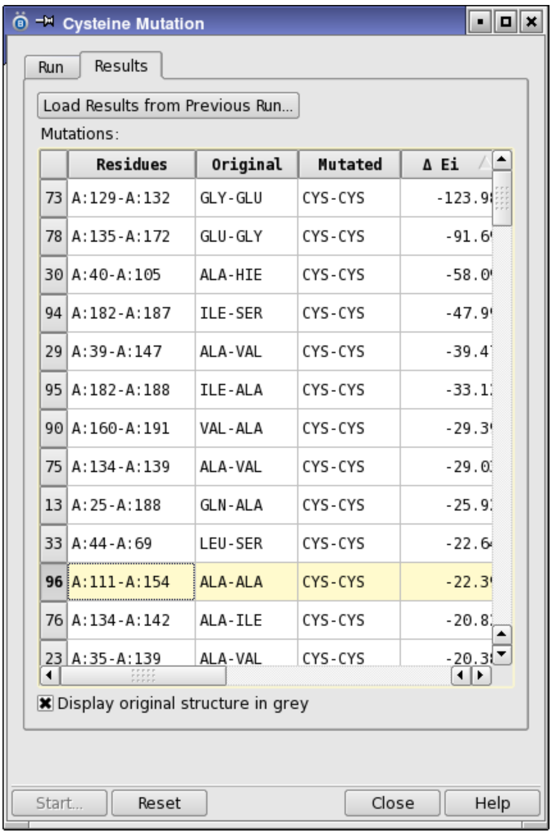


Figure 14.2. The Results tab of the Cysteine Mutation panel.

Table 14.2. Columns of the Mutations table.

Column	Description
Residues	Chain name, residue number and insertion code of both residues in the mutated pair.
Original	3-letter names of original residues in the pair.
Mutated	3-letter names of mutated residues.
Δ Ei	Change in interaction energy between the residue pair and the rest of the protein on mutation.
Δ Strain E	Change in strain energy on mutation. The strain energy is the difference in internal energy between the state of the residue pair in the protein and the relaxed state of each residue or of the disulfide in the gas phase.
Δ Ei+Strain E	Change in the sum of interaction energy and strain energy on mutation, equal to Δ Ei + Δ Strain E.
Quality	Classification of the quality of a predicted disulfide bond into Good, Medium, Bad, based on the changes in interaction energy and strain energy on mutation. Good Δ Ei < 0 and Δ Strain E < 20. Bad Δ Ei > 40 or Δ Strain E > 40 or Δ Ei > 20 and Δ Strain E > 20 Medium Any other values

14.6 Workflow Summary

1. Display the protein you want to analyze in the Workspace. You should ensure that it is properly prepared (with the Protein Preparation Wizard).
2. If you want to analyze only part of the protein, select the residues that you want to analyze, and select Analyze only selected Workspace residues.
3. Click Analyze Workspace.

   When the analysis finishes, the Residue pairs for mutation table is populated with the residue pairs that were found.
4. Filter the list to show only the mutations you are interested in:
5. Use the display options to display the mutation types that you are interested in.
6. Use the menus below the table to select the residues that you want to mutate to or from. The table is updated to show only those residues.
7. Use the cutoff to filter out pairs for which the distance is too large to form a disulfide bond.

8. Select the pairs that you want to mutate in the table. Each pair is mutated independently to produce a structure and an energy evaluation.
9. Decide whether you want to optimize residues near to the mutated residues by selecting from the Gas phase optimization shell options. The mutated residues are always optimized.
10. Click Start to run the mutation job. You can set the job name and decide where to send the output in the dialog box that opens.
11. When the job finishes, go to the Results tab. If the results are not listed in the Mutations table, click Load Results from Previous Run, navigate to and open the output Maestro file from the job.

Antibody Modeling

The main task of antibody modeling in BioLuminate is to construct a homology model based on a database of antibody templates. Once the model is constructed, it can be humanized if required. A database of antibody templates is provided with BioLuminate, based on a new analysis of antibodies in the PDB from 2010 [2], which you can modify or add to, or create your own database. Antibody-antigen docking is covered in Section 10.5 on page 53.

15.1 Modeling an Antibody Structure

Modeling of the Fv region of an antibody involves prediction of both the framework (FR) region and the variable loop (CDR) region. These regions are identified automatically and their structure predicted by homology, based on known antibody structures from the PDB, or from your own database of antibodies. You can also use input coordinates for some parts of the structure and predict the rest. Finally, the H3 loop can be refined after the initial structure is generated.

The modeling is performed using the Antibody Prediction panel, which you open by choosing Tasks → Antibody Modeling → Prediction in the main window.

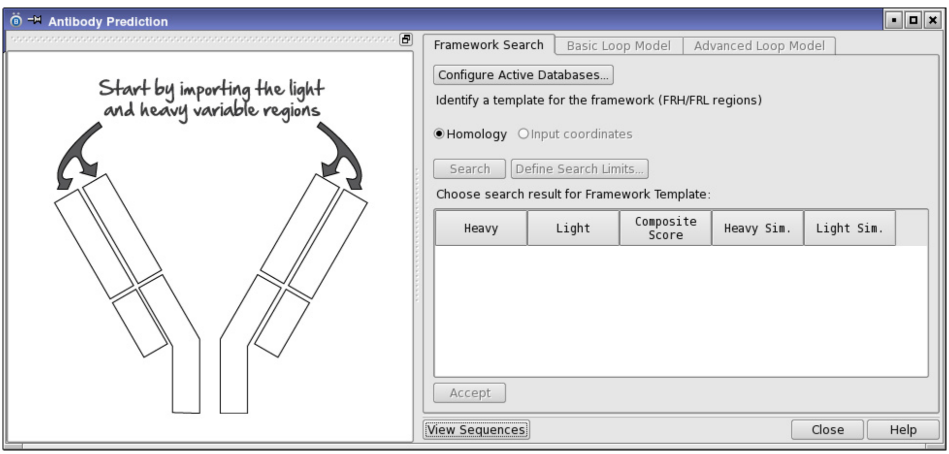


Figure 15.1. The Antibody Prediction panel, initial view.

15.1.1 Importing the Antibody Sequence

The first step in the modeling process is to import the antibody sequence, and if you want to re-predict only a part of the antibody, the existing structure.

The left part of the panel displays a diagram of the Fv region of an antibody. Clicking on the light variable or heavy variable region in the diagram displays a menu, from which you can choose the source of the sequence for this region. The choices are:

- Browse for File—Open a file browser in which you can navigate to the desired location and select the file that contains the sequence.
- From Workspace—Use the sequence for the structure that is displayed in the Workspace. Only one structure must be displayed.
- From Selected Entries in the Project Table—Use the sequence from the entry that is selected in the Project Table. Only one entry must be selected.
- From PDB ID—Use the sequence from the specified PDB ID. Opens the Enter PDB ID dialog box, in which you can enter the PDB ID of the sequence. The sequence is retrieved from a local copy of the PDB if it is available, or from the RCSB web site, depending on the preference set for PDB retrieval.
- Enter/Paste New Sequence—Type or paste in the sequence. Opens the Sequence Editor dialog box, in which you can name the sequence and type it or paste it in, as a string of single-letter codes.

