

# Prime 4.0

## User Manual

Prime User Manual Copyright © 2015 Schrödinger, LLC. All rights reserved.

While care has been taken in the preparation of this publication, Schrödinger assumes no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

Canvas, CombiGlide, ConfGen, Epik, Glide, Impact, Jaguar, Liaison, LigPrep, Maestro, Phase, Prime, PrimeX, QikProp, QikFit, QikSim, QSite, SiteMap, Strike, and WaterMap are trademarks of Schrödinger, LLC. Schrödinger, BioLuminate, and MacroModel are registered trademarks of Schrödinger, LLC. MCPRO is a trademark of William L. Jorgensen. DESMOND is a trademark of D. E. Shaw Research, LLC. Desmond is used with the permission of D. E. Shaw Research. All rights reserved. This publication may contain the trademarks of other companies.

Schrödinger software includes software and libraries provided by third parties. For details of the copyrights, and terms and conditions associated with such included third party software, use your browser to open [third\\_party\\_legal.html](#), which is in the docs folder of your Schrödinger software installation.

This publication may refer to other third party software not included in or with Schrödinger software ("such other third party software"), and provide links to third party Web sites ("linked sites"). References to such other third party software or linked sites do not constitute an endorsement by Schrödinger, LLC or its affiliates. Use of such other third party software and linked sites may be subject to third party license agreements and fees. Schrödinger, LLC and its affiliates have no responsibility or liability, directly or indirectly, for such other third party software and linked sites, or for damage resulting from the use thereof. Any warranties that we make regarding Schrödinger products and services do not apply to such other third party software or linked sites, or to the interaction between, or interoperability of, Schrödinger products and services and such other third party software.

May 2015

# Contents

---

Document Conventions .....	ix
Chapter 1: Introduction .....	1
1.1 Using Prime from Maestro .....	1
1.2 Prime Documentation .....	3
1.3 Running Schrödinger Software .....	4
1.4 Starting Jobs from the Maestro Interface .....	6
1.5 Citing Prime in Publications .....	7
Chapter 2: Using Prime–Structure Prediction .....	9
2.1 Prime Runs and Maestro Projects .....	9
2.2 General Panel Layout .....	10
2.3 The Menu Bar .....	12
2.3.1 The File Menu .....	12
2.3.2 The Edit Menu .....	13
2.3.3 The Display Menu .....	13
2.3.4 The Step Menu .....	16
2.4 The Prime Toolbar .....	16
2.5 The Prime Sequence Viewer .....	17
2.6 Prime Job Options and Control .....	20
2.6.1 The Job Options Dialog Box .....	20
2.6.2 Setting Up and Launching Jobs .....	20
2.6.3 Monitoring Jobs .....	21
2.6.4 Adding Structures to the Project Table .....	21
Chapter 3: Structure Prediction: Initial Steps .....	23
3.1 Input Sequence Step .....	23
3.2 Find Homologs Step .....	25
3.2.1 Importing Homologs .....	25
3.2.2 Searching for Homologs .....	25

3.2.3 Find Homologs Search Options .....	27
3.2.4 Search Results.....	28
3.2.5 Multiple Templates and Structure Alignment.....	31
3.2.6 Continuing to the Next Step .....	32
3.2.7 Saving Runs With Different Templates.....	32
<b>Chapter 4: Structure Prediction—Edit Alignment.....</b>	<b>33</b>
<b>4.1 Overview of the Edit Alignment Step.....</b>	<b>33</b>
<b>4.2 Using Existing Alignments .....</b>	<b>33</b>
<b>4.3 Generating an Alignment .....</b>	<b>35</b>
4.3.1 Generating SSPs .....	35
4.3.2 Importing and Exporting SSPs .....	35
4.3.3 Adjusting SSPs .....	36
4.3.4 Running the Alignment.....	37
4.3.5 Align Program Technical Details .....	38
<b>4.4 Manually Editing an Alignment.....</b>	<b>39</b>
<b>4.5 Error Messages.....</b>	<b>39</b>
<b>Chapter 5: Structure Prediction—Build Structure .....</b>	<b>41</b>
<b>5.1 The Build Process .....</b>	<b>41</b>
<b>5.2 Choosing the Model Type .....</b>	<b>41</b>
5.2.1 Building a Composite Model .....	42
5.2.2 Building a Consensus Model .....	43
5.2.3 Building Multimers.....	43
<b>5.3 Preparing to Build .....</b>	<b>44</b>
5.3.1 Including Ligands and Cofactors.....	44
5.3.2 Setting Constraints.....	44
5.3.3 Build Options.....	45
<b>5.4 Building the Structure .....</b>	<b>46</b>
<b>5.5 Technical Notes for the Build Structure Step.....</b>	<b>47</b>

