Protein Preparation Guide

Schrödinger Suite 2012 Update 2

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

The typical structure file from the PDB is not suitable for immediate use in molecular modeling calculations. A typical PDB structure file consists only of heavy atoms and may include a cocrystallized ligand, water molecules, metal ions, and cofactors. Some structures are multimeric, and may need to be reduced to a single unit. Because of the limited resolution of X-ray experiments, it can be difficult to distinguish between NH and O, and the placement of these groups must be checked. PDB structures may be missing information on connectivity, which must be assigned, along with bond orders and formal charges.

Schrödinger has therefore assembled a set of tools to prepare proteins in a form that is suitable for modeling calculations. This manual describes these tools and their use in preparing proteins for various applications. The tools are combined in the Protein Preparation Wizard, which is described in Chapter 2. Procedures for manual adjustment and preparation of proteins are described in Chapter 3.

1.1 The Protein Preparation Process

The preparation of a protein involves a number of steps, which are outlined below. The procedure assumes that the initial protein structure is in a PDB-format file, includes a cocrystallized ligand, and does not include explicit hydrogens. The result is refined, hydrogenated structures of the ligand and the ligand-receptor complex, suitable for use with other Schrödinger products. In many cases, not all of the steps outlined below need to be performed.

1. Import a ligand/protein cocrystallized structure, typically from the Protein Data Bank, into Maestro.

2. Locate any waters you want to keep, then delete all others.

    These waters are identified by the oxygen atom, and usually do not have hydrogens attached. Generally, all waters (except those coordinated to metals) are deleted, but waters that bridge between the ligand and the protein are sometimes retained. If waters are kept, hydrogens are added to them in the preparation process.

3. Simplify multimeric complexes.

    - Determine whether the protein-ligand complex is a dimer or other multimer containing duplicate binding sites and duplicate chains that are redundant.

    - If the structure is a multimer with duplicate binding sites, remove redundant binding sites and the associated duplicate chains.

4. Adjust the protein, metal ions, and cofactors.

   • Fix any serious errors in the protein. Incomplete residues are the most common errors, but are relatively harmless if they are distant from the active site. Structures that are missing residues near the active site should be repaired.

   • Check the protein structure for metal ions and cofactors.

   • If there are bonds to metal ions, delete the bonds, then adjust the formal charges of the atoms that were attached to the metal as well as the metal itself.

   • Set charges and correct atom types for any metal atoms, as needed.

   • Set bond orders and formal charges for any cofactors, as needed.

   • Fix the orientation of any misoriented groups (such as amide groups of Asn and Gln).

5. Adjust the ligand bond orders and formal charges.

   If you are working with a dimeric or large protein and two ligands exist in two active sites, the bond orders have to be corrected in both ligand structures.

6. Adjust the ionization and tautomerization state of protein and ligand, if necessary.

7. Refine the structure.

   This step relieves any strain from the adjustments, and can also reorient groups.

8. Review the prepared structures.

   • Examine the refined ligand/protein/water structure for correct formal charges and protonation states and make final adjustments as needed.

   • Check the orientation of water molecules and other groups, such as hydroxyls, amides, and so on.

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 Maestro interface with the following command:

```
$SCHRODINGER/maestro &
```

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

Windows:

The primary way of running Schrödinger applications on a Windows platform is from a graphical interface. To start the Maestro interface, double-click on the Maestro icon, on a Maestro project, or on a structure file; or choose Start → All Programs → Schrodinger-2012 > Maestro. You do not need to make any settings before starting Maestro 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 Maestro interface, click its icon on the dock. If there is no Maestro 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 Maestro from the command line in the same way as on Linux. The default working directory is then the directory from which you start Maestro. 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 the Protein Preparation Wizard in Publications

The use of the Protein Preparation Wizard should be acknowledged in publications as:

Schrödinger Suite 2012 Protein Preparation Wizard; Epik version 2.3, Schrödinger, LLC, New York, NY, 2012; Impact version 5.8, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012.

If you optimize the H-bond networks, you should also cite PROPKA, as follows:

Protein pK_a predictions were performed using PROPKA. 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.

The Protein Preparation Wizard

Many of the protein preparation tasks can be performed automatically with the Protein Preparation Wizard. The Protein Preparation Wizard panel allows you to take a protein from its raw state, (which may be missing hydrogen atoms and have incorrect bond order assignments, charge states, or orientations of various groups) to a state in which it is properly prepared for use by Schrödinger products such as Glide, Prime, QSite, Liaison, and MacroModel.

The Protein Preparation Wizard panel has three tabs, which contain tools for the stages of protein preparation. In the Import and Process tab, you can import a protein and perform the basic tasks for fixing the structures. The Review and Modify tab allows you to delete unwanted chains and waters, and fix or delete het groups. In the Refine tab you can optimize orientations of hydrogen-bonded groups and minimize the structure. For an unprocessed protein, such as one imported from the PDB, you should run through all three stages: import and process, review and modify, then refine. If your protein has no missing atoms and has the bond orders assigned already, you can start with the review and modify stage or the refine stage.

Because automatic procedures cannot cover all possible cases, it is very important to check the correctness of the structure before using it for other applications. The Wizard helps you to do this by allowing you to step through various parts of the system and check their correctness. For each task completed, the Wizard adds a Boolean property to the structure, so you can track what has been done. Procedures for manual checking and correction of structures are given in Chapter 3.

To open the Protein Preparation Wizard panel, choose Workflows → Protein Preparation Wizard or Tasks → Protein Preparation in the main window, or click the Prep Wiz button on the Project toolbar.

You can also select the option Open Protein Preparation Wizard after importing PDB file in the Import panel, to open the Protein Preparation Wizard panel automatically when you import a PDB structure.

The shorter tasks performed by the Wizard are run in the foreground, but longer tasks are run under Job Control. You can provide a prefix for these jobs in the Job prefix text box, and choose a host to run them on from the Host option menu.
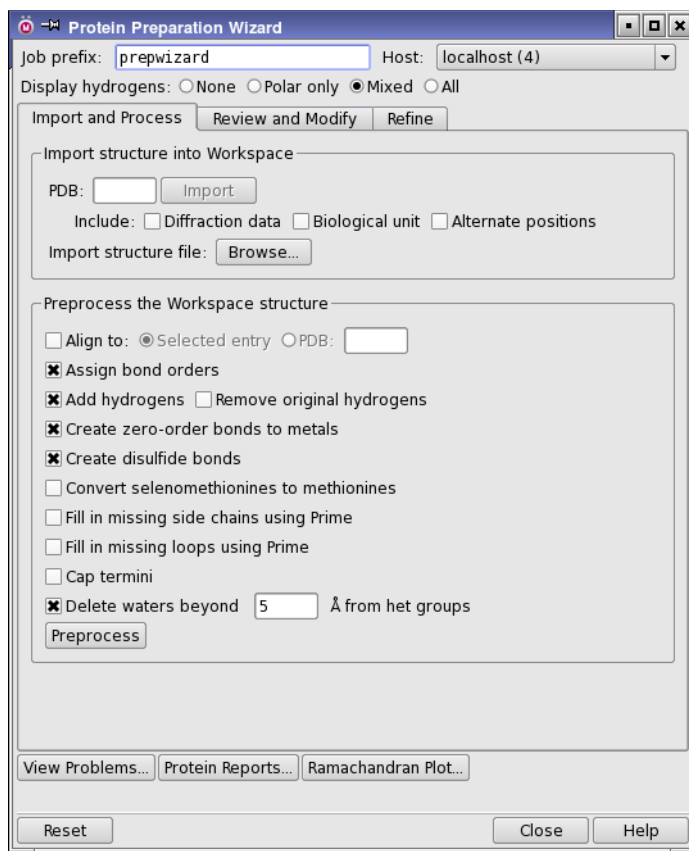
Figure 2.1. The Import and Process tab of the Protein Preparation Wizard panel.

2.1 Protein Structure Input

The Protein Preparation Wizard uses the structure in the Workspace as its input. The Workspace structure may be an included entry that exists in the Project Table, or one that is imported using the Protein Preparation Wizard. The import controls in the Import structure into Workspace section of the Import and Process tab are provided so you can conveniently load a structure from an external source. The Wizard allows you to obtain a structure from one of two sources: an external file, or the RCSB Protein Data Bank (PDB) web site.

Importing from a file: To import a structure from a file, click Browse. This button opens the Import panel, in which you can browse to the file that contains the protein structure. Select PDB from the Files of type option menu to load a PDB file. Click Options to ensure that you have Replace Workspace selected, so that on import, the structure is displayed in the Workspace.

However, you can always include the structure in the Workspace from the Project Table. If the file has multiple protein structures, you may have to use the Project Table to select the desired structure from the file.

Importing from the PDB: To import a structure from the PDB, enter the 4-letter ID in the PDB text box, then click Import. The structure is imported as an entry into the Project Table and included in the Workspace. If you want to download the biological unit, select Biological unit before importing. All structures of the biological unit are merged into a single entry. Downloading the biological unit is done from the RCSB web site, and requires internet access.