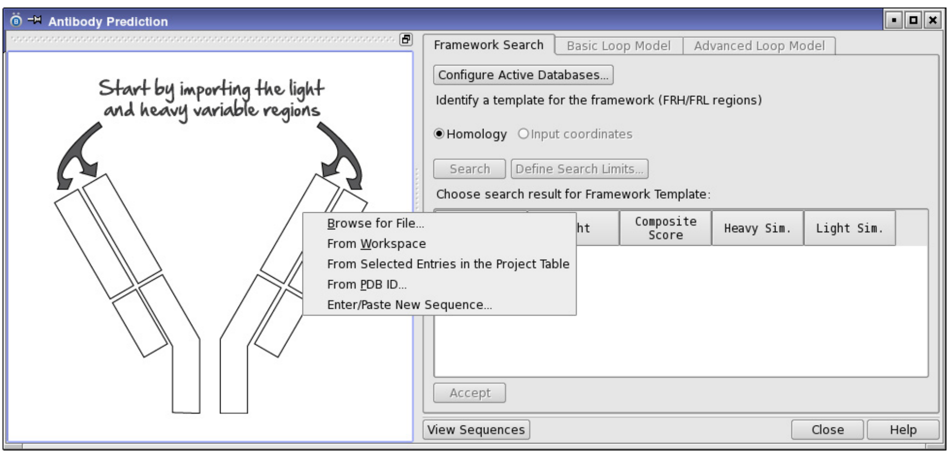


Figure 15.2. The Antibody Prediction panel showing the import menu.

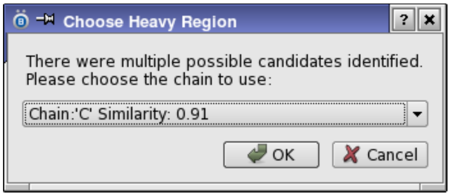


Figure 15.3. The Choose Heavy Region dialog box

If you intend to model only part of an antibody and use the input structure for the rest, you should ensure that you import a sequence that has an associated structure. This means that you can choose any of the menu items except the last.

When the protein is read in, it is analyzed to find the chains and the loops. If there is more than one possible chain that could be used (for example, in a dimer), a dialog box opens, in which you can choose one of the chains. When the analysis finishes and you have chosen a chain if requested, the region in the diagram is colored to indicate that the chain is assigned. When both chains have been chosen, the text prompting you to import the two chains is no longer displayed.

You can view the sequences in a sequence viewer at any time by clicking View Sequences.

15.1.2 Choosing a Database

The model of the antibody is built using a database of antibodies, each of which has been analyzed and curated in terms of framework and hypervariable loop regions. This database is used to model antibody structures from a sequence. A curated database obtained from the PDB is provided with the software, but you can add your own structures to customize it, or build your own databases—see Section 15.3 on page 110.

You can choose one or more databases to use when modeling an antibody. The default is the database in the installation. If you want to select the databases, click Configure Active Databases, to open the Database Configuration dialog box. This dialog box has a table of databases that you can choose from. You can do the following:

- Add a database to the list, by clicking Add Database and navigating to the database in the file selector that opens.
- Select a database to use, by checking the check box in the Active column.
- Remove a database from the list, by clicking the button in the Actions column.

When you have finished modifying the list and choosing the active databases, click OK.

15.1.3 Selecting the Coordinates for the Framework Region

You can choose between two sources for the coordinates of the framework region: the input coordinates, or a homology model.

If you want to use input coordinates, select Input coordinates. The sequences that you imported for the light and heavy regions must of course be associated with a structure, and an error message is posted if there is no structure available.

If you want to use a homology model, select Homology. To run the search, click Search. When the search finishes, the table in the lower part of the Framework Search tab is filled in with the results, in order of their score. The table columns are described in Table 15.1. You can select a single result in the table to use as the template. The table columns are described below.

Table 15.1. Description of search results table columns.

Column	Description
Heavy	PDB ID of the homolog for the heavy framework region
Light	PDB ID of the homolog for the light framework region
Composite Score	Average of the Heavy Sim. and Light Sim. scores. This score is used to order the homologs in the table.
Heavy Sim.	Sequence similarity of the entire variable domain sequence (framework and CDR) for the heavy chain. The similarity is the number of matching residues divided by the total number of residues, where “matching” means that the two residues have a positive score in the BLOSUM62 matrix.
Light Sim.	Sequence similarity of the entire variable domain sequence (framework and CDR) for the light chain. The similarity is the number of matching residues divided by the total number of residues, where “matching” means that the two residues have a positive score in the BLOSUM62 matrix.

When you have selected a template or chosen to use input coordinates, click Accept to accept the choice and move on to the next stage. The Basic Loop Model tab is displayed automatically, after a short pause.

15.1.4 Filtering the Database by Property

By default, the homology search scans all of the structures in the databases you have chosen. If you want to restrict the search to only the structures that have certain properties, you can set up a filter on these properties before conducting the search. To set up the filter, click Define Search Limits, which opens the Limit Criteria Definition dialog box.

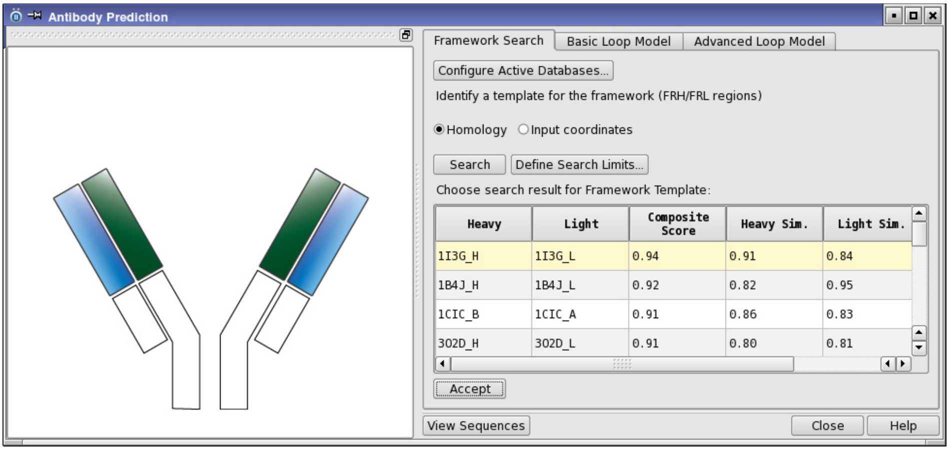


Figure 15.4. The Antibody Prediction panel after searching for templates.

First, choose a property from the Available properties list at the top of the panel. The list shows the property name, the family (category) it belongs to, and the range of values, for numeric properties. You can limit the list to a particular property family by choosing from the Show family option menu. If you type in the Property text box, a completion list is displayed below it, from which you can choose a property.