---

<b>Chapter 6: Refining Protein Structures</b> .....	<b>49</b>
<b>6.1 Preparing Structures for Refinement</b> .....	<b>50</b>
<b>6.2 The Refinement Panels</b> .....	<b>53</b>
<b>6.3 Refining Loops</b> .....	<b>53</b>
<b>6.4 Predicting Side Chains</b> .....	<b>56</b>
<b>6.5 Minimizing Structures</b> .....	<b>57</b>
<b>6.6 Minimizing Parts of Structures as Rigid Bodies</b> .....	<b>58</b>
<b>6.7 Hybrid Monte Carlo Conformational Sampling</b> .....	<b>60</b>
<b>6.8 Refining Protein-Ligand Complexes</b> .....	<b>60</b>
<b>6.9 Calculating the Energy</b> .....	<b>62</b>
<b>6.10 Setting Refinement Options</b> .....	<b>62</b>
6.10.1 Common Options .....	62
6.10.2 Loop Refinement Options .....	64
6.10.3 Minimization Options.....	66
6.10.4 Hybrid Monte Carlo Options .....	67
6.10.5 Protein-Ligand Complex Refinement Options.....	68
6.10.5.1 Local Optimization Options.....	68
6.10.5.2 Covalent Ligand Sampling Options .....	68
6.10.5.3 Minimization Monte Carlo Options .....	68
6.10.5.4 Hierarchical Optimization Options .....	70
<b>6.11 Setting Up an Implicit Membrane</b> .....	<b>70</b>
<b>6.12 Applying Constraints</b> .....	<b>73</b>
<b>6.13 Technical Details for Refinement</b> .....	<b>76</b>
6.13.1 Loop Refinement.....	76
6.13.2 Cooperative Loop Refinement .....	77
6.13.3 Side-Chain Prediction .....	78
<b>6.14 Refinement with Nonstandard Residues</b> .....	<b>79</b>
<b>6.15 Output of Refinement Jobs</b> .....	<b>80</b>

Chapter 7: Protein Structure Alignment.....	81
7.1 The Protein Structure Alignment Panel.....	81
7.2 The Align Binding Sites Panel .....	83
Chapter 8: Prime MM-GBSA.....	87
8.1 Running Prime MM-GBSA from Maestro .....	87
8.2 Running Prime MM-GBSA from the Command Line.....	89
8.2.1 Command Input File.....	90
8.2.2 ASL Expressions.....	90
8.3 Prime MM-GBSA Output .....	90
Chapter 9: Visualizing Prime Energy Terms.....	93
Chapter 10: Command Syntax.....	97
10.1 blast—Run Blast Searches .....	97
10.2 prime—Build and Refine Protein Structures .....	99
10.2.1 Command Input File.....	100
10.2.2 Constraint Syntax.....	110
10.2.3 Restarting an Incomplete Loop Refinement Job.....	111
10.2.4 Environment Variables .....	111
10.2.5 Examples .....	112
10.2.5.1 Model Building .....	112
10.2.5.2 Energy calculation.....	112
10.2.5.3 Side-Chain Prediction .....	113
10.2.5.4 Minimization .....	113
10.2.5.5 Loop Prediction.....	113
10.3 ska—Protein Structure Alignment .....	114
10.4 ssp—Secondary Structure Prediction.....	119
10.5 sta—Single Template Alignment .....	120
10.6 tfm—Tertiary Folding Module .....	123

---

Chapter 11: Command-Line Utilities .....	127
11.1 Multiple Template Alignment: <code>structalign</code> .....	127
11.2 Format Conversion: <code>seqconvert</code> .....	129
11.3 Structure Alignment: <code>align_binding_sites</code> .....	129
11.4 Fast Loop Building: <code>buildloop</code> .....	129
11.5 Database Update Utilities .....	130
11.5.1 <code>update_BLASTDB</code> .....	131
11.5.2 <code>rsync_pdb</code> .....	131
Appendix A: Third-Party Programs .....	133
A.1 Location of Third-Party Programs .....	133
A.1.1 BLAST/PSI-BLAST, Pfam, and the PDB .....	133
A.1.2 Secondary Structure Prediction .....	134
A.2 Third-Party Program Use Agreement .....	134
Appendix B: Environment Variables .....	135
Appendix C: File Formats .....	137
C.1 Input File Formats .....	137
C.2 Output File Formats .....	137
Appendix D: Error Messages and Warnings .....	139
D.1 Input Sequence .....	139
D.2 Find Homologs .....	139
D.2.1 Error Messages .....	139
D.2.2 Warnings in Homologs Table .....	140
D.3 Edit Alignment .....	141
D.4 Build Structure .....	141
D.5 Refinement .....	141

**Contents**

---

Getting Help .....	143
Glossary.....	147
Index.....	149



---

# 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	<code>\$SCHRODINGER/maestro</code>	File names, directory names, commands, environment variables, command input and output
Italic	<i>filename</i>	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

Prime 4.0 is a highly accurate suite of programs for protein structure prediction. The Homology Modeling workflow incorporates the complete protein structure prediction process from template identification, to alignment, to model building. Refinement can then be done from a separate set of panels, and includes side-chain prediction, loop prediction, binding-site refinement, and minimization. The MM-GBSA method can be used for prediction of the binding energy and strain energy of a set of ligands in a protein-ligand complex. The structural alignment tools can be used to align two or more proteins so that the similar regions are aligned.

The Prime interfaces were designed to accommodate both novice and expert users, while the underlying programs were designed to produce superior results in a variety of applications, including:

- High-resolution homology modeling
- Refinement of active sites
- Induced-fit optimization
- Structure-based functional annotation
- Generation of alternate loop conformations
- Ligand binding energies
- Protein structural alignment

## 1.1 Using Prime from Maestro

Maestro is the graphical user interface (GUI) for all of Schrödinger's products. Maestro has several "profiles" that present different layouts of the interface. For protein modeling the main profiles are the Maestro profile and the BioLuminate profile.

Prime provides the tools for three main tasks: homology modeling, structure refinement, and binding energy estimation, all of which are accessible from the Maestro interface.

**Homology modeling** can be done through one of two interfaces, the Structure Prediction wizard, and the Multiple Sequence Viewer. In both you can perform the required steps leading from a protein sequence to the construction of a 3D structure. For cases where the homology is high and adjusting the alignment is not necessary, the Homology Model interface is available. This interface is documented in [Chapter 12](#) of the *BioLuminate User Manual*.

To open the Structure Prediction wizard interface:

From Maestro:

- Choose Applications → Prime → Structure Prediction or Tasks → Homology Modeling in the main window, then click the Structure Prediction Wizard icon in the Homology Modeling panel.



From BioLuminate:

- Choose Tasks → Homology Modeling → Advanced Homology Modeling in the main window.

To open the Multiple Sequence Viewer interface:

From Maestro:

- Choose Applications → Prime → Structure Prediction or Tasks → Homology Modeling in the main window, then click the Multiple Sequence Viewer icon in the Homology Modeling panel.



From BioLuminate or Maestro:

- Choose Tools → Multiple Sequence Viewer in the main window.

For information on using this interface see the [Multiple Sequence Viewer](#) document.

To open the Homology Model interface:

From Maestro:

- Choose Applications → BioLuminate → Homology Modeling → Simple Homology Modeling in the main window.

From BioLuminate:

- Choose Tasks → Homology Modeling → Simple Homology Modeling in the main window.

**Structure refinement** is a facility for refining protein structures that have been imported into Maestro or added to the Project Table from the Homology Modeling workflow or another Schrödinger application.

To open the Refinement panel:

From Maestro:

- Choose Applications → Prime → Refinement or Tasks → Protein Refinement in the main window.

From BioLuminate:

- Choose Tasks → Loop and Side-Chain Prediction in the main window.

**Binding energy estimation** can be done with Prime MM-GBSA which is a facility for calculating ligand binding energies within the MM-GBSA continuum solvation model.

To open the Prime MM-GBSA panel:

From Maestro:

- Choose Applications → Prime → MM-GBSA or Tasks → Binding Energy Estimation → MM-GBSA in the main window.

From BioLuminate:

- Choose Tasks → Ligand Tasks → Binding Energy Estimation → MM-GBSA in the main window.

## 1.2 Prime Documentation

The *Prime User Manual* describes the Prime suite, including information about:

- Essentials of using the Prime–SP module: runs and projects, job control, and the layout and common features of step panels
- Running a structure prediction: the Input Sequence, Find Homologs, Edit Alignment, and Build Structure steps
- Using the Prime–Refinement module
- Running Prime MM-GBSA calculations
- Aligning protein structures
- Running Prime jobs from the command line

This document expands on the information provided in the Maestro online help. In addition to describing steps, features, and options, this document suggests ways to determine which choices are the most useful, describes results, and provides some background and technical detail. Terms pertinent to Prime are defined in the [Glossary](#).

For a tutorial introduction to Prime, see the [Prime Quick Start Guide](#). For information on installing Prime and the related third-party products, see the [Installation Guide](#).

Maestro online help includes topics for Prime. See [Getting Help](#) for a discussion of help in Maestro, including tool tips (Balloon Help) and technical support contacts. For a fuller description of the help facility, see [Chapter 18](#) of the *Maestro User Manual*.

### 1.3 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"
```

To launch Structure Prediction jobs from Maestro, you must have access to the PDB, either in a local installation or to the web site. You must set an option in the Jobs tab of the Preferences panel in Maestro to use web-based searching—see [Section 14.13](#) of the *Maestro User Manual*. If you are using a local installation and the PDB database is not installed into `$SCHRODINGER`, you must set the `SCHRODINGER_PDB` environment variable to the path to the PDB installation.

For information on how to set environment variables, see [Appendix A](#) of the *Installation Guide*.

### 1.4 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*.

The Structure Prediction panel does not have a Job toolbar, as it has its own buttons for running the jobs that are part of the workflow.

## 1.5 Citing Prime in Publications

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

Prime, version 4.0, Schrödinger, LLC, New York, NY, 2015.



# Using Prime–Structure Prediction

This chapter presents the essentials of using Prime–Structure Prediction (Prime–SP):

- Starting, naming, and saving Prime–SP runs in Maestro projects
- Navigating within and between runs
- Job control, job monitoring, and Prime–SP job options
- Using Prime–SP step panels, including their common layout and features
- Using the Prime menu bar and Prime toolbar

## 2.1 Prime Runs and Maestro Projects

A single execution of the Prime structure prediction workflow using a particular set of choices of templates, paths, and settings is called a *run*.

Prime runs are stored within Maestro projects. Structures generated in completed Prime runs are added to the Maestro Project Table. If you are unfamiliar with Maestro projects and the Maestro Project Table, you may want to review the Project Table and Project Facility online help topics. For more information, see the *Maestro User Manual*.

Maestro always has a project open. If you begin to work without selecting or naming a project, Maestro creates a *scratch* (unnamed) project.

Prime always has a run open, which belongs to the current project (whether it is a named project or a scratch project). If you begin the Prime workflow without specifying an existing run, the default name for the current run is `run1`. Multiple runs can be performed within a project, but you must save the project in order to save the data in the runs.

Each project can be displayed in the Project Table panel. In Prime, Project Table entries are usually finished model structures with their properties. Prime stores data in the project as each step in a run is performed, but most of this data is not displayed in the Project Table, which can remain empty until a run is completed. For this reason you should save scratch projects in which you have done work, even if the Project Table is empty.

Two Maestro projects can be merged into one. If you merge two Maestro projects, each of which contains one or more Prime runs, all the runs are included in the merged project. If two runs have the same name, Maestro automatically makes the names unique by appending `-mrgr1` to one of them.

Maestro projects containing multiple runs are useful for comparing the results of workflows in which different selections and options were chosen: for example, a different template selected for building, or a job performed with a different set of options. Creating these multiple runs may involve navigating to an earlier step in your current run, making different selections, and then navigating forward again in a new run. For a description of the features used in navigation (the Next button, the Back button, and the Guide), see [Section 2.2 on page 10](#).

If you return to an earlier step in your current run and make different choices as you proceed forward, you will be asked whether you want to overwrite your existing data, ignore changes, or cancel. If you want to use the new choices but keep the data in your current run for comparison rather than overwriting it, you can start a new run. The commands for saving the old run, starting a new one, and switching between Prime runs are on the File menu.

The File menu commands for managing runs are described in [Section 2.3.1 on page 12](#).

### **To backtrack and try a different set of choices without overwriting your current workflow:**

1. Choose Save As from the File menu.

This saves the current run under the old name (`run1` is the default for the first run in a new project) and creates a copy under the new name.

You are now in a new workflow identical to the first except for its name.

2. Navigate to the step where you want to make a different choice. If you want to go back more than one step, click the step name in the Guide.

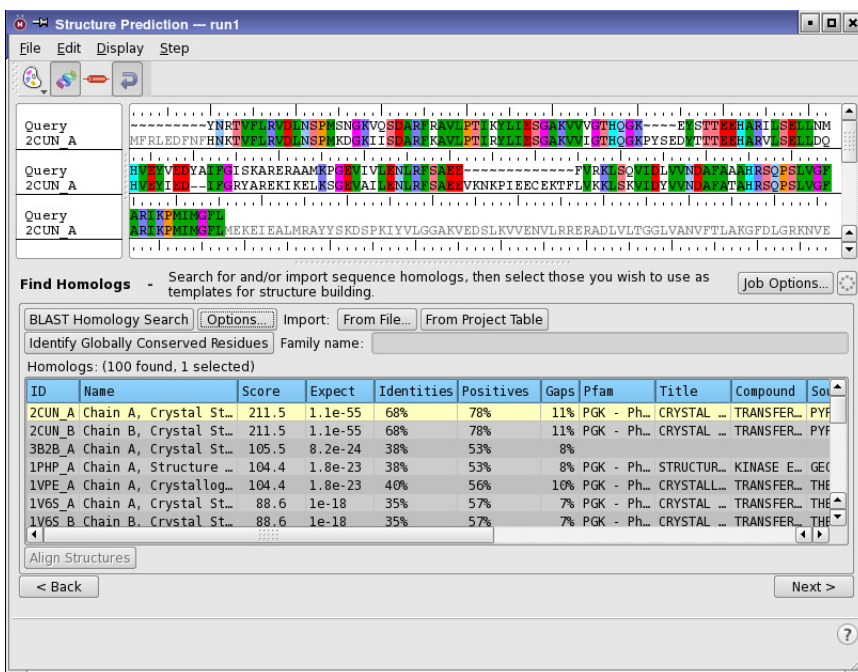
## 2.2 General Panel Layout

Each step in the Structure Prediction panel has a common set of features as well as features that are unique to the step. This section discusses these common elements, their location, and their function. An example of a Prime–SP panel is shown in [Figure 2.1](#).

In the upper part of the panel are the menu bar, the toolbar, and the Prime sequence viewer. These are described in the following sections.

The center of the panel contains the Step View, which includes the step name and a brief description of the step, followed by a row of buttons, and data relevant to the step.

The step usually includes some action buttons, one of which runs the main action of the step, which may involve launching a job. See [Section 2.6 on page 20](#) for more information on running Prime jobs. An Options button may be located to the right of the action buttons. Most buttons, including those in the Prime toolbar, the Prime Guide, and the Maestro toolbar, have tool tips that you can view by pausing the cursor over the button.



**Figure 2.1. The Find Homologs step of the Structure Prediction panel.**

The job status icon to the right of the step description spins when you launch a job. When the icon stops spinning, the job is finished. Click the icon to open the Monitor panel. For more information on job control, see [Section 2.6 on page 20](#).

The data relevant to the step is usually shown in table format. You can view the complete text of a truncated table entry as a tool tip by pausing the pointer over the text. To resize a table column, drag a column border. You can sort the tables in the Step View by clicking the column heading, once for ascending order, twice for descending order. You can also export data from any table in CSV or HTML format from the table shortcut menu.

Below the Step View are the Back and Next buttons, the primary way to navigate between steps. You can perform a simple workflow by clicking Next after completing each step. However, it is sometimes useful to go back to previous steps, make different choices, and proceed forward again using the Next button or the Guide, explained next.

The Guide, below the Back and Next buttons, is a diagram of the Prime workflow as a series of steps, each step represented by a button. It is hidden by default. To display the Guide, select the Guide check box in the Step menu.

The Guide has two purposes: to show your current location in the Prime–SP process, and to let you navigate back and review or change the input at previous steps in the process. The current step is highlighted in the Guide diagram. All previously accessed steps, forward or backward, are available. You can go to any of these steps by clicking them in the Guide.

Clicking on an available step button in the Guide displays the panel for that step. If a step is not yet available, its button is dimmed. