PrimeX 2.8

User Manual

Schrödinger Press

May 2015

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 About PrimeX

PrimeX performs refinement of molecular replacement solutions in X-ray crystal structure determination. Rigid-body refinement, individual coordinate and B-factor refinement, and grouped occupancy refinement can be performed in reciprocal space, with maximum likelihood and least-squares targets. With real-space refinement, loops can be built and optimized, side chains fitted, and ligand, solvent, and water molecules placed. In addition to many of the standard features, PrimeX has some unique capabilities:

- Extended polypeptides can be built into electron density. Loops (stretches of amino acids with two anchor points) of up to 40 residues in length and tails (stretches of amino acids with one anchor point) of up to 20 residues in length can be built without manual intervention. Polypeptides are built from the target sequence, not from the sequence of the initial model.
- Multiple side chains can be placed cooperatively: the correct conformer can be found for each side chain even though they may occupy the same volume in the starting model.
- Point mutations can be performed automatically to bring the sequence of the initial model into agreement with the target sequence.
- All topology and parameter generation for ligands and modified side chains are handled automatically.
- Application of the OPLS force field for geometric restraints produces a finished structure already prepared for use with other Schrödinger computational tools.

PrimeX also has access to the wide range of structure and surface manipulation and management tools that are a part of Maestro.

An overview of the PrimeX process is given in Chapter 2, and a brief description of the implementation is given in Appendix B.

PrimeX can handle up to 3000 amino acids and 5000 waters on a 64-bit machine (using the `linux-x86_64` architecture).

1.2 Running Schrödinger Software

Schrödinger applications can be run 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 the Maestro interface. This directory then becomes the 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-2015-2 → 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.

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 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-2015-2. 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 SchrodingerSuite2015-2 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. To set other variables, on OS X 10.7 use the command

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

and on OS X 10.8, 10.9, and 10.10 use the command

`launchctl setenv` *variable* `"`*value*`"`

For more information on starting Maestro see Section 2.1 of the *Maestro User Manual*.

Jobs:

Schrödinger software is run under Schrödinger's Job Control facility, which manages the submission of jobs to remote hosts and retrieval of the results. To submit jobs to remote hosts some configuration is required. See Chapter 7 of the *Installation Guide* for information.

1.3 Starting Jobs from the Maestro Interface

To run a job from the Maestro interface, you open a panel from one of the menus (e.g. Tasks), make settings, and then submit the job to a host or a queueing system for execution. The panel settings are described in the help topics and in the user manuals. When you have finished making settings, you can use the Job toolbar to start the job.

You can start a job immediately by clicking Run. The job is run on the currently selected host with the current job settings and the job name in the Job name text box. If you want to change the job name, you can edit it in the text box before starting the job. Details of the job settings are reported in the status bar, which is below the Job toolbar.

If you want to change the job settings, such as the host on which to run the job and the number of processors to use, click the Settings button. (You can also click the arrow next to the button and choose Job Settings from the menu that is displayed.)

You can then make the settings in the Job Settings dialog box, and choose to just save the settings by clicking OK, or save the settings and start the job by clicking Run. These settings apply only to jobs that are started from the current panel.

If you want to save the input files for the job but not run it, click the Settings button and choose Write. A dialog box opens in which you can provide the job name, which is used to name the files. The files are written to the current working directory.

The Settings button also allows you to change the panel settings. You can choose Read, to read settings from an input file for the job and apply them to the panel, or you can choose Reset Panel to reset all the panel settings to their default values.

You can also set preferences for all jobs and how the interface interacts with the job at various stages. This is done in the Preferences panel, which you can open at the Jobs section by choosing Preferences from the Settings button menu.

Note: The items present on the Settings menu can vary with the application. The descriptions above cover all of the items.

The icon on the Job Status button shows the status of jobs for the application that belong to the current project. It starts spinning when the first job is successfully launched, and stops spinning when the last job finishes. It changes to an exclamation point if a job is not launched successfully.

Clicking the button shows a small job status window that lists the job name and status for all active jobs submitted for the application from the current project, and a summary message at the bottom. The rows are colored according to the status: yellow for submitted, green for launched, running, or finished, red for incorporated, died, or killed. You can double-click on a row to open the Monitor panel and monitor the job, or click the Monitor button to open the Monitor panel and close the job status window. The job status is updated while the window is

open. If a job finishes while the window is open, the job remains displayed but with the new status. Click anywhere outside the window to close it.

Jobs are run under the Job Control facility, which manages the details of starting the job, transferring files, checking on status, and so on. For more information about this facility and how it operates, as well as details of the Job Settings dialog box, see the *Job Control Guide*.

1.4 Citing PrimeX in Publications

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

PrimeX, version 2.8, Schrödinger, LLC, New York, NY, 2015.

The PrimeX Process

PrimeX has two unique features that change the way that this program is employed in refining a molecular replacement crystal structure solution, as compared to other refinement programs. Although the workflow discussed here for PrimeX will seem familiar to protein crystallographers, the use of the OPLS force field and the availability of strong real-space refinement tools affect workflow in ways that deserve special mention.

First, because PrimeX uses the all-atom OPLS force field throughout, hydrogens must be part of the refinement. Hydrogen atoms only participate in the calculation of geometric restraints, and not in X-ray calculations. An option in the Calculations Settings dialog box determines whether a structure is returned from refinement with or without hydrogens, a choice to be made depending on whether you wish to explicitly refine hydrogen positions or not. Early in the refinement of a structure, you might wish to ignore hydrogens. In that case, hydrogen atoms are added to the structure in an intelligent manner when it is submitted for any type of refinement. After the refinement job finishes, the hydrogens are deleted so that their temporary presence is of no direct concern to the crystallographer. Late in a structure refinement, you might consider refining hydrogen positions over multiple cycles of refinement in order to focus on the details of hydrogen bonding or catalytic mechanism. In this situation, hydrogen coordinates are returned with the structure by selecting that option.

Second, PrimeX depends on two real-space refinement tools, loop refinement and side-chain prediction. Loop refinement builds into an omit map loops of up to 40 residues in length, and tails (polypeptides anchored only at one end) of up to 20 residues in length. Side-chain prediction can fit multiple side chains into their electron density during a single application of this tool, so that side chains that might occupy the same volume of space in the model can be correctly placed in the electron density.

The sample workflow presented here is representative of how you might proceed with the refinement of a molecular replacement solution using PrimeX. Any such workflow can be affected drastically by such factors as the high resolution limit of the data and the sequence identity between the starting model and the target sequence. This workflow is relatively simple compared to the possible complications that can develop during any given refinement.

Starting with a molecular replacement solution obtained from other software, the first PrimeX operation is most likely to be rigid body refinement. Single or multiple groups may be defined according to the complexity of the contents of the asymmetric unit. Early rounds of refinement should normally apply anisotropic B-factor scaling. The mask method for bulk solvent correc-

tion is usually the best choice, although one might consider using the Babinet method while the boundary between protein and solvent is still especially uncertain. At these early stages of refinement, you can avoid having hydrogens returned with a refined structure by clearing the Include H-atoms in structure output option in the Calculation Settings dialog box.

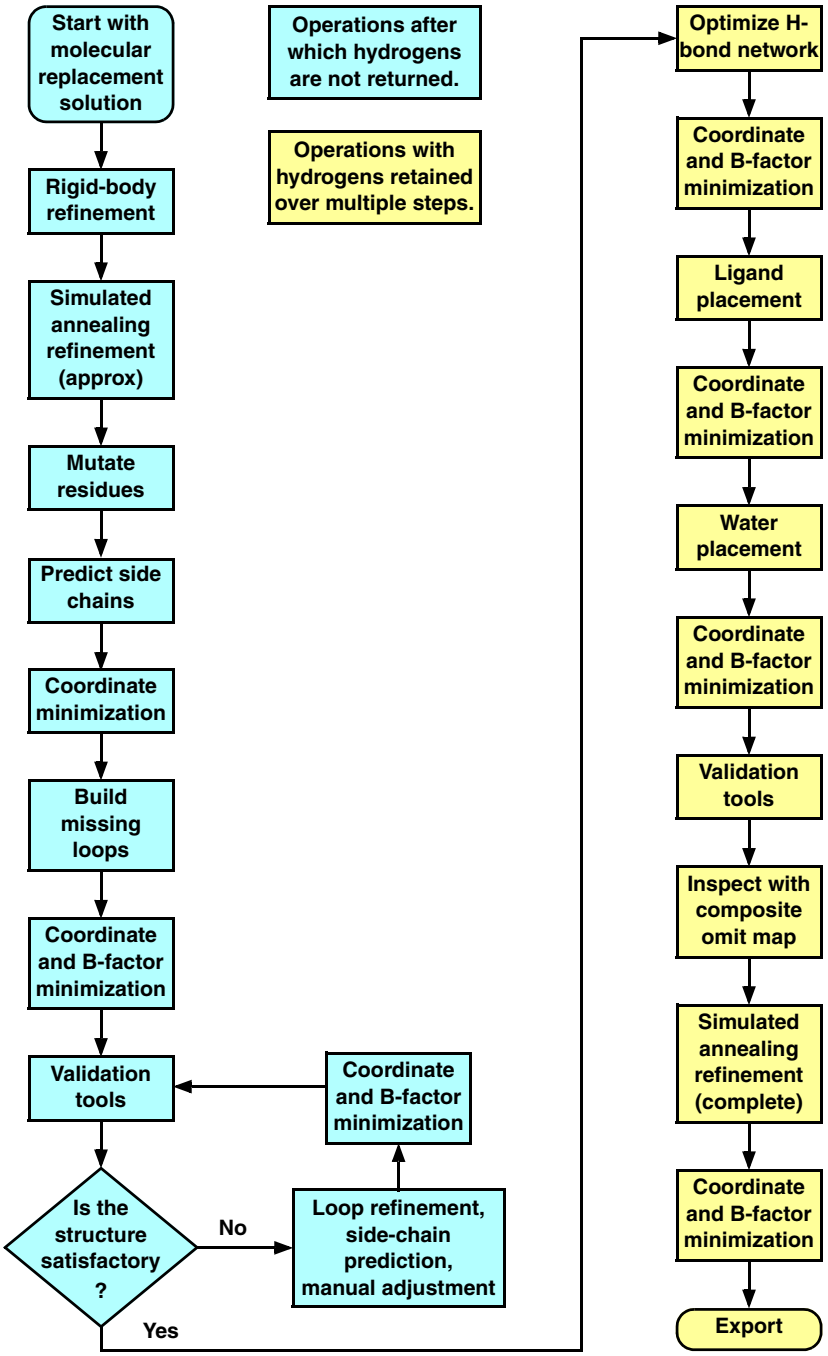
At some point in the workflow, the individual isotropic temperature factors from the original model should be reset to a constant value to help reduce the bias of the target toward the starting model. For simple molecular replacement solutions, this operation can occur in the Input Data panel. This function is also available in the General tab of the Reciprocal-Space Refinement panel for use at any point in the workflow. Note that resetting B-factors is likely to raise the R-factors to the extent that they have some resemblance to the correct B-factors for the target. Difficult molecular replacement solutions may be made even more tenuous if B-factors are reset too soon. Refinement of individual B-factors is often left until the target structure is substantially complete with respect to the protein backbone. Initial B-factor refinement is usually highly restrained, and may be followed late in the refinement with less-highly restrained B-factor refinement. The appropriate course of action will depend strongly on the resolution of the structure, as well as other details of a particular refinement problem.

Simulated annealing refinement is often helpful to lower the R-factors for a molecular replacement model at this stage, although use of this function may not be required where the starting model and target are already very similar. The approximate energy model in molecular dynamics refinement is intended to be used in this situation where faster results are more important than strict adherence to the most accurate energy model. Default values should be suitable for most situations.

Point mutation differences between the starting model and the target sequence may be resolved at this time by using the Mutate Model to Sequence tool, which replaces side chains in reasonable conformations without reference to the electron density. The Predict Side Chain tool can fit multiple side chains to electron density in one job, and should immediately follow the mutation of one or more residues. Missing residues may be built into electron density with the correct sequence by using the Loop refinement tool at this point or in later stages of the refinement. Both of these tools are available in the Real-Space Refinement panel.

Reciprocal-space minimization is frequently used to optimize changes made in the structure introduced by other methods such as side chain prediction. How and when these methods are applied very much depends on details of the structure determination, in ways that are beyond the scope of this software manual. The same good refinement practices that are to be used with other refinement packages also apply to PrimeX.

Figure 2.1. (opposite) Flow diagram for a simple typical workflow.


Start with molecular replacement solution
Rigid-body refinement
Simulated annealing refinement (approx)
Mutate residues
Predict side chains
Coordinate minimization
Build missing loops
Coordinate and B-factor minimization
Validation tools
Is the structure satisfactory ?
No
Yes
Loop refinement, side-chain prediction, manual adjustment
Coordinate and B-factor minimization
Operations after which hydrogens are not returned.
Operations with hydrogens retained over multiple steps.
Optimize H-bond network
Coordinate and B-factor minimization
Ligand placement
Coordinate and B-factor minimization
Water placement
Coordinate and B-factor minimization
Validation tools
Inspect with composite omit map
Simulated annealing refinement (complete)
Coordinate and B-factor minimization
Export

Validation tools should help to pinpoint problem areas of the structure that are incorrect and need remedial action beyond what reciprocal-space refinement is able to accomplish. The validation tools available in PrimeX are a Density Fit table, the Protein Reports pages and a Ramachandran plot. The power of real-space loop refinement and side chain prediction make these functions the first choice for improving problematic stretches of residues. If those tools cannot resolve the observed issues, manual adjustment of the model in Maestro may be necessary.

Once the model is substantially correct, one should begin to consider some of the more subtle issues in structure determination, such as the orientation of the His, Gln and Asn side chains in the context of hydrogen-bonding partners. The correct order of operations is to first select Include H-atoms in structure output in the Calculation Settings dialog box, if not already set, and then treat the structure with reciprocal-space minimization. At that point in the workflow, hydrogen bonding networks can be considered. The Optimize H-bond Networks tool can automatically orient these side chains and all polar hydrogens for optimized hydrogen-bond interactions. You can also manually select and fix the orientation of any hydrogen bond or side chain.

Ligand placement is best performed when the refinement of the binding site is as complete as possible, before the addition of solvent atoms or water. Once a ligand pose is chosen and refined, water addition can begin using the Water placement tool.

A final round of low-temperature simulated annealing may help position atoms, and especially hydrogens. Since hydrogen positions are influenced strongly by electrostatic interactions, the complete energy model should be used at this stage of the refinement. Final reciprocal-space minimization of the structure will help produce the best possible R-factors and a structure suitable for immediate use in various applications of computational chemistry.

PrimeX and Maestro Basics

The graphical interface for all Schrödinger products, including PrimeX, is Maestro. Maestro provides many features for storing, managing, manipulating, creating, and modifying structures, surfaces and associated data. Full details on the general capabilities of Maestro can be found in the *Maestro User Manual*. This chapter describes the Maestro features that are particularly relevant to PrimeX, in the context of performing common tasks.

3.1 The Maestro Main Window

At the top of the Maestro main window are the title bar and the menu bar, followed by the toolbars. In the center is the Workspace, where structures and surfaces are displayed. At the bottom is the Auto-Help area, which displays information about the current task. Several optional components can be displayed at the bottom of the window: the sequence viewer, the status bar, the command input area, and the clipping planes window. These components are described in detail in Chapter 2 of the *Maestro User Manual*. Some of these are displayed by default. For PrimeX, it is useful to display the sequence viewer and the status bar, which you can do from the Window menu.

There are several shortcut menus in the main window, which you display by right-clicking and holding over an atom, an empty part of the Workspace, or the sequence viewer.

The toolbars contain buttons that perform an action immediately, and button menus, from which you can select an action to perform. The menus open when you click the arrow to the right of the button. The toolbars are configurable, from the toolbar shortcut menu or the Maestro menu. Toolbars are displayed from the Manager toolbar, which has buttons for showing or hiding each toolbar. For PrimeX, you will want to display the Proteins toolbar.

Maestro is capable of displaying structures in stereo, and can use perspective and fogging to enhance the stereo effect. For information on these capabilities, see Section 4.7 of the *Maestro User Manual*.

3.2 Maestro Projects

Molecular structures, surfaces and data are managed in a Maestro *project*. When you start Maestro, a project is opened; if none is specified, the project is a *scratch* project, which can be saved as a named project. All project data is automatically saved to disk.

Structures are stored in project *entries*, and may consist of multiple molecules. As well as the structure, entries can include properties of the structure as a whole (such as crystal parameters) and of the atoms in the structure (such as B-factors). When you run a task, the updated structure is returned as a new project entry, or several entries, if more than one structure is returned.

The primary interface to the project is the Project Table panel. This panel lists the entries and their properties, and provides a range of tools for selecting, filtering, sorting and organizing entries and properties. To open the panel, click the Table button on the Project toolbar, or press CTRL+T.

Controls for display of surfaces (maps) are available in the Manage Surfaces panel. To open this panel, choose Workspace → Surface → Manage Surfaces, click the S button in the Title column of the Project Table or in the Entry List panel, or click the Surf Table button in the PrimeX panel.

3.3 Manipulating Structures

This section describes the tools available in Maestro for manipulating structures in the Workspace. You can change the orientation or location of the entire Workspace contents, or of any part of a structure. The coordinate system is set up so that the *x* axis is horizontal, the *y* axis is vertical, and the *z* axis is perpendicular to the plane of the screen. The displayed structures are clipped along the *z* axis, so that (depending on the zoom) you only see a "slice" of the structure. You can adjust the clipping in the Clipping Planes window, which can be displayed from the Window menu. You can also make use of perspective and view the Workspace in stereo.

3.3.1 Changing the View (Global Transformations)

Changing the view of the Workspace in Maestro is done with the middle and right mouse buttons. To translate the entire contents of the Workspace, drag with the right mouse button. To rotate the Workspace contents, drag with the middle mouse button for rotation about the *x* and *y* axes, and hold down the CTRL key and drag with the middle mouse button for rotation about the *z* axis. You can change the coordinate origin for rotation by "spot-centering": right-click on the atom you want to use as the new origin, or right-click on a map to center the display on that point. These operations are also called *global transformations*, and only change the view of the structures, not the coordinates. If you want to use different mouse actions for these and other operations, you can customize them with Maestro → Customize → Mouse Actions.

There are several other operations that you can do to change the view of the structure:

- To zoom in on (or out of) part of a structure, use the mouse wheel, or drag to the left or right with the middle and right mouse buttons.
- To fit the structure to the Workspace, click the Fit button on the View toolbar.

- To fit a selection of atoms to the Workspace, select the atoms first, then click the Fit to Workspace toolbar button. This action also zooms in in the clipping planes window by default; the default can be changed in the Workspace – Fitting tab of the Preferences panel.
- To restore the original view, click the Reset button on the View toolbar.

- To center the view on an atom, right-click the atom.
- To center the view on a particular part of a density map, right-click on one of the displayed vertices of the map.
- To rotate by 90° around the *x* axis or the *y* axis, click the View toolbar buttons.

- To align the view on a particular axis or plane, choose Edit → Align.

Rotation, translation, zooming, adjustment of clipping, and other view changes can also be done with keystrokes. See Section 2.9 of the *Maestro User Manual* for more information.

When you are performing rotations, the rotation center can be fixed on a part of the structure and move when you translate the structure, or it can be set in the center of the screen regardless of how you translate the structure. The behavior is controlled from the Preferences panel, by setting the Keep center of rotation fixed while translating option in the Workspace – Rotation tab. Selecting this option ensures that the center of rotation remains in the same location regardless of translation of the structure. This location is not necessarily the center of the Workspace: it is the last location used. If the option is deselected, the center of rotation moves with the structure during translation. Centering the view on an atom or fitting to screen changes the rotation center. There is also an option to display the center of rotation (in the same panel and tab).

3.3.2 Local Transformations

If you have multiple structures in the Workspace and want to reorient only one of them, or if you want to move only part of a structure, you can perform a *local transformation*. When you do, the coordinates of the structure you move are changed. You can also use local transformations to move a set of atoms as a rigid body.

To perform a local transformation:

1. Make a choice from the Transform button menu on the Edit toolbar.

2. Pick the parts of the Workspace that you want to move.
3. Drag with the middle mouse button to rotate, or the right mouse button to translate.

An alternative method is the following:

1. Pick the parts of the Workspace that you want to move using the Select tool.

2. Right-click in the Workspace and choose Transform from the shortcut menu.
3. Drag with the middle mouse button to rotate, or the right mouse button to translate.

While you are performing a local transformation, you can temporarily switch to global transformation mode by holding down the space bar. When you release the space bar, you return to local transformation mode.

Rotations are performed about the centroid of the picked parts of the structure. You can set the center of rotation in the Advanced Transformations panel, which you open by choosing Advanced from the Local transformation toolbar button. The center of rotation is also updated when you "spot-center" on an atom, by right-clicking on the atom. If you want the center of rotation to stay in the same place, and not move when you translate the structure, open the Preferences panel from the Maestro menu, and under Workspace – Rotation, select Keep center of rotation fixed while translating. You can also select Display center of rotation in the same tab if you want a marker of the rotation center to be displayed.

If the atoms you select for local translation include atoms with alternate positions, you can translate both the main and the alternate coordinates or just the main coordinates. This choice is made in the Workspace – Transformations tab of the Preferences panel.

A special kind of local transformation, structure sculpting, is available from the Tools menu and also from the Sculpt button on the Proteins toolbar:

In structure sculpting, a local transformation is followed by a limited force-field minimization in the region around the moved atoms. Once you enter this mode (by selecting the menu item), you can select atoms and drag them to a new location with the left mouse button. When you release the mouse button, the minimization starts. The selected atoms are restrained. Atoms in the same residues have no constraints placed on them; atoms in residues that are within 5 Å of the moved residues are restrained, and all other atoms are frozen. The minimization proceeds for a fixed number of cycles or until there is no significant change. During the minimization you can continue to work on the structure. The minimization stops if you exit the mode, for example by choosing some other action that involves picking atoms, or change the content of the Workspace. If you select some more atoms and drag them while the minimization is proceeding, it stops and a new minimization begins.

3.3.3 Structure Adjustments

You can change individual bond lengths, bond angles, and dihedral angles (torsions), by using the toolbar buttons and the mouse. On the Proteins toolbar, the Quick Torsion tool allows you to pick a bond, then drag with the left mouse button or use the mouse wheel to rotate around that bond. The smaller part of the structure that the bond joins is rotated.

The Adjust button menu on the Edit toolbar offers a wider range of adjustment tools, arranged in several groups.

The first group includes the Quick Torsion tool that is also on the Proteins toolbar, along with an item to display contacts between atoms (see Section 10.2.2 of the *Maestro User Manual*).

The second group contains tools for rotating parts of the backbone structure of a protein.

- Rotate Peptide Plane—rotate peptide link about the line between the alpha carbons.
- Rotate Carbonyl Plane—rotate the plane of the carbonyl in a peptide link
- Convert Cis/Trans—Convert the peptide link from cis to trans or trans to cis.

To use one of these tools, select the menu item, click on a peptide link, and drag with the left mouse button or use the mouse wheel.