Once you have chosen a property, it is displayed in the Property text box. You can then use the text boxes and menus to the right to define the restrictions on the values of this property. Click Add to add the filter to the Filtering definitions and criteria list. You can add multiple criteria or definitions, and each of them is applied to the databases. The number of structures in the database that match the filters is reported at the end of the list. Property values are case-sensitive, so for example 1FSK does not match 1fsk.

If you want to filter on multiple properties, you can choose another property, set up the restrictions on its value, and click Add to add the filter to the list. The filters are cumulative (implicit AND), so the resulting structures are those that match both filters.

As an example, to include structures with properties in a specified range (all values from a minimum to a maximum inclusive):

1. Choose >= from the first menu.
2. Enter the minimum in the next text box.
3. Choose AND from the second menu.
4. Choose <= from the third menu
5. Enter the maximum value in the final text box.

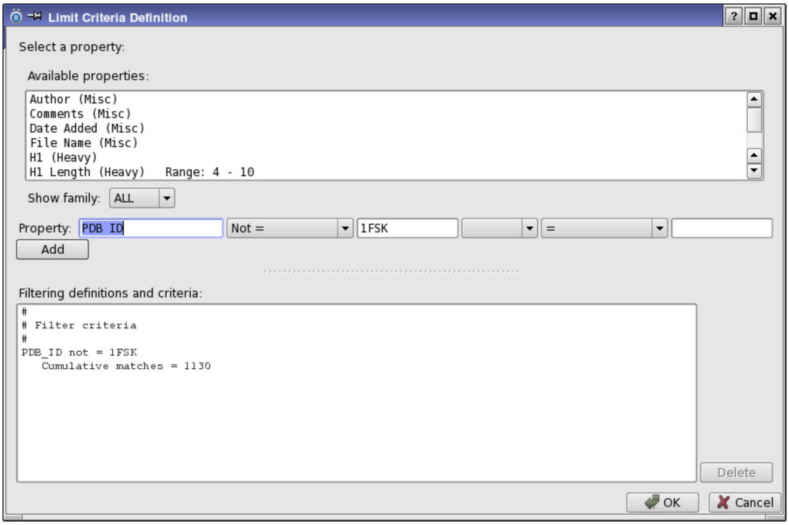


Figure 15.5. The Limit Criteria Definition dialog box.

As another example, to filter our structures for which a property has values in a given range:

1. Choose < from the first menu
2. Enter the minimum in the next text box
3. Choose OR from the second menu
4. Choose > from the third menu
5. Enter the maximum value in the final text box.

15.1.5 Generating the Loop Model

With the framework region defined, you can now generate a model for the six loops. As for the framework region, you have the choice of using input coordinates for a loop or predicting the loop from the database.

The six CDR loops are listed in a table that provides check boxes for selecting the source of input coordinates. To choose the source, check the appropriate box. By default all loops are predicted from the database. Accepting the default loop assignments is usually the best option.

If you want to choose individual clusters for one or more loops, click View Clusters. This opens the Loop Clusters dialog box, which shows you the clusters for each loop, and allows you to select a cluster to use in the model. It also allows you to set a similarity cutoff for the loop to be used in the model. The procedure is described in more detail in Section 15.1.6 on page 98.

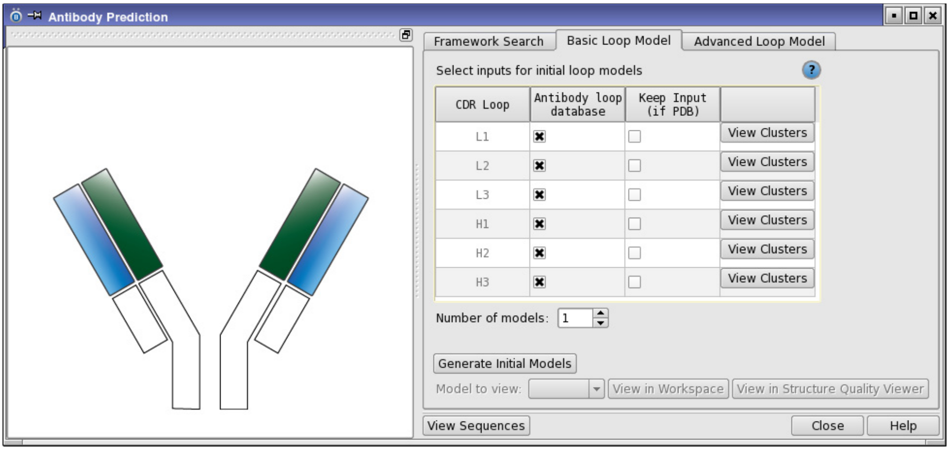


Figure 15.6. The Antibody Prediction panel, Basic Loop Model tab.

Otherwise, the selection of the loop is done by sorting the clusters by the size of the cluster, then locating the largest cluster in the list that has a member whose loop similarity to that of the query is greater than the loop similarity cutoff. The cluster member with the greatest similarity to the query is the one that is used to build the loop.

You can build more than one model for the structure, by setting the desired number in the Number of models text box. If more than one model is requested, a series of diverse models is returned. The first model returned is usually the most likely to be correct.

To generate the loop model or models, click Generate Initial Models. These are called "initial models" because you might want to refine the models, particularly the H3 loop. To assess the quality of a generated model, you can examine it in the Workspace (click View in Workspace), or analyze it in the Structure Quality Viewer (click View in Structure Quality Viewer). If you have multiple models, you can choose them from the Model to view option menu.

When you view a model in the Workspace, it is colored by residue with the following color scheme:

- Blue—residues for which the full residue conformation was copied from the template.
- Cyan—residues for which the residue backbone conformation was copied from the template, and the side chain was modeled (because there was a residue mutation in the template relative to the query).
- Red—residues for which both the backbone and the side chain were modeled.
- Maroon—residues in the CDR loops.

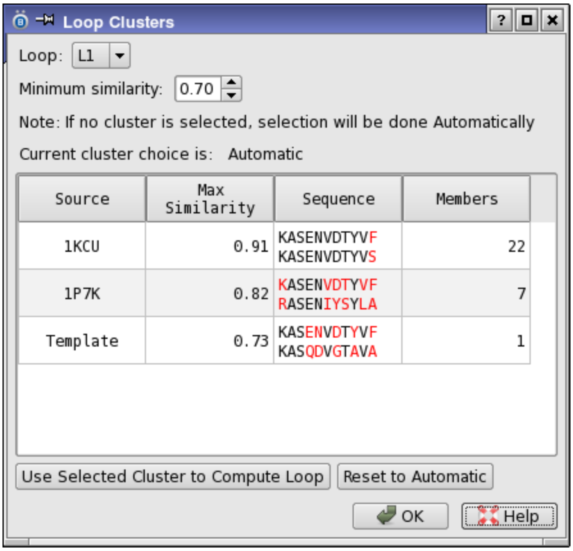


Figure 15.7. The Loop Clusters dialog box.

When you view the models in the Protein Structure Quality Viewer, all models are listed in the table at the top of the viewer, and the structures are colored in the Workspace according to the Ramachandran plot regions.