When importing the structure, the Protein Preparation Wizard first looks for a local installation of the PDB, and if one is not found then the structure is downloaded from the RCSB web site. In order to use your local installation of the PDB, one of the following conditions must be met:

- The PDB installation is part of the Schrödinger software installation.
- The SCHRODINGER_PDB environment variable[1] points to the PDB installation.

To circumvent your local PDB installation and download from the RCSB web site, use the Get PDB utility (Project → Get PDB in the main window) and include the structure in the Workspace before running the Protein Preparation Wizard. For more information on Get PDB see Section 3.1.5 of the *Maestro User Manual*.

If you want to download the X-ray diffraction data as well as the structure, select Diffraction data. The diffraction data is useful for checking or refining the structure with PrimeX, for example. Likewise, select Alternate positions if you want to make use of alternate positions of atoms in the PDB file.

2.2 Displaying Hydrogens

The many hydrogen atoms in an all-atom PDB structure can clutter the display, making it difficult to see what is important. The Protein Preparation Wizard provides three options for display of hydrogen atoms:

- None—Do not display hydrogens
- Polar only—Display only the polar hydrogens
- Mixed—Display all ligand hydrogens, but only the polar hydrogens on the receptor
- All—Display all hydrogens

These options correspond to the options on the Display button menus on the main toolbar. If hydrogens are added later, they are displayed according to the options selected here. It is useful to display only the polar hydrogens when checking hydrogen bonds, for example.

---

1. For information on how to set an environment variable, see Appendix A of the *Installation Guide*.

2.3 Preprocessing the Structure

After importing a structure, the first stage of the process is to address the main structural issues: assignment of bonds and bond orders, addition of hydrogens, filling in missing loops or side chains, capping uncapped termini, adjusting bonds and formal charges for metals, and correcting mislabeled elements. You can also delete water molecules that are outside a certain distance from the het groups, and you can align the protein structure to another protein structure. This preprocessing is necessary for further structure preparation actions, such as generating het states, H-bond assignment, and minimization, as well as being required for use by other applications.

To preprocess the structure, select the desired options, then click Preprocess. The options are as follows:

- Align to—Align the protein structure to that of another protein. You can choose the other protein by selecting an entry in the Project table (Selected entry) or by specifying a PDB ID for a structure from the PDB. The alignment is done with the Protein Structure Alignment tool that is on the Tools menu—see Section 10.3.3 of the *Maestro User Manual*.

- Assign bond orders—This option selects the assignment of bond orders, and performs the same task as Assign Bond Orders on the Tools menu.

- Add hydrogens—This option adds hydrogens to all atoms in the structure that lack them. The hydrogens are added by the utility `applyhtreat`.

- Remove original hydrogens—This option removes the original hydrogens before hydrogens are added, and is only available if Add hydrogens is selected. It ensures that any problems with H atoms are fixed, including nonstandard PDB atom names, which is important for the H-bond optimization tool.

- Create zero-order bonds to metals—This option breaks bonds to metals, replacing them with zero-order bonds. and adjusts the formal charge on the metal and the neighboring atoms. Sulfurs that interact with metals have their hydrogens removed, if necessary, and are assigned a negative charge. The force fields usually treat metals formally as an ion, without bonds to their ligands.

- Create disulfide bonds—This option detects and adds bonds between sulfur atoms that are within 3.2 Å of each other. CYS residues are renamed to CYX when the bond is added. This option is selected by default.

- Convert selenomethionines to methionines—This option converts selenomethionines (MSE) to methionines (MET), and is not selected by default. If you want to use the OPLS_2001 force field, you should select this option, as OPLS_2001 does not have parameters for Se, but OPLS_2005 does.

- Fill in missing side chains using Prime—This option allows you to add and optimize these atoms by running a Prime structure refinement job. See Section 6.5 of the *Prime User Manual* for more information. If you do not select this option, you can fill in the missing side chains later, using the process described below. (Not available from Maestro Elements.)

- Fill in missing loops using Prime—Fill in missing loops from the SEQRES records in the PDB file, using Prime. The resulting loop may not be of high quality, and a Prime loop refinement should be performed to obtain higher quality. See Section 6.4 of the *Prime User Manual* for information on refining loops with Prime. If the missing residues are far from the site of interest, it might be sufficient to cap them by selecting Cap termini. (Not available from Maestro Elements.)

- Cap termini—This option adds ACE (N-acetyl) and NMA (N-methyl amide) groups to uncapped N and C termini. These termini include breaks in the chains where there are missing residues. If the chain breaks are far from the region of interest, it might be sufficient to cap them. If you want to fill in the chain breaks rather than cap them, select Fill in missing loops using Prime.

- Delete waters beyond *N* Å from het groups—This option deletes waters that are more than the specified distance (in angstroms) from any het group. It is mainly useful for retaining waters that are important for ligand binding, while deleting all other waters. You can also delete selected waters in the Review and Modify tab, and delete waters that do not form hydrogen bonds with non-waters in the Refine tab.

When you click Preprocess, a new entry is created for the corrected structure, and this entry is used for all subsequent operations. The progress of the structure correction is displayed at the foot of the panel. When all operations are finished, the tables in the Review and Modify tab are filled in, and the structure is colored by element.

Note: You should always check that the structure is correct after any automatic procedure is run. In particular you should ensure that the bond orders for the het groups are correct. If any part of the structure is incorrect, you can correct it manually using the procedures in Section 3.4 on page 26.

If the protein has residues for which atoms are missing (often the side chains), atom types are misassigned, atoms are overlapping, or atoms have alternate positions, the Protein Preparation - Problems dialog box opens. This dialog box lists the atoms and residues that have these problems. You can click on any row in the tables of this dialog box to select the atom or residue listed in that row in the Workspace and zoom in on it. This makes it easy to inspect the structure and fix it. You can perform any Maestro operation while this dialog box is open. After fixing the structure, click Update to reanalyze the Workspace and display an updated list of problems. You can reopen the dialog box at any time by clicking View Problems.
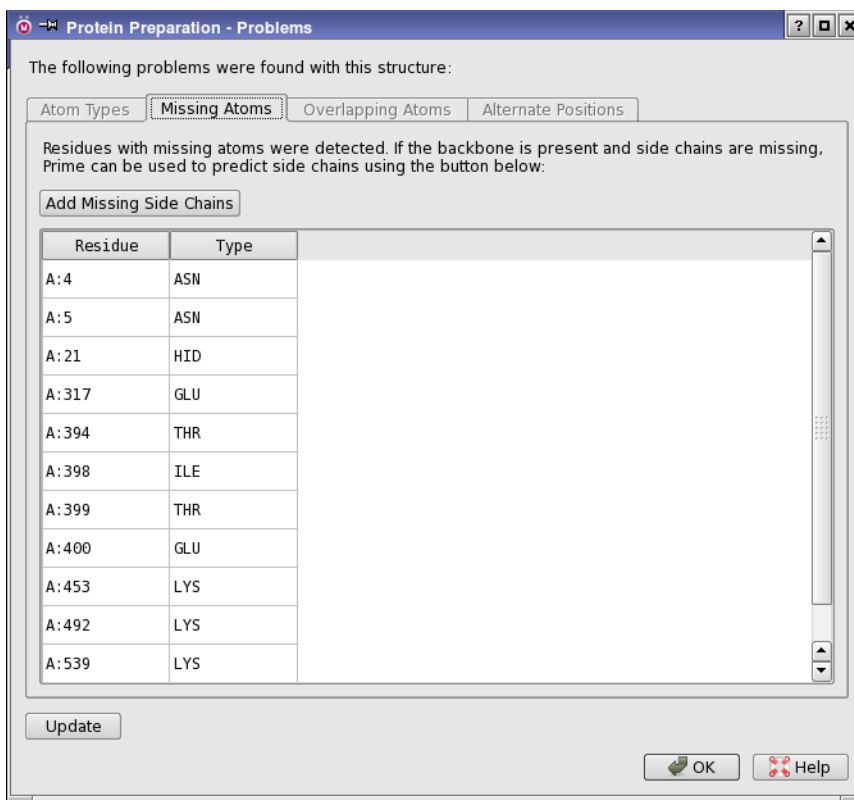
Figure 2.2. The Protein Preparation - Problems dialog box.

To fix mistyped atoms, you can use the tools in the Build panel. See Section 5.11 of the *Maestro User Manual* for more information.

To add missing side-chain atoms (if you did not select Fill in missing side chains using Prime when preprocessing), you can run a Prime side-chain prediction. These residues are colored red when you import the protein structure into Maestro. Clicking the Add Missing Side Chains button automatically starts a Prime job with all the residues in the table. The panel is unavailable until the job finishes. If you want to run this job independently, or for selected residues, you can use the following procedure:

1. Choose Applications → Prime → Refinement in the main window, or click Add Missing Side Chains in the Problems dialog box.

   The Prime Refinement panel opens.