The third group contains tools for adjustment of distances, angles, dihedrals, and chirality. To use these tools, select the menu item, then click on the appropriate number of atoms (two for distances, three for angles, and four for dihedrals). When all atoms are selected, you can then drag with the left mouse button or use the mouse wheel to make the adjustment. For a chirality inversion at a center, first pick the chiral center, then the two atoms that will remain in place. The conversion is done when the third atom is picked. These tools are also available from the Adjust panel, which you open from the Edit menu, and which allows you to specify distances or angles by entering a numerical value. You can also perform chirality inversions or flip ring substituents with the Build toolbar button:

The last item on the menu, Rotamers, allows you to select a side-chain rotamer from among the common rotamers for each residue. To use this tool, first select the residues you want to adjust, then select the menu item. The Rotamers dialog box opens, in which you can select one of the standard rotamers for the residue, or, if you have a map displayed, select the rotamer that best fits the density, by clicking Choose Best Fit. The best fit excludes rotamers that have clashes with other parts of the structure. The map must be a 2Fo-Fc or 3Fo-2Fc map. A measure of the density fit is given in the Fit column, and a + sign in the Clash column indicates that there are clashes with other atoms.

You can navigate the residue list with the left and right arrow keys, and the rotamers list with the up and down arrow keys. You can sort the rotamers table by the values in any column, by clicking the column heading—for example, you can sort by density fit. If you decide that you prefer the original rotamer, you can select it in the table or click Restore to Original Conformation.

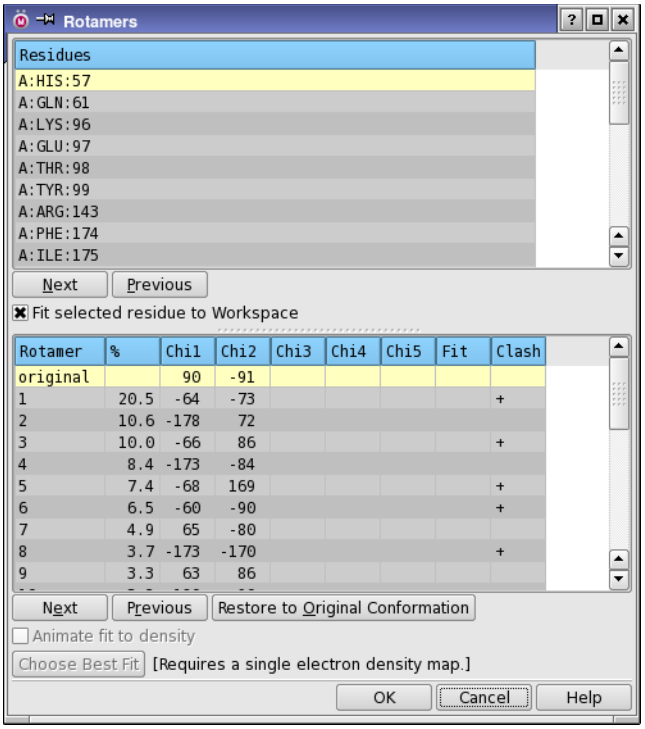


Figure 3.1. The Rotamers dialog box.

3.4 Identifying Structural Features

Maestro provides a variety of ways to identify structures or features in a structure. This section summarizes the use of color schemes, labeling, displaying atoms and bonds, and feedback information. For complete information, see the *Maestro User Manual*.

3.4.1 Using Color Schemes

Color schemes are useful for identifying parts of a structure by the values of a property. The schemes can be applied to some or all of the atoms in the structure displayed in the Workspace. The color scheme is also applied to the residues in the sequence viewer.

To color all atoms with a particular scheme:

- Choose the scheme from the Color Scheme button menu on the Representation toolbar.

To color selected atoms with a particular scheme:

1. Open the Atom and Bond Coloring panel from the Workspace menu.
2. Select Color Scheme and then select the scheme from the option menu.
3. Use the Apply color scheme to atoms picking tools to select the atoms in the Workspace that you want to color. For more information on the picking tools, see Chapter 6 of the *Maestro User Manual*.

There are several color schemes that are particularly useful for PrimeX:

- Atom PDB B factor —This scheme uses a blue-white-red color ramp for the temperature, with blue at the low end and red at the high end. The actual mapping of the colors to the B-factor values can be displayed in the Color Scheme tab of the Atom Coloring panel.
- Residue type—This scheme colors residues according to their identity. This enables you to quickly pick out a Gln, for example.
- Entry—This scheme applies a unique color to each entry in the Workspace, and is useful when you are examining the results of loop prediction or ligand placement and have several predictions in the Workspace.
- Density Fit—This scheme applies a green-yellow-white color ramp to residues according to the accuracy of the density fit based on real-space R-factors, with green for the best fit and white for the worst. The density fit has to be precalculated using the Density Fit panel.

3.4.2 Labeling Atoms

Adding labels to atoms or residues can be used to identify them, but can also be used to provide information about their properties. For a protein, it may not be useful to apply labels to all atoms, because doing so can make Workspace operations slow. You can apply labels selectively, either by picking or by using the Atom Selection dialog box. Labeling tools are available from the Labels toolbar and from the Atom Labels panel, which you open from the Workspace menu. For details, see Section 8.1 of the *Maestro User Manual*.


3.4.3 Displaying Atoms and Bonds

Maestro includes a range of tools that you can use to selectively display atoms in a structure and to change the appearance of atoms and bonds.

When you want to focus on a particular region of a molecule—for example, a loop that has just been refined—it is useful to display only the atoms of interest. The Maestro toolbar contains five tools for displaying all or selected atoms. Three of these provide shortcuts for protein backbones, side chains, and waters, among others. For more information, see the description of the Display Atoms toolbar in Section 7.5 of the *Maestro User Manual*.

Maestro has five representations of molecules that you can use to distinguish different parts of the structure—for example, distinguishing the ligand or waters from the protein. The five representations are wire, ball and stick, CPK, tube, and thin tube, and can be applied from the Representation toolbar. For more information, see the description of the toolbar in Section 7.1 of the *Maestro User Manual*, and Section 7.3 of the *Maestro User Manual* for details.

You can also display hydrogen bonds in the Workspace from the Measurements toolbar, either within a molecule (such as a protein) or between molecules (such as between a protein and a ligand). The parameters that define a hydrogen bond can be adjusted in the Preferences panel, which you open from the Maestro menu. For more information, see Section 14.8.1 of the *Maestro User Manual*.

3.4.4 Workspace and Sequence Viewer Feedback Information

When you pause the cursor over an atom in the Workspace, information on the atom is displayed in the status area immediately below the Workspace. This information includes the chain name, residue name and number, and atom name. The display of information is a preference that can be set in the Workspace tab of the Preferences panel, which you open from the Maestro menu.

Likewise, in the sequence viewer, pausing the cursor over a residue displays the residue name and number in a tooltip.

You can also display information on a single unique entry that is displayed in the Workspace—the only entry in the Workspace, the only selected entry in the Project Table that is also displayed in the Workspace, or the only non-fixed entry in the Workspace. The display of this information is also a preference that can be set in the Preferences panel.

3.4.5 Finding Structural Components in the Workspace

If you want to find an atom, molecule, residue, chain, or other group of atoms in the Workspace, you can do so with the Find toolbar, which you open from the Edit menu or by pressing CTRL+F (⌘F). All instances of the object that match are located. This toolbar is described in Section 2.10 of the *Maestro User Manual*.

If you have a single residue selected in the Workspace (or atoms from a single residue), you can go to the next or previous residue by pressing N or P.

3.5 Modifying Structures

When you modify a structure in the Workspace, the changes are also made in the project entry to which the structure belongs. This is the default behavior. You can change this behavior in the Project tab of the Preferences panel, under Synchronize Workspace changes, so that changes in the Workspace are only propagated to ("synchronized with") the entries when you request it. This is convenient when you want to experiment with a structure, for example. However, you should be careful to reset the preference when you want changes to be saved automatically again. When there are unsaved changes in the Workspace, a button for saving changes is displayed in the status bar.

An alternative to changing the preference is to create a new entry with the Workspace contents when you want to experiment with a structure. You can do this with the Checkpoint button in the PrimeX panel. To do so you must have a reflection file as well as a structure, since checkpointing associates the structure with the reflections as well as creating a new entry. Any changes already made to the structure are associated with the new entry if the Workspace contents have not been synchronized. The checkpoint function can thus act as a multi-step undo function, by deleting entries from the latest one saved.

Maestro provides a number of other tools for modifying structures, many of which are available from the Build panel, which you open from the Edit menu. The basic building tools are displayed on the Build toolbar.

One of the useful features available from the Build panel is the ability to mutate the standard residues, which you might want to do on an individual basis rather than use the Mutate Model to Sequence panel.

To mutate a residue:

1. Choose Amino acids from the Fragments option menu in the Fragments tab of the Build panel.

   The text area below the option menu is populated with buttons for the standard 22 amino acids plus four protonated or deprotonated forms, and the Mutate option appears in the top right of the tab.

2. Select Mutate.
3. Ensure that Place is selected.
4. Click on the button for the amino acid you want to mutate to.
5. Click on an atom in the amino acid you want to mutate in the Workspace.

   You can click on the residue in the sequence viewer so that it is highlighted in the Workspace, or use the feedback in the status bar to identify the residue.

An alternative procedure, which you can perform in the Workspace, is as follows:

1. Select the residue in the Workspace or in the sequence viewer.

   You can use the feedback displayed in the status bar to identify the residue in the Workspace, or the tooltips in the sequence viewer.

2. Choose the new residue from the Mutate submenu of the selection shortcut menu (which is displayed when you right-click and hold in the Workspace).

Another common task that you can do from the Build panel is to place a solvent molecule or ion.

To place a solvent molecule or ion:

1. Choose Protein Ions & Solvents from the Fragments option menu.
2. Ensure that Place is selected.
3. Click on the button for the molecule or ion you want to place.
4. Click in the Workspace at the location you want to place the molecule or ion.

   Make sure you don't click on an atom: this performs a fragment mutation. If you have a map displayed you can click on the corresponding density peak. The centroid of the ion is placed at the pointer position, with a z coordinate of zero (the z axis is perpendicular to the screen). You will probably need to adjust the positioning of the fragment that you placed and optimize its location.

3.6 Tracking Changes

PrimeX tasks return a structure as a new entry in the Project Table (except for R-factor calculation, which adds to an existing entry). To track the changes made to a structure as tasks are performed, you can use any of the following properties to track information about the progress of the refinement:

- Each entry has a title that is displayed and that can be modified freely, by editing the Title cell in the Project Table
- The job name for each PrimeX job is also recorded as a property in the Project Table. This name is used to name input and output files.
- A History property is added to each entry that records the task performed, and a Parent ID property records the entry ID of the structure used to start the job.
- You can create your own properties to record text information.

3.7 The PrimeX Panel

The PrimeX panel provides access to all the tools that are available for protein X-ray structure refinement. In the process of refining a structure, you will use the tools in this panel multiple times, so it is useful to leave it open. By default it is docked into the Workspace, but you can undock it if you wish. To open the panel, choose Tasks → Protein X-Ray Refinement or Applications → PrimeX.

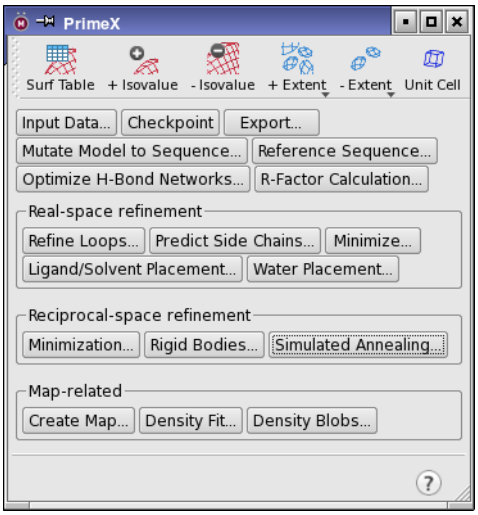


Figure 3.2. The PrimeX panel.

The main part of the panel has a set of buttons for opening task panels, or performing tasks directly. The buttons are grouped by function, and are described below:

- Input Data—Opens the Input Data dialog box so you can read structure and X-ray data into the project. See Chapter 5 for details.
- Checkpoint—Save the contents of the Workspace as a new entry in the Project Table. This creates a "checkpoint" of the current structure. A reflection file is required as well as a structure (and a sequence) for this operation.
- Export—Export the completed structure to a PDB file. See Section 11.4 on page 81.
- Mutate Model to Sequence—Mutate the model structure to match the sequence. Opens the Mutate Model to Sequence panel, which is described in Section 11.1 on page 77.
- Reference Sequence—Opens the Reference Sequence dialog box so you can examine the reference sequence.
- Optimize H-Bond Networks—Optimize the hydrogen bonding network, including terminal amide and histidine flips. Opens the Interactive H-bond Network Optimizer panel, which is described in Section 11.2 on page 78.
- R-Factor Calculation—Calculate R factors. See Section 11.3 on page 81.
- Refine Loops—Run a real-space refinement of protein loop structures. See Section 7.1 on page 60.
- Predict Side Chains—Run a real-space prediction of protein side chains. See Section 7.2 on page 61.
- Minimize—Run a real-space minimization. See Section 7.3 on page 63.
- Ligand/Solvent Placement—Place the ligand or solvent molecules by docking into a region of unassigned density. See Section 8.1 on page 65.
- Water Placement—Place water molecules in regions of unassigned density. See Section 8.2 on page 69.
- Minimization—Run a reciprocal-space minimization of the R-factors, B-factors, grouped B-factors, or grouped occupancies in the Reciprocal Space Refinement panel. See Section 6.4 on page 51.
- Rigid Bodies—Run a rigid-body refinement of selected parts of the structure in the Reciprocal Space Refinement panel. See Section 6.5 on page 53.
- Simulated Annealing—Perform a simulated annealing refinement in the Reciprocal Space Refinement panel. See Section 6.6 on page 55.

- Create Map—Opens the Create Map panel so you can create an electron density map. See Chapter 10.
- Density Fit—Opens the Density Fit panel, to display the fit to the density for each residue. See Section 12.1 on page 83 for more information on this panel.
- Density Blobs—Section 11.4 on page 81.

Each task panel has a Calculation Settings button, which opens the Calculation Settings dialog box, where you can set various general parameters for calculations. This dialog box is described in detail in Section 5.4 on page 39.

The PrimeX panel has a toolbar at the top that contains some map and display tools that are not on the Proteins toolbar. The buttons are described below.

Surf Table
Opens the Manage Surfaces panel. When you click on the arrow, a menu is displayed from which you can select a map type from the available maps to be displayed. See Section 12.4 of the *Maestro User Manual* for more information on this panel.

+ Isovalue
Increase the isovalue for the displayed surface (this makes the surface smaller). Each click on the surface increases the isovalue by 0.1 σ. Using control-click increases the isovalue by 1.0 σ for each click.

- Isovalue
Decrease the isovalue for the displayed surface (this makes the surface larger). Each click on the surface decreases the isovalue by 0.1 σ. Using control-click decreases the isovalue by 1.0 σ for each click.

+ Extent
Increase the extent of the density map that is displayed. Each click on the button increases the cube dimensions by the amount that has been selected from the button menu. (Click on the arrow to open this menu.) The default value is 1 Å. The available values range from 0.5 Å to 16 Å, and are synchronized with the values on the Decrease density extent button menu.

- Extent
Decrease the extent of the density map that is displayed. Each click on the button decreases the cube dimensions by the amount that has been selected from the button menu. (Click on the arrow to open this menu.) The default value is 1 Å. The available values range from 0.5 Å to 16 Å, and are synchronized with the values on the Decrease density extent button menu.

Unit Cell
Displays the unit cell boundaries in the Workspace.

You will probably also want to use the Proteins toolbar in the main window, which has the following buttons.

Quick Torsion
Rotate part of a structure around a bond. Pick the bond in the Workspace, and drag with the left mouse button to adjust the torsional angle.

Sculpt
Select and move atoms, followed by a short force-field-only minimization of the nearby atoms (Structure Sculpting).

Protein Reports
Opens the Protein Reports panel so you can view information about the protein structure. See Section 12.2 on page 84 for more information on this panel.

Ramachandran
Opens the Ramachandran Plot panel so you can view a Ramachandran plot of the protein dihedrals. See Section 12.3 on page 86 for more information on this panel.

Crystal Mates
View all or part of symmetry-related molecules that are within van der Waals contact of the asymmetric unit. If you click on the arrow, a menu is displayed, from which you can choose a distance from the displayed structure within which the crystal mates will be displayed. Performs the same action as Crystal Mates on the Workspace menu.

Regularize
Perform a short, real-space coordinate minimization for the selected residues, using only the force field (no X-ray terms).

Partial Occ
Displays three items:
Partial Occupancy—opens the Partial Occupancy dialog box.
Display Alternate Positions—displays the B conformations as ghost atoms.
Switch Alternate Positions—switches between the A and B conformations.

PrimeX for Non-Crystallographers

The purpose of this chapter is to provide additional background and guidance to users of PrimeX who have limited experience with protein crystallography. One popular application by non-crystallographers is the preparation of a structure refined with other programs for use in computational chemistry. This chapter includes some step-by-step instructions combined with enough additional information to use PrimeX with confidence.

This chapter assumes that you are familiar with basic operations in PrimeX and Maestro, such as how to use the Project Table. You may wish to review the Maestro documentation for help with such issues. Additional information on the PrimeX operations in this chapter is available in the *PrimeX Quick Start Guide*, as well as the remainder of this manual.

4.1 R-factors and Refinement

Protein crystal structures are molecular models created and refined to maximize agreement with X-ray diffraction data. The vast majority of protein crystal structures are solved in a resolution range where the coordinates of individual atoms cannot be discerned directly from the diffraction experiment (2.8 – 1.6 Å). Nevertheless, atoms can be placed in a model with errors on the order of 0.1 Å, given the known shape of amino acids and the observed shape of the electron density. Reciprocal-space refinement is typically used to optimize the fit of the atoms to the electron density more precisely than could be achieved manually.

Reciprocal-space refinement works through an approximate mathematical relationship that relates the shift of atom positions or of B-factors (see B-factor discussion below) to the differences between the observed diffraction pattern, and the one calculated from the current model. Because this relationship is only valid for very small shifts, optimization requires multiple minimization steps to converge.

The match between the observed diffraction pattern and the calculated diffraction pattern is measured by a relative-error statistic called the crystallographic R-factor. The R-factor is defined in the following equation, where F_{obs} and F_{calc} are the observed structure factor amplitudes and the predicted structure factor amplitudes calculated from the model, respectively.

$$R = \sum \left| F_{obs} - F_{calc} \right| / \sum F_{obs}$$

Reciprocal-space refinement is problematic because of the number of parameters to be optimized. At low resolution, three coordinates and one isotropic B-value for each atom can be as

numerous as the reflections that are observed. As with any operation where the number of parameters and observations are roughly equal, over-fitting of data (i.e., fitting to noise) is a danger. In order to confirm that changes in the structure are indeed improving the fit to real information, the statistical concept of cross-validation is used. A small portion of the observed data, typically 5 to 10%, is set aside and not used in the fitting. A special R-factor is calculated using only these test data. This quantity is known as R-free. A large R-free or an increasing difference between R and R-free suggests that alterations to the model have more to do with fitting to noise than increasing the accuracy of the model.

4.2 Molecular Geometry, Restraints, and Hydrogen Atoms

Certain aspects of molecular geometry are more accurately known than can be obtained from protein crystal diffraction. Bond lengths, bond angles, group planarity and non-bonded contacts can be derived from small molecule crystal structures, which are typically known to a much higher resolution than protein structures. Similar information is also available from quantum mechanical calculations or from microwave spectroscopy. This geometric information is applied during the refinement of protein crystal structures as restraints on the relative positions of atoms. The benefit of using this information in refinement is that false minima far from the correct solution are avoided. Another way of looking at this use of geometric restraints is that they bring additional information to the fitting procedure to compensate for the lack of sufficient diffraction data.

While this use of molecular geometry is easy to understand, the correct weighting of these two types of information during optimization is not always obvious. Protein geometry is usually quantified on the basis of energy terms while crystallographic results are in terms of electrons per $Å^3$. A weight factor is required to set the two types of terms to similar levels. Many crystallographers choose the weight that produces the lowest R-free consistent with good geometry. Fortunately, the R-free varies very slowly with changes in the weight near the minimum, so that a change in the weighting by a factor of 2 is barely significant near the optimal weight. In practice, a constant weight of 1 in PrimeX is often an adequate choice. Most structures will require a weight between 5 and 0.05. (See page 46 for information on setting this weight.)

In the resolution range of most protein crystal structures, no information about the position of hydrogen atoms is obtained from the diffraction experiments. Thus, crystallographic models of proteins usually do not contain coordinates for hydrogen atoms. Even if hydrogen coordinates can only be inferred indirectly from the positions of the other atoms, hydrogen positions are often needed for various calculations and for the interpretation of various chemical phenomena. Hydrogens may be added to protein crystal structures after the completion of refinement, but PrimeX operates with a different philosophy. Refinement of a protein model

with hydrogens present helps assure that all of the atomic coordinates are consistent with the observed data and known chemical geometry standards.

Geometric restraints leave their mark on crystallographic models. Of special concern are non-bonded interactions. All popular protein crystal structure refinement programs, except PrimeX, use relatively weak restraints to try to avoid unrealistic close approach of atoms in forming non-bonded contacts. In molecular mechanics, these close contacts are evaluated as high energy conformations that mask the contribution of other terms in energy calculations. Crystallographers are familiar with the chemical considerations that make these close non-bonded contacts improbable. However, they do not strictly restrain these interactions because of the concern that, if the restraints are expressed as their full extent, atoms may be locked in a false, local minimum and not move past one another during refinement of the structure. Some crystallographers also make the argument that since protein crystallography presents a time-averaged picture of the protein structure, the most highly populated positions for two atoms might be in close contact even though at any given instant, the atoms would never approach that close. PrimeX works with the philosophy that the same rigor should be applied to non-bonded interactions as other known geometric restraints. In practice, enforcement of plausible non-bonded contacts in PrimeX produces molecular models with equally good R and R-free as those with close contacts.

4.3 B-factors and Refinement

X-ray diffraction data contains information about atomic motions as well as their positions. The isotropic B-factor (also known as the isotropic temperature factor) is an atomic quantity proportional to the root-mean-square displacement of that atom in space due to thermal motion. However, errors in determining the B-factors at the typical resolutions of protein crystal structures make interpretation of the refined B-factors problematic.

That these B-factors are isotropic, which means the same in all directions, deserves some comment. Of course, one would not expect atomic motions to be independent of direction, but a more complete description of molecular motions (anisotropic temperature factor) would mean an additional six parameters per atom in a situation where the ratio of adjustable parameters to data points is already excessive.

When B-factors are adjusted in refinement, restraints are also used. In PrimeX, the assumption is made that atoms that are bonded to one another or that define a bond angle should have similar B-factors. Although this assumption may be questioned, it works as well as other models to restrain B-factors from varying more than is plausible.

4.4 Input Data

PrimeX requires three types of data. First, you need to have a list of coordinates for an existing model, usually available as a PDB file. After the selection of the model that you wish to work on, you will be asked about setting the isotropic B-factors. At this stage, you should set them all to a constant value, with 20 being a good starting point.