15.1.6 Selecting Clusters for a Loop

If you want to manually select the loops that are used to build the model, you can do so in the Loop Clusters dialog box, which you open by clicking View Clusters in the Basic Loop Model tab. In this dialog box you can examine the loop clusters for each of the CDR regions, filtered by similarity, and select a cluster for each loop. You can also set the minimum similarity that is used in selecting a cluster, both manually and in the default automatic procedure.

When this dialog box is first opened, the antibody databases are searched for loops of the same length as those in the query sequence for each of the six loops, and the loops are clustered structurally. The dialog box may therefore take a short while to open. Subsequently these loops and clusters are reused, so opening is much faster.

The clusters for the loop chosen from the Loop option menu are listed in the table, in order of decreasing similarity to the query sequence. The list is restricted to the clusters that have a loop whose similarity to the query loop is greater than the cutoff specified in the Minimum similarity box. The table columns are described in Table 15.2.

You can set the minimum acceptable similarity between a loop from the database and the query loop in the Minimum similarity box. Only the clusters that contain at least one loop whose similarity to the query is greater than this threshold are used in building the model, whether in the default automatic procedure or by manual selection of a cluster.

If the minimum cluster similarity is changed, this change applies to all loops. If no cluster is found with a loop that has the minimum similarity, then the program automatically uses the template loop in the database that has the greatest similarity to the query, no matter which cluster it belongs to.

Table 15.2. Columns of the cluster table.

Column	Description
Source	PDB ID of the member of the given cluster that has the highest similarity to the query. The template used for the framework region is also included in the list.
Max Similarity	Similarity of the member of the cluster that has the highest similarity to the query.
Sequence	Sequence of the query for this loop (top) and of the most similar cluster member (bottom). Residues that differ are marked in red.
Members	Number of members of the cluster.

To choose a loop cluster for a particular loop:

1. Select the loop from the Loop option menu.
2. Select the cluster in the table.
3. Click Use Selected Cluster to Compute Loop.

The cluster choice is shown above the table.

To clear a loop cluster selection:

1. Select the loop from the Loop option menu.
2. Click Reset to Automatic.

15.1.7 Refining Loops

Frequently, the model based on homology is adequate, and no further prediction is needed. However, in some cases, prediction of certain loops (especially H3) can be problematic, and advanced methods can result in better predictions.

The H3 loop, however, is often difficult to predict, and may need more advanced refinement. You can carry out advanced loop refinement using the Advanced Loop Model tab. From this

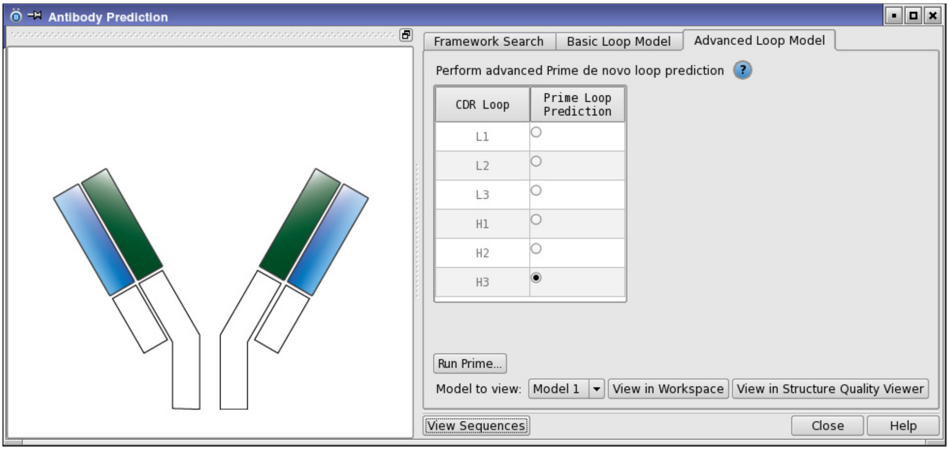


Figure 15.8. The Antibody Prediction panel, Advanced Loop Model tab.

tab, you can use Prime protein refinement to re-predict selected loops. If you generated more than one model, you should select the model you want to use from the Model to view option menu before leaving the Basic Loop Model tab.

Some general guidelines for when to refine a loop with Prime in the Advanced Loop Model tab are given below:

- If the loop is long (9 residues or more) and the homology to the template is good (similarity above about 80%) then refinement with Prime is not usually necessary.
- If the loop sequence similarity is less than 40%, the basic model quality usually is very poor and a Prime refinement is recommended.
- The quality of the Prime refinement is greater for shorter loops.
- When building a new H3 loop with new sequences on the native (crystal) structure of an antibody, Prime refinement is recommended.

Accurate and detailed loop predictions using Prime can take hours for each loop. It is not usually necessary to run an advanced loop prediction for any of the loops except the H3 loop. The H3 loop is selected by default for an advanced prediction, as it is the hardest loop to predict. The results are returned to the Project Table as new structures.

The input structure for the loop prediction is always the structure that came from the Basic Loop Model tab. This means that you cannot do sequential loop predictions in this panel, but you can use the Refinement panel to predict more than one loop (Tasks → Loop + Sidechain Prediction). See Chapter 6 of the *Prime User Manual* for more information on refinement tasks.

To choose the loop for prediction, select it in the table. Click Run Prime to run a Prime loop prediction job for the selected loop. A dialog box opens, in which you can name the job, select a host, and specify the maximum number of processors to use. The single, best loop prediction is returned. The Prime loop prediction algorithm is automatically selected based on the length of the loop.

The same controls as in the Basic Loop Model tab are present for viewing the structure.

15.1.8 Summary

The basic procedure for running a prediction using a homology model is as follows:

1. Import the sequences for the light and heavy chains: Click the part of the diagram for the region you want to import, and choose a source from the list that is displayed.
2. Set up the antibody databases that you want to use to search for homologs for the framework region and the loops:
3. (Optional) Click Configure Active Databases in the Framework Search tab to add databases and select them in the table. The default database is the one from the installation.
4. (Optional) Click Define Search Limits to filter the databases by properties of the structures in the databases.
5. Select Homology and run the homology search by clicking Search.
6. Select the homolog in the results table that you want to use for the framework region, and click Accept.
7. (Optional) In the Basic Loop Model tab, select the loop cluster you want to use for each loop in the model by clicking View Clusters, and choosing a cluster in the Loop Clusters panel. By default a cluster is chosen automatically, based on cluster size and a minimum similarity criterion.
8. (Optional) Set the number of models of the antibody that you want to generate.
9. Click Generate Initial Models to generate the models.
10. (Optional) If you think that the H3 loop needs further refinement, select it in the Advanced Loop Model tab, and click Run Prime.

The models are added to the Project Table.

You can also use coordinates from existing structures instead of homology models, for example if you want to vary just one of the loops. The procedure is similar to that given above.