2. Choose Predict side chains from the Task menu.

3. Choose the residues to refine.

   If you click Select Residues with Missing Atoms, the ASL text box in the Residues for side chain refinement section is filled in with an expression that selects all the residues with missing atoms.

4. Click Start, select job options in the Start dialog box, and click Start.

When the job finishes, the new structure is incorporated and displayed in the Workspace. You can continue the preparation process with this structure. You might want to delete unwanted groups, as outlined in the next section, before running the Prime side-chain prediction. If you do, you can open the Problems dialog box again to run the job.

If the structure has overlapping atoms, you might want to inspect these regions. Ensure that you have all hydrogens displayed first. If the structure is not tangled, the Impref minimization should move these atoms away from each other. You can perform a manual minimization by selecting a table row and pressing CTRL+M with the pointer in the Workspace. If there are duplicate het groups, minimization will not fix the problem. In this case, the atoms are on top of each other, which makes it impossible to find by visual inspection. The problem report alerts you to the existence of the duplicate. You can delete one copy of the het group in the Review and Modify tab, once you have closed the Protein Preparation - Problems dialog box.

If the protein has missing residues that are critical to the intended use of the protein structure, you might need to run a Prime loop refinement to fill in the missing residues. If you selected Fill in missing loops using Prime, you should still refine these loops because they are not opti-mized. This task can be run from the Prime Structure Prediction panel, and can take some time, depending on the length of the missing section. If the missing residues are not critical, it is probably sufficient to cap them, or use the inexpensive loop prediction that is run when you select Fill in missing loops using Prime.

If the input structure has alternate positions, the residues involved are listed in the Alternate Positions tab. These positions are legitimate for some applications, like PrimeX, but for others, a single set of positions is required. This tab allows you to choose one of the alternate positions for each residue.

To choose a position for a residue, first select the residue in the table. The view zooms in to the residue, and the default position is marked. Click Next Position to switch between positions. The alternate position is When the position that you want to keep is displayed, click Commit. The atoms in this position are kept, and the alternate atoms are deleted.

If you keep alternate positions, the rest of the protein preparation is done only for the main positions. To prepare a protein for different alternate positions, you should set up and refine a copy of that protein for each set of alternate positions.
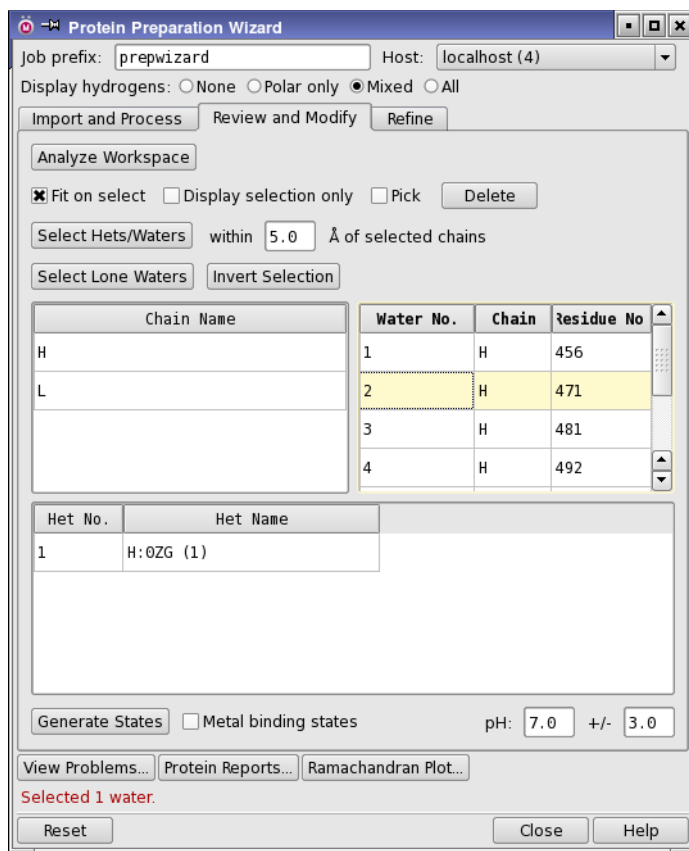
Figure 2.3. The Review and Modify tab of the Protein Preparation Wizard panel.

2.4 Inspecting the Structure and Deleting Unwanted Groups

After fixing any structural defects, the next stage is to inspect the structure and delete unwanted parts of the structure. The first part of the process of inspection is done in the Protein Preparation - Problems dialog box, which is described in the previous section. The next part of the inspection is done in the Review and Modify tab.

If you are starting the process at this stage, without preprocessing, display the structure in the Workspace, and click Analyze Workspace. The tables are then populated with information on the chains, waters, and het groups.

The three tables in this tab display a list of chains, a list of water molecules, and a list of het groups, which you can select, examine, and delete. Het groups are everything that is not a water or a protein residue, and include ligands, metal ions, and cofactors. Chains are defined by the chain label, and can include waters and het groups.

- You can select multiple rows in a table using shift-click and control-click, and you can make a selection from more than one table at a time.

- You can move the selection up and down a table with the UP ARROW and DOWN ARROW keys.

Above the tables are a number of options and buttons that can be used for selection, controlling the display, and deleting the selected objects.

- You can zoom in on the selected objects by selecting Fit on select, and you can limit the display to the selected objects by selecting Display only selected.

- You can select a table row by picking in the Workspace. To do so, select Pick and pick an atom. If you pick an atom in a water or a het group, the water or het group row is selected; if you pick any other atom, the chain is selected.

- You can select all hets and waters within a specified distance of the chains that are selected by entering the distance in the within *N* Å of selected chains text box, and clicking Select Hets/Waters. This feature is useful for reducing a multimer to a monomer: select the chain, select the het groups/waters, invert the selection and then delete.

- You can select waters that have nothing other than waters within 3.5 Å by clicking Select Lone Waters.

- You can invert the selection (select the objects that are not selected, deselect those that are selected) by clicking Invert Selection. This is useful if you want to select the objects to keep and delete the rest: make your selection, click Invert Selection, then click Delete.

When you select a chain, water, or het group, the Element (Green Carbons) color scheme is used for the selected atoms, and the color scheme of the unselected atoms is changed to use darker colors, so that the selected atoms stand out. The selection also computes and displays hydrogen bonds between the selected and the unselected atoms.

Before inspecting the structure, it is advisable to delete unwanted parts of the structure. If the protein is a multimer and you want to simplify it, you should delete the chains for the duplicate structural units. When you delete the chains, the waters and het groups that are labeled with that chain label are also deleted. Otherwise you must delete them separately. See Section 3.2 on page 23 for information on determining what to do with multimeric complexes.

To delete chains, waters, or het groups, select the table rows, then click Delete.

When you have finished deleting unwanted parts of the structure, it is advisable to inspect the remaining parts and correct any structural problems manually. However, you might want to examine the possible ionization and tautomeric states of the het groups first, as described in the next section.

2.5 Ionizing and Tautomerizing Het Groups

For het groups, you can use Epik to predict ionization and tautomeric states of the het groups in the pH range specified in the pH text boxes. You can also request Epik to generate additional states that are suitable for binding to metals in metalloproteins, by selecting Metal binding states. To generate the states, click Generate States. An Epik job is run, which may take a few minutes. When the job finishes, Sn (state) columns are added to the het groups table and are populated with buttons for the selection of the states. You can select a state by clicking its button, and you can step through the states with the LEFT ARROW and RIGHT ARROW keys (provided that only one het group is selected). The states are sorted by increasing state penalty, which includes a reward for the number of hydrogen bonds formed.

When you select a state, the structure is displayed in the Workspace, with markers to indicate the atoms that differ between states. The status area at the foot of the panel displays the state penalty, the tautomer probability, and the charge.

It may be useful to display formal charges on the atoms when examining the states. The charged atoms are automatically labeled with the charge, but if they are not, choose Formal Charge from the Label All button menu on the Labels toolbar.

Before proceeding to optimization of the structure, you should select the desired ionization states for the het groups. The project entry is changed to use the structures that you select.

2.6 Optimizing the Hydrogen Bonding Network

The next stage of protein preparation is to optimize the hydrogen-bonding network by reorienting hydroxyl and thiol groups, water molecules, amide groups of asparagine (Asn) and glutamine (Gln), and the imidazole ring in histidine (His); and predicting protonation states of histidine, His, aspartic acid (Asp) and glutamic acid (Glu) and tautomeric states of histidine. These optimizations are necessary because the orientation of hydroxyl (or thiol) groups, the terminal amide groups in asparagine and glutamine, and the ring of histidine cannot be determined from the X-ray structure. Flipping the terminal amide groups and the histidine ring can improve charge-charge interactions with neighboring groups as well as improving hydrogen bonding. The 180° flips preserve the heavy-atom placement deduced from the X-ray electron

density. In addition, the protonation state of histidine, aspartic acid, and glutamic acid are varied to optimize hydrogen bonding and charge interactions. If waters are included with the protein structure, their orientations are also varied to optimize hydrogen bonding. This stage does not include a full energetic optimization, which can be done in the refinement step.

Each hydrogen bond donor, His ring, and Asn and Gln terminal amide is considered a separate orientable species. Optimizing the orientation of the various groups is an iterative process, which passes over all the groups whose H-bonds need to be optimized multiple times. For information on the algorithm used, see Appendix B.

This task is run from the H-bond assignment section of the Refine tab, which provides both an automated optimization, in which the optimization runs on all H-bonds (with or without the inclusion of waters), and an interactive optimization, in which clusters of hydrogen-bonded species can be selectively optimized. You can run the automatic optimization first, then use the interactive optimization to check the results, or you can proceed directly to the interactive optimization.

2.6.1   Automated Optimization

The automated optimization provides a limited amount of control over the process. You also select Minimize hydrogens of altered species to minimize the energies of these hydrogen atoms once the hydrogen bonding network is optimized.

On some occasions, you may want to manually orient the water molecules, then use the automated procedure for the other orientations. You can exclude the orienting of water molecules by deselecting Sample water orientations, which is selected by default.

Note: If you have a lot of water molecules in the structure, this process can take a long time. You should ensure that you have deleted unwanted waters before you start this process.

The hydrogen bonding network depends on the protonation states of the protein as well as the ligand, which are determined at a given pH. The protonation states of the ligand are selected in the Review and Modify tab. You can predict the pK_a values of the protein by selecting Use PROPKA, and entering the target pH value in the pH text box. If you want to label protein atoms with their pK_a values, select Label pKas.

You can also use a simpler model for the protonation states of the protein by selecting one of the Use simplified rules options. These rules are the fallback if for some reason PROPKA fails.

- Very low—protonate Asp, Glu, His
- Low—protonate His
- Neutral—normal biological states
- High—deprotonate Cys
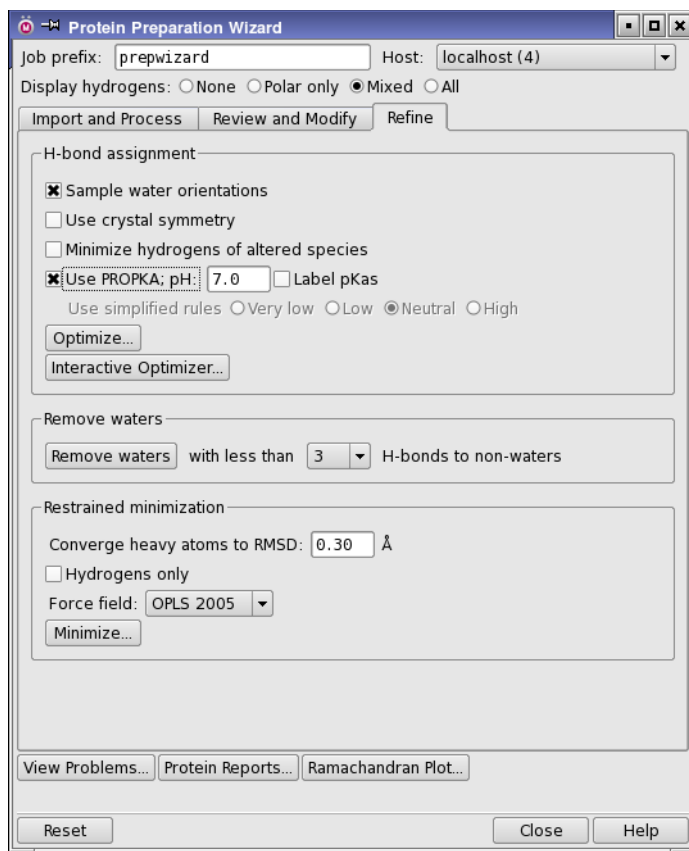
Figure 2.4. The Refine tab of the Protein Preparation Wizard panel.

When you have selected the desired options, click Optimize. The job starts on the host selected at the top of the panel, with a with a job name determined from the job prefix. You can monitor the job in the Monitor panel.

When the job finishes, the results are incorporated as a new entry. You should check the results carefully, to ensure that they are all correct. The amino acid "flips" are labeled for easy identification. To inspect one of these residues, you can use the following procedure:

1. Choose Residue from the Select button menu on the Edit toolbar.

2. Pick an atom in the residue.

3. Click the Fit button on the Workspace toolbar.

You can also examine the results using the interactive optimizer.

2.6.2   Interactive Optimization

The interactive optimization is performed in the Interactive H-bond Network Optimizer panel, which you open by clicking Interactive Optimizer. This panel is used to interactively perform the assignments of the H-bond network.

The first task is to analyze the system into hydrogen-bonded clusters, which you do in the Analysis section by clicking the Analyze Network button. The analysis is dependent on the following choices:

- pH—select the pH range for the protonation of protein residues. This choice affects the charge states used for certain residues.

- Use PROPKA—Use PROPKA [1–4] to predict the pK_a of protein residues and determine the protonation states at the specified pH. If you deselect this option, the protonation state is determined by simple rules according to the pH range:

  - very low—protonate Asp, Glu, His
  - low—protonate His
  - neutral—normal biological states
  - high—deprotonate Cys

- Include current orientations—Include the current orientations of the hydrogen-bonded species among the orientations to be sampled.

- Only analyze Workspace selection —Restrict the scope of the analysis to the atoms that are selected in the Workspace. This allows you to choose the part of the structure that you want to analyze and optimize.

- Use crystal symmetry—When analyzing the Workspace structure, use crystal symmetry to include in the analysis any H-bonds to atoms in neighboring cells. This can be important if only part of the biologically-relevant structure is present in the asymmetric unit. Using this option can significantly increase the time taken in the optimization.

The analysis identifies the species, clusters them and generates their states, which are then listed in the table. You must perform the analysis when you first open the panel, or if you add or delete bonds or atoms in the Workspace or include a new structure in the Workspace. The analysis can take a minute or so, depending on the size of the system.
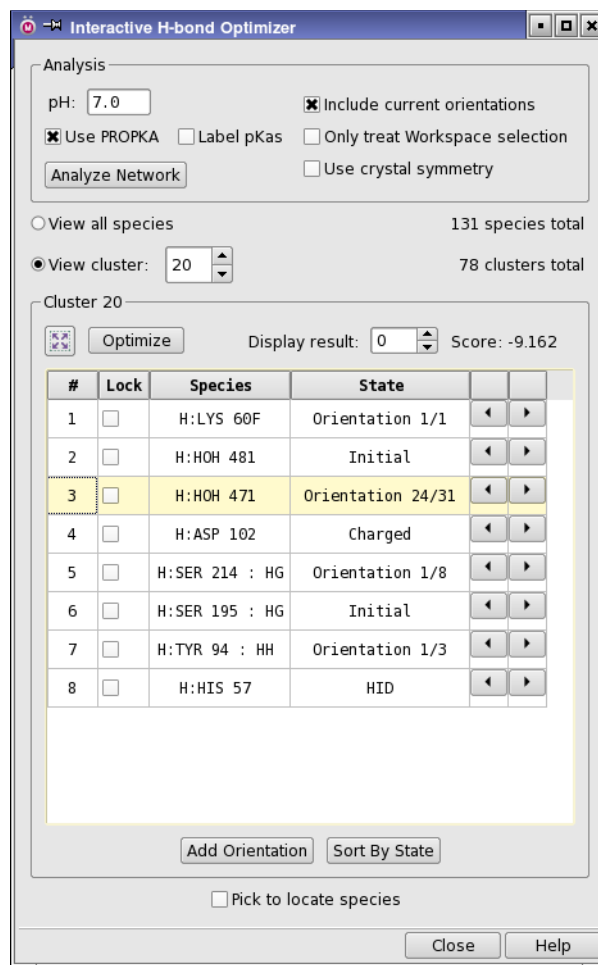
Figure 2.5. The Interactive H-bond Optimizer panel.

Once the analysis is complete, you can view the results by cluster, or list the entire set of species identified. For each cluster, the species are listed in a table. You can then optimize the H-bond network for the cluster by clicking Optimize. The optimization does not vary any species for which a lock is set by selecting the Lock option for that species in the table.

The optimization works as follows. An attempt is made to find a single solution that alleviates all serious clashes within the network. A Monte Carlo algorithm is then used to sample orientations in each cluster.

When you optimize a cluster, multiple results are returned, ordered by score. Each result corresponds to a particular combination of states for the species in the cluster. The result with the

best score is displayed in the table by default. You can view the other results using the Display result control. The score is updated as you change the result that is displayed.

When you choose a cluster or a species in the table, the Workspace view zooms in on the cluster, the cluster carbons are colored green, and altered species are labeled. The H-bonds in and to the cluster are also shown in the Workspace.

To change the orientation of a particular species, select it in the table and use the arrow buttons to step through the possible states. You can also pick a species in the Workspace, by selecting Pick to locate species (at the foot of the panel) and picking an atom in the Workspace. The view zooms in to the species that is picked. You can step through species with the UP ARROW and DOWN ARROW keys. If you want to sort the species by state, click Sort By State. The sorted order remains in effect until you select a different cluster.

The panel also provides a manual component of the optimization. You can use the Workspace adjustment tools to reorient any rotatable hydrogen or water manually, and add this orientation to the table with the Add Orientation button. The states so added are labeled User *n*, where *n* is an index that starts from 1.

2.7 Retaining Structural Waters

Once the hydrogen-bonding network has been refined, you can choose to retain or remove waters based on the number of hydrogen bonds they form with parts of the structure that are not other waters. This allows you to keep waters that have significant binding to the receptor, for example, forming bridges, and remove the rest of the waters. To do so, choose the minimum number of hydrogen bonds from the Remove waters with less than N H-bonds to non-waters option menu, and then click Remove Waters. This is an alternative to keeping waters based on the proximity to the ligand, or manually removing waters.

2.8 Minimizing the Structure

The final step in the preparation process is to refine the structure. A minimization is run in which the heavy atoms can be restrained, so that strain can be relieved but the final result does not deviate too much from the input geometry. Hydrogen atoms are not restrained at all, which allows the optimized H-bond network from the previous step to be refined. This task is initiated in the Restrained minimization section of the Refine tab, and is performed by the `impref` utility (see Section A.1 on page 39). Before you start the job, you can specify the RMSD of the atom displacement for terminating the minimization in the Converge heavy atoms to RMSD text box, choose the force field from the Force field option menu (default OPLS2005), and select Hydrogens only if you want to optimize only the hydrogen atom positions and leave heavy atoms in place.

When you have made your selection of options, click Minimize. The job starts on the host selected at the top of the panel, with a with a job name determined from the job prefix. You can monitor the job in the Monitor panel. Once the job is started, you can use the panel for another protein—you do not need to wait until it has finished.

When the minimization job finishes, the protein preparation process is complete, and you can use the structure for other Schrödinger applications.

2.9   Checking the Final Structure

You should always check your structures before using them for an application. While a great deal of effort has been invested to make the Protein Preparation Wizard robust and cover a wide variety of cases, it cannot catch every problem in every structure.

To assist you in checking your structures, the panel has three buttons that open panels that contain information on the protein structure.

- View Problems—Opens the Protein Preparation - Problems dialog box, which contains tables of residues that have missing atoms, overlapping atoms, and atoms that are incorrectly typed. This dialog box is described on page 9.

- Protein Reports—Opens the Protein Reports panel. This panel contains extensive information on protein properties, and is described in detail in Section 10.6.2 of the *Maestro User Manual*.

- Ramachandran Plot—Opens the Ramachandran Plot panel. This tool is described in detail in Section 10.6.3 of the *Maestro User Manual*.

Always check your structures before proceeding!

2.10 Starting a New Protein Preparation

If you want to start a new protein preparation after you have completed a preparation, you can click Reset to reset all the values and options in the panel to their defaults, and clear all data. If you want to keep settings that you have already made, simply import the new protein and start the preparation process.

Manual Protein Preparation

For the most part, protein preparation can be performed with the Protein Preparation Wizard. However, there are occasions when you may need to make corrections manually. This chapter lays out some procedures that you can use for manual preparation and correction of a complex.

3.1 Deleting Unwanted Waters

Water molecules in the crystallographic complex are generally not used unless they are judged critical to the functioning of the protein–ligand interaction. When waters are used, they are later included in the protein as "structural" waters.

3.1.1 Locating Structural Waters

Structural waters can be located by displaying atoms within a certain distance of the ligand. The procedure below displays water molecules as balls, which makes them easier to identify in the Workspace.

To locate structural waters in the protein structure:

1. Choose Molecule Number from the Color Scheme button menu on the Representation toolbar.

   The ligand should be clearly distinguishable. The water molecules are represented only by the oxygen atoms, because a united-atom atom type is in use.

2. Choose Molecule from the Display Sel button menu on the Display Atoms toolbar.

3. Click on an atom in the ligand.

   The ligand is displayed and all other atoms are undisplayed.

4. Choose a distance from the Within button menu on the Display Atoms toolbar.

The ligand plus all atoms (including water oxygens) within the chosen distance of the ligand are now displayed.

5. Choose Select from the Ball & Stick button menu on the Representation toolbar.

The Atom Selection dialog box opens.

6. In the Residue tab, select Residue Type.

7. Select the water residue type, HOH, click Add, then click OK.

The water oxygens, assuming no hydrogens have been added, are displayed as balls.

3.1.2   Deleting All Water Molecules

If you decide to delete all waters, choose Waters from the Delete button menu on the Edit toolbar.

All water molecules are deleted.

3.1.3   Deleting Distant Water Molecules

If you want to keep one or more waters, it is a good idea to begin by removing those that are farther than a specified distance from the ligand.

1. Choose Select from the Delete button menu on the Edit toolbar.

The Atom Selection dialog box opens.

2. In the Molecule tab, choose Molecule number.

3. Click on a ligand atom in the Workspace, then click Add.

Workspace selection automatically supplies the molecule number.

4. Click Proximity. In the Proximity dialog box:

    a. Select Beyond, enter a distance in the text box, and select Angstroms.

    b. Under Fill, select Residues and Exclude source.

       This keeps the ligand itself from being deleted.

    c. Click OK to exit the Proximity dialog box.

5. In the Residue tab, choose Residue Type, and select HOH.

6. Click Intersect.

   Most of the water oxygens are marked in the Workspace.

7. Click OK to delete the selected water molecules.

3.1.4   Deleting Remaining Unwanted Waters

After deleting water molecules beyond a certain distance from the ligand, examine the Workspace and delete any remaining water molecules you do not want to keep:

1. Choose Molecules from the Delete button menu on the Edit toolbar:

2. Click on a water oxygen to delete that water molecule.

3.2   Simplifying a Protein Complex

For computational efficiency it is often desirable to keep the number of atoms in the complex structure to a minimum. If the complex is a multimer and the important interactions do not span more than one unit of the multimer, you can delete the others.

3.2.1   Determining Whether the Complex Is a Multimer

To determine whether the ligand-receptor complex is a multimer, compare the chains that appear in the sequence viewer. If there are two or more chains with identical sequences, the complex may be a multimer. If this is the case, there may be duplicate copies of the binding site of interest.

To see whether two duplicate chains are involved with the active site, undisplay the protein's amino acid residues:

1. Choose Protein from the Undisplay button menu on the Display Atoms toolbar.

Ligands, cofactors, metal ions, and water oxygens remain visible. If two or more identical ligands or ligand/cofactor groups are present, then the complex is most likely a multimer, and the redundant groups and the duplicate chains associated with them can be deleted.

3.2.2   Removing Unwanted Subunits

If the protein complex structure is a multimer with duplicate binding sites, it can be truncated by deleting all but a single ligand binding site and the associated receptor subunits. The procedure below assumes that the protein is initially not displayed. If it is displayed, follow the instructions in the previous section to undisplay it.

To remove redundant subunits or receptor sites of a multimer:

1. Delete all but one ligand or ligand/cofactor pairing:

    a. Choose Molecules from the Delete button menu on the Edit toolbar.

    b. Click on any atom in a molecule to delete that molecule.

2. Display the ligand or ligand/cofactor pair in CPK:

    a. Choose Molecules from the CPK button menu on the Representation toolbar:

    b. Click on an atom in the ligand to display it in CPK.

    c. If there is a cofactor, click on an atom in that molecule as well.

    d. Click the toolbar button a second time to leave the Draw atoms in CPK pick state.

The purpose of this step is to make the ligand or ligand/cofactor pair visually distinct. You can choose any representation that fulfils this purpose, not just CPK.

3. Choose Protein Backbone from the Also Display button menu on the Display Atoms tool-bar.

The protein backbone is redisplayed. Making just the backbone visible will provide enough information without unduly cluttering the Workspace.

4. Choose Chain Name from the Color Scheme button menu on the Representation toolbar.

5. Delete duplicate protein chains:

   a. Choose Chains from the Delete button menu.

   b. Click on a backbone atom in each protein chain you want to delete.

6. Choose All from the Display Only button menu on the Representation toolbar to redisplay the rest of the protein:

7. (Optional) Put all atoms, including the ligand and any cofactors, back into wire-frame representation:

   Double-click the Wire button on the Representation toolbar.

3.3 Selecting Residues by PDB Conversion Status

When you are fixing problems detected in the conversion of the structure from a PDB file, it is useful to select residues based on the conversion status. This status is stored as a Maestro prop-erty, and has an associated color scheme that can be applied at any time, by choosing PDB Conversion Status from the Color Scheme button menu on the Representation toolbar.

To select residues with a given PDB conversion status:

1. Click Select in the relevant set of atom selection tools, or choose Select from the relevant menu.

   The Atom Selection dialog box opens.

2. In the Residues folder, choose PDB Conversion Status from the property list.

   The available statuses are displayed in the list in the center.

3. Choose the relevant status from the PDB Conversion Status list.

4. Click Add, then click OK.

   The Atom Selection dialog box closes, and the residues are selected.

3.4 Correcting Bonds, Atom Types, and Charges

In some situations, you may need to correct bond orders, atom types, or formal charges. These operations are described in this section. The situations in which you may have to perform these actions are described in later sections.

If the complex structure contains any bonds from the ligand or a cofactor to a protein metal, they must be deleted, and the charge adjusted to represent an ionic state. For example, bonds from a porphyrin ring to a metal must be deleted, the nitrogen atoms given a charge of –1, and the metal charge incremented by +2. The OPLS-AA force fields model such interactions as a van der Waals plus electrostatic interaction, not a covalent interaction.

The MacroModel atom types for metal ions are sometimes incorrectly translated into dummy atom types (Du, Z0, or 00) when metal-protein bonds are specified in the input structure. Furthermore, isolated metal ions may erroneously be assigned general atom types (GA, GB, GC, etc.).

Some of the operations described below are made from the Build toolbar; other operations are made from the Build panel. If the Build toolbar is not already displayed, click Build on the Manager toolbar, or choose Window → Toolbars → Build. To open the Build panel, choose Edit → Build → *tab name*.

3.4.1 Display Tasks

When making corrections, it is useful to display information that is relevant to the task: for example, to display formal charges when correcting the charge or the atom type. This section presents some procedures for displaying information.

To display formal charges for all atoms:

 • Choose Formal Charge from the Label All button menu on the Labels toolbar.

To display element labels and formal charges for all atoms:

1. Choose Composition from the Pick to Label button menu on the Labels toolbar.

   The Atom Labels panel opens at the Composition tab.

2. Click Deselect All Properties.

3. Select Element and Formal charge from the Atom properties list.

4. Click All in the Label atoms section.

   All atoms in the Workspace are labeled.

5. Click Close.

To clear all labels:

1. Choose Labels from the Delete button menu on the Edit toolbar.

To undisplay the protein:

1. Choose Protein from the Undisplay button menu on the Display Atoms toolbar.

This is useful for locating metal ions, cofactors, and waters.

To display only the cofactor:

1. Choose Select from the Display only button menu.

   The Atom Selection dialog box is displayed.

2. In the Residue tab, choose Residue Type.

3. Click the residue type of the cofactor, which will be near the end of the list.

   The cofactor is highlighted.

4. Click Add, then click OK.

The cofactor is displayed. Because the cofactor was chosen by residue type and not molecule number, this method works even if the cofactor is covalently bonded to another residue. If there is more than one cofactor, you can select each cofactor and click Add before clicking OK.

3.4.2   Correction Tasks

The basic tasks for correcting bond orders, atom types, and formal charges are given below. It is a good idea to display formal charges for both atom type changes and formal charge changes, since atom typing can change the formal charge, and Maestro corrects the formal charges of surrounding atoms to compensate, if necessary.

To delete bonds use one of the following methods:

1. Choose Bonds from the Delete button menu on the Edit toolbar.

2. Click on the bonds to be deleted.

or

• Right-click on the bond and choose Delete from the shortcut menu.

To set or change bond orders:

1. On the Build toolbar, click the - Bond Order or + Bond Order button, as appropriate.

2. Click on bonds as necessary to set the bond order.

or

• Right-click on the bond and choose the order from the Order submenu of the shortcut menu.

To set or correct the formal charge on any atom:

1. On the Build toolbar, click the + Formal Chg or -Formal Chg button, as appropriate.

2. Click on an atom whose formal charge must be changed, as many times as necessary.

To correct atom types:

1. Open the Build panel at the Atom Properties tab.

2. Select Atom Type (MacroModel) from the Property option menu.

3. Select the correct atom type from the list.

   Atom types include both element name and formal charge. Atom type numbers are in parentheses.

4. Click on the atom to be changed.

3.5   Correcting PDB Atom Names

Prime requires that the PDB atom names of all atoms are correct. In the imported structure, residues that have incorrect PDB atom names are colored red or blue. You can select these residues using the procedure given in Section 3.3 on page 25.

To correct PDB atom names:

1. From the Label All button menu on the Labels toolbar, choose PDB Atom Name.

2. In the Atom Properties folder of the Build panel, choose PDB Atom Name from the Property option menu.

3. Ensure that Pick is selected in the Apply PDB atom name section, and that Atoms is chosen from the Pick option menu.

4. For each incorrectly-named atom, enter the correct PDB atom name in the PDB atom name text box, then click on the atom.

   You may need to zoom in to pick the atom, which you can do with the mouse wheel.

3.6 Fixing Incomplete Residues

Incomplete residues are common errors in PDB structures, but may be relatively harmless if they are distant from the active site. Structures that have incomplete residues near the active site should be repaired. Often, incomplete residues are missing side chains.

When PDB structures are imported into Maestro, a color coding system is used to classify residues that cannot be fixed with the standard templating procedure (see Section 3.1.6 of the *Maestro User Manual* for details). Red is used for residues that are missing atoms.

If you have access to Prime, you can use the side-chain prediction in the Prime Refinement panel to fill in the missing side chains.

To fix incomplete residues using Prime Refinement:

1. Fix any bond orders, formal charges, and misnamed atoms in the ligand, and add hydrogens to the ligand.

   The orange residues are usually ligand residues, so fixing the orange residues first will generally address this concern.

2. Double-click the Add hydrogens button on the Edit toolbar.

   +H

   Hydrogen atoms are added to all atoms as needed to complete the valence.

3. Choose Applications → Prime → Refinement in the main window.

   The Prime Refinement panel opens.

4. Choose Predict side chains from the Task option menu.

5. Click Select Residues with Missing Atoms.

   The ASL text box in the Residues for side chain refinement section is filled in with an expression that selects the residues with missing atoms.

6. Click Start, select job options in the Start dialog box, and click Start.

If you do not have access to Prime, you can use the procedure below to fill in the residues and adjust the structure. This procedure uses the Mutate option in the Build panel to mutate the residue, but in this case it is mutated to the same residue rather than a different one.

When you have fixed the structure, you should perform a minimization calculation on the corrected residues. In some cases, you might want to perform a conformational search on the

side chain (with the bulk of the protein frozen, for speed) using MacroModel, to obtain the best orientation of the side chain.

To fix incomplete residues manually:

1. From the Color Scheme button menu on the Representation toolbar, choose PDB Conversion Status.

Residues with missing atoms are colored red.

2. Right-click on an atom in a red residue, choose Mutate from the option menu, then choose the residue from the list.

   The side chain is added to the residue. To check which residue you are mutating, look in the status bar. Even though formally you are mutating a residue, you are not actually changing the residue type.

The side chain might not be added in the correct geometry, and may have bad contacts with other residues. The following procedure can help to adjust the side-chain geometry to a more reasonable location. The procedure involves displaying bad and ugly contacts, and selecting from a set of known rotamer states for the residue.

To display contacts between the residue and the rest of the complex:

1. Choose Residue from the Select button menu on the Edit toolbar.

2. Pick an atom in the side chain of the residue you want to adjust.

   The side chain is used to ensure that you get all the atoms. If you already have the backbone selected, the selection might not change. The residue should be marked in yellow.

3. Choose Tools → H-Bonds and Contacts.

4. In the Contacts tab, ensure that the following options are selected:

   - Display contacts
   - Bad contacts
   - Ugly contacts
   - Exclude 1-4 interactions

   Hydrogen bonds are excluded from consideration as contacts by default. You can set the behavior in the Preferences panel—click Preferences to open this panel.

5. Under Atom set 1, click the Atom Selections button and choose Workspace Selection.

The ASL text box shows `workspace_selection`.

6. Click the Atom Selections button in the Atom set 2 section, then click Select.

   The Atom Selection dialog box opens.

7. Click Selection, then click Invert.

   The selection is added then inverted, so that the expression covers all atoms that are *not* in the selection.

8. Click OK.

Contacts are now displayed between the side chain and the rest of the complex.

To adjust the side-chain geometry using rotamers:

1. Ensure that the residue of interest is selected.

   This should be the case if you are continuing on from the display of contacts procedure.

2. From the Adjust button menu on the Edit toolbar, choose Rotamers.

   The Rotamers dialog box is displayed. It contains a list of the most common rotamers for the selected residue.

3. Select rotamers in turn from the table, until you find one whose geometry looks reasonable.

4. Click OK.

   The Rotamers dialog box closes.

To adjust the side-chain geometry manually:

1. From the Adjust button menu on the Edit toolbar, choose Display contacts and Quick Torsion.

2. Click on the bond you want to rotate around.

   An arrow appears on the bond pointing towards the group that will be rotated.

3. Use the mouse wheel or drag with the left mouse button to adjust the torsional angle.

You can use the contacts to determine when the side chain is in a better position.

You might also want to use the other tools on the Adjust button menu to adjust the structure and thus reduce the number of bad contacts. You can also do a minimization of the adjusted residue, as follows.

To minimize the selected residue:

1. Choose Residue from the Select button menu on the Edit toolbar.

2. Pick an atom in the side chain of the residue you want to adjust.

3. Press CTRL+M (⌘M).

   A minimization of the atoms in the residue is performed.

3.7   Adding Missing Residues

Some structures are missing entire residues, even though the sequence may be complete. The residues where the break occurs are not color-coded on import into Maestro, and the residues are not displayed in the sequence viewer. If your structure has missing residues that could be important, you should consider doing a Prime structure prediction to add the missing residues.

To check for missing residues:

1. If the sequence viewer is not displayed in the main window, display it:

   - **Maestro:** Windows → Sequence Viewer.
   - **Elements:** Edit → Toolbars → Sequence Viewer.
   - **BioLuminate:** Edit → Settings → Show Sequence Viewer.

2. Right-click in the sequence name section of the sequence viewer, and choose Chain Breaks from the Color Scheme submenu of the shortcut menu.

   N-terminal residues are colored blue, and C-terminal residues are colored red. The break is indicated by a pair of residues, one red and one blue, in the same chain.

3. Right-click in the sequence name section of the sequence viewer, and choose Align by residue number from the shortcut menu.

   The gaps in the sequence are now displayed in the sequence viewer. Note that choosing this option alone is not sufficient to detect breaks, because there might be a numbering break for which there was no break in the chain.

   Another method of displaying breaks is as follows:

1. From the Color scheme button menu, choose Molecule number.

   If the chain has breaks, each piece will be treated as a separate molecule.

2. Scan the chains in the sequence viewer for changes in color.

To predict the structure of the missing residues with Prime:

1. Choose Applications → Prime → Structure Prediction.

   The Homology Modeling panel opens.

2. Click the Structure Prediction Wizard button.

   The Prime Structure Prediction panel opens.

3. In the Input Sequence step, click From File, and import the sequence from the PDB file.

   You must import the sequence directly from the PDB, not from the Workspace, because the PDB sequence is complete and the Workspace sequence is not. Using the Workspace sequence would result in deletions.

4. Proceed to the Find Homologs step.

5. Click Import, and import the PDB file as the template.

6. Proceed with the workflow (see Chapter 4 of the *Prime User Manual*).

3.8   Adjusting Protonation and Tautomerization

You should look for inconsistencies in hydrogen bonding to see whether a misprotonation of the ligand or the protein might have left two acceptor atoms close to one another without an intervening hydrogen bond. One or more residues may need to be modified to resolve such an acceptor-acceptor or donor-donor clash. In addition, you should check for the correct protonation and tautomerization of HIS residues, and for protonation of ASP, GLU, LYS, and ARG.

To protonate a neutral atom:

1. Change the formal charge on the atom to the correct value.

2. From the Add H button menu on the Edit toolbar, choose Atoms.

   +H

3. Click on the atom that you want to protonate.

4. Check that the formal charge is now correct by labeling the atoms, as described above.

To deprotonate an atom:

1. From the Delete button menu on the Edit toolbar, choose Atoms.

   ✕

2. Click on the hydrogen you want to remove.

3. Click the - Formal Chg button on the Build toolbar.

4. Click on the atom to which the hydrogen was attached.

5. Check that the formal charge is now correct by labeling the atoms, as described above.

To change the tautomerization:

1. Delete the hydrogen atom that needs to be moved (use the Delete button menu on the main toolbar).

2. Change the bond orders between the atoms to reflect the new tautomer.

3. Change any formal charges to reflect the new tautomer.

4. Add the hydrogen to its new location (use the Add hydrogens button).

   +H

3.9 Correcting Orientations

If the orientation of water molecules is incorrect, reorient the molecules with the following procedure:

1. Choose Advanced from the Transform button menu on the Edit toolbar.

   The Advanced Transformations panel opens.

2. Under Atoms For Transformation, use the picking controls to select the entire water molecule you want to reorient.

3. Under Center For Transformation, use the picking controls to select the oxygen atom of the water molecule.

   This ensures that the rotation is about the oxygen atom, rather than about the centroid of the water molecule atoms, which will shift the oxygen as well as the hydrogens. If you had chosen Molecule from the Local transformation button menu and rotated the water molecule, the rotation center would have been the centroid of the water atoms.

4. Under Rotation/Translation Scope, select Local.

5. Use the middle mouse button to change the orientation of the water hydrogens.

6. Close the Advanced Transformations panel.

   Transformations should now be global again.

You might also have to adjust the geometry of some groups in the active site. Check particularly for GLN and ASN: the N and O atoms can look the same in the X-ray structure determination, so they might need to be flipped at their terminal dihedral if there is a poor hydrogen-bonding pattern or steric clashes. Likewise, the ring in HIS might need to be flipped. You can do this by rotating the dihedral angle by 180°.

To rotate a dihedral angle:

1. Choose Quick Torsion from the Adjust button menu on the Edit toolbar.

2. Pick the bond around which you want to rotate.

   An arrowhead is placed on the bond, pointing to the group that will be rotated. If this is not the correct group, click on the bond again.

3. Use the mouse wheel or drag horizontally (with the left mouse button) to change the angle.

If you want to be more precise in the rotation, use the following steps:

1. Choose Dihedral from the Adjust button menu on the Edit toolbar.

2. Pick four atoms to define the dihedral angle.

   The fourth atom should be one of the atoms in the group you want to flip. Markers and the value of the angle are displayed.

3. Drag horizontally with the left mouse button or scroll with the mouse wheel until the angle has changed to the appropriate value.

Command-Line Tools

A.1   impref

The impref utility is run by the Protein Preparation Wizard to perform the refinement stage of protein preparation. It uses Impact to perform minimizations in which the heavy atoms are restrained with a harmonic potential of 25 kcal mol^{-1} Å^{-2} and unrestrained hydrogens. There is little need to run impref directly. The syntax of the command is as follows:

    $SCHRODINGER/utilities/impref [*options*] *input-file*

The input file must be a Maestro file, and can be uncompressed (.mae) or compressed (.mae.gz, .maegz). The output file has the same compression as the input file. The command options are given in Table A.1.

Table A.1. Options for the impref command.

Option	Description
-f *version*	Specify OPLS force field version (2001, 2005, 2.0)
-fixheavy	Do not change heavy-atom coordinates: refine only the hydrogens.
-fixlig	Do not change ligand heavy-atom coordinates. Requires -l and -s.
-fixprot	Do not change protein heavy-atom coordinates. Requires -l and -s.
-forcesave	Return first structure generated if no structure satisfies the requested rmsd.
-h	Print usage message and exit.
-i[ter]	Set maximum number of refinement iterations. Default: 4.
-k	Keep Impact minimization *.inp, *.log, and *.mae* files.
-l *ligfile*	Read ligand from file *ligfile*, instead of *input-file.* If this option is used, *input-file* must contain the protein structure alone. If this option is not used, *input-file* must contain the protein/ligand complex.
-r *rmsd*	Specify maximum RMSD allowed. Default: 0.35.
-s[eparate]	Write out refined protein and ligand structures separately, rather than in one combined structure. Requires -l *ligfile*.
-v	Print version number and exit.

A.2   prepwizard

If you know what treatment is needed in advance, you can run the Protein Preparation Wizard from the command line, using the utility prepwizard. The syntax of the command is as follows:

$SCHRODINGER/utilities/prepwizard [*options*] [*job-options*] *input-file output-file*

The options are given in Table A.2. Standard Job Control options are also accepted, as described in Table 2.1 of the *Job Control Guide*, and in addition the job options given in Table 2.2 of the *Job Control Guide* except for -INTERVAL. One further job option is –NJOBS *n*, which can be used to distribute preparation of multiple proteins. The number of jobs should not be more than the number of proteins in the input file.

The input file can be a single-structure or multi-structure Maestro file (.mae, .maegz, or .mae.gz) or a single-structure PDB file (.pdb or .ent); the output file should be in Maestro format. If you supply a multi-structure Maestro file as input, the script processes all structures in the file.

Note: If you run the Protein Preparation Wizard from the command line, it is especially important that you verify the correctness of the results. Information on the changes made is given in the log file.

Table A.2. Options for the prepwizard command.

Option	Description
-align *structure*	Align the protein to the specified structure, which can be specified by a file or a PDB ID.
-c[aptermini]	Cap termini with NMA or ACE.
-delwater_hbond_cutoff *int*	When deleting far waters, do not delete waters that make at least this number of H-bonds to non-waters. By default, this feature is not used.
-disulfides	Create bonds to proximal sulfurs, deleting hydrogens as needed. CYS residues are renamed to CYX when the bond is formed.
-f *ffield*	Specify the OPLS force field version to user for impref. Allowed values are 2001, 2005, 2008. Default: 2005.
-fillloops	Fill in missing loops with Prime. This is not a full loop refinement, which should be run separately if needed.
-fasta_file *filename*	If filling missing residues, specify the custom FASTA file to use for the sequence. If not specified, the FASTA file will be generated from the sequence that was stored in the input structure when it was converted from the PDB format.