Second, you need sequence information for the protein. The sequence information should normally be the sequence for the intact protein in fasta format. The utility program `$SCHRODINGER/utilities/seqconvert` can translate sequence data between several formats. Alternatively, PrimeX can also read the sequence from a PDB file if this source of information is correct and more convenient.

Finally, you need an X-ray diffraction reflection file in one of two common formats, the CNS format (usually designated with a `.cv` suffix) or the CCP4 format (designated with an `.mtz` suffix). An `.mtz` file may contain multiple data sets, and some additional detective work may be required to find the correct columns in the file for use with PrimeX. Since the names used for data within `.mtz` files are only loosely standardized, you should consult with the source of the data file. If you are provided with data in another format (such as the PDB cif format), use the utility `$SCHRODINGER/utilities/refconvert` (see Section 13.9.2 on page 104) to translate the file into the CNS format. After specifying the reflection file, you will be asked about the test set. Accept the current test set, unless the number of test reflections is less than 200 or no test set is defined in the data that you are using. A test set comprising 5% of the reflections is usually adequate.

If you suspect that the sequence of the protein model differs from the reference sequence that you entered, you can click Mutate Model to Sequence in the PrimeX panel to open a panel that shows the discrepancy, if any. Normally one would not expect any differences at this stage. If you do find any differences, see the *PrimeX Quick Start Guide* for more information about dealing with this situation.

4.5 Where to Start: R-Factors

To calculate the R-factors, click R-Factor Calculation in the PrimeX panel (see Section 11.3 on page 81). Default values are almost always appropriate for this calculation, although a number of options are available in the Calculation Settings panel. After the calculation is complete, find the place in the Project Table where R and R-free are displayed.

Although the values of R and R-free provide a measure of the agreement of the model with the observed data, the interpretation of these values requires some experience and care. For most crystal structures in the typical resolution ranges, R and R-free values lower than 0.20 and 0.25, respectively, are very good results. Several other issues may play a role in determining if

R and R-free reach values that low, but the quality of the diffraction data is the most important. Resolution plays a big role relative to quality. The worse the resolution, the more difficult the task can be to produce a good model and to obtain good statistics. In the lower resolution range for which PrimeX is usually employed (3.0-2.6 Å), an R-factor less than 0.25 and R-free less than 0.3 can be a good result.

Some statistics that measure the quality of a diffraction data set are the percent completeness in the highest resolution range (how much data is missing up to the resolution limit), the redundancy of measurement (on average, how many times each data point was measured) and I/σ_I (signal to noise ratio) for the highest resolution data as well as the entire data set. Assessing these statistics is difficult, but obviously large amounts of missing data (completeness less than 0.90) and low I/σ_I (less than 5 overall) are warning signs that the data set may be lacking.

A poor data set is usually not the fault of the crystallographer. Often the biggest challenge in data collection is to obtain crystals that are large enough to provide quality data. One approach is to ask the crystallographer who collected the data (or if he is unavailable, any crystallographer) for an assessment of the data quality and the quality of the starting model that you have.

We assume that you are starting your work with a structure that is substantially correct. The approach will be to add hydrogens and other missing atoms, and then to refine the structure to remove bad contacts and bring the structure into agreement with the OPLS-AA force field.

4.6 Electron Density Maps

Calculation of an electron density map is a good next step. Click the Create Map button in the PrimeX panel to open the panel. For the map coefficients, select 2FoFc. All other parameters can be left at the defaults. After running the job, the Manage Surfaces panel opens, which allows you to customize the map display. You may wish to experiment with the map settings and the four icons on the PrimeX panel toolbar that control map size and contour level. Right-clicking on a residue centers the display and redraws the map.

The 2FoFc map formulation is standard for observing the electron density. The FoFc map, sometimes called a difference map, provides a direct indication where the model and the X-ray data do not agree well. Blue contours indicate volumes where too little electron density (possible missing atoms) is observed in your atomic model. Red contours indicate volumes where too much electron density is present in the model.

4.7 Missing Residues

Crystallographers have three main ways of designating which residues are not well defined in the electron density. The least common practice is to set the B-factors of the atoms that are not observed to a very high value (80 $Å^2$ or more) to indicate that that portion of the structure is disordered in some manner. Another method is to set the value of the occupancy for those atoms to zero. The occupancy is a quantity that ranges from 0 to 1 and specifies the fraction of the asymmetric units in the crystal in which an atom is present. However, the usual practice is to omit any atoms from the coordinate list for whose position essentially no information is observed in the electron density. This practice can be a problem for computational chemists since chemical information is usually available to show that those atoms are present, even if they are disordered and not observed in the electron density.

If you wish to place missing atoms in the structure, first use the Protein Reports panel (Tools menu) to find them. Go to the site of a residue with missing atoms to see if you can rationalize the choice made by the original crystallographer. Compare the map contours for another similar side chain that is complete. To add the missing atoms, build that part of the structure using real-space loop refinement, followed by side chain placement. Both of these operations are available in the Real-Space Refinement panel, which you open from the PrimeX panel. More information on using these two tools is available in Chapter 7 and in the *PrimeX Quick Start Guide*. This is the more rigorous approach.

If a single residue is missing atoms, you can right-click on the residue and use the Mutate tool to replace the residue with the same residue again (this time with all atoms). To place the side chain into the electron density, right-click on the residue and select Rotamers. In the Rotamers dialog box, you can click Choose Best Fit to allow the software to place the residue. Alternatively, you can step through the various common rotamers in the table to see which best fits the electron density. If you are not satisfied with any of these alternatives, you can use the Quick Torsion tool (PrimeX toolbar) to adjust the side chain conformation.

Water and other small molecules are probably correctly placed in your starting structure. Nevertheless, you should be aware of one error in a few crystal structures, where water is placed into electron density that belongs to part of the protein. Inappropriately-placed water molecules may interfere with placing missing atoms. Take care to examine the electron density at the sites where you are adding atoms to be sure that electron density for those atoms was not misinterpreted as bound water molecules. When building in one or more residues, any water molecules that might conflict with the new residue should be removed, and then replaced only after you are satisfied that they do not have unreasonably close contacts with newly-placed atoms.

One other complication should be considered about protein crystal structures and placing new atoms in them. Although the coordinates for one unique set of atoms are displayed, because of the periodic nature of crystals these atoms are surrounded by symmetry-related copies. To interpret electron density on the surface of the protein, viewing the symmetry-related copies is often useful. To view symmetry-related atoms, choose Within 10Å from the View Crystal Mates toolbar button menu.

4.8 Reciprocal-Space Minimization of Coordinates

When you are satisfied that all previously missing atoms in the model have been adequately placed into the electron density, reciprocal-space minimization is a logical next step. Choose Reciprocal-Space Refinement from the Select a task button menu. In the panel that opens choose Minimization as the method, and choose Optimal for the minimizer. Use the default settings for all other options. In the Map tab, select 2FoFc with otherwise default settings. In the General tab there is one important setting, the weight on the X-ray terms. Select a constant weight with a value of 1.0. We will discuss this setting in more detail below.

The NCS (non-crystallographic symmetry) tab requires some detailed explanation. Non-crystallographic symmetry only applies if you are working at low resolution (worse than 2.5 Å) and if you have more than one chain in the coordinate set with the same sequence. By restraining the two chains to be similar in structure, the chance of over-fitting is reduced. The Groups table allows you to define which chains should be considered similar. Clicking Add should automatically enter the basic information. The only additional parts that you need to enter are those residues to be excluded from consideration. NCS requires that the exact same atoms be present in the parts of the structures that are being considered similar. For example, if one chain has one more N-terminal residue than the other chain, you must exclude that residue by listing it in the NCS table. The default weights are adequate for most situations.

Once you have made settings, start the minimization. After the results are incorporated into the Project Table, you should note how the rms deviations for bond length and bond angle have changed, as well as how the R-factors have changed. If both R and R-free increase significantly during the course of minimization, delete that entry from the Project Table, increase the constant weight in the General tab by a factor of 2, and run the minimization again. If the R-factor decreases, but the R-free value increases, decrease the constant weight by a factor of 2. Finally, if the rms bond deviation increases above 0.03 or the rms bond angle increases above 3.0 degrees, decrease the constant weight by a factor of 2 and run the minimization again. Although you may need to change the weight a few times, once you find a good value you probably won't need to change it again for a given protein.

4.9 Optimizing H-Bond Networks

You will note that PrimeX has added hydrogens to your model. Hydrogens are essential to refinement in PrimeX. Further, it is important that the position of hydrogens be optimized to form the most likely hydrogen bonds. At this point, choose the Optimize H-Bond Network task from the PrimeX task button menu. Although this panel will allow you to view and optimize individual hydrogen-bonding clusters with manual control over all changes, we will use this task in the fully automatic mode. First, click Analyze Network. The network is analyzed and the table is populated. Next, ensure that View all species is selected and then click Optimize. The total number of hydrogen-bonding clusters and the progress in optimizing these clusters will be shown. You can close this panel when the task is finished.

4.10 Additional Minimization and B-Factor Refinement

You should minimize the atomic positions at least one more time, as described in Section 4.8, before using the model for additional calculations. Pay attention to the value used for the X-ray weight to derive the best balance between the force field and the X-ray data. If this minimization job resulted in a significant decrease in R-free, you might want to consider one more round of minimization. Additional minimization will do no harm, and may do some good.

B-factor minimization should now be performed, using the Minimization task in the Reciprocal-Space Refinement panel. In the Minimization tab, select Individual B-factors under Minimize, and in the Isotropic B-factor refinement section, set the B-factor restraint weighting factor to 1.0. Start the minimization. Observe the behavior of R and R-free due to B-factor minimization. Now, decrease the B-factor restraint weight by a factor of between two to five, and run the minimization again. Continue to reduce the restraint until R-free stops decreasing. At this point, two final rounds of coordinate and B-factor minimizations will help reduce R and R-free as low possible.

4.11 A More Complex Remediation Workflow

If you have reason to believe that the crystal structure that you are working with requires more substantial remediation, a possible workflow is outlined here. The *PrimeX Quick Start Guide* is a useful resource for additional information on the new operations not previously mentioned in this section.

1. Remove all waters and ligands.
2. Check for problems in the sequence using the Mutate Model to Sequence panel.
3. Add missing atoms as necessary.

4. Minimize with reciprocal-space refinement.
5. Use the Density Fit table to determine parts of the structure that may not fit the electron density well. Check Protein Reports for regions with unusual molecular geometry.
6. Consider fitting questionable parts of the structure using real-space loop fitting and side chain placement. Use manual adjustment of the structure if necessary.
7. Minimize with reciprocal-space coordinate refinement and optimize hydrogen-bond networks.
8. Minimize with simulated annealing using the complete energy model, and then additional reciprocal-space coordinate refinement.
9. Add ligands using the Ligand/Solvent Placement task.
10. Minimize with reciprocal-space coordinate and B-factor refinement.
11. Add solvent molecules using the Water Placement task.
12. Cycle through reciprocal-space coordinate minimization, reciprocal-space B-factor minimization and water placement at progressively lower peak heights until R-free no longer decreases.

Setting Up a PrimeX Refinement

The first step in the PrimeX refinement process is to specify the source of the model structure, the sequence, the reflection data, optional maps, and set the space group and unit cell parameters. These tasks are performed in the Input Data dialog box.

To open the Input Data dialog box, click the Input Data button in the PrimeX panel.

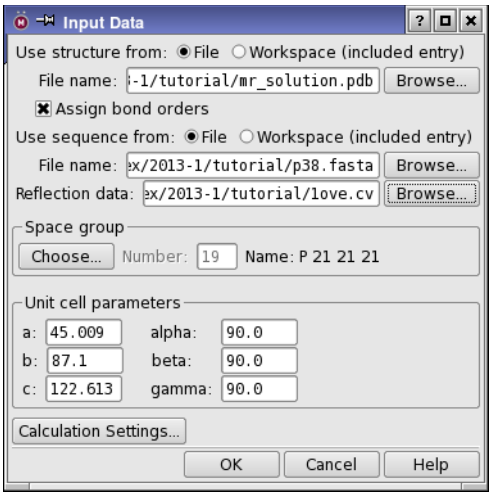


Figure 5.1. The Input Data dialog box.

5.1 Specifying the Initial Structure and Reference Sequence

The initial structure is specified with the Use structure from options. You can use one of two sources:

- File—Read the initial structure from a file. You can specify the file name in the File name text box, or click Browse and navigate to the desired file. The file must be in PDB or Masetro format. The space group and cell parameters are also read from the file if possible, and replace any existing definitions.
- Workspace—Use the structure in the Workspace as the initial structure.

There are some preparation steps that are necessary before a structure can be used in PrimeX, in particular, bond orders must be assigned for the ligand and other nonstandard groups, bonds to metals must be broken and formal charges changed to reflect a formally ionic metal. The treatment of metals is performed automatically. Assignment of bond orders is performed when the structure is read if you select Assign bond orders. You should also ensure that the residue names in the initial structure are properly assigned—see Appendix A for more information.

The protein sequence is specified with the Use sequence from options. This sequence is also known as the reference sequence. The two options are:

- File—Read the sequence from a file. You can specify the file name in the File name text box, or click Browse and navigate to the desired file. The file must be in PDB, Maestro, or FASTA format, and there must be a 1:1 correspondence between the chains in the file and the chains in the input structure.
- Workspace—Use the sequence of the structure in the Workspace as the reference sequence.

5.2 Specifying the Reflection Data

To specify the file containing the reflection data, enter the file name in the Reflection data text box, or click Browse and navigate to the file. In addition to the reflections, the space group and cell parameters are read from the file if possible, and replace any existing definitions.

If the reflections are read from an `.mtz` file, the properties for Fobs, Sigma, and the test set must be assigned from the properties in the file. When you click OK in the file selector, a dialog box opens that allows you to select a property for Fobs, Sigma, and the test set. For the test set, you can choose None to ensure that a test set is not read from the reflection file. By default, the test set from the reflection file is used.

If the reflection data file does not contain a test set, or if the test set defined in the input data is insufficient, the Input Data – Test Set dialog box opens. This dialog box displays the number of reflections, and allows you to select a random test set, by entering the percentage or number of reflections and clicking New. Note that this selection is made before rejecting any reflections, so it is possible that reflections in the test set could be discarded.

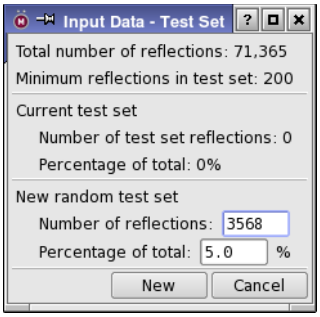


Figure 5.2. The Input Data - Test Set dialog box.

PrimeX can also make use of phased reflections to display maps and for use in real-space calculations. See Section 5.4.9 on page 43 and Section 10.1 on page 73.

5.3 Specifying Crystallographic Parameters

The two sections in the lower part of the Input Data dialog box provide a place to specify the space group and the unit cell parameters.

In the Space group section, you can enter the group number in the text box, or click Choose and choose the space group from the list provided in the Choose Space Group dialog box. If the structure file contained information on the space group, this information is automatically displayed, but can be changed to the relevant space group.

In the Unit cell parameters section you can specify the unit cell parameters. The parameters are automatically read from the reflection data file, and you can adjust them if you wish.

Note: Adjustments to the unit cell parameters must be made in this dialog box. Although in principle you could edit the values in the Project Table, this does not change the parameters used by PrimeX.

5.4 Specifying Calculation Settings

In addition to specifying the basic input data, you can set a number of calculation settings prior to running calculations. These settings are made in the Calculation Settings dialog box, which you open by clicking the Calculation Settings button. They include setting resolution limits on and filtering the reflection data, choosing a bulk solvent correction, controlling B-factor scaling, and including H atoms in structure output.

The Calculation Settings button is available in many of the PrimeX panels, so you can change the settings as needed in the various stages of refinement.

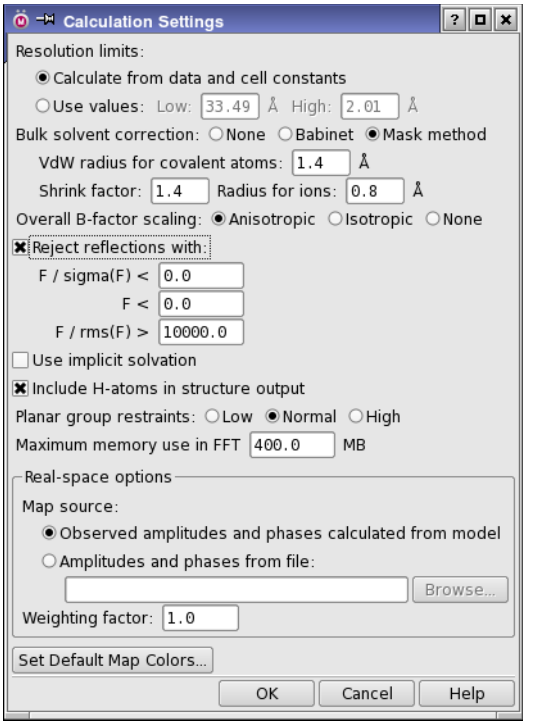


Figure 5.3. The Calculation Settings dialog box.

5.4.1 Resolution Limits

The Resolution limits options allow you to either calculate the limits from the data, by selecting Calculate from data and cell constants, or to specify the limits, by selecting Use values and entering the limits in the Low and High text boxes. Only the reflection data that lie between these two limits are used in the calculations. The initial values are set from the reflection data after input is completed. The limits can be changed for different stages of the refinement process. You might, for example, want to set the high-resolution limit to 4 Å for the rigid-body refinement, then change it back for the following stages.

5.4.2 Bulk Solvent Corrections

There are three options for bulk solvent correction: None, Babinet, and Mask method. The Babinet method is less accurate, but does not depend on having a good estimate of the structure. The Mask method relies on a definition of the boundary between the protein and the solvent, and is controlled by the following three parameters:

- VdW radius for covalent atoms—Van der Waals radius of covalently bound atoms. This radius is larger than normal van der Waals radii, to eliminate small regions that should not be considered as bulk solvent regions.
- Shrink factor—This factor is applied to reduce the covalent radii once the solvent regions are identified.
- Radius for ions—Van der Waals radius for atoms in ions. These are explicitly identified ions, which are usually in the first solvation shell.

The default values of these parameters are usually satisfactory.

5.4.3 Overall B-Factor Scaling

There are three options for overall B-factor scaling:

- Anisotropic—A single anisotropic scaling factor is optimized. Both isotropic and anisotropic components of this scale factor are applied, to the isotropic B-factors and to Fo, respectively This option is useful in most circumstances, and is also the default.
- Isotropic—Overall scaling is applied only to isotropic B factors.
- None—No B-factor scaling is performed.

5.4.4 Reflection Filtering

You can filter reflection data that are either too weak or too strong by selecting Reject reflections with and specifying the cutoff criteria. This option enables the rejection of reflections based on cutoffs on their F values. When you change any cutoff, the number of reflections rejected is displayed to the right of the text box. The three criteria are:

- F/sigma(F) < *cutoff* —Reject reflections for which the ratio of the F value to the standard deviation of the F values is less than the cutoff. This criterion eliminates reflections with large relative errors.
- F < *cutoff*—Reject reflections for which the F value is less than the cutoff. This criterion eliminates reflections with small amplitudes.
- F/rms(F) > *cutoff*—Reject reflections for which the ratio of the F value to the root-mean-square F value is greater than the cutoff. This option eliminates reflections with high relative amplitude, which might dominate the Fourier series and introduce large amounts of noise if they are in error.

5.4.5 Hydrogen Placement

Proper handling of hydrogens is critical in PrimeX. The presence of hydrogens is essential in any of the calculations that use force fields, because PrimeX uses the OPLS all-atom force fields. Once you have added hydrogens, and optimized their positions, you may retain them for the rest of the calculation. Although the hydrogens can be added back after they have been removed, their placement will not be optimal. The Include H-atoms in structure output option allows you to include the positions of hydrogen atoms in the output structure.

When PrimeX adds hydrogens, it does so in an intelligent manner during preprocessing, with consideration of possible hydrogen bonding partners. However, hydrogens already present in the structure are not optimized during this procedure, on the assumption that their positions have already been optimized to some degree. Hydrogen bonding networks are additionally optimized using a Monte Carlo procedure in the task Optimize H-bond Networks. In this task, additional orientations of asparagine, glutamine, and histidine side chains, the charge state of ionic groups, and the tautomeric state of histidines, are all considered while optimizing hydrogen-bond formation.

5.4.6 Force Field Solvation Model

When performing real-space refinements of any kind or reciprocal-space minimizations and rigid-body refinements, you can use a surface generalized Born (SGB) implicit solvation model for the solvent regions of the crystal, by selecting Use implicit solvent.

5.4.7 Restraints on Planar Groups

The force fields used in PrimeX real-space refinement permit nominally planar groups to deviate from planarity. Such deviations can be an indication of the need for more refinement. However, you can also choose to restrain these groups more closely to planarity by including additional restraints on the improper torsions. There are three options:

- Low—Allow the improper torsions to be determined solely by the force field, without further restraint.
- Normal—Apply a moderate restraint to improper torsions over what is already in the force field, typical of that used in other refinement programs.
- High—Apply a high level of restraint to improper torsions over what is already in the force field.

5.4.8 Memory Use in FFT

You can set the amount of memory used in the fast Fourier transform by entering a value in the Maximum memory use in FFT text box. Larger values will result in faster calculations, but this value should not exceed the available physical memory. The default value is a safe choice for all but the most limited of computer hardware.

5.4.9 Real-Space Options

There are a number of options that apply only to real-space refinement tasks.

For density maps, you can choose to use the phases from the model, or make use of phased reflection data. The options are:

- Observed amplitudes and phases calculated from model—phases are calculated from the model, using the observed amplitudes. This is the default option.
- Amplitudes and phases from file—phases are obtained from a file that contains phased reflections. When you select this option, the Browse button becomes available, and you can enter the file name in the text box or click Browse and navigate to the file in a file chooser.

Real-space refinements use a weighted combination of force-field terms and X-ray terms as the function to be minimized. PrimeX determines a fairly optimal weighting of these two components. However, you might want to weight one or the other more heavily. You can specify the relative weight of the X-ray terms in the Weighting factor text box.

5.4.10 Map Colors

To change the default map colors, click Set Default Map Colors. You can then choose colors for each of the map types in the Set Default Map Colors dialog box. Default map colors can also be set in the Preferences panel, in the Surfaces – Creation section. Colors for individual maps can be changed from the Manage Surfaces panel.

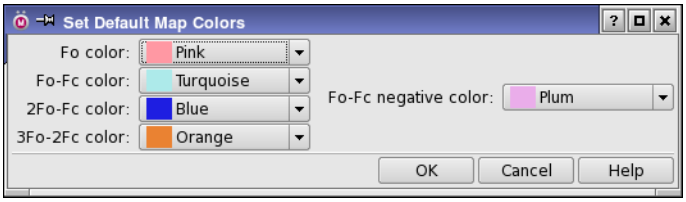


Figure 5.4. The Set Default Map Colors dialog box.

5.5 Initial Structure and Sequence Display

When you click OK in the Input Data dialog box, the initial structure is displayed in the Workspace. If you have the sequence viewer displayed in the main window, the sequence of the initial structure is displayed in the sequence viewer. The Project Table panel opens with the initial structure selected.

The reference sequence is not displayed by default. To view the reference sequence, click the Reference Sequence button in the PrimeX panel. The Reference Sequence panel opens, and displays the sequence. Each chain is listed separately, and the number of residues in the chain is given. You can find a residue or a sequence of residues by entering the sequence in the Find text box and clicking Previous or Next.

5.6 Preparing the Protein Structure

The basic preparation of the protein structure (assignment of bond orders and metal treatment) are performed on import when you select Assign bond orders. If you did not select this option when you imported the structure, you can assign bond orders by choosing Assign Bond Orders from the Tools menu. Metal treatment is always done on import.

You can also do the preparation in the Protein Preparation Wizard panel, using the following procedure with the structure displayed in the Workspace.

1. Choose Applications → Protein Preparation Wizard or Tasks → Protein Preparation Wizard.

   The Protein Preparation Wizard panel opens.

2. In the Import and Process tab, ensure that only Assign bond orders is selected.

   Hydrogens will be added later with a more sophisticated algorithm.

3. Click Preprocess.

   The selected tasks are performed, and a new entry is created for the fixed structure, which is displayed in the Workspace.

4. Click Close to close the Protein Preparation Wizard panel.

Reciprocal-Space Refinement

Often, the first step in a refinement is the overall positioning of the protein, which is done by treating the protein as a rigid body. Simulated annealing is generally used in the middle stages of the refinement, after the larger changes in positions have been completed with rigid-body and real-space refinement. Reciprocal-space minimization is useful at all stages of refinement, and is often used in the final stages of the refinement.

In PrimeX, you can perform all of these tasks in the Reciprocal-Space Refinement panel. To open the panel, click the button for the task in the Reciprocal-Space Refinement section of the PrimeX panel.

- Minimization
- Rigid Bodies
- Simulated Annealing

After each refinement job is done, maps are automatically generated. You can select the type of map and set up the parameters of the map in the Maps tab.

The panel has four tabs, which contain controls for the refinement. Three of these are common to all refinement methods, and are described in later sections. The fourth tab changes according to the method chosen.

Below the tabs is a Calculation Settings button, which opens the Calculation Settings dialog box. This dialog box is described in Section 5.4 on page 39.

At the bottom of the panel is the Job toolbar (see Section 1.3 on page 3 for details). The Settings button opens a Job Settings dialog box, in which you can make job settings and start the refinement job. To start the job with the current job settings, click Run.

6.1 General Setup

In the General tab, you specify the target for refinement and the relative weighting of X-ray and energy terms in the target function, and choose how to set the isotropic B factors.

There are two choices of refinement target: Maximum likelihood and Least squares. If you choose Least squares, the Unweighted option in the Weighting options group of the Maps tab is selected and the controls are then disabled so that you cannot change the option. This is done because you cannot create a weighted map with the least squares target. These options are enabled again if you choose Maximum likelihood.

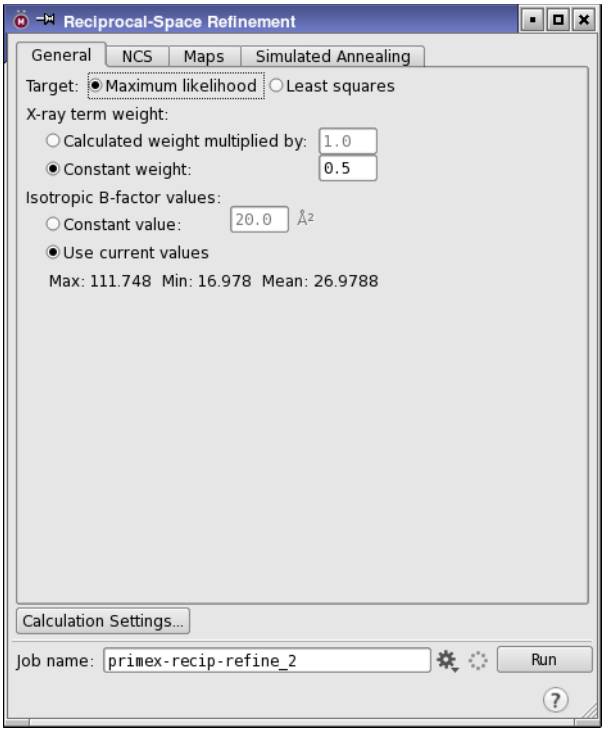


Figure 6.1. The General tab in the Reciprocal-Space Refinement panel.

There are two options for setting the isotropic B-factors: Use current values and Set to constant value of N Å^2. The current maximum, minimum, and mean values are listed below the option.

The two X-ray term weight options provide a way of controlling the weighting of the gradient of the electron density fit relative to the energy gradient in the refinement. The options are:

- Calculated weight multiplied by—The optimal weight calculated by PrimeX can be multiplied by a factor that weights the X-ray terms (density fit). A value of 1.0 selects the calculated weight.
- Constant weight—Specify the weight of the X-ray terms.

Weighting of the X-ray terms against geometric restraints is a complicated balancing act. Choices must be made based on the stage of the refinement process, the deviation of molecular geometry from expected values, and the current values for R and R-free. The complexity of these choices in PrimeX can be greatly reduced with the application of the following guidelines.

The Constant weight option with a value of 2.0 is usually an appropriate starting point for reciprocal-space minimization and simulated annealing refinement for most molecular replacement cases. If the divergence between R and R-free is becoming unacceptably high, or if the molecular geometry deviates too far from expected values, then this value should be decreased to 1 or lower. Values less than 0.1 are rarely productive. Insufficient progress in refinement can sometimes be improved with a larger constant weight. Also, if the values for R and R-free both increase, a higher weight is indicated. The weight will not need to be adjusted very often once an appropriate value is found.

Although PrimeX estimates an appropriate value for weighting the X-ray term upon selection of the Calculated weight option, this value tends to be conservatively low, especially for early stages of refinement. The same considerations described above apply to adjusting the multiplier with this option.

6.2 Setting Up Maps

The Maps tab provides options for defining the parameters and extent of the maps that will be created with a refinement job. These maps are created automatically after the refinement is completed. The options are synchronized with those in the Create Map panel.

Four choices are provided for the coefficients to be used in the map: Fo, Fo-Fc (difference map), 2Fo-Fc, and 3Fo-2Fc. You can select any number of these options to create multiple maps.

If you are using the maximum likelihood target, you can specify the weighting model used. The two options are SigmaA (see [1]) and Unweighted. If you are using least squares as the target, you can only create an unweighted map.

Two preset options for the grid spacing are provided: 0.33 Å and 0.25 Å. If you want to specify your own grid spacing, select Other and enter a value in the text box.

You can scale the maps to use units of the standard deviation of the electron density, rather than the default, by selecting Scale map to sigma units. The units of the default map depends on the units of Fobs.

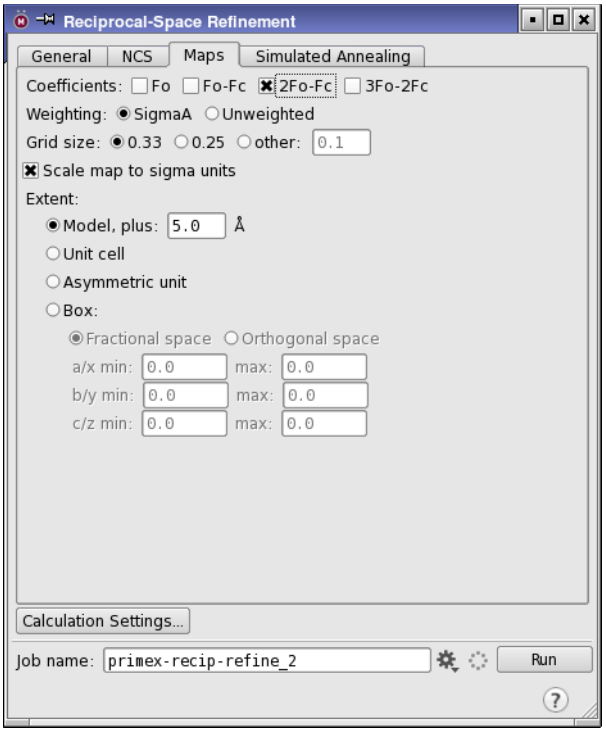


Figure 6.2. The Maps tab in the Reciprocal-Space Refinement panel.

The Extent options allow you to specify the extent of the map. There are four options:

- Molecule, plus *N* Å—Limit the map to the region within the specified distance of the selected molecule.
- Unit cell—Calculate the map for the entire unit cell.
- Asymmetric unit—Calculate the map for the asymmetric unit.
- Box—Calculate the map for the box defined in the text boxes. The values in the text boxes are given in terms of the space selected:
  - Fractional space—Define the map as a fraction of the unit cell parameters. You can enter the fractional values in the min and max text boxes for each crystallographic direction.
  - Orthogonal space—Define the extent of the map in Cartesian (*x*, *y*, *z*) coordinates. The coordinate origin is that of the structure in the Workspace.

6.3 Using Noncrystallographic Symmetry Restraints

In the NCS tab, you can define restraints on the optimization to satisfy noncrystallographic symmetry.

Noncrystallographic symmetry (NCS) is imposed at the chain level. To impose this symmetry, you create groups of chains or chain segments that are related by NCS. The parts of the chains that are considered to be related must have identical atom composition, so you must omit residues that differ between the chains, including differences in protonation or in residue name. The terminal residues of each chain are unique in PrimeX, and must be excluded from the restraint system. Restraints can be applied optionally to just the backbone as well as to whole residues.

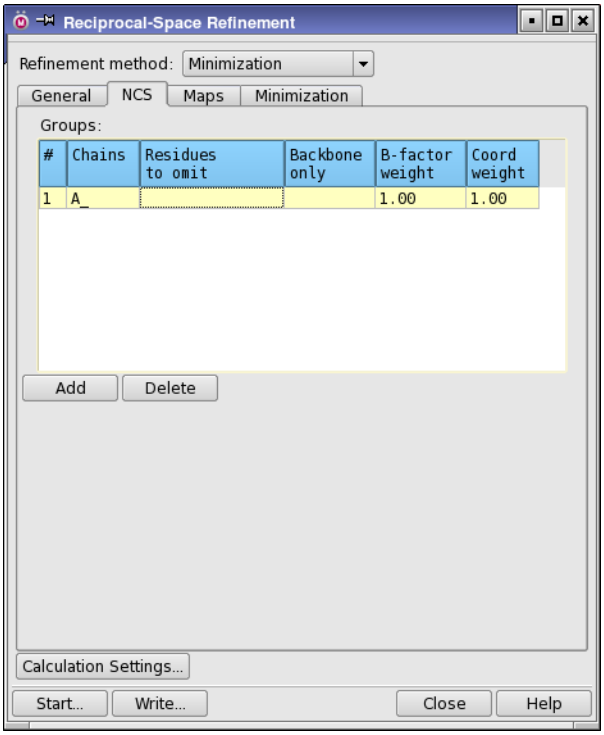


Figure 6.3. The NCS tab of the Reciprocal-Space Refinement panel.

To set up a chain group:

1. Click Add.

   A new group is added to the Groups table. If groups have been defined previously, the group is initialized with the chains that have not yet been used. If all chains have been used in previous groups, the group is initialized with all chains, so that you can make other connections between chains than have already been defined.

2. Edit the Chains cell in the table to include only the desired chains.
3. Add as many groups as are needed to define the symmetry-related chains or segments.
4. Check whether the chains or segments in each group have identical residues.

   You can use the sequence viewer in the Workspace to compare the chains, and you can click on residues to select them in the Workspace. You should check that the residues have the same atoms: if they do not, they must be omitted from the group.

5. List the residues to be excluded from the group in the Residues to omit cell of the table.

   The list is a comma-separated list of residue numbers (including insertion codes) or ranges, such as 223-226a.

6. Click the box in the BB only column if you want to apply restraints only to the backbone.
7. Set weights of the restraints for the B-factors and coordinates if desired.

Table 6.1. Groups table columns

Column	Description
#	Group number
Chains	Chains that are included in the group. The chains need not be identical, but if they are not, you must omit the residues that differ.
Residues to omit	Comma-separated list of residue numbers or ranges (such as 12-=128) to be omitted from the specified chains. If the chains are not identical, you must omit residues that differ between chains or are missing from some chains in the group.
Backbone only	If the box is checked, apply restraints to the backbone only, not to the side chains.
B-factor weight	Weight of the restraint on the B-factors
Coord weight	Weight of the restraint on the coordinates

6.4 Minimization

PrimeX allows you to minimize the R factors for the coordinates, the B-factors, grouped B-factors, and grouped occupancies. Settings for a minimization are made in the Minimization tab.

Three minimizers are offered in the Minimizer option menu, Quasi-Newton (LBFGS), Truncated Newton, and Conjugate gradient. A fourth choice, Optimal, selects the conjugate gradient minimizer for the calculation if the initial gradients are large, otherwise it selects the quasi-Newton method. Minimizations are performed in cycles, each consisting of a number of minimization steps. At the beginning of each cycle, second derivatives and scaling factors are computed. Cycling reduces the chance of accumulated error building up to the point at which it affects the results. You can specify the maximum number of cycles and the maximum number of steps per cycle to take in the Number of cycles and Maximum steps per cycle text boxes. The defaults are 3 cycles and 8 steps per cycle.

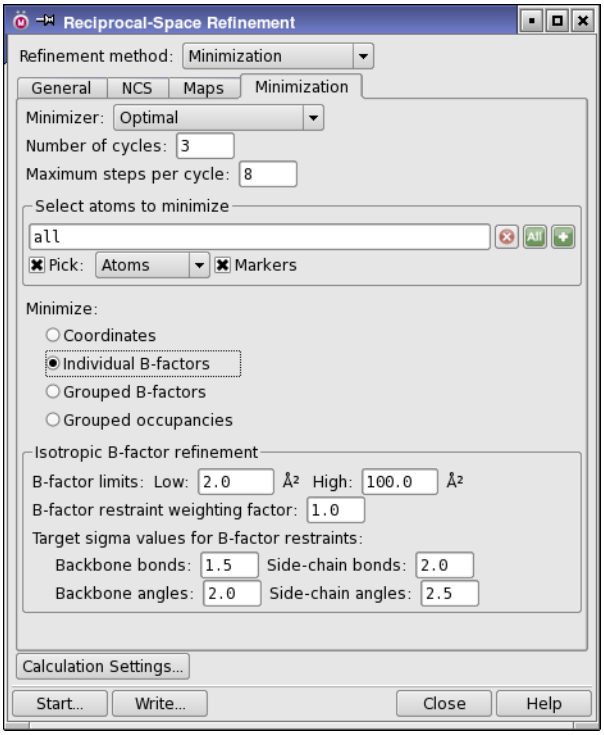


Figure 6.4. The Minimization tab of the Reciprocal-Space Refinement panel.

Minimizations can be performed on selected atoms. To choose which atoms to minimize, you use the standard picking controls in the Select atoms to minimize section. These controls are described in Chapter 6 of the *Maestro User Manual*. The default is to minimize all atoms. If you have calculated a density fit (with the Density Fit panel—see Section 12.1 on page 83), you can select residues whose density fit exceeds a certain threshold, by clicking Select and using the Atom tab of the Atom Selection dialog box.

Next, you need to choose the quantity whose error (represented by the target specified in the General tab) is to be minimized. You can choose only one of the Minimize options for a given minimization job. The options are:

- Coordinates—Minimize the error in the coordinates.
- B-factors—Minimize the error in the isotropic B-factors for each atom. B-factor minimization is usually performed with anisotropic overall B-factor scaling.
- Grouped B-factors—Minimize the error in the isotropic B-factors for groups of atoms. Each residue consists of two groups, the side chain atoms and the backbone atoms (with the exception of glycine, which has only one group). A common B factor is used for the atoms within a group. This approach allows you to refine the B-factors at low resolution where otherwise the number of variables being optimized might exceed the number of data points.
- Grouped occupancies—Minimize the error by optimizing the fractional occupancies of groups of atoms. You must select the atoms for this task, in a single occupancy group, in the Select atoms to minimize section. The atoms with partial occupancy can be alternate conformations, or a ligand or other molecule that is not present in saturating amounts. In the first case, the partial occupancies must sum to one; in the second the partial occupancy must be less than one.

  Occupancy refinement should be applied only after the atomic positions and B-values have been thoroughly refined, near the end of the refinement process. Reciprocal-space refinement refines the coordinates of both of a pair of alternate conformations. Note also that only at high resolution is enough information available for refinement to converge at independent values for both B-factors and occupancy.

If you choose to refine B-factors, you can set parameters that control the refinement in the Isotropic B-factor refinement section. The parameters are:

- B-factor limits—Specify the smallest and largest permissible B-factor values in the Low and High text boxes.
- B-factor restraint weighting factor—Specify the weight of the restraint of B-factor values to the target sigma values.

- Target sigma values for B-factor restraints—Specify the target standard deviation of the B-factors for geometric parameters in each residue. The restraints are applied to each residue separately. The four geometric parameter types are Backbone bonds, Backbone angles, Side-chain bonds, and Side-chain angles.

There are also some settings that you should make in the General tab: Use current values should be selected under Isotropic B-factor values, and the X-ray term weight setting should be set to Constant weight, with a value of 1.0. With this weight for the X-ray term and a B-factor restraint weight of 1.0, the restraints on the B-factors will be very strong. This setting is appropriate for the beginning of individual isotropic B-factor refinement of a moderate-resolution structure. Approaching the end of the refinement, you might reduce this restraint weight by factors of 2 or more, until R-free no longer decreases.

6.5 Rigid-Body Refinement

Often, the first step in a refinement is the overall positioning of the protein, which is done by treating the protein as a rigid body. PrimeX provides the means to define multiple rigid bodies for the refinement.

To set up a rigid-body refinement, choose Rigid body from the Refinement method option menu. When you do, the Rigid Bodies tab is added to the panel. In this tab you can define one or more rigid bodies for rigid-body refinement. Each rigid body is defined by a group of atoms, which you can select at will. You can also decide what to do with atoms that are not part of any explicitly chosen rigid body.

Three minimizers are offered in the Minimizer option menu, Quasi-Newton (LBFGS), Truncated Newton, and Conjugate gradient. A fourth choice, Optimal, selects the conjugate gradient minimizer for the calculation if the initial gradients are large, otherwise it selects the truncated Newton method. The truncated Newton minimizer is recommended for rigid-body refinement.

Minimizations are performed in cycles, each consisting of a number of minimization steps. At the beginning of each cycle, second derivatives and scaling factors are computed. Cycling reduces the chance of accumulated error building up to the point at which it affects the results. You can specify the maximum number of cycles and the maximum number of steps per cycle to take in the Number of cycles and Maximum steps per cycle text boxes. The defaults are 3 cycles and 8 steps per cycle.

The main task is to define the rigid bodies, which you do in the Define rigid body group section. In this section you can define the content of a rigid body group, using the standard Maestro picking controls (see Chapter 6 of the *Maestro User Manual*). You can pick objects in the Workspace, or use the Atom Selection dialog box (opened by the Select button) to define the atoms in the rigid body group.

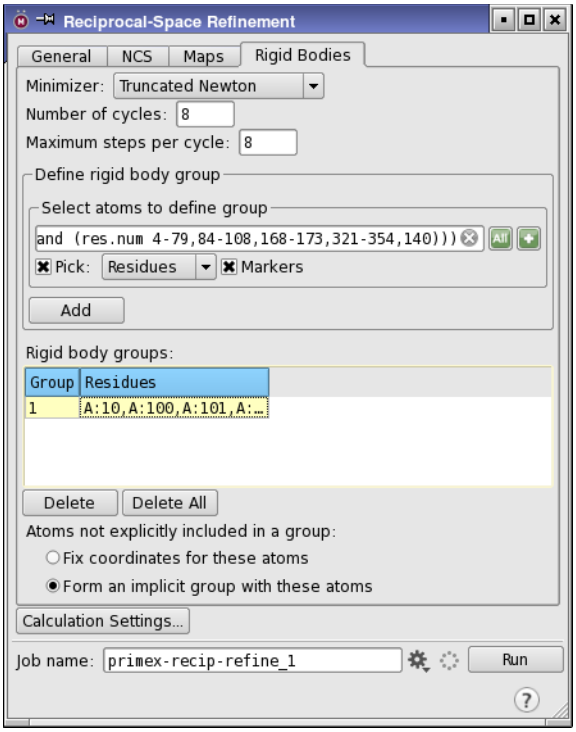


Figure 6.5. The Rigid Bodies tab of the Reciprocal-Space Refinement panel.

When you have made your selection, click Add to create a rigid body group with the selected atoms. The new group is then listed in the Rigid body groups table. This table lists the rigid body groups that have been defined. The Group column lists the index of the group, and the Residues column lists the residues that belong to the group.

To delete groups, select the groups and click Delete. To delete all groups, click Delete All.

Once you have defined the groups, you must then decide what to do with the atoms that are not explicitly included in a group. There are two choices:

- Fix coordinates for these atoms
- Form an implicit group with these atoms

Rigid-body refinement is usually performed with only the low-resolution data. Before you submit the job, you might therefore want to increase the high-resolution cutoff for the reflection data in the Calculation Settings dialog box.

6.6 Simulated Annealing

Simulated annealing is generally used after the larger changes in positions have been completed with rigid-body refinement. Usually the initial temperature is decreased with successive runs. This method is useful for moving across low energy barriers.

Simulated annealing can be performed on selected atoms. To choose the atoms, use the standard picking controls in the Select atoms to minimize section. These controls are described in Chapter 6 of the *Maestro User Manual*. If you have calculated a density fit (with the Density Fit panel—see Section 12.1 on page 83), you can select residues whose density fit exceeds a certain threshold, by clicking Select and using the Atom tab of the Atom Selection dialog box. The default is to minimize all atoms.

The molecular mechanics energies computed in the simulated annealing can be approximated, or computed exactly. These choices are available as Energy model options:

- Approximate—Neglect solvation and electrostatic terms. Runs faster, and useful for early stages of refinement.
- Complete—Use all molecular mechanics terms. Slower, but safer for later stages of refinement, and more appropriate if you wish to use simulated annealing to help improve hydrogen positions.

The parameters that can be set for the simulated annealing process are:

- Initial minimization steps—Number of steps to take in the initial minimization (at the initial temperature).
- MD energy scale estimation steps—Number of steps to use in estimating the MD energy scale.
- Temperatures—The initial temperature, the high temperature for the start of the annealing process, and the final temperature. The high temperature should be at least 500 K for sampling of conformations outside the local minimum.
- Heating steps—Number of steps used to heat the system to the high temperature.
- Cooling steps—Number of steps used to cool the system from the high temperature to the final temperature.
- Final minimization steps—Number of steps used for the final minimization.

Finally, you can set the MD time step in ps and the cutoff for nonbonded interactions in angstroms in the Molecular dynamics section.

When you have made all the desired settings, click Start to make job settings and start the job.