1. Import the sequences for the light and heavy chains: Click the part of the diagram for the region you want to import, and choose a source from the list that is displayed.
2. Set up the antibody databases that you want to use to search for homologs for the framework region and the loops:
3. (Optional) Click Configure Active Databases in the Framework Search tab to add databases and select them in the table. The default database is the one from the installation.
4. Select Input coordinates and click Accept.
5. (Optional) In the Basic Loop Model tab, select the loop cluster you want to use for each loop in the model by clicking View Clusters, and choosing a cluster in the Loop Clusters panel. By default a cluster is chosen automatically, based on similarity, then cluster size.
6. (Optional) Check the boxes for the loops for which you want to use input coordinates.
7. (Optional) Set the number of models of the antibody that you want to generate.
8. Click Generate Initial Models to generate the models.
9. (Optional) If you think that the H3 loop needs further refinement, select it in the Advanced Loop Model tab, and click Run Prime.

15.2 Humanizing Antibodies

If you have an antibody model that is not based on a human species, you can humanize it by mutating residues to create a more human-compatible protein. Appropriate sites for mutation can be found by comparing the antibody to homologs and identifying residues that satisfy various criteria, such as solvent accessibility and maximum number of side-chain interactions with the antigen or with the antibody itself. Multiple mutations can be done at each site, but only one site is mutated in each mutant structure.

The mutations can be set up and run in the Humanize Antibodies panel, which you open by choosing Tasks → Antibody Modeling → Humanize in the main window. The panel has two tabs, one for setting up the criteria for choosing residues to mutate, and one for selecting the residues and defining the mutants. When both these tasks are done, you can start the job to mutate the residues.

15.2.1 Analyzing the Antibody

The first step in the process is to analyze the antibody to locate the antibody regions and calculate solvent-accessible surface areas (SASA). By default, the entire Workspace is analyzed, but if you want to limit the analysis, you can select Analyze only selected Workspace residues to limit the analysis to the Workspace selection. This is useful if you want to analyze the effect of the mutations on the binding of the antibody to an antigen, which can be done when the mutations are performed. You can include the entire complex in the Workspace and select only the antibody chains to analyze. If you want to examine binding, you must include the entire structure in the Workspace at this stage.

Once the analysis is complete, the sequence is shown in the panel's sequence viewer, colored by residue type, with its secondary structure assignment and disulfide bond annotation. You can choose which chain to display by using the Show in viewer options. This sequence is called the "parent" sequence.

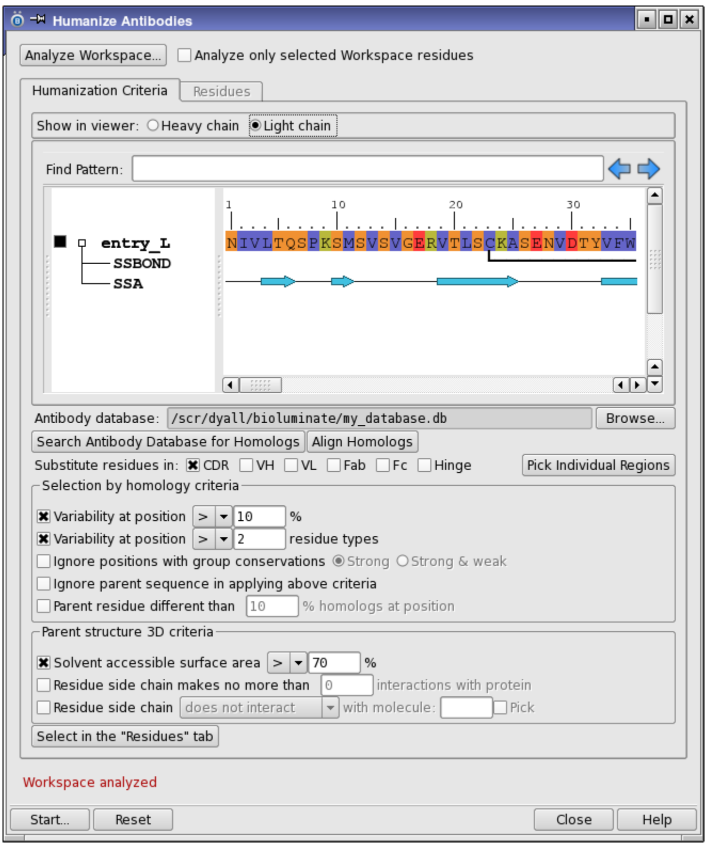


Figure 15.9. The Humanize Antibodies panel, Humanization Criteria tab.

15.2.2 Finding Homologs

The next step is to search for homologs in an antibody database. These homologs are used to identify residues that could be mutated, based on a comparison of the homologs with the parent sequence.

The path to the database that will be used for the search is displayed in the (noneditable) Antibody database text box. If you want to change the database, click Browse and navigate to the database. The default database is a database of human antibody data.

To find the homologs, click Search Antibody Database for Homologs. The progress of the search is shown in the status area at the bottom of the panel. When the search is done, the homologs must be aligned to the parent sequence, so that selection of residues for mutation can be done on the basis of matching residue positions. Click Align Homologs to perform the multiple sequence alignment of the homologs to the parent.

15.2.3 Selecting Regions to Mutate

If you want to limit the mutations to specified regions of the antibody, you can do so by selecting one or more of the Substitute residues in options. The regions can be selected by group: CDR, VH, VL, Fab, Fc, or Hinge. These options are available by default. For finer control, you can select individual regions. To display the options for the individual regions (which are hidden by default), click Pick Individual Regions. You can then pick the individual regions of the antibody for mutation. Note that you can only pick by group or pick individual regions: there is no connection between the two.

15.2.4 Setting Up Residue Selection Criteria

The next task is to set up criteria for selecting the residues to mutate. There are two kinds of criteria that you can use: those based on homology, and those based on the 3D structure of the parent.

Criteria that are based on homology use the variations in the residues at each residue position among the homologs, or between the homologs and the parent. The variations found are used as a basis for choosing default mutations. The criteria that you can set are:

- Variability at position >/< *N* %—Filter residues based on the percentage variability at the residue position. Choose whether to apply a minimum or a maximum variability from the option menu, and specify the percentage threshold in the text box.
- Variability at position >/< *N* residue types—Filter residues based on the residue type variability at the residue position. Choose whether to apply a minimum or a maximum variability from the option menu, and specify the threshold for the number of residue types

that can vary in the text box.

- Ignore positions with group conservations—Ignore (do not select) residues that are strongly conserved or that are either strongly or weakly conserved. Select the appropriate option for strong conservation or both strong and weak conservation.
- Ignore parent sequence in applying above criteria—Ignore the parent (query) sequence when applying the variability or conservation criteria. Only the variability in the homologs is considered.
- Parent residue different than *N* % homologs at position—Select residues for which the parent residue is different from more than the specified percentage of homologs at the residue position.

The criteria based on the 3D structure of the parent antibody include solvent-accessible surface area (SASA) and interactions between residues. Interactions are determined by a distance cutoff: any residue that has atoms within 4 Å of a given side chain is considered to interact with it.

- Solvent accessible surface area—Select this option to filter residues by their solvent-accessible surface area (SASA) relative to an isolated residue of the same type, and set a threshold for the maximum or minimum allowed relative SASA. This option is useful for locating surface (or buried) residues.
- Residue side chain makes no more than *N* interactions with protein—Select this option to filter out residues whose side chains make multiple interactions with the protein, and set the maximum number of such interactions.
- Residue side chain interacts/does not interact with molecule *N*—Select this option to filter residues by their interaction with a selected molecule. Choose whether to allow or disallow the interaction from the option menu, and specify the molecule in the text box, or select Pick and pick the molecule in the Workspace.

15.2.5 Selecting the Residues and Their Mutations

When you have finished setting up the selection criteria, click Select in the Residues Tab. The residues that match the criteria you provided in the Humanization Criteria tab are now selected in the Residues tab.

At the top of the Residues tab is a table that lists all the residues in the Workspace. The first table column contains check boxes that you can use to select residues for mutation. The residues that were selected in the Humanization Criteria tab are already selected for mutation. The second column identifies the residues. The third column specifies the mutations of the residues that are selected for mutation.

To select other residues for mutation, you can select the check box in the Mutate column, or you can select table rows and click Check Mutation Box for Selected Residues.

To define the mutations for a given residue, click in the Mutations column. A menu is displayed in the column, from which you can select one or more residues to mutate to, or select groups of residue types. When you make a selection from the list, the residue is checked in the list and is added to the text box at the top of the menu. To close the list, click the arrow button at the right of the text box, or click in some other location. To hide the menu, click somewhere else in the table (for example, in the Residue column). The residue is then selected for mutation if it is not already selected.

To define a common set of mutations for multiple residues, first select the residues in the table. Then choose the residues from the For all selected rows, set mutations to menu, and click Apply. This menu works in the same way as the menu in the Mutations column of the table.

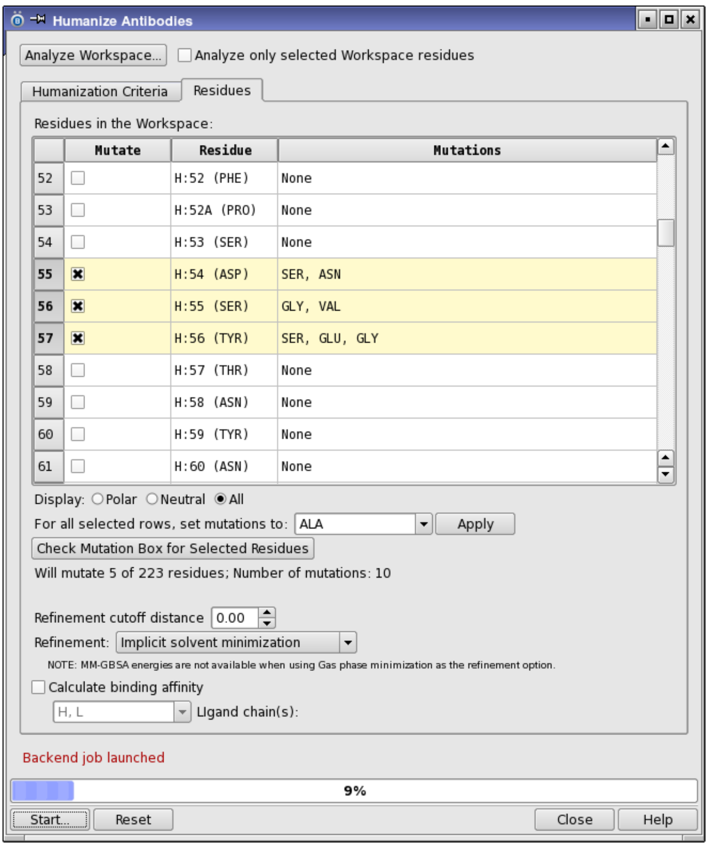


Figure 15.10. The Humanize Antibodies panel, Residues tab.

By default, a set of mutations based on the variations seen in the homologs is chosen for the residues you selected in the Humanization Criteria tab.

The number of mutations is reported below the Check Mutation Box for Selected Residues button.

15.2.6 Setting Refinement Options for Mutated Residues

In the job that performs the mutations, the structure of the mutated residue is optimized, but you can also refine the structures of adjacent residues. The residues that are included in the refinement are selected by distance from the mutated residue. Any residue that has an atom within the specified distance of a hypothetical Arg residue at the mutation site is included in the refinement. A hypothetical Arg residue is used to ensure that the set of residues refined is identical regardless of the initial or mutated residue identities.

There are several methods available for refinement of the mutated (and other) residues, on the Refinement menu:

- Gas-phase minimization—Minimize the energy of the residues using the internal force-field minimizer. This is the fastest method, but does not include solvation effects.
- Implicit solvent minimization—Minimize the energy of the residues with an implicit (continuum) solvation model, using Prime.
- Side-chain prediction—Prior to minimization, a thorough exploration of possible side chain conformations is performed.
- Side-chain prediction (cbeta)—Prior to minimization, a thorough exploration of possible side chain conformations, including the CA-CB torsion, is performed.
- Side-chain prediction (bb)—Prior to minimization, a thorough exploration of both potential side chain and potential backbone conformations is performed.

With the exception of the first, the refinements are run with the Prime protein modeling program.

As part of the refinement, you can calculate the binding affinity of selected chains to the rest of the Workspace structure, by selecting Calculate binding affinity. The calculation is performed with the MM-GBSA method in the Prime program. This option is not available with gas phase minimization.

The chains are treated as the ligand (which in this situation can be a structure of any size, such as an antibody). By default, the heavy and light chains of the antibody are selected, so the binding affinity of the antibody to the rest of the Workspace is calculated, such as an antigen. A choice of chains to treat as the “ligand”, derived from the analysis of the Workspace structure, is offered in the option menu.

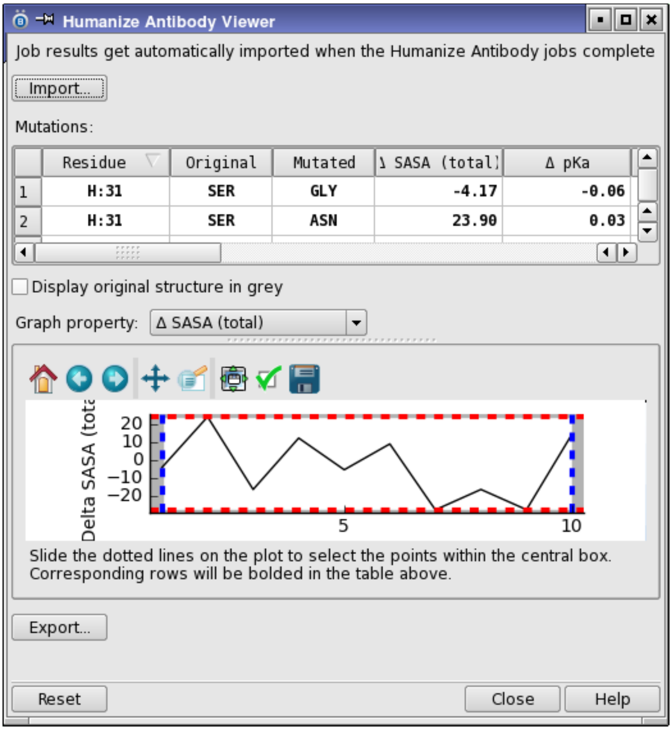


Figure 15.11. The Humanize Antibody Viewer panel.

15.2.7 Running the Mutation Job

When you have finished making settings, click Start. A dialog box opens, allowing you to choose to add the mutated structures to the project (Append new entries) or to leave the structure file in the working directory (Do not incorporate). You can also name the job. Click Start in this dialog box to start the job, which runs locally.

15.2.8 Analyzing the Results

When the job finishes, the Humanize Antibody Viewer opens automatically, with the results loaded. You can also open this panel by choosing Tasks → Antibody Modeling → Humanization Results, and then clicking Import in the panel to locate and load a set of results.

The Mutations table lists all the mutations that were generated, along with the changes in a range of properties as a result of the mutation. The properties include SASA (total, polar, and nonpolar), pK_a, hydropathy, number of rotatable bonds, energy, potential energy, and stability. These quantities are defined in Table 13.1 on page 79. You can sort the table columns by clicking on the column heading. You can plot any of these properties against the mutation (table row) by choosing the property from the Graph property option menu. If you want to export the table data as a CSV file, click Export, and navigate to a location and name the file.

You can select a region in the graph using the horizontal and vertical dashed lines, which can be dragged to create the selection. The rows corresponding to the selected region of the graph are highlighted in the table above, and the residues are highlighted in the Workspace.

If you select a table row, the view zooms in to the mutated residue. To display the original structure, select Display original structure in grey. The parent antibody is displayed and colored grey. You can then see how the mutation is positioned in relation to the original residue.

15.2.9 Summary

1. Display the structure in the Workspace and analyze it, to identify residues and antibody features (Analyze Workspace).
2. Load an antibody database that is used to search for homologs.
3. Run the search (Search Antibody Database for Homologs) and align the homologs (Align Homologs).
4. (Optional) Select an option for the chain to show, and examine the alignment for that chain.
5. Choose the regions that you want to substitute residues in.
6. Specify the criteria for automatic selection of residues, in the Selection by homology criteria and Parent structure 3D criteria sections.
7. Click Select in the Residues Tab to select the residues that meet the criteria.
8. In the Residues tab, make any changes to the residues marked for mutation.
9. Choose the mutations for these residues by clicking in the Mutations column of the table and selecting from the option menu that is displayed.

   A default selection is included for the automatically selected residues that is based on the variations seen in the homologs.
10. Click Start, make job settings, and start the job.

    The mutated structures are incorporated into the project as new entries, and the Humanize Antibody Viewer opens.

15.3 Antibody Databases

Modeling antibodies is done with the help of a database of prepared antibody structures. A database is supplied with the distribution, and is the default database. You can create and manage your own databases in the Antibody Database Management, which you open by choosing Tasks → Antibody Modeling → Database Management in the main window.

Structures that are added to a database are automatically characterized and curated. Antibody structures and the light and heavy chains are identified using a similarity search against known antibodies, and structures that do not meet the similarity criteria are rejected. Then the constituent regions of the antibody chains, including the framework region, as well as the six hypervariable loops, are identified and annotated for use in subsequent predictions.

When you first open the panel, the default database is loaded. Normally this database is installed by an administrator and you do not have write privileges, so it is opened read-only. You can import this database into your own database if you want, as described below. It is only necessary to do this if you want to modify the data in some way, because you can specify multiple databases for modeling.

To open a database, or to create a new database, click Open/Create. A file selector opens, and you can navigate to the desired location. If you want to open a database, select the database from the list of files. It should have the extension `.db`. If you want to create a new database, enter the name in the File name text box.

The structures in the database are shown in the antibody table. You can restrict the structures that are listed in the table by searching for strings in the table and only showing the rows that contain the string. By default, all visible text is searched, but you can change it by clicking

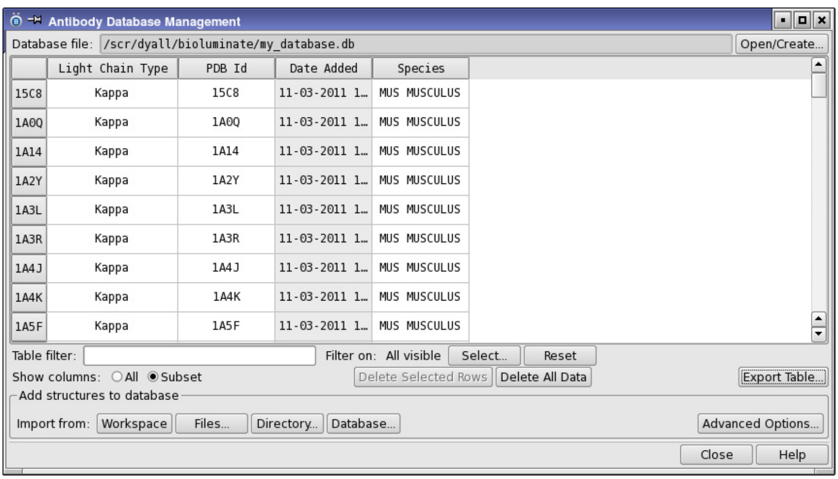


Figure 15.12. The Antibody Database Management panel.

Select and choosing the columns you want to search on. To return to sorting all visible columns, click Reset. The text box has a tool tip that explains the syntax for the search string, which can include relational operators, wild cards, regular expressions, and some special terms.

The table shows only a few columns by default. If you want to see all the columns, by Show columns, select All. The full set of columns includes the identity (residue range) and length of each loop, and a range of information from the originating PDB structure. If you want to export the information in the columns to a CSV file, click Export Table, and navigate to a location and name the file in the file selector that opens.

To delete structures from the database, select them in the table and click Delete Selected Rows. If you want to clear the entire database, click Delete All Data. You should exercise care when deleting rows or all data, as these functions are not reversible.

Structures can be added to the database from several sources, represented by buttons in the Add structures to database section. In each case, the imported data is automatically processed for you: the antibody chains are identified, the constituent regions of the chains are determined, and all the pertinent information required for subsequent modeling is saved in a rapidly accessed format. You need only supply the antibody structures in PDB (or Maestro) format.

- Workspace—import structures from the Workspace.
- Files—import structures from PDB files or Maestro files. Opens a file selector, in which you can navigate to and select multiple files for import.
- Directory—import all structures from a directory. Opens a directory selector, in which you can navigate to the directory to import from.
- Database—import structures from an existing database. Opens a file selector, in which you can navigate to and select the database (.db) to import from. Import progress is displayed in a bar at the bottom of the panel.

The structures are filtered to ensure that only those structures that have characteristics of antibodies are included in the database. You can alter the criteria for filtering structures in the Antibody Database - Advanced Options dialog box, which you open by clicking Advanced Options. You can also select the file types that are presented when importing structures from files.

In the Minimum sequence identity section, you can specify cutoffs for determining whether a structure should be included in the database, based on percentage sequence identity. If no chain in the structure passes the tests, it is not included in the database. A chain must meet the FR threshold and either the VL or VH threshold to be included in the database. If a chain meets only one of the VL or VH thresholds, it is only included as a light or heavy variable chain.

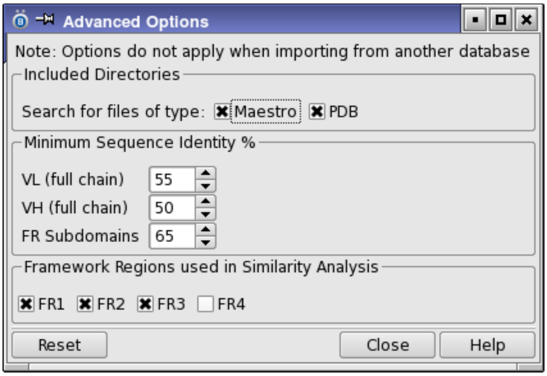


Figure 15.13. The Antibody Database Management Advanced Options panel.

For the framework regions, you can specify which of the four framework regions to use in the similarity analysis, by selecting the options for FR1, FR2, FR3, or FR4 in the Framework regions used in similarity analysis section.

To revert to the default options, click Reset.

References

1. Olsson, M. H. M.; Søndergard, C. R.; Rostkowski, M.; Jensen J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a predictions. *J. Chem. Theor. Comput.*, **2011**, *7*, 525–537.

2. North, B.; Lehmann, A.; Dunbrack, R.L., Jr. A new clustering of antibody CDR loop conformations. *J. Mol. Biol.* **2011**, *406*, 228.

3. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**, *20*, 45–50.

4. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res.* **2004**, *32*, W96–W99.

5. Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **2006**, *65*, 392–406.

6. Kyte, J.; Doolittle R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

Getting Help

Information about Schrödinger software is available in two main places:

- The `docs` folder (directory) of your software installation, which contains HTML and PDF documentation. Index pages are available in this folder.
- The Schrödinger web site, http://www.schrodinger.com/, particularly the Support Center, http://www.schrodinger.com/supportcenter, and the Knowledge Base, http://www.schrodinger.com/kb.

Finding Information in Maestro

Maestro provides access to nearly all the information available on Schrödinger software.

To get information:

- Pause the pointer over a GUI feature (button, menu item, menu, ...). In the main window, information is displayed in the Auto-Help text box, which is located at the foot of the main window, or in a tooltip. In other panels, information is displayed in a tooltip.

  If the tooltip does not appear within a second, check that Show tooltips is selected under General → Appearance in the Preferences panel, which you can open with CTRL+, (⌘,). Not all features have tooltips.
- Click the Help button in a panel or press F1 for information about a panel or the tab that is displayed in a panel. The help topic is displayed in your browser.
- Choose Help → Online Help or press CTRL+H (⌘H) to open the default help topic in your browser.
- When help is displayed in your browser, use the navigation links or search the help in the side bar.
- Choose Help → Manuals Index, to open a PDF file that has links to all the PDF documents. Click a link to open the document.
- Choose Help → Search Manuals to search the manuals. The search tab in Adobe Reader opens, and you can search across all the PDF documents. You must have Adobe Reader installed to use this feature.

For information on:

- Problems and solutions: choose Help → Knowledge Base or Help → Known Issues → *product*.
- Software updates: choose Maestro → Check for Updates.
- New software features: choose Help → New Features.
- Scripts available for download: choose Scripts → Update.
- Python scripting: choose Help → Python Module Overview.
- Utility programs: choose Help → About Utilities.
- Keyboard shortcuts: choose Help → Keyboard Shortcuts.
- Installation and licensing: see the *Installation Guide*.
- Running and managing jobs: see the *Job Control Guide*.
- Using Maestro: see the *Maestro User Manual*.
- Maestro commands: see the *Maestro Command Reference Manual*.

Contacting Technical Support

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail:	help@schrodinger.com
USPS:	Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone:	(503) 299-1150
Fax:	(503) 299-4532
WWW:	http://www.schrodinger.com
FTP:	ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information:

- All relevant user input and machine output
- BioLuminate purchaser (company, research institution, or individual)
- Primary BioLuminate user
- Installation, licensing, and machine information as described below.

Gathering Information for Technical Support

This section describes how to gather the required machine, licensing, and installation information, and any other job-related or failure-related information, to send to technical support.

For general enquiries or problems:

1. Open the Diagnostics panel.
   - **Maestro:** Help → Diagnostics
   - **Windows:** Start → All Programs → Schrodinger-2012 → Diagnostics
   - **Mac:** Applications → Schrodinger2012 → Diagnostics
   - **Command line:** `$SCHRODINGER/diagnostics`
2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.
3. Attach the file specified in the dialog box to your e-mail message.

If your job failed:

1. Open the Monitor panel in Maestro.

   Use Applications → Monitor Jobs or Tasks → Monitor Jobs.
2. Select the failed job in the table, and click Postmortem.

   The Postmortem panel opens.
3. If your data is not sensitive and you can send it, select Include structures and deselect Automatically obfuscate path names.
4. Click Create.

   An archive file is created in your working directory, and an information dialog box with the name of the file opens. You can highlight and copy the name of the file.
5. Attach the file specified in the dialog box to your e-mail message.
6. Copy and paste any log messages from the window used to start Maestro (or the job) into the email message,or attach them as a file.
   - **Windows:** Right-click in the window and choose Select All, then press ENTER to copy the text.
   - **Mac:** Start the Console application (Applications → Utilities), filter on the application that you used to start the job (Maestro, BioLuminate, Elements), copy the text.

If Maestro failed:

1. Open the Diagnostics panel.
   - **Windows:** Start → All Programs → Schrodinger-2012 → Diagnostics
   - **Mac:** Applications → Schrodinger2012 → Diagnostics
   - **Linux/command line:** `$SCHRODINGER/diagnostics`
2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.
3. Attach the file specified in the dialog box to your e-mail message.
4. Attach the file `maestro_error.txt` to your e-mail message.

   This file should be in the following location:
   - **Windows:** `%LOCALAPPDATA%\Schrodinger\appcrash`
     (Choose Start → Run and paste this location into the Open text box.)
   - **Mac:** `Documents/Schrodinger`
   - **Linux:** Maestro's working directory specified in the dialog box (the location is given in the terminal window).
5. On Windows, also attach the file `maestro.EXE.dmp`, which is in the same location as `maestro_error.txt`.

120 West 45th Street
17th Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

245 First Street
Riverview II, 18th Floor
Cambridge, MA 02142

8910 University Center Lane
Suite 270
San Diego, CA 92122

155 Gibbs St
Suite 430
Rockville, MD 20850-0353

Dynamostraβe 13
D-68165 Mannheim
Germany

Zeppelinstraβe 73
D-81669 München
Germany

Potsdamer Platz 11
D-10785 Berlin
Germany

Quatro House
Frimley Road
Camberley GU16 7ER
United Kingdom

8F Pacific Century Place
1-11-1 Marunouchi
Chiyoda-ku, Tokyo 100-6208
Japan

No. 102, 4th Block
3rd Main Road, 3rd Stage
Sharada Colony
Basaveshwaranagar
Bangalore 560079, India

SCHRÖDINGER.