Table A.2. Options for the prepwizard command. (Continued)

Option	Description
`-fillsidechains`	Fill in missing side chains with Prime `refinestruct`.
`-fix`	Fix heavy atoms and sample hydrogens only, in `impref`.
`-force` *list*	Force a particular residue to adopt the specified state. See -f option in Table A.3 for details.
`-h[elp]`	Display usage message and exit.
`-keepfarwat`	Do not delete waters that are far from het groups.
`-label_pkas`	Label residues with their pK_a values as predicted by PROPKA.
`-metal_binding`	Generate metal-binding states with Epik.
`-ms` *nstate*	Number of states per het group to keep. The states are sorted by number of hydrogen bonds first, then by state penalty, before the selection is made.
`-mse`	Convert selenomethionine residues to methionine.
`-nobondorders`	Do not assign bond orders to het groups.
`-noepik`	Do not run Epik to generate states for het groups.
`-nohtreat`	Do not add hydrogens.
`-noimpref`	Do not run `impref` (see Section A.1 on page 39).
`-nometaltreat`	Do not treat metals by breaking bonds and adjusting formal charges.
`-nopropka`	Do not use PROPKA for protonation states, but use the simple rules specified with `-pH`.
`-noprotassign`	Do not run `protassign` to optimize hydrogen bonding.
`-pH` *range*	Set target pH value for `protassign`. See -pH option in Table A.3 for details.
`-propka_pH` *value*	Specify the pH value to use in PROPKA for prediction of protonation states.
`-r[msd]` *rmsd*	Set RMSD cutoff for `impref`. Default: 0.3 Å.
`-reference_st_file` *filename*	File name of reference structure to align protein to.
`-reference_pdbid` *pdbid*	PDB ID of reference structure to align protein to.
`-rehtreat`	Remove and re-add hydrogens. This action resets PDB atom names.
`-s[amplewater]`	Sample waters when optimizing H-bonding with `protassign`.

Table A.2. Options for the prepwizard command. (Continued)

Option	Description
`-v[ersion]`	Display the program version and exit.
`-watdist` *dist*	Distance threshold for "far" waters. Default: 5 Å.
`-xtal`	Use crystal symmetry in `protassign`.

A.3 protassign

The `protassign` script optimizes the hydrogen-bonding network, rotating hydroxyl and thiol hydrogens (including waters), generating appropriate protonation and tautomerization states of His, and performing chi flips in Asn, Gln, and His residues. For details of the methods used, see Appendix B. The syntax of the command is as follows:

$SCHRODINGER/utilities/protassign [*options*] *input-file output-file*

The input and output files must be Maestro files (`.mae`, `.maegz`, `.mae.gz`). The options are described in Table A.3. The standard Job Control options, listed in Table 2.1 of the *Job Control Guide*, are supported, and the `-WAIT`, `-LOCAL` (use local disk for remote jobs), `-NOLOCAL` (use scratch disk for local jobs) and `-NOJOBID` (run without Job Control) options are also supported.

The log file reports an overall score for the structure, which you can use to compare different assignments.

Table A.3. Options for the protassign command.

Option	Description
-asl *expr*	ASL expression defining the residues to sample. Must be enclosed in quotes. Default: `all`.
-d	Print debugging information.
-f *res state*	Force a particular residue *res* to adopt the specified state *state*. Multiple `-f` specifications are permitted. *res* is specified as *chain*:*residue* where residue is the residue number and insertion code (if any). *state* is one of the following:

For HIS:

HID	delta form
HIE	epsilon form
HIP	protonated form
"Flip HID"	flipped delta form
"Flip HIE"	flipped epsilon form
"Flip HIP"	flipped protonated form

For ASN/GLN:

Flip	flipped form
"No Flip"	unflipped form

For ASP/GLU:

Charged	ionized form
"Neutral 1"	protonated on OE1
"Neutral 2"	protonated on OE2

Option	Description
-h[elp]	Display usage message.
-j *jobname*	Job name. Default is the input file name minus the extension.
-label_pkas	Label residues with their pK_a values as predicted by PROPKA.
-minimize *boolean*	Minimize hydrogens that were sampled. Allowed values: `true`, `false`.
-noflip	Do not sample ASN/GLN/HIS chi flips.
-nopropka	Do not use PROPKA for protonation states, but use the simple rules specified with `-pH`.
-noprot_asl *asl*	ASL expression for residues whose protonation state should not be changed.
-nowater	Do not sample water orientations.
-pH *range*	Specify the pH range to be used in determining the protonation state of protein residues. Allowed values are `very low` (protonate ASP, GLU, HIS), `low` (protonate HIS), `neutral`, `high` (deprotonate CYS)

Table A.3. Options for the protassign command.

Option	Description
-propka_pH *value*	Specify the pH value to use in PROPKA [1–4] for prediction of protonation states.
-t	Print execution time for each cluster.
-xtal	Use crystal symmetry in the optimization process.

H-Bond Optimization Technical Notes

While X-ray structures are invaluable starting points in modeling studies, they have a critical weakness: due to the relative lack of electron density around hydrogens, it is nearly impossible to accurately resolve their locations in the structure. While many hydrogen positions are easily estimated from simple geometric considerations, the problem is more difficult in the case of hydroxyl and thiol hydrogens, where the electrostatic environment must be taken into account. In addition, the lack of hydrogen coordinates leads to other ambiguities in the structure. For example, without knowledge of hydrogen locations, the protonation state of residues such as His cannot be directly determined from experiment. More subtly, without knowledge of the location of the hydrogens, it is generally not possible to distinguish the oxygen and nitrogen in the amides of Asn and Gln residues. A 180° flip about the relevant chi dihedral angle, transposing the oxygen and nitrogen, will often produce an alternate structure that is equally consistent with the electron density. A similar ambiguity exists with His, with respect to the carbon and nitrogen atoms of the imidazole ring.

In order to make the best use of X-ray structures in modeling studies, it is clearly important to resolve these three types of structural ambiguities. The purpose of the Protein Assignment utility (`protassign`) is to select the most likely:

- position of hydroxyl and thiol hydrogens (including any present on ligands, waters, and cofactors)
- protonation states and tautomers of His residues
- chi "flip" assignments for Asn, Gln and His residues
- orientations of water molecules
- protonation states of Asp, Glu, and Tyr residues
- protonation states and orientation of Lys residues

In order to optimize performance, localized clusters of hydrogen-bonding species are identified. Two such species are considered to be in the same cluster if their heavy atoms are within 3.5 Å of one another. Within a given cluster, all possible combinations of assignments for the species are then enumerated. In the case of hydroxyl and thiol torsions and water orientations, where countless possible assignments exist, a smaller set of possibilities is selected based on the local environment. Potential hydrogen-bond donors and acceptors are identified, and the rotatable hydrogen is directed towards each potential acceptor, as well as 120° away from each potential donor (to simulate hydrogen-bonding to the lone pair of its corresponding heavy atom). Positively charged metal species are treated as virtual hydrogen-bond donors.

The algorithm for optimization is a sophisticated algorithm that involves iterative sampling techniques, a kind of genetic algorithm for combining ensembles of orientations, and a kind of simulated annealing technique for improving the orientations.

Once all possible assignments for a given cluster are identified, each possibility is scored to determine the quality of the hydrogen-bond network (among the species in the cluster itself as well as with the surrounding environment). The nature of the scoring function is loosely based on simple electrostatic considerations. The intent is to score the existence of hydrogen-bonding networks, rather than estimate the actual electrostatic energy of the system.

- **Hydrogen-Bond Scoring:** The core of the scoring function involves an evaluation of the quality of the hydrogen-bonding network. In addition to identifying the number of hydrogen-bonds created, the quality of those bonds are also taken into account based on their geometries relative to an idealized hydrogen-bond.

- **Hydrogen Clash Penalty:** Assignments that place two polar hydrogen atoms too close to one another are given a high enough penalty to effectively eliminate them from consideration. The actual distance varies with the type of hydrogen.

- **Protonation Reward/Penalty:** The protonation penalty for unlikely protonation states is determined from the pK_a as predicted by PROPKA. If this program is not used, the following set of rules determine the protonation. A histidine residue is ruled to be protonated if it is either in close proximity to a negatively charged species, or if both its delta and epsilon nitrogens are in close proximity to hydrogen bond acceptors. Similarly, aspartic and glutamic acid residues are ruled to be neutralized if one or both of their carboxylate oxygens are in close proximity to a hydrogen bond acceptor.

References

1. Li, H.; Robertson, A. D.; Jensen J. H. Very Fast Empirical Prediction and Interpretation of Protein pK_a Values. *Proteins*, **2005**, *61*, 704–721.

2. Bas, D. C.; Rogers, D. M.; Jensen J. H. Very Fast Prediction and Rationalization of pK_a Values for Protein-Ligand Complexes. *Proteins*, **2008**, *73*, 765–783.

3. 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.

4. Søndergard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theor. Comput.*, **2011**, *7*, 2284–2295.

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
- Schrödinger Suite purchaser (company, research institution, or individual)
- Primary Schrödinger Suite 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.

SCHRÖDINGER®