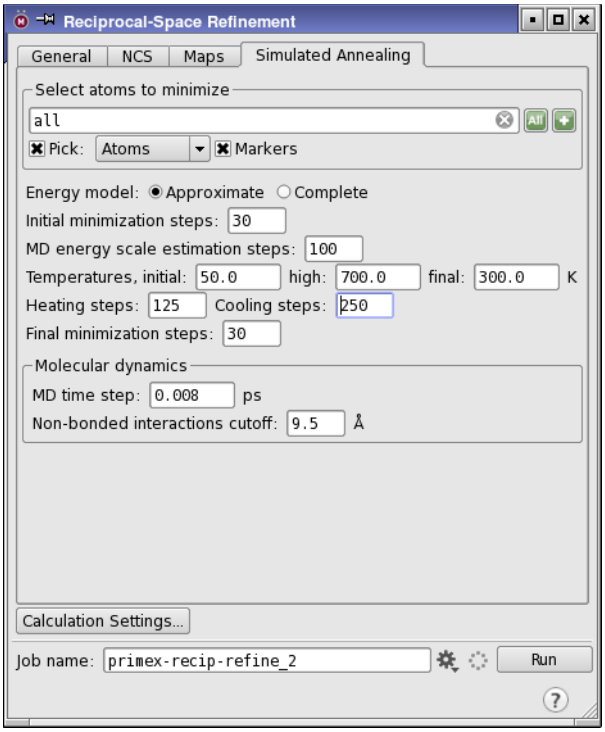


Figure 6.6. The Simulated Annealing tab of the Reciprocal-Space Refinement panel.

When a simulated annealing job fails with the error message "Too many unreasonably large velocities detected", the cause of the failure can be that the X-ray weight is too large. This problem is most likely to be seen with the approximate energy model. The error message suggests other possible solutions: minimizing the starting structure or reducing the MD time step. Early in the refinement of a structure, either reducing the X-ray weight or performing a minimization is probably the most useful solution. Later in the refinement, reducing the X-ray weight or decreasing the time step is probably the most useful solution. Reducing the time step to 0.006 usually fixes the problem.

6.7 Summary of Molecular Geometry

During post-processing after every reciprocal space refinement job, four statistics are calculated that summarize the quality of the molecular geometry. These quantities are:

- Steric clashes: The number of interactions of non-hydrogen atoms where the distance between atom centers is less than 0.85 of the sum of the atoms' van der Waals radii. If the two atoms are involved in a hydrogen-bonding interaction, close contacts to within 0.75 of the sum are allowed.
- Bond Lengths: The rms deviation from the ideal bond lengths for proteins as tabulated by Engh & Huber [11].
- Bond Angles: The rms deviation from the ideal bond angles for proteins as tabulated by Engh & Huber [11].
- Improper torsions: The rms deviation from the target improper torsion angles required for a set of perfectly planar side chain groups.

See also the section on the Protein Reports panel, Section 12.2 on page 84.

Real-Space Refinement

Real-space refinement in PrimeX makes use of Prime technology with OPLS force fields for minimization, loop building and side-chain placement, but with the addition of X-ray terms to the force field terms in the function to be optimized. The balance between the two can be set in the Calculation Settings dialog box, which is described in Section 5.4 on page 39. For information on Prime methodology, see the *Prime User Manual*.

Real-space refinement calculations are performed from the Real-Space Refinement panel To open the panel, click the task button for the relevant task in the Real-Space Refinement section of the PrimeX panel.

There are three choices of real-space refinement tasks. For each task, a different set of controls is displayed in the panel. These sets of controls are described in the following sections. The three tasks are:

- Refine loops
- Predict side chains
- Minimize

These tasks are described in the following sections of this chapter.

In addition to the task-specific controls, there is a Grid spacing text box in which you can specify the grid spacing (in angstroms) for the electron density map used in the refinement, and an option to request the setting of atom B-factors, with a text box for the value. These two items are included under the Manual build mode for loop refinement. Setting the B-factors to a low value improves the convergence of real-space refinements if the refined isotropic B-factors are high, which can happen if the B-factors of the atoms were refined when the coordinates contained substantial error.

The Calculation Settings dialog box also has settings for the source of the map, which can be taken from a phased reflections file, and for the B-factors of the atoms to be fit in the real-space refinement. These settings are described in Section 5.4.9 on page 43.

7.1 Loop Refinement

The controls for loop refinement consist of a set of options for setting the grid spacing and B-factors, a table that lists the loops, text boxes for the loop limits, a button for loading the loops table from the Workspace structure and an option to bias the loop prediction towards the existing structure. Only one loop can be operated on for each loop refinement job run.

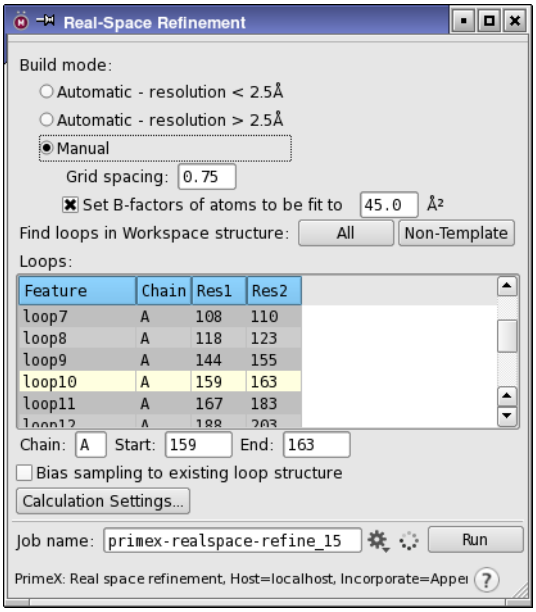


Figure 7.1. The Real-Space Refinement panel showing loop refinement controls.

The Build mode options provide automatic setting of the grid spacing and B-factors for the atoms to be refined.

- Automatic - resolution < 2.5Å—Select this option for automatic setting of the grid spacing and B-factors if the high-resolution limit is lower than 2.5 Å. The grid spacing is set to 1.0 and the B-factor for the atoms to be fit is set to 5.0.
- Automatic - resolution > 2.5Å—Select this option for automatic setting of the grid spacing and B-factors if the high-resolution limit is higher than 2.5 Å. The grid spacing is set to 0.75 and the B-factor for the atoms to be fit is set to 45.0.
- Manual—Set the grid spacing and B-factors manually. The Grid spacing text box and Set B-factors of atoms to be fit to *N* Å option and text box become available when you select this option.

The Loops table lists all the loops, along with the starting and ending residue numbers for each loop. The table columns are described in Table 7.1. To choose the loop you want to refine, select the table row for the loop. The chain name, and the starting and ending residue numbers are loaded into the Chain, Start, and End text boxes. To change the limits of a loop, change the residue numbers in the Start and End text boxes. If the structure in the Workspace changes, you can reanalyze the loops by clicking Load from Workspace.

Table 7.1. Columns in the Loops table

Column	Description
Feature	Loop label
Chain	Chain name.
Res1	First residue in loop.
Res2	Last residue in loop.

Loops whose residues are missing from the structure are not listed in the table. To predict the structure of such a loop, you can select any loop and change the starting and ending residue numbers to span the loop you are interested in. For an example of how to handle such a loop, see Section 2.15 of the *PrimeX Quick Start Guide*.

The Bias sampling to existing loop structure option allows you to sample conformational space more heavily around the existing structure, with the result that the refined loop is more likely to resemble the input loop.

Note: To run a loop refinement, the model sequence must match the reference sequence. If this is not the case, you should mutate the model first (see Section 11.1 on page 77).

7.2 Side-Chain Prediction

Side-chain prediction is a necessary follow-up to mutation of the model to match the input sequence (which you must do before you can run a side-chain prediction). The mutation is performed automatically by the Mutate Model to Sequence task. To perform this task, first click Mutate Model to Sequence in the PrimeX panel, then in the Mutate Model to Sequence panel, click Mutate to Match Sequence. You can then proceed to side-chain prediction by clicking Place Side Chains, which opens the Real-Space Refinement panel with Predict side chains tools displayed.

You can also perform residue mutation manually, using the procedure described on page 21, and adjust the side chain manually, using the Rotamers dialog box as described on page 16. However, this manual procedure only selects certain dihedrals, whereas the side-chain prediction varies the dihedrals in the side chain to find the best fit.

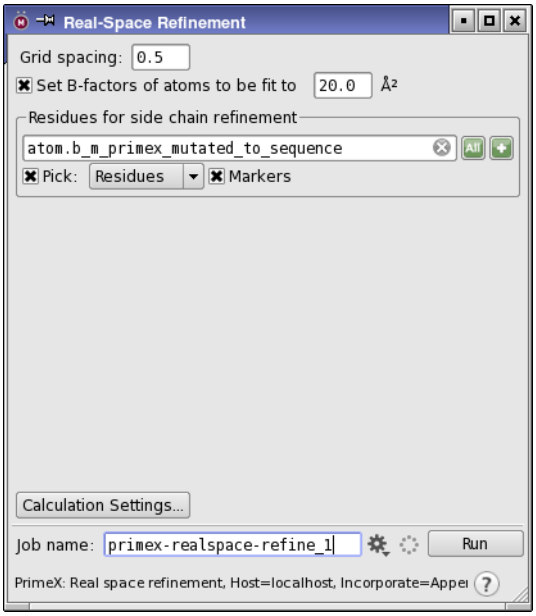


Figure 7.2. The Real-Space Refinement panel showing side-chain prediction controls.

Side-chain prediction may be necessary after loop building, since the building process for the main chain does not comprehensively optimize the positions of the side chains. In addition, it can be useful at any other time during refinement to help improve side-chain placement in electron density.

The side-chain prediction task places all the selected side chains in a single run: there is no need to optimize them independently. Side chains are placed using a single omit map. Omission of more than 5% of the atoms in the map calculation can reduce the quality of the map and hence the accuracy of the side-chain placement. Thus, multiple side chains should be optimized at the same time, but selection of a large number of side chains carries some risk.

The controls for the prediction of side chains consist of a set of standard picking controls, with which you can pick the residues whose side chains are to be refined. If you open this panel from the Mutate Model to Sequence panel, the residues that were mutated are selected already, and you only need to start the job. If you have calculated a density fit (with the Density Fit panel—see Section 12.1 on page 83), you can select side chains for refinement by clicking in the Refine column for the desired residues, then clicking Refine Selected Sidechains to open this panel with the residues already selected. You can also select residues whose side-chain real-space R factor exceeds a certain threshold, by clicking Select and using the Atom tab of the Atom Selection dialog box.

7.3 Minimization

Real-space minimization performs a standard geometric variation of the coordinates to minimize the energy, or in PrimeX, the functional that includes both the force-field energy and the density fit.

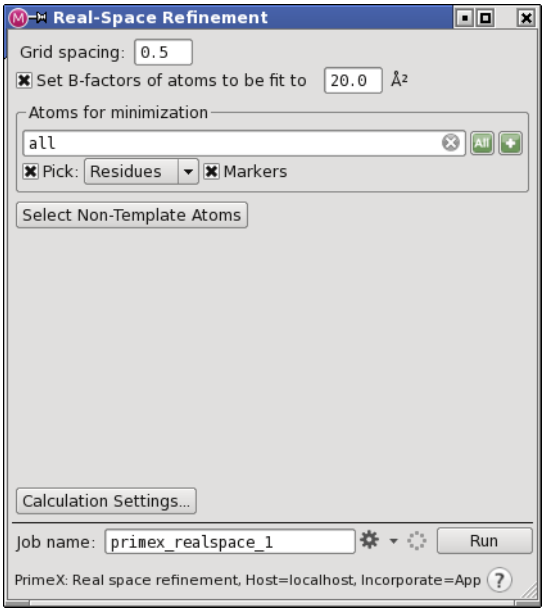


Figure 7.3. The Real-Space Refinement panel showing minimization controls.

If you have alternate conformations in your structure, you can only perform a minimization on one of the conformations at a time. Once you have minimized one, you can switch to the other and minimize it.

The controls for minimization consist of a set of standard picking controls, with which you can pick the residues to be refined. If you have calculated a density fit (with the Density Fit panel—see Section 12.1 on page 83), you can select residues whose density fit exceeds a certain threshold, by clicking Select and using the Atom tab of the Atom Selection dialog box.

Real-space minimizations can sometimes result in nonplanar structures for groups that are nominally planar, such as peptide links, other amides, and unsaturated ring systems. The nonplanarity may appear to be large, but is usually not more than a few degrees. These small deviations from planarity may arise because of interactions with other parts of the structure or noise in the X-ray data. Other refinement programs enforce planarity for these groups quite strongly. Strictly enforcing planarity is in fact unphysical: these groups are not rigid, but the energetic cost for deviating from planarity is fairly high. PrimeX allows small deviations from

planarity in an effort to provide the physically correct degree of restraint. Further refinement in PrimeX can sometimes alleviate the unfavorable interactions that resulted in the nonplanarity and return the groups to as near planar as is physically reasonable. If the cause of the deviation from planarity is due to noise in the X-ray data, decreasing the weight on the data relative to the restraining force fields will improve the results.

If you want to enforce planarity more strictly than is done by the force field, you can use the Planar group restraints options in the Calculation Settings dialog box—see Section 5.4.7 on page 42 for details.

Placing Ligand and Solvent Molecules

Once the protein is reasonably well refined, you can proceed to place the ligand, solvent molecules, and water molecules.

8.1 Placing the Ligand and Solvent Molecules

Ligand and solvent molecule placement works by docking the molecule with Glide to a protein site that is defined by a region of electron density. This procedure is especially useful if the ligand or the solvent molecule can adopt more than one conformation. Glide performs a rapid conformational search of the most likely conformations, and docks each conformation using SP docking mode. The results are determined by both the standard energetic terms from Glide and the fit to the density, with the fit to the density more highly weighted by default. The best-scoring conformations are returned.

The density regions are identified by analyzing a difference density map, which must be generated and displayed before you start (see Section 10.1 on page 73). Connected density peaks are combined into density "blobs" and listed in a table. You can then select the density blob you want to fit to and the molecules or ligands you want to fit into the region defined by the blob, and run the job. The blob is used to define the Glide enclosing box and also the map that is used to fit the ligand. The centroid of the blob is the centroid of the box.

You can place multiple molecules in the same density blob, to see which of them fits best. When the job finishes, the best poses for all molecules are returned as separate entries in the Project Table.

Setting up a ligand or solvent molecule placement job is done in the Ligand/Solvent Placement panel. Placement of water molecules is done separately, in the Water Placement panel.

To open the Ligand/Solvent Placement panel, click the Ligand/Solvent Placement button in the PrimeX panel.

To place a ligand or solvent molecule:

1. Generate a difference density map (Fo-Fc), and display it in the Workspace along with the structure.
2. Open the Ligand/Solvent Placement panel.

   If you do not have a difference density map displayed when you open the panel, you can generate one, then click Update in the panel to analyze the peaks and locate the blobs.

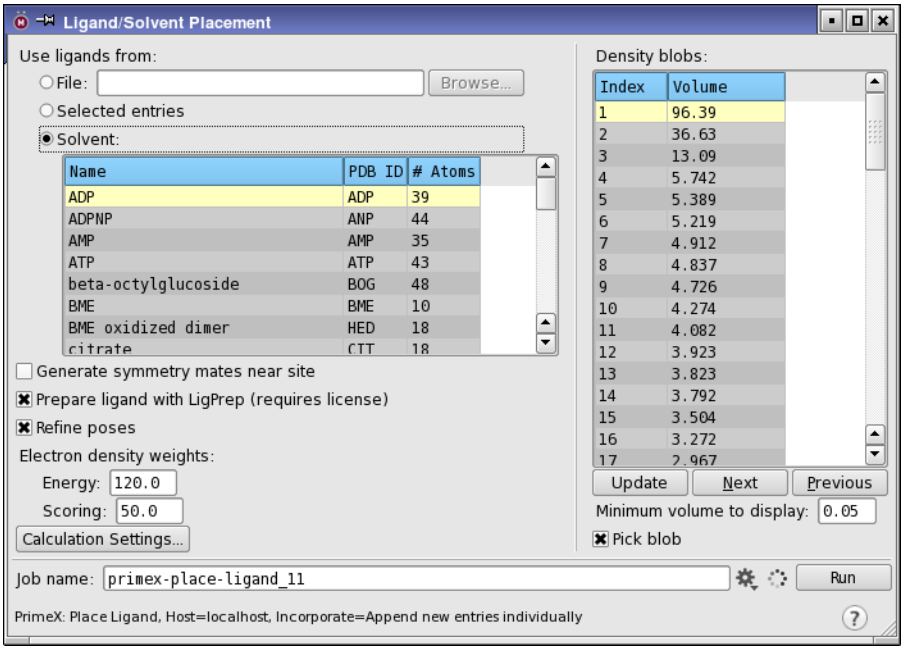


Figure 8.1. The Ligand/Solvent Placement panel.

3. Select the source of the ligands, under Use ligands from.
   - If you selected File, enter the name of the Maestro file that contains the ligands in the File name text box, or click Browse and navigate to the file.
   - If you selected Selected entries, ensure that the desired entries are selected in the Project Table.
   - If you selected Solvent, select the desired solvent molecules in the table.
4. Select the density blob you want to fit the molecule to.

   You can select the blob in the Density blobs table or click on the blob in the Workspace. Blobs are marked with green colored spheres; the selected blob is colored red. To use the Workspace, the Pick blob option must be selected. The minimum volume for the blobs that are listed in the table can be changed in the Minimum volume to display text box.

   A box is drawn around the selected blob, which is the enclosing box for Glide, and is also used to generate a map for Glide docking.

   If there are several adjacent blobs that might contribute to the ligand, choose the one that is closest to the center of the desired ligand location. Sometimes decreasing the contour level of the maps (see Section 10.2 on page 75) can join two blobs into one. After making

this adjustment, click Update in the Ligand/Solvent Placement panel to obtain the new blob values.

5. Select options for the calculation.

   These include ligand preparation, generation of symmetry mates near the site, and refinement of poses after docking.

6. Click Start.

   The Start dialog box opens, in which you can make job settings and start the job.

The job returns a number of "poses" (orientation and position with respect to the protein site) of the placed molecules. The number and ranking of poses returned depends on the Refine Poses option (see below).

After placing molecules in one location and refining the results, you can generate a new difference map and place molecules in another location. Maps can be generated using the Create Map panel (see Section 10.1 on page 73).

The various controls in the panel are described below, with some details of their operation.

Use ligands from options

Select one of these options to specify the source of the ligands to be placed in the selected density blob. The three options are:

- File—Select this option to use ligands from a Maestro file. You can enter the name of the file in the File name text box, or click Browse and navigate to the file.
- Selected entries—Use the entries that are selected in the Project Table for the ligands.
- Solvent—Use the solvent molecules that are selected in the table below this option for the ligands. The table lists common solvent molecules. The columns display the name, PDB ID, and the number of atoms in the molecule.

Generate symmetry mates near site option

Select this option to generate crystal mates near the active site. This option is useful if symmetry-related atoms that are part of another protein monomer are close to the active site, and could influence the ligand placement. These symmetry-related atoms are then included in the grid generation for the Glide docking job.

Prepare ligand with LigPrep option

Select this option to run LigPrep on the ligand before placing it. The LigPrep job is run as the first part of the ligand placement job. You must have a LigPrep and an Epik license to use this option. For more information on LigPrep, see the *LigPrep User Manual*. The preparation includes generation of ionization and tautomeric states, with retention of the original tautomer and a maximum of 4 structures generated.

Refine poses option

Select this option to run a post-docking real-space minimization of the ligand in the field of the receptor. This option improves the geometry of the docked ligand and the fit to the electron density. When selected, three additional properties related to the real-space refinement are returned to the Project Table:

- Real-space R-factor.
- Coverage—This statistic presents the percent of atoms in the ligand covered by electron density when a Fo-Fc map is contoured at 3 sigma.
- LigandScore—This property is calculated from the other two values, and provides a way of comparing the electron density fit of two different ligands in the same site.

All likely poses are ranked by the real-space R-factor and then the five molecule poses with the best ranking are returned, with no consideration of whether they are the same or different molecules from the input list.

With the Refine poses option cleared, a variable number of poses are returned, with at least eight poses for every molecule in the input list. Two properties are returned to the Project Table from the ligand/solvent placement job, labeled glide gscore and glide denscore. The first of these is the GlideScore, which is a measure of the ligand binding affinity. The second is a parameter that measures the fit to the density, which is calculated from the sum of the electron density over all atoms in the ligand. In both cases, lower (more negative) values indicate a better fit. The results are ranked and sorted in the Project Table by the denscore.

Electron density weights text boxes

These text boxes specify the weight of the density terms used to mix into the Coulomb/van der Waals energy (Energy) and into the rough scoring function (Score). The default values are 120 for Energy and 50 for Score (arbitrary units); allowed values range from 1 to 9999. A weight of 10 produces results that weight the density fit and the molecular interaction terms approximately equally. Thus, the larger default values have been optimized to produce poses typical of those found in the Protein Data Bank.

Calculation Settings button

This button opens the Calculation Settings dialog box, which is described in Section 5.4 on page 39. Of particular interest are the settings for the source of the map, which can be taken from a phased reflections file. These settings are described in Section 5.4.9 on page 43.

Density blobs table

This table lists the density blobs computed from the current map, in order of volume. The volume is displayed in the table. You can select only one blob at a time. The selected blob is used for solvent or ligand placement. A blob that is picked in the Workspace is also selected in the table. When you select a blob, the Workspace view centers on the blob, and the sphere that marks the blob is colored red.

Update button

Recompute the density blobs from the map and update the table. This button is useful if you generate a difference density map after opening the panel, for example, or make changes in the contour level.

Next and Previous buttons

Select the next blob or the previous blob in the table.

Minimum volume to display text box

The value in this text box sets the cutoff for the volume of the blobs that are displayed in the Density blobs table. Blobs with volumes smaller than the value set are not displayed.

Pick blob option

Select this option to pick a density blob in the Workspace, represented by a sphere. The selected blob is colored red; unselected blobs are green. When a blob is picked, it is also selected in the Density blobs table.

8.2 Placing Water Molecules

In the Water Placement panel you can set options and run a job to place water molecules. The options in this panel ensure that water molecules are placed in reasonable locations.

To open the Water Placement panel, click the Water Placement button in the PrimeX panel.

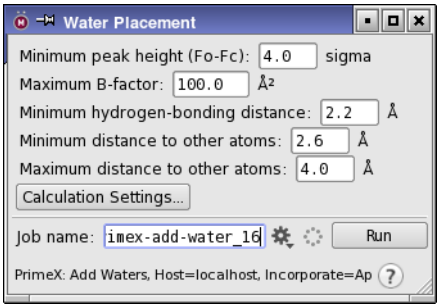


Figure 8.2. The Water Placement panel.

Water molecules are placed in all locations that match the criteria specified in the text boxes. When you have set these values, click Start to open the Start dialog box, in which you can make job settings and run the water placement job. You can also click Write to write the input file, and run the job from the command line. Note that the Calculation Settings dialog box has settings for the source of the map, which can be taken from a phased reflections file. These settings are described in Section 5.4.9 on page 43.

Once the job is run, it is a good idea to check the placements using the Density Fit panel

The text boxes in the panel are described below:

Minimum peak height (Fo-Fc) text box

Specify the minimum height of any peak that will be used to place a water molecule.

Maximum B-factor text box

Specify the maximum B-factor for any water molecule to be accepted as placed. Water molecules with larger B-factors are rejected as “noise”.

Minimum hydrogen-bonding distance text box

Specify the minimum distance between the oxygen atom and a hydrogen-bond donor or acceptor. Water molecules will not be placed in locations where this distance is less than the value specified.

Minimum distance to other atoms text box

Specify the minimum distance between the oxygen atom and any other atom. Water molecules will not be placed in locations where this distance is less than the value specified.

Maximum distance to other atoms text box

Specify the maximum distance between the oxygen atom and any other atom. Water molecules will not be placed in locations where this distance is greater than the value specified.

Partial Occupancy

PrimeX can perform refinements on systems with partial occupancies—for example, where only a fraction of the proteins in the crystal have bound ligands, or where ligands (and the nearby active site residues) occupy different conformations. The refinement of the occupancy values can be set up as a minimization calculation in the Reciprocal Space Refinement panel. The alternate conformations themselves must be set up using the Partial Occupancy dialog box. PrimeX supports multiple sites that have partial occupancy, but for each site there must be only two occupancies, which are labeled A and B.

To open the Partial Occupancy dialog box, choose Partial Occupancy from the Partial Occ button menu on the Proteins toolbar.

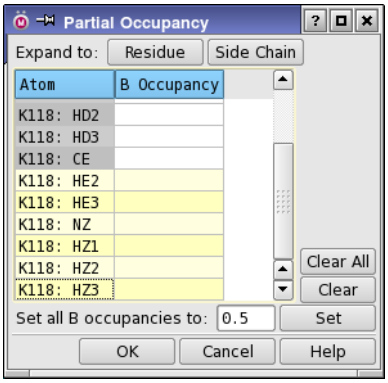


Figure 9.1. The Partial Occupancy dialog box.

To set partial occupancies for a set of atoms:

1. Select the atoms in the Workspace for which you want to assign partial occupancy.

   You can fill the selection out to complete residues or side chains in the dialog box.

2. Open the Partial Occupancy dialog box.

   The atoms you selected are listed in the occupancy table.

3. (Optional) Select rows in the occupancy table and click Residue or Side Chain to expand the atom list to the residues or side chains containing the selected atoms.

4. Enter an occupancy value in the Set all B occupancies to text box, and click Set.
5. Click OK.

You can repeat this procedure for as many sets of atoms as you like. If you have multiple sites where there are partial occupancies, you can select all the atoms at one time, and set the occupancies. You can also change the B occupancy value for individual atoms by editing the table cells. (The A occupancy is 1 minus the B occupancy.)

You can clear the occupancies for multiple atoms by selecting the atoms and clicking Clear, or you can clear all occupancies by clicking Clear All. When you close the dialog box, these atoms no longer have partial occupancy.

Once you have set the partial occupancies for a set of atoms, Maestro creates a duplicate set of coordinates for the B state for these atoms. Initially, these coordinates are the same as for the A state. By default, the A state is displayed (or "active") in the Workspace. To switch between the two states, select the atoms, then right-click and choose Switch Alternate Positions from the shortcut menu, or choose Switch Alternate Positions from the Partial Occupancy button menu. You can then select atoms in the B state and adjust their positions.

For example, if the atoms are part of a side chain, you can choose Rotamers from the Adjust button menu on the main toolbar

(or from the Adjust submenu of the Edit menu) and select the desired rotamer state. If you have a density map, you can click Choose Best Fit to select the rotamer that has the best fit to the density.

To display both positions at the same time, choose Alternate Positions from the Workspace menu or choose Display Alternate Positions from the Partial Occupancy button menu. Both states are displayed, but only one can be selected. The other is displayed as a "ghost".

If you want to move a set of atoms with a local transformation for which you already have alternate coordinates, you can either transform both sets of coordinates, or transform only the "main" set, by selecting or deselecting Include alternate positions in local transformations in the Workspace – Transformations tab of the Preferences panel.

Maps

10.1 Creating Maps

Maps can be created automatically at the end of reciprocal-space refinement jobs, but there may be other times when you want to create a map. In addition to regular maps, you may want to create omit maps. You can set up a job to create a map in the Create Maps panel.

To open the Create Map panel, click the Create Map button in the PrimeX panel.

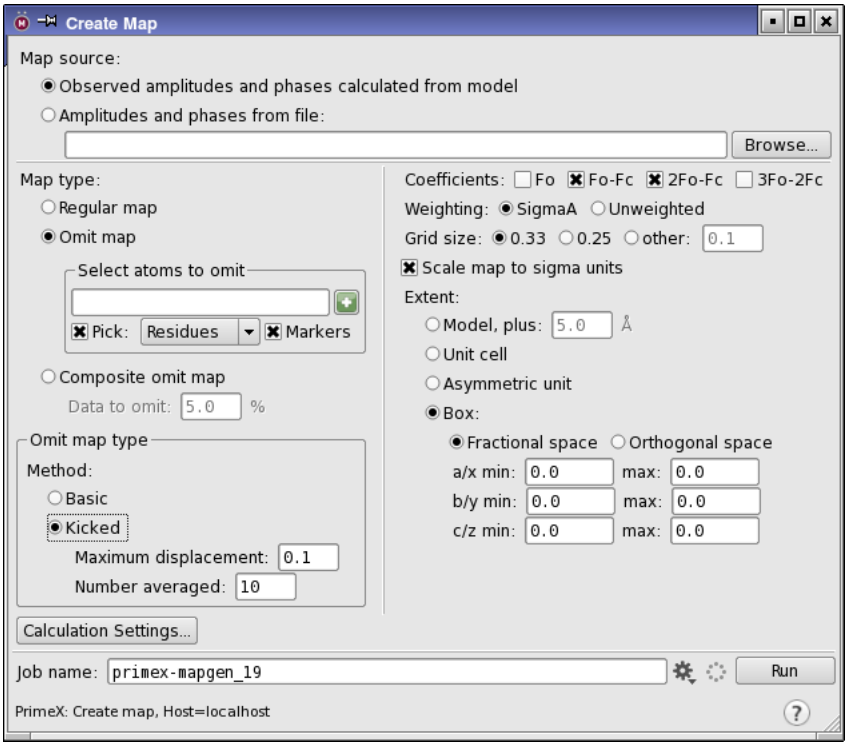


Figure 10.1. The Create Map panel.

The features on the right side of the Create Maps panel provide choices of the type of map and the data that will be used for the map.

The Map source section provides two choices for the source of amplitudes and phases for the map:

- Observed amplitudes and phases calculated from model—With this option, phases are calculated from the model to be used with the observed amplitudes. This is the default option.
- Amplitudes and phases from file—With this option, phases are obtained from a file that contains phased reflections. When you select this option, the Browse button becomes available, and you can enter the file name in the text box or click Browse and navigate to the file in a file chooser.

PrimeX supports three types of maps, which you select with the options in the Map type section:

- Regular map—Creates a full density or density difference map, according to the selection in the lower portion of the panel. This is the default map, and the only kind of map that can be automatically generated following a refinement job.
- Omit map—Creates a map in which selected atoms are omitted. When you select the option, the picking controls in the Select atoms to omit section become available, and you can pick the atoms to be omitted.
- Composite omit map—Creates a map that is constructed by systematically omitting a percentage of atoms. When you select this option, the Data to omit text box becomes available, and you can enter the percentage of atoms to omit.

For omit maps, you can choose between two types of maps:

- Basic—The default type of omit map, in which the atoms are simply omitted.
- Kicked— After all of the atoms to be omitted are removed the calculation, the model is modified by adding a small displacement of random size and direction to each coordinate before phases are calculated [15]. This process is repeated multiple times and the multiple maps are averaged. You can set the maximum displacement along any axis in the Maximum displacement text box, and the number of maps that is averaged to produce the final output, in the Number averaged text box. The maximum distance that an atom can move is $\sqrt{3}$ times the value set in the text box.

The right side of the panel contains options for setting up the map. These options are synchronized with those in the Reciprocal Space Refinement panel.

Four choices are provided for the coefficients to be used in the map: Fo, Fo-Fc (difference map), 2Fo-Fc, and 3Fo-2Fc. You can select any number of these options to create multiple maps. If you read amplitudes and phases from file, an Fo map calculation is implied since the amplitudes from the file are used without modification.

If you are using the maximum likelihood target, you can specify the weighting model used. The two options are SigmaA (see Ref. 1) and Unweighted. If you are using least squares as the target, you can only create an unweighted map. If you read amplitudes and phases from file, no additional scaling is performed that was not already part of the calculation that created the file.

Two preset options for the grid spacing are provided: 0.33 Å and 0.25 Å. If you want to specify your own grid spacing, select Other and enter a value in the text box.

The Extent options allow you to specify the extent of the map. There are four options:

- Model, plus *N* Å—Limit the map to the region within the specified distance of the model.
- Unit cell—Calculate the map for the entire unit cell.
- Asymmetric unit—Calculate the map for the asymmetric unit.
- Box—Calculate the map for the box defined in the text boxes. The values in the text boxes are given in terms of the space selected:
  - Fractional space—Define the extent of the map as a fraction of the unit cell parameters. The fractional values can be entered in the min and max text boxes for each crystallographic direction.
  - Orthogonal space—Define the extent of the map in Cartesian (*x*, *y*, *z*) coordinates. The coordinate origin is that of the structure in the Workspace.

10.2 Displaying Maps

From the perspective of Maestro, maps are simply another kind of volume that can be used to generate a surface. Maps are associated with project entries and are stored with the project.

The controls for display of surfaces are available in the Manage Surfaces panel. To open this panel, choose Workspace → Surface → Manage Surfaces, click the S button in the Surf column of the Project Table, or click the Surf Table button on the PrimeX toolbar.

Each map is listed separately in the table. The Volume Name column shows the name of the map, which is constructed from the job name used to generate the map. If the name is truncated, you can see it in the tool tip that is displayed when you pause the pointer over the table cell. The Surface Type column shows the type of map. The name in the Surface Name column is constructed from the surface type, and made unique for the originating entry if necessary. The sigma value is listed in the last column.

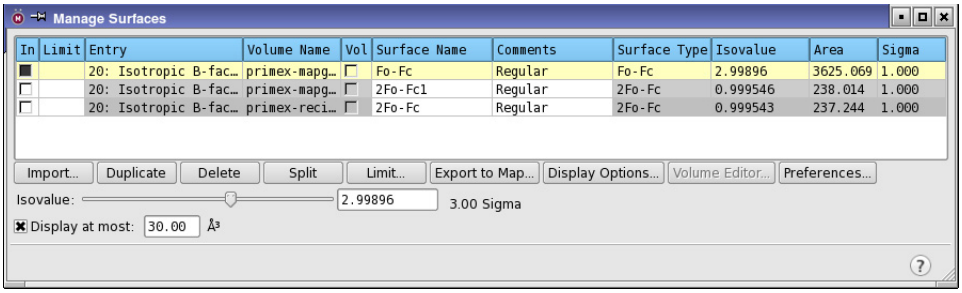


Figure 10.2. The Manage Surfaces panel.

The isovalue is given in terms of sigma for PrimeX maps, and can be adjusted by using the Isovalue slider, by entering a value in the text box, or by using the Increase Isovalue and Decrease Isovalue buttons on the PrimeX toolbar.

The isovalue increases or decreases by 0.1 sigma for each click.

Three other controls in the Manage Surfaces panel are of particular relevance for PrimeX.

The Display at most text box allows you to change the volume of the map that is displayed. The dimension given is the dimension of the side of the box (rather than the box volume). The box is centered on the map origin by default, and changes when you spot-center the structure (right-click on an atom), or choose Center Electron Density from the selection shortcut menu, to center it on the selected atoms. When you rotate or translate the structure in the Workspace, however, the map does not move relative to the structure, but is rotated or translated along with the structure.

The Export to Map button allows you to export the map from Maestro to a standard CNS map, formatted or unformatted, or to a CCP4 map. The button opens a file selector in which you can choose the format, navigate to the location, and name the map.

The Import button allows you to import a map, and associate it with an entry in the Project Table. After choosing the entry for the map, the Map Type dialog box is displayed, so that you can select the map type from among the common types. The identification of the map type is important for operations that expect a particular type of map.

Other Tasks and Tools

There are several other PrimeX tasks that have not been described or have only been mentioned in the preceding chapters. These tasks are described here. As for the other tasks, you can initiate a task in the PrimeX panel.

11.1 Mutate Model to Sequence

The mutate model to sequence task enables you to automatically mutate all residues in the model to match those in the sequence. This task is performed by the Maestro builder. The Alignment to reference sequence text area in the Mutate Model to Sequence dialog box shows the alignment of the reference sequence to the model. The residues that differ are highlighted in the model sequence. Residues that are missing at the beginning of the model are marked with a tilde.

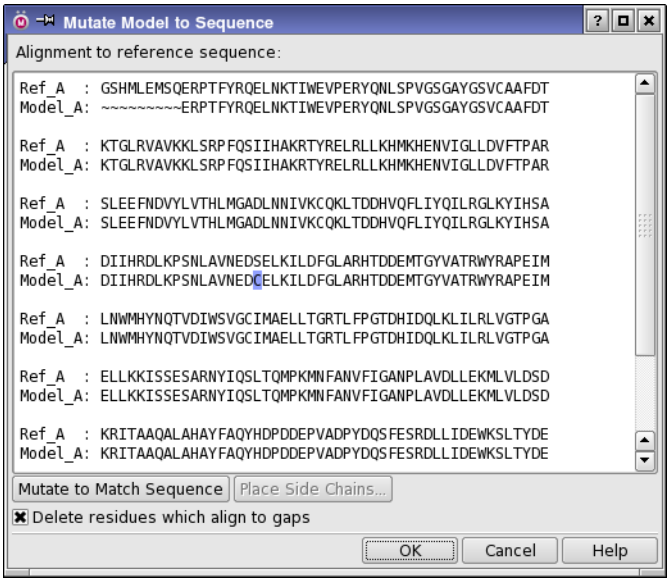


Figure 11.1. The Mutate Model to Sequence panel.

Before mutation, only the Mutate to Match Sequence button is available. When you click this button, the mutation is performed, and a new entry is created for the mutated structure. After a few seconds, the Place Side Chains button becomes available, indicating that the mutation is finished. The sequence is not updated unless you re-enter the dialog box.

The mutation places the side chains of the residues in a standard orientation. Although you could adjust these manually using the Rotamers dialog box, and even fit the best of the orientations provided to the density, if you have a map, it is better to do a full optimization of the side chains. Clicking Place Side Chains opens the Real-Space Refinement panel with the Predict side chains task selected and the mutated residues selected for refinement.

If the model lacks residues at either of the termini, some effort will be required to refine the newly-placed residues. You may want to perform manual adjustment on these residues, then run a coordinate refinement, for example.

If the reference sequence has gaps, this means that the original model has insertions compared to the protein you are refining. You can choose to delete the residues that align to these gaps when the mutation is performed. To do so, ensure that Delete residues which align to gaps is selected. You will then have to perform a loop refinement to build the shorter loop correctly according to the electron density and join the gap where the residues were deleted.

11.2 Optimize H-Bond Networks

Most X-ray experiments cannot distinguish between the N and the O of terminal amides in Gln and Asn, and cannot determine the orientation or ionization state of the ring in His, or the orientation of water molecules. The Optimize H-Bond Networks task analyzes the network of hydrogen bonds in the structure, including those with water molecules, and allows you to either optimize the orientations automatically or to manually select the orientation of each group or water molecule. The orientations are optimized on the basis of energetics of interaction with other groups, which includes optimization of hydrogen bonding. Choosing this task opens the Interactive H-bond Network Optimizer panel. For detailed information on this panel, see Section 2.6.2 of the *Protein Preparation Guide* or the online help.

You do not need to have hydrogens in the structure in order to optimize the H-bond networks. If the hydrogens are missing, they will be added when the new entry is created for the structure, and removed when the optimization is complete.

As with any automated procedure of this type, you should check the orientations after the assignment, to ensure that they are correct, or at least reasonable. The Interactive H-bond Network Optimizer panel provides tools to perform manual checking and adjustments.

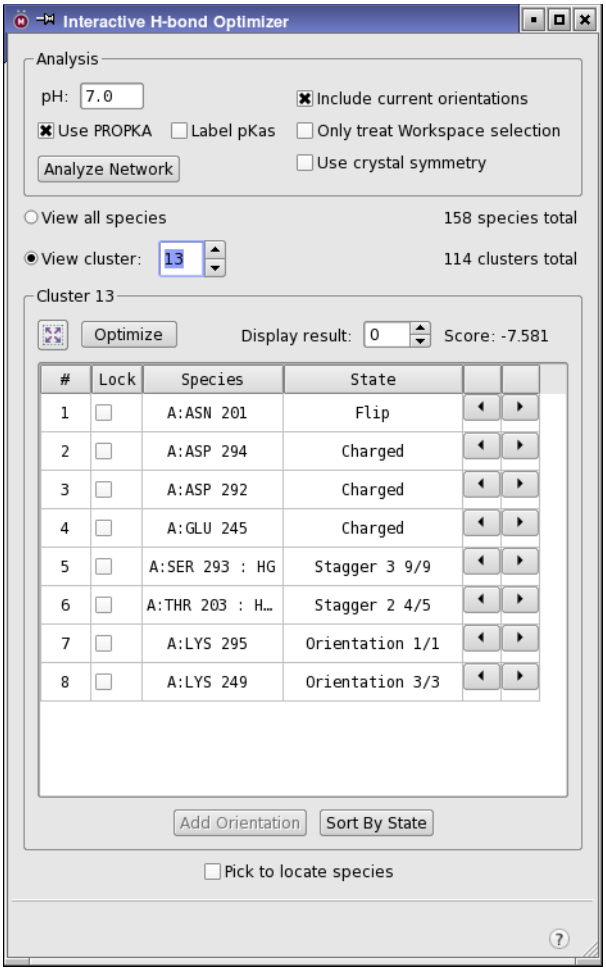


Figure 11.2. The Interactive H-bond Optimizer panel.

This tool can take account of crystal symmetry when performing the optimization. Including hydrogen bonds between asymmetric units takes much longer than without consideration of crystal symmetry. The only time that you should clearly use crystal symmetry is when the asymmetric unit contains less than one biologically-relevant unit. This choice must be made by selecting Use crystal symmetry before clicking Analyze Network.

Two other options are available for the network analysis. Include current orientations should normally be selected. This option ensures that the starting conformation is considered during optimizations, and is especially important when more than one serial optimization is applied.

Selecting Only analyze Workspace selection limits the analysis to the atoms that are selected in the Workspace, and is a useful means of limiting the scope of the analysis.

You can also control the charge states of certain residues by setting the pH. This is useful if the crystal structure was obtained at a pH other than the physiological value of 7. To set the pH value, enter the value in the pH text box.

In the second phase of hydrogen bond optimization, each hydrogen bond cluster can be investigated by selecting View cluster. You can step through the clusters by using the arrow buttons next to the View cluster option. The relevant groups are displayed in the Workspace. Clicking Optimize button starts a Monte Carlo procedure that searches for the best hydrogen bonding network for that cluster. The Degree of sampling control sets the amount of optimization for complex clusters. The table lists each group ("species") in the cluster being examined. Selecting a row allows you to explore the options for that group with the arrow keys on that row. Selecting the checkbox in the Lock column specifies that the current configuration of that group should remain unchanged during a subsequent optimization.

As an alternative to optimizing individual clusters, you can select View all species. The Optimize All button is then displayed. Clicking this button optimizes hydrogen bonding over all species, limited for complex clusters of hydrogen bonds by the Degree of sampling chosen.

You might want to display crystal mates when you check the orientation of the residues near the surface of the protein. You can do this by choosing a range from the View crystal mates button menu on the Proteins toolbar:

To check the hydrogen bonding, you can use the HBonds button menu on the Measurements toolbar.

Choose All from this button menu. The hydrogen bonds are displayed as dotted yellow lines. You can change the criteria for what is considered a hydrogen bond, in the Non-bonded ineractions – Criteria section of the Preferences panel, which you open from the Maestro menu. For more information, see Section 14.8.1 of the *Maestro User Manual*.

11.3 R-Factor Calculation

Most tasks include an R-factor calculation, so there is usually no need for a separate calculation of R factors. At the beginning of a refinement, however, the R factors are not likely to have been calculated. You can calculate them with the R-Factor Calculation task. The R-Factor Calculation panel contains a Calculation Settings button, which opens the Calculation Settings dialog box (see Section 5.4 on page 39). If you are satisfied with the settings, click Run. The job takes only a few minutes.

11.4 Locating Density Blobs

The Ligand/Solvent Placement panel makes use of density "blobs" to define the location of the ligand or solvent molecule. Density blobs are defined as collections of density peaks from a difference map that are part of a continuous region of density difference, with some minimum value set for determining the continuity of the region. The ability to display density blobs may be useful in other contexts. You can locate and display density blobs in the Density Blobs panel, which you open by clicking Density Blobs in the PrimeX panel.

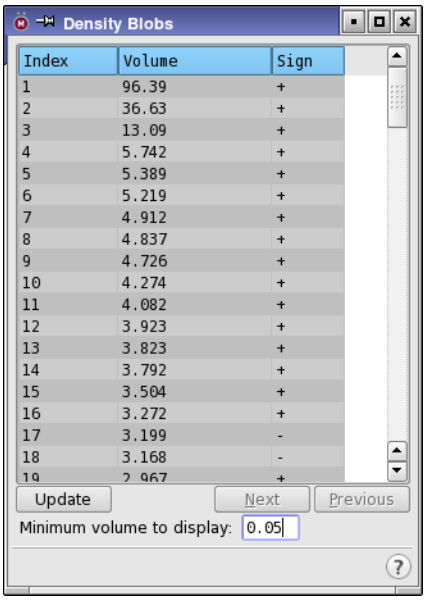


Figure 11.3. The Density Blobs panel.

Before opening the panel, ensure that an Fo-Fc map is displayed in the Workspace. The map is analyzed for density peaks, and peaks that are connected to other peaks by a saddle point that has a density above a predefined threshold are joined into blobs. If you display the map after

opening the panel, display a different map, or make changes in the Workspace, you can click Update to recompute the blobs.

The blobs are listed in order of volume, which is displayed in the Blobs table along with the sign of the difference density. You can select only one blob at a time. When you select a blob in the table, the Workspace view centers on the blob. You can set the minimum volume for which blobs are listed in the Minimum volume to display text box.

11.5 Exporting a Structure

When you are ready to generate a file with the completed structure (or to transfer a structure to some other location), you can run the Export task. This task uses the calculation settings and runs a job to generate a Maestro file. The job includes an R-factor calculation. When you click Export in the PrimeX panel, a file chooser opens so that you can name the file. When you click Export in the file selector, the Start dialog box opens, in which you can choose a title for the structure, and start the job that writes the file. To convert the file to PDB format, use the `pdbconvert` utility, and manually add any required information from the log file of the export task to the PDB file.

Validation Tools

PrimeX includes several validation tools that you can use to check the predicted structure. Two of these tools, Ramachandran plots and protein reports, are a part of the Maestro tool set and have wider applicability. The third, density fit, is specific to PrimeX. All three tools are described here.

12.1 Density Fit

The Density Fit panel displays a measure for each residue of how well the electron density predicted by the current structure fits the current electron density map. This panel can be used to identify residues that should be targeted for further refinement, based on the fit to the electron density map.

To open the Density Fit panel, click the Density Fit button in the PrimeX panel.

The table shows the a list of residues with the density fit value for the residue, the backbone, and the side chain. The values are based on the real-space R-factors. You can export the data from the table to a plain text file by clicking Export. The values are also stored as atom-level properties, which you can use in the Atom tab of the Atom Selection dialog box to select residues for refinement.

The last column in the table, Refine, contains a set of check boxes that you can use to select residues for side-chain refinement. When you have selected the desired residues, click Refine Selected Sidechains to open the Real-Space Refinement panel, with Side chains chosen from the Method option menu, and the side chains selected in the Residues for side chain refinement section.

You can sort the table by the values in a column by clicking the column heading. The values are initially in order of residue number. This feature is particularly useful to bring the residues with the worst fit to the top of the table for consideration.

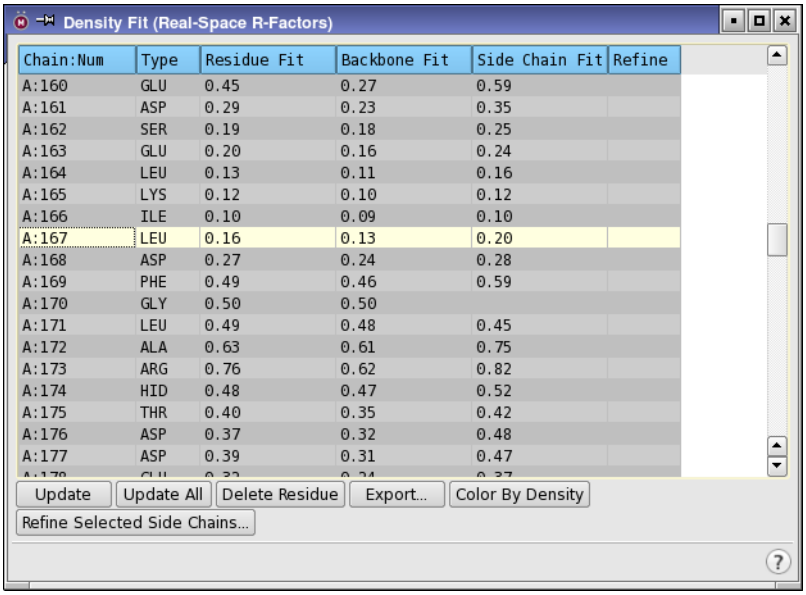


Figure 12.1. The Density Fit panel.

When you select a row in the table, the Workspace view zooms in to the selected residue. You can delete the selected residue from the Workspace structure by clicking Delete Residue. This action is mainly useful for deleting waters that are misplaced. If you delete a protein residue, you will have to build it back in again.

The residue density fit values can be applied to the structure in the Workspace as a color scheme, by clicking Color By Density. The colors range from dark green for the best fit through yellow to white for the worst fit. Residues colored white are therefore good candidates for further refinement. Once you have calculated the density fit, this color scheme is available as a standard Maestro color scheme, and can be reapplied after using some other scheme.

If you display a new structure in the Workspace, you can recalculate the density fit by clicking Update All. If you make a manual adjustment to a residue, you can update the density fit for just that residue by selecting it in the table and clicking Update.

12.2 Protein Reports

Reports on various properties of the protein that is in the Workspace are presented in the Protein Reports panel, which you open by choosing Tools → Protein Reports. The Display option menu lists the various protein properties for which reports can be generated. When you choose an item from the menu, the table is updated with the relevant report.

You can also open this panel by clicking the Protein Reports button on the Proteins toolbar.

The table displays the label for each residue or residue pair in the report in the first column, and the report data in the remaining columns. Clicking a row marks the residue or the relevant part of the residue in the Workspace, and zooms in on the residue. Clicking a column heading sorts the table by the data in the column. Repeated clicking cycles through ascending order, descending order, and original order.

Table 12.1. Display menu items in the Protein Reports panel

Item	Description
Steric Clashes	Ratio of the interatomic distances to the sum of the atomic (van der Waals) radii from the force field. The threshold is 0.85.
Bond Length Deviations	Deviation from the ideal value derived from Engh and Huber. The RMSD is reported at the bottom.
Bond Angle Deviations	Deviation from the ideal value derived from Engh and Huber. The RMSD is reported at the bottom.
Backbone Dihedrals	G-factors for backbone dihedrals (see below for definition), along with dihedral angles. Less negative (closer to zero) means that the combination of angles is more probable.
Sidechain Dihedrals	G-factors for side-chain dihedrals (see below for definition), along with dihedral angles. Less negative (closer to zero) means that the combination of angles is more probable.
G-Factor Summary	G-factors for backbone dihedrals, side-chain dihedrals, and the sum of the two.
Average B-Factors	Average B-factors (temperature factors) for the backbone and the side chain for each residue, and their standard deviations.
Gamma-Atom B-Factor	B-factor for the gamma atom in each residue.
Peptide Planarity	RMSD of the atoms in the peptide linkage from the plane that minimizes the RMSD.
Sidechain Planarity	For side chains that have nominally planar groups, RMSD of the atoms in the planar groups from the plane that minimizes the RMSD.
Improper Torsions	For side chain atoms that are nominally planar, RMSD of the improper torsion for these atoms.
C-alpha Stereochemistry	Stereochemistry of the C-alpha atoms.
Missing Atoms	Residues for which atoms are missing from the structure.

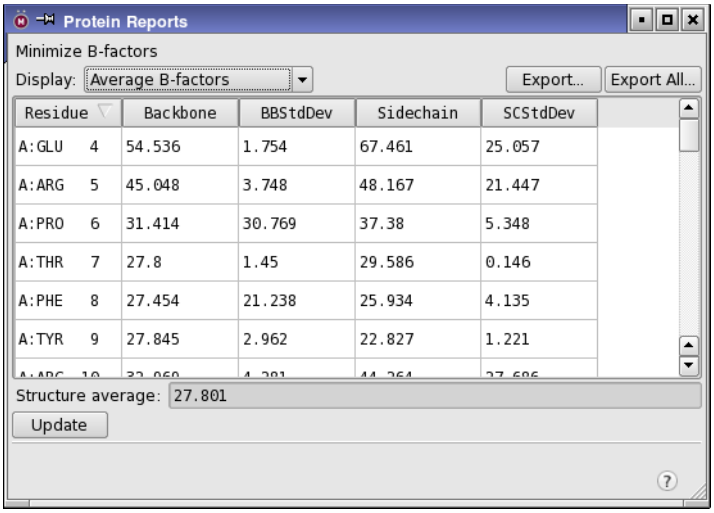


Figure 12.2. The Protein Reports panel.

G-factors are calculated by binning the dihedral angles from a collection of high-resolution structures into 10-degree ranges, and calculating the probability of a pair of angles lying in a given range. The G-factor is the logarithm of this probability. If there are no values in a given range, the G-factor is reported as "disallowed".

12.3 Ramachandran Plots

The Ramachandran Plot panel displays a plot of the dihedral angles ϕ and ψ for each residue in the protein that is displayed in the Workspace. To open the panel, choose Tools → Ramachandran Plot. This panel is useful for identifying residues that fall in disallowed regions of protein dihedrals, so that adjustments can be made to the protein geometry. After adjusting the geometry, click Refresh to redisplay the plot. You can only display a plot for one entry at a time.

You can also open this panel by clicking the Ramachandran Plot button on the Proteins toolbar.

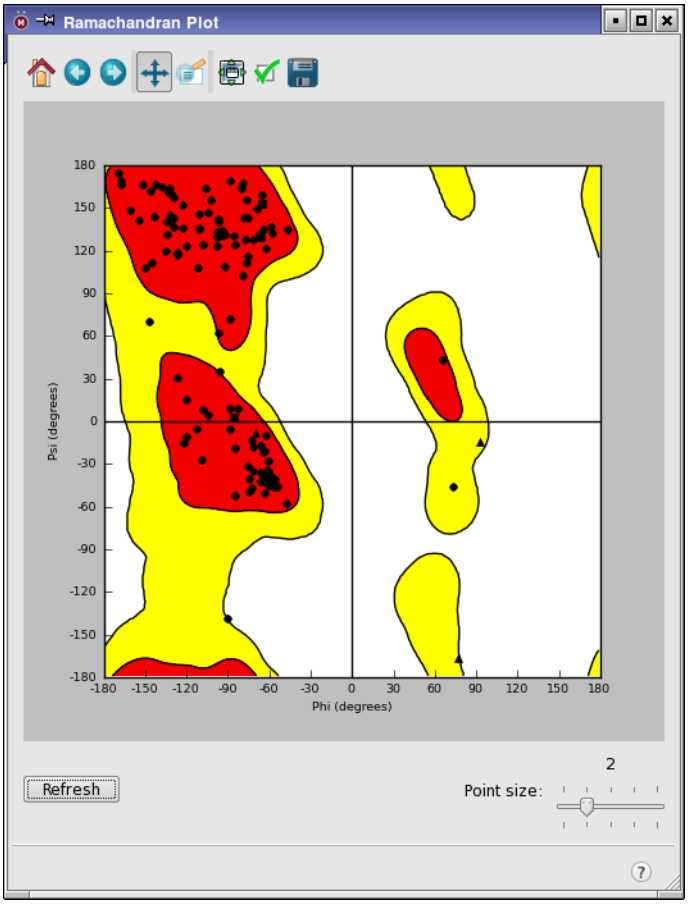


Figure 12.3. The Ramachandran Plot panel.

The plot area displays a plot of protein dihedrals for all residues in the protein. The plot is organized as follows:

- Glycine is plotted as triangles, proline is plotted as squares, all other residues are plotted as circles.
- The orange regions are the "favored" regions, the yellow regions are the "allowed" regions, and the white regions are the "disallowed" regions.

The following actions can be performed on the plot:

- Clicking on a point zooms the Workspace image in to that residue. The point is displayed as an outline instead of solid black. The residue information is displayed at the top of the panel and the residue is highlighted in the Workspace and the sequence viewer.

- Pausing the pointer over a point displays information for that residue at the top of the panel, and highlights the residue in the Workspace and the sequence viewer. When you move the pointer off the point, the display returns to that of the selected point, if there is one.

The size of the points on the plot can be adjusted using the Point size slider.

The panel has a toolbar that you can use to configure the plot or to save an image of the plot. 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 by dragging with the left mouse button, zoom by dragging 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.

Copy to clipboard
Copy an image of the plot to the clipboard. You can then paste it into another application. This button is only available in some panels.

Running PrimeX Tasks from the Command Line

PrimeX tasks can be run from the command line by creating a PrimeX input file with the appropriate keywords, and executing the following command:

`$SCHRODINGER/primex` [*job-options*] *input-file*

The only options for this command are the standard Job Control options, represented by *job-options*. For details on these options, see Section 2.3 of the *Job Control Guide*. These options include specification of the host, the user name, location of temporary storage space, and so on. For more information on Job Control, see the *Job Control Guide*.

PrimeX uses a common input file for all job types. The job type is specified by a keyword in the input file. The input file contains a list of keyword-value pairs; the keyword and value are separated by spaces. Keywords are required unless a default is given, or the dependency on the presence of other keywords is stated.

Several groups of keywords are used for multiple job types. These groups are described in the next section. In the following sections, the keyword groups required for each job type are listed, along with the keywords that are job-specific. The input file for each job type must begin with the two general keywords, which define the calculation type and the structure file.

Although it is possible to create an input file entirely by hand, in most cases you will probably want to write the input file from Maestro, then modify it to suit your needs. Writing the file from Maestro ensures that all the keywords that are required for a particular job type are included in the input file.

13.1 Input File Global Keywords

General keywords are listed in Table 13.1. Each input file must contain these keywords.

Table 13.1. General keywords.

Keyword	Setting
PRIMEX_TYPE	Type of job to run. Allowed values are: ADD_WATER Water placement ANNEAL Simulated annealing (reciprocal space) CALC_RFACT R-factor calculation LIGAND Ligand/solvent placement LOOP_BLD Loop prediction (real space) MAP_GEN Map generation PDB_EXPORT Export to PDB file REAL_MIN Real-space minimization RECIP_MIN Reciprocal-space minimization RIGID_BODY Rigid-body refinement (reciprocal space) SIDE_PRED Side-chain prediction (real space)
SEQ_FILE	Complete sequence in FASTA format. Not needed for map generation jobs.
STRUCT_FILE	Name of Maestro or PDB format file that contains the starting coordinates.
OPLS_VERSION	Version of OPLS force field used for dynamic atom typing (ligands, nonstandard and modified residues). Allowed values: 2001, 2005. Default: 2005.

13.1.1 Calculation Settings

Keywords for calculation settings are given in Table 13.2. These keywords correspond to the settings in the Calculation Settings dialog box, with the exception of OPLS_VERSION.

Table 13.2. Keywords for calculation settings.

Keyword	Description
BFAC_RES	Resolution cutoff for B-factor scaling. Allowed values: 6.0– 8.0. Default: 6.0.
BFAC_SCALE	Select type of B factor for scaling. Allowed values: none (no scaling), iso (scale isotropic B factors only), aniso (scale anisotropic and isotropic B factors). Default: none.
INCLUDE_H	Keep H atoms in the output structure. Allowed values: yes, no. Not required for map generation jobs (PRIMEX_TYPE set to MAP_GEN).

Table 13.2. Keywords for calculation settings. (Continued)

Keyword	Description
ION_RAD	Van der Waals radius for atoms in ions. Only required if SOLV_MOD is set to mask.
LIMIT_REFL	Cutoff for rejecting reflections based on absolute value. Reject reflections for which the F value is less than the cutoff. If not included, reflections are not rejected on this criterion.
LIMIT_RMS	Cutoff for rejecting reflections based on the RMS F value. Reject reflections for which the ratio of the F value to the root-mean-square F value is greater than the cutoff. If not included, reflections are not rejected on this criterion.
LIMIT_SIGMA	Cutoff for rejecting reflections based on sigma. Reject reflections for which the ratio of the F value to the standard deviation of the F values is less than the cutoff. If not included, reflections are not rejected on this criterion.
MAX_FFT_MEM	Maximum memory to use in the FFT, in MB. If omitted, no limit is imposed.
PLANARITY_RESTRAINT	Level of restraint applied in the force field for planar groups to ensure planarity. Allowed values: low, normal, high. Default: normal.
SHR_FACT	Factor applied to reduce the covalent radii once the solvent regions are identified. Only required if SOLV_MOD is set to mask.
SOLV_MOD	Bulk solvent correction method. Allowed values: none, babinet, mask. Default: none.
USE_SGB	Use implicit solvation for real-space refinement jobs other than simulated annealing, Allowed values: yes, no. Default: no.
VDW_RAD	Van der Waals radius of covalently bound atoms. Only required if SOLV_MOD is set to mask.
XRES_HIGH XRES_LOW	High-resolution and low-resolution limits for reflection data, in angstroms.

13.1.2 Map Settings

These parameters describe both the extent of the map itself and some aspects of the model and method used to create it. For jobs that produce maps, multiple maps can be specified and the keywords MAP_COEFF_0, MAP_COEFF_1, and so on, can be used, each with a single value, such as 1_1 for an Fo-Fc (difference) map.

Keywords for map settings are given in Table 13.3.

Table 13.3. Keywords for map settings.

Keyword	Description
BOX_COORD	Set the coordinate type as either a fraction of the unit cell, or Cartesian. Required only if MAP_EXTENT is set to box. Allowed values: frac, cart.
GRID_SIZE	Fraction of the resolution to use for the grid spacing.
KICK_DIST	Maximum displacement to use in a kicked omit map. Values over 1 Å are not recommended. Used only if OMIT_TYPE is set to kicked. Default: 0.1 Å.
KICK_NUM	Number of kicked omit maps to average. Used only if OMIT_TYPE is set to kicked. Default: 10.
MAP_BOX_*i*_*j*	Minimum and maximum coordinate values of map box. *i* can take values X, Y, and Z, and *j* can take values MIN and MAX, e.g. MAP_BOX_X_MIN. Required only if MAP_EXTENT is set to box.
MAP_BUFFER	Size of buffer around selected molecule, in angstroms. Required only if MAP_EXTENT is set to mol. Default: 5.0.
MAP_COEFF_*n*	Map coefficients for map *n*, where *n* is a serial number starting from 0. The value of the keyword takes the form *i*_*j*, where *i* is the coefficient of Fo and *j* is the coefficient of Fc: for example, Fo-Fc is 1_1.
MAP_COEFFA_*n*	Name of column for amplitudes for map *n*. Only valid when MAP_FORM is set to mtz-refl. Commonly, when MAP_COEFF_*n* is 1_0 the value is FOWT, for 1_1 it is DELFWT, for 2_1 it is FWT, and for 3_2 it is 3FO2FCWT.
MAP_COEFFP_*n*	Name of column for phases for map *n*. Only valid when MAP_FORM is set to mtz-refl. Commonly, when MAP_COEFF_*n* is 1_0 the value is PHFOWT, for 1_1 it is PHDELWT, for 2_1 it is PHFWT, and for 3_2 it is PH3FO2FCWT.
MAP_EXTENT	Determines how the map extent is defined. Allowed values: asym Asymmetric unit unit Unit cell box Box of given dimensions, either in (*a*,*b*,*c*) or (*x*,*y*,*z*). mol Model structure plus buffer zone, specified by MAP_BUFFER.

Table 13.3. Keywords for map settings. (Continued)

Keyword	Description
MAP_FORM	Map output format. Allowed values: cns-form CNS formatted (default) cns-bin CNS unformatted cns-refl CNS phased reflections (not valid with omit maps) ccp4 CCP4 unformatted mtz-refl MTZ phased refelctions
MAP_TYPE	Type of map—regular map, omit map, or composite omit map. Allowed values: reg, omit, comp. Atom selection data keywords (see Table 13.5) are required if MAP_TYPE is set to omit. If omitted, no map is generated.
MAP_WEIGHT	Weighting model. Allowed values: sigma, unweighted.
OMIT_DATA	Percentage of data to omit. Used only if MAP_TYPE is set to comp.
OMIT_TYPE	Type of omit map to generate. Allowed values: bhat-cohen, kicked. Used only if MAP_TYPE is set to omit or comp. Default: bhat-cohen.
SCALE_MAP	Scale map to sigma units. Allowed values: yes, no. Default: no.

13.1.3 Reflection Data Settings

Reflection data settings are listed in Table 13.4.

Table 13.4. Keywords for reflection data.

Keyword	Value
CELL_*i*	Cell parameters a, b, c, α, β, and γ. *i* can take the values A, B, C, ALPHA, BETA, and GAMMA. Used if the data is not otherwise available.
ISO_BFAC	Isotropic B-factor value. Default: 20.0. Used if the data is not otherwise available.
REFLECT_FILE	File name of the reflection file.
REFLECT_FOBS	Column name for F_{obs} in the reflection file. Default: FOBS.
REFLECT_FORM	Format of the reflection file. Allowed values: cns, dtr, hkl, mmcif, mtz, scalepack.
REFLECT_SIGF	Column name for σ in the reflection file. Default: SIGF.
REFLECT_TEST	Column name for the test set flag in the reflection file. Default: TEST.
PHASE_FILE	File name of the phased reflection file.
PHASE_FORM	Format of the phased reflection file. Allowed values: cns, dtr, hkl, mmcif, mtz, scalepack.

Table 13.4. Keywords for reflection data. (Continued)

Keyword	Value
PHASE_MAP_COEFF	Column name for cns-style complex map coefficients in phased reflection file. Incompatible with PHASE_MAP_COEFFA and PHASE_MAP_COEFFP.
PHASE_MAP_COEFFA	Column name for mtz-style map amplitudes in phased reflection file. Incompatible with PHASE_MAP_COEFF.
PHASE_MAP_COEFFP	Column name for mtz-style map phases in phased reflection file. Incompatible with PHASE_MAP_COEFF.
SPACE_GRP	Space group name. Used if the data is not otherwise available.

13.1.4 Atom Selection Keywords

These keywords can be used to specify an atom selection, where the interface uses picking tools. The keywords are listed in Table 13.5.

Table 13.5. Keywords for specifying atom selections.

Keyword	Values
ATOM_FILE	Name of file containing a list of atoms. Each line in the file must be of the format *chainID*:*resnum*:*atomID*, as described for the ATOM_*i* keyword.
ATOM_*i*	Atom specification. The format of this specification is *chainID*:*resnum*:*atomID*, where *chainID* is the single-letter chain identifier, *resnum* is the residue number, and atomID is the PDB atom name, with underscores replacing blanks—for example, A:162:_C__. *i* is an index that starts from zero.
RESIDUE_FILE	Name of file containing a list of residues. Each line in the file must be of the format *chainID*:*resnum*, as described for the RESIDUE_*i* keyword.
RESIDUE_*i*	Residue specification. The format of this specification is *chainID*:*resnum*, where *chainID* is the single-letter chain identifier, and *resnum* is the residue number—for example, A:132. *i* is an index that starts from zero.
SELECT	Atom selection type. If an atom selection is required by another keyword, this keyword is required. Allowed values: all(select all atoms), pick (pick atoms to use). If set to pick, exactly one of the other types of ASL data keywords should be used.

13.1.5 Keywords for Noncrystallographic Symmetry

Keywords for use of noncrystallographic symmetry are listed in Table 13.6. NCS is implemented by defining groups of chains that are related by NCS. Within each group, there can be multiple symmetry-related (equivalent) units. Each such unit can be composed of a number of residues, which do not need to be contiguous. There must be a 1:1 correspondence of atoms between each group, so you must omit any residues from the group that do not match the corresponding symmetry-related residue.

Harmonic restraints are applied to impose the symmetry. These restraints can be applied to the backbone only, or to all atoms in the group. The weighting factor for the restraints can be set for each group.

Table 13.6. Keywords for use of noncrystallographic symmetry.

Keyword	Description
`NCS_`*i*`_`*j*`_`*k*`_0` `NCS_`*i*`_`*j*`_`*k*`_1`	Beginning (`_0`) or ending (`_1`) residue of structural segment *k* in equivalent unit *j* of group *i*. All three indices start from zero. The value for these keywords is a residue specification in the format *chainID*`:`*resnum*, for example `A:118`.
`NCS_FILE`	Name of a file that specifies the NCS structural segments. This is an alternative to using `NCS_`*i*`_`*j*`_`*k*`_0` and `NCS_`*i*`_`*j*`_`*k*`_1`. Each line in the file specifies a segment in the following format: *i*`:`*j*`:`*k* *chainID*`:`*resnum* [*chainID*`:`*resnum* ...]
`NCS_WEIGHT_`*i*	Weighting factor for the restraint for group *i*; *i* starts from zero.
`NCS_SIGB_`*i*	B-factor sigma for group *i*; *i* starts from zero.
`NCS_BB_`*i*	Apply restraints for backbone only for group *i*; *i* starts from zero. Allowed values: `yes`, `no`. If omitted, restraints are applied to all the atoms in the group.

13.2 Map Generation

The keywords required for a map generation job are:

- `PRIMEX_TYPE MAP_GEN`
- Map settings keywords
- Reflection data keywords
- Calculation settings keywords

The `INCLUDE_H` keyword is not required for this job type.

13.3 Real-Space Refinement

The keyword groups below must be included for all real-space refinements.

- Reflection data keywords
- Calculation settings keywords

In addition, the general keywords for real-space refinement, listed in Table 13.7, must be included.

Table 13.7. General keywords for real-space refinement.

Keyword	Description
BFAC_FIT	Value of B-factor to use for atoms being fit. Default: 0.0, which means "use the values from the input file".
DENSITY_WT	Scaling factor applied to default density scoring term weight. Increasing the value increases the weight of the density fit relative to the force field terms. Default: 1.0.
RECIP_MIN	Option to perform reciprocal-space minimization after real-space refinement. Allowed values: yes, no. Default: no.
SAMPLING_GRID	Map grid spacing for real-space sampling, in angstroms. The allowed range is in the interval [0.5,1.0]. Default:0.25*XRES_HIGH.
SEQ_FILE	Name of file containing the complete sequence in FASTA format.

13.3.1 Loop Prediction

The following keywords are needed for loop predictions, in addition to the general real-space refinement requirements.

- PRIMEX_TYPE LOOP_BLD
- Loop prediction keywords, given in Table 13.8.

13.3.2 Side Chain Prediction

The following keywords are needed for side chain predictions, in addition to the general real-space refinement requirements.

- PRIMEX_TYPE SIDE_PRED
- Atom selection data keywords, to define the residues for which side chains are to be predicted. Either an atom specification or a residue specification can be used to specify a residue.

Table 13.8. Loop prediction keywords.

Keyword	Description
USE_BIAS	Bias sampling of loops to existing loop structure. Conformational space is sampled more heavily around the existing structure. Can only be used if the loop is complete (no missing residues) Allowed values: yes, no.
LOOP_FILE	Name of file containing definition of loop. Must contain lines with the format *chain0*:*resnum0* *chain1*:*resnum1*, specifying the beginning and end of the loop.
LOOP_0_RES_0 LOOP_0_RES_1	Beginning (_0) or end (_1) of loop. The value for these keywords is a residue specification in the format *chainID*:*resnum*, for example A:118.

13.3.3 Minimization

The following keywords are needed for minimizations, in addition to the general real-space refinement requirements.

- PRIMEX_TYPE REAL_MIN
- Atom selection data keywords to define the residues to minimize. Either an atom specification or a residue specification can be used to specify a residue.

13.4 Reciprocal-Space Refinement

The keyword groups below must be included for all reciprocal-space refinements.

- Reflection data keywords
- Calculation settings keywords
- Map settings keywords. Only keywords for MAP_TYPE reg are allowed.
- General reciprocal-space refinement keywords, as listed in Table 13.9. (Note that there are defaults for both target keywords, so in practice this group can be omitted).

Table 13.9. General reciprocal-space refinement keywords.

Keyword	Description
SEQ_FILE	Name of file containing the complete sequence in FASTA format.
TARGET_TYPE	Refinement target type. Allowed values are MLF (maximum likelihood) and LSF (least-squares). Default: MLF.
TARGET_WEIGHT	Relative weight of X-ray scoring term. Positive if used as a scaling factor for the optimal weight, and negative if used as a constant multiplier of the weight. (In the latter case, the absolute value is what is actually used.) Default: 1.0.

13.4.1 Rigid-Body Refinement

The following keywords are needed for rigid-body refinements, in addition to the general reciprocal-space refinement requirements.

- PRIMEX_TYPE RIGID_BODY
- NCS data keywords
- Rigid-body refinement keywords, as listed in Table 13.10.

Table 13.10. Rigid-body refinement keywords.

Keyword	Description
DEFAULT_STATUS	Treatment of atoms that are not explicitly included in a group. Allowed values: fixed Keep atom coordinates fixed at their present values group Treat atoms as an implicit group
GROUP_FILE	Name of file defining rigid body groups. Each line in the file has the format *i*:*j* *chainID*:*resnum0* *chainID*:*resnum1* where *i* and *j* are indices of the group and group segment, and *resnum0* and *resnum1* are the beginning and ending residue number for the segment. This is an alternative to using the GROUP_*i*_*j*_0 and GROUP_*i*_*j*_1 keywords. The following example specifies residues 1 through 41 as segment 0 of group 0. 0:0 A:1 A:41
GROUP_*i*_*j*_0 GROUP_*i*_*j*_1	Specify the residue that begins (_0) or ends (_1) segment *j* of rigid body group *i*. *i* and *j* are indices that start from zero. The value for each keyword must be a residue specification in the format *chainID*:*resnum*. By defining segments, a group can be composed of discontiguous portions of a chain. For example, GROUP_0_0_0 A:1 GROUP_0_0_1 A:41 specifies residues 1 through 41 as segment 0 of group 0.
MAX_CYC	Number of cycles of minimization. Each cycle can use MAX_STEP minimization steps. Scaling factors and second derivatives are recalculated between cycles. Default: 1.
MAX_STEP	Minimization steps per cycle. Default: 0.
MINIM_METHOD	Minimization method. Allowed values: tn (truncated Newton), cg (conjugate gradient), qn (quasi-Newton, LBFGS), opt (cg when initial gradients are large, tn otherwise). Default: tn.

13.4.2 Simulated Annealing

The following keywords are needed for simulated annealing refinements, in addition to the general reciprocal-space refinement requirements.

- `PRIMEX_TYPE ANNEAL`
- Atom selection keywords
- Simulated annealing keywords, as listed in Table 13.11.

Table 13.11. Simulated annealing keywords.

Keyword	Description
`MD_ENERGY`	Number of MD energy scale estimation steps.
`MD_FINAL`	Number of MD steps used to equilibrate the system at the final temperature.
`MD_STEP`	Molecular dynamics time step, in ps.
`NB_CUTOFF`	Non-bonded interactions cutoff, in angstroms.
`NMIN_INIT` `NMIN_FINAL`	Number of initial and final minimization steps.
`SIM_MODE`	Energy model used for simulation. Allowed values: `approx` (neglect solvation and electrostatic terms), `complete` (use all force field terms).
`STEPS_HEAT` `STEPS_COOL`	Number of heating and cooling steps.

`T_INIT`
`T_HIGH`
`T_FINAL`
	Initial, high, and final temperatures for the simulated annealing calculation.

13.4.3 Reciprocal-Space Minimization

The following keywords are needed for minimizations, in addition to the general reciprocal-space refinement requirements.

- `PRIMEX_TYPE RECIP_MIN`
- ASL data
- NCS data
- Reciprocal-space minimization keywords, as listed in Table 13.12.

Table 13.12. Reciprocal-space minimization keywords.

Keyword	Description
BFAC_GROUP	Specify how to group B-factors for group refinement. All atoms in a given group have the same B-factor. Allowed values: residue, chain, side, back, sideback (one for side chains, one for backbone).
BFAC_MIN BFAC_MAX	Minimum and maximum allowed B-factor values.
BFAC_OPT	Specify which B-factors to optimize. Only isotropic B-factor optimization is currently supported. Allowed values: iso.
BFAC_WEIGHT	Weight of the restraint of B-factor values to the target sigma values.
HBSET	Set behavior for the assignment of B-factors for hydrogen atoms. The allowed values are: parent—Set the temperature factor of each hydrogen atom to that of the parent plus an offset that is defined by HBVALUE (Default). constant—Set the temperature factor of each hydrogen atom to the value given by HBVALUE. mean—Set the temperature factor of each hydrogen atom to the mean of the non-hydrogen atoms (Former default). no—Do not change hydrogen atom temperature factor values that are obtained as a result of PrimeX jobs.
HBVALUE	Value or offset of the B-factor for hydrogen atoms. Used when HBSET is set to constant or parent. Default: 2.00.
MAX_CYC	Number of cycles of minimization. Each cycle can use MAX_STEP minimization steps. Settings such as scaling are updated between cycles. Default: 1.
MAX_STEP	Minimization steps per cycle. Default: 0.
MINIM_BFAC	Specify B-factor minimization. Follows coordinate minimization, if both are specified. Allowed values: yes, no. Default: no.
MINIM_COORD	Specify coordinate minimization. Precedes B-factor minimization, if both are specified. Allowed values: yes, no. Default: yes.
MINIM_METHOD	Minimization method. Allowed values: tn (truncated Newton), cg (conjugate gradient), qn (quasi-Newton, LBFGS), opt (cg when initial gradients are large, qn otherwise). Default: qn.
MINIM_OCC	Specify grouped partial occupancy minimization. Allowed values: yes, no. Default: no.
RESTR_BB_BOND RESTR_BB_ANGLE RESTR_SC_BOND RESTR_SC_ANGLE	Target standard deviation of the B-factors for backbone and side chain bond lengths and bond angles.

13.5 Ligand and Solvent Placement

The following keywords are needed for ligand and solvent molecule placement.

- PRIMEX_TYPE LIGAND
- Reflection data keywords
- SEQ_FILE
- Calculation settings keywords
- Ligand and solvent placement keywords, as listed in Table 13.13.

Table 13.13. Ligand and solvent placement keywords.

Keyword	Description
ENERGY_WEIGHT	Weight of density terms used to mix into the Coulomb/van der Waals energy. Allowed values: 1–9999. Default: 10.
GEN_MATES	Generate crystal mates near the active site. Essential if crystallographically-related atoms that are not explicitly in the current protein structure are close enough to the ligand or solvent molecule position to affect its interactions. Allowed values: yes, no. Default: no.
LIGAND_FILE	Name of Maestro format file that contains the ligand or solvent molecule structure.
MAX_POSES	Maximum number of poses returned from the entire job.
NUM_POSES	Number of ligand poses to return from Glide docking. If postminimization is used, this is the number of poses that are refined. Default: 10.
POCKET_CENTER_*i*	Coordinates of center of density pocket. *i* can be X, Y, or Z.
POCKET_EXTENT_*i*	Dimensions of the grid generation box and the map used for Glide docking of the ligand. This box is centered on the point defined by the POCKET_CENTER_*i* keywords. *i* can be X, Y, or Z. The protein is trimmed beyond the box to reduce the time taken to generate the grid and dock the ligand, so it is important that the box is not made too small. See Section 4.3 of the *Glide User Manual* for information on how the grid generation box is set up.
POST_REFINE	Refine the ligand or solvent molecule structure after docking with Glide (Glide docking postminimization). Allowed values: yes, no. Default: no.
SAMPLE_RING_CONF	Vary ring conformations. Allowed values: yes, no. Default: yes.
SCORING_WEIGHT	Weight of density terms used in the rough scoring function. Allowed values: 1–9999. Default: 10.
USE_LIG_PREP	Prepare the ligand with LigPrep. This should be done if the ligand is not a 2D, all-atom structure with reasonable geometry and charge state. Allowed values: yes, no. Default: no.

13.6 Water Placement

The following keywords are needed for ligand and solvent molecule placement.

- PRIMEX_TYPE ADD_WATER
- Calculation settings keywords
- Reflection data keywords
- Water placement keywords, as listed in Table 13.14.

Table 13.14. Water placement keywords.

Keyword	Description
PEAK_HT	Minimum height of a peak in the difference map used for placement, in terms of sigma.
BFAC_MAX	Maximum B factor, for any water molecule to be accepted as placed. Water molecules with larger B-factors are rejected as noise.
DIST_MIN DIST_MAX	Minimum and maximum distances between the water oxygen atom and any other atom. Water molecules will not be placed in locations where this distance is outside the specified range.
DIST_HB	Minimum distance between the water oxygen atom and a hydrogen-bond donor or acceptor. Water molecules will not be placed in locations where this distance is less than the value specified.
LIG_SCREEN	Screen out ligand-sized blobs when identifying water blobs, using additional criteria. Allowed values: yes, no. Default: no.
ERAD_CUT	Distance from a peak to search for possible neighboring peaks that might be part of the same blob. Must be greater than the value of BOND_UP. Valid only with LIG_SCREEN yes.
BOND_LOW	If the distance between two peaks is less than this value, the peaks are treated as connected. Valid only with LIG_SCREEN yes.
BOND_UP	If the distance between two peaks is less than or equal to this value, the peaks are treated as connected if the electron density at the midpoint is greater than sigma times the value given by MID_CUT. If the distance is greater, the peaks are treated as separate. Valid only with LIG_SCREEN yes.
MID_CUT	Threshold density for connection of two peaks, in units of sigma. If the value at the midpoint is greater than this value, the peaks are treated as connected. Used in conjunction with BOND_UP. Valid only with LIG_SCREEN yes.
EDGE_CUT	Cutoff for inclusion of grid points around peaks in the current blob, in units of sigma. Valid only with LIG_SCREEN yes.
WATER_RAD	Water radius, used to define the volume of a water molecule. Valid only with LIG_SCREEN yes.

13.7 R-Factor Calculation

The following keywords are needed for R-factor calculations.

- `PRIMEX_TYPE CALC_RFACT`
- Calculation settings keywords
- Reflection data keywords

13.8 Exporting a Structure to a File

The following keywords are needed to export a structure to a PDB file.

- `PRIMEX_TYPE PDB_EXPORT`
- Calculation settings keywords
- Reflection data keywords
- Export keywords, as listed in Table 13.15.

Table 13.15. Export keywords.

Keyword	Description
`SEQ_FILE`	Name of file containing the complete sequence in FASTA format.
`OUTPUT_NAME`	File name of file to be exported. Must be in Maestro format. Either a relative path or an absolute path can be specified.

13.9 Utilities

Utilities for all products can be found in `$SCHRODINGER/utilities`. For PrimeX use, the utilities described below may be useful.

13.9.1 mtzprint

This utility reads a binary `.mtz` (CCP4) file and writes out a text version of the file. The command syntax is as follows:

`mtzprint -i` *input-file* `[-o` *output-file*`] [-nref` *number-of-reflections*`]`

If the output file is not specified, the file is written to standard output. The default number of reflections is all reflections. This utility may be of use for interpreting the column labels: when reading an `.mtz` file, an assignment of the Fobs and Sigma F properties is required.

13.9.2 refconvert

This utility converts between various reflection file formats. The command syntax is:

`refconvert` [*options*] [*MTZ-options*] `-i`*informat input-file* `-o`*outformat output-file*

The format string defines the formats. There must be no spaces between the format string and the `-i` or `-o` option. For information on the options, run the command with the `-h` option.

13.9.3 ccp42cns

This utility converts a CCP4 map to a CNS map. The command syntax is:

`ccp42cns -i` *ccp4-map-file* `-o` *cns-map-file*

13.9.4 Protein Preparation

To run the preparation steps that are necessary before a structure can be used by PrimeX, you can use the `prepwizard` utility, as follows:

```
$SCHRODINGER/utilities/prepwizard -keepfarwat -nohtreat
```

These options perform the same treatment as is done from the Input Data panel when you select Assign bond orders. The `-keepfarwat` option ensures that waters are kept, and `-nohtreat` ensures that hydrogen atoms are not added.

For more information on this utility, see Section A.2 of the *Protein Preparation Guide*.

13.9.5 Writing Parameter Files for Ligands

If you want to write out parameter (topology) files for your ligands, you can use the utility `hetgrp_ffgen`:

```
$SCHRODINGER/utilities/hetgrp_ffgen {2001|2005} input-file -write_cif
    [-jobname jobname]
```

where the first argument is the OPLS force field version, and the input file contains the ligands in Maestro format. If you need to convert another file format to Maestro format for input to `hetgrp_ffgen`, you can use the utility `structconvert`. Output is generated in two forms. The first is a file that contains all the ligands with the definitions (restraints, parameters, topology) used to define the ligand geometry, *jobname*`-prime.cif`. This is a plain text file in mmCIF format, which is documented at http://mmcif.rcsb.org. The second is a set of REFMAC dictionary files, one per ligand, named *jobname*`-`*lig*`-refmac.cif`. Here, *jobname* is the stem of the input file name by default, but can be specified with the `-jobname` option; and *lig* is the 3-character PDB identifier for the ligand.

Parameter files are automatically generated for ligands during ligand placement and saved in the default (working) directory.

Reserved Residue Names

PrimeX uses residue names to classify residues as protein residues, and hence determine whether chains are protein chains or not. In addition to the standard names for the 20 standard amino acids, the following names are also recognized as protein residues:

ACE	CYX	HIP	PTR
AS1	GL1	LYN	SEP
ASH	GLH	MSE	SRO
ASN	HID	NMA	TPO
CYT	HIE	NH2	TYO

You should ensure that these residue names are only used for protein residues. In addition, HOH is recognized as water. Internally, PrimeX uses the names XXX, YYY, and ZZZ, and any other names starting with X. You should therefore ensure that none of these internal names is used in your model.

Implementation

Map generation in PrimeX is based on the sigmaA weighting scheme of Reed [1], a data treatment that is shown to decrease the bias in electron density maps. Optionally, an unweighted map may be calculated. Treatment of omit maps is by the kicked method [15].

Reciprocal-space refinement with maximum likelihood target using the formulation of Pannu and Reed [5] has been implemented following the general concepts developed in the publications of Axel Brünger and coworkers [6,7]. A maximum likelihood target has been shown to improve the convergence of refinement and reduce the effects of model bias [8]. Optionally, a more traditional least-squares target may be selected.

Geometric restraints are based on the OPLS force field [9], which has been successfully employed in the modeling of protein structure [10]. PrimeX uses a slightly modified version of OPLS that provides very good correspondence to the expected values for bond lengths and angles reported by Engh and Huber [11]. Ligand placement into electron density is augmented with the use of a Glide score, which is derived from the very successful ligand docking program Glide [12].

Simulated annealing was implemented within IMPACT [13], employing well-validated concepts [14].

References

1. Read, R. J. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Cryst.* **1986**, *A42*, 140–149.

2. Bhat, T.N.; Cohen, G.H. OMIT MAP: An Electron Density Map Suitable for the Examination of Errors in a Macromolecular Model. *J. Appl. Crystallog.* **1984**, *17*, 244–248.

3. Bhat, T.N. Calculation of an OMIT map. *J. Appl. Crystallog*. **1988**, *21*, 279–281.

4. Vellieux, F. M. D.; Dijkstra, B. W. Computation of Bhat's OMIT Maps with Different Coefficients. *J. Appl. Crystallog.* **1997**, *30*, 396–399.

5. Pannu, N. S.; Read, R. J. Improved Structure Refinement through Maximum Likelihood. *Acta Cryst.* **1996**, *A52*, 659–668.

6. Brünger, A. T. A Memory-efficient Fourier Transformation algorithm for Crystallographic Refinement on Supercomnputers. *Acta Cryst.* **1989**, *A45*, 42–50.

7. Brünger, A. T.; Adams, P. D.; Clore, G. M.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, N.; Panu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallograpy & NMR System (CNS): A New Software System for Macromolecular Structure Determination. *Acta Cryst.* **1998**, *D54*, 905–921.

8. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by Maximum-Likelihood Method. *Acta Cryst.* **1997**, *D53*, 240–255.

9. Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

10. Zhu, K.; Shirts, M. R.; Friesner, R. A. Improved Methods for Side Chain and Loop Predictions via the Protein Local Optimization Program: Variable Dielectric Model for Implicitly Improving the Treatment of Polarization Effects. *J. Chem. Theory Comput.* **2007**, *3*, 2108–2119.

11. Engh, R.A.; Huber, R. Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement. *Acta Cryst.* **1991**, A47, 392–400.

12. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K. Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J. Med. Chem. **2004**, 47, 1739–1749.

13. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program. Applied Chemical Theory. (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752–80.

14. Adams, P. D.; Pannu, N. S.; Read, A. T.; Brünger, A. T. Cross-validated Maximum-likelihood Enhances Crystallographic Simulated Annealing. Proc. Natl. Acad. Sci. USA **1997**, *94*, 5018–5023.

15. Guncar G.; Klemencic, I.; Turk, B.; Turk, V.; Karaoglanovic-Carmona, A.; Juliano, L; Turk, D. Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. *Structure* **2000**, *8*, 305.

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/, In particular, you can use the Knowledge Base, http://www.schrodinger.com/kb, to find current information on a range of topics, and the Known Issues page, http://www.schrodinger.com/knownissues, to find information on software issues.

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 the lower right corner of 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 the Help panel. The button may have text or an icon:

- Choose Help → Online Help or press CTRL+H (⌘H) to open the default help topic.
- When help is displayed in the Help panel, use the navigation links in the help topic or search the help.
- Choose Help → Documentation Index, to open a page that has links to all the 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*.
- New software features: choose Help → New Features.
- 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.

Web:	http://www.schrodinger.com/supportcenter
E-mail:	help@schrodinger.com
Mail:	Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone:	+1 888 891-4701 (USA, 8am – 8pm Eastern Time)
	+49 621 438-55173 (Europe, 9am – 5pm Central European Time)
Fax:	+1 503 299-4532 (USA, Portland office)
FTP:	ftp://ftp.schrodinger.com

Generally, using the web form is best because you can add machine output and upload files, if necessary. You will need to include the following information:

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

Gathering Information for Technical Support

The instructions below describe how to gather the required machine, licensing, and installation information, and any other job-related or failure-related information, to send to technical support. Where the instructions depend on the profile used for Maestro, the profile is indicated.

For general enquiries or problems:

1. Open the Diagnostics panel.
   - **Maestro:** Help → Diagnostics
   - **Windows:** Start → All Programs → Schrodinger-2015-2 → Diagnostics
   - **Mac:** Applications → Schrodinger2015-2 → 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. Upload the file specified in the dialog box to the support web form.

   If you have already submitted a support request, use the upload link in the email response from Schrödinger to upload the file. If you need to submit a new request, you can upload the file when you fill in the form.

If your job failed:

1. Open the Monitor panel, using the instructions for your profile as given below:
   - **Maestro/Jaguar/Elements:** Tasks → Monitor Jobs
   - **BioLuminate/MaterialsScience:** Tasks → Job Monitor
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, and an information dialog box with the name and location of the file opens. You can highlight and copy the name of the file.
5. Upload the file specified in the dialog box to the support web form.

   If you have already submitted a support request, use the upload link in the email response from Schrödinger to upload the file. If you need to submit a new request, you can upload the file when you fill in the form.

6. Copy and paste any log messages from the window used to start the interface or the job into the web form (or an e-mail 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-2015-2 → Diagnostics
   - **Mac:** Applications → SchrodingerSuite2015-2 → 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. Upload the file specified in the dialog box to the support web form.

   If you have already submitted a support request, use the upload link in the email response from Schrödinger to upload the file. If you need to submit a new request, you can upload the file when you fill in the form.
4. Upload the error files to the support web form.

   The files should be in the following location:

- **Windows:** `%LOCALAPPDATA%\Schrodinger\appcrash`
  (Choose Start → Run and paste this location into the Open text box.)
  Attach `maestro_error_`*`pid`*`.txt` and `maestro.exe_`*`pid_timestamp`*`.dmp`.
- **Mac:** `$HOME/Library/Logs/CrashReporter`
  (Go → Home → Library → Logs → CrashReporter)
  Attach `maestro_error_`*`pid`*`.txt` and `maestro_`*`timestamp_machinename`*`.crash`.
- **Linux:** `$HOME/.schrodinger/appcrash`
  Attach `maestro_error_`*`pid`*`.txt` and `crash_report_`*`timestamp_pid`*`.txt`.

If a Maestro panel failed to open:

1. Copy the text in the dialog box that opens.
2. Paste the text into the support web form.

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

Dynamostraße 13
D-68165 Mannheim
Germany

Zeppelinstraße 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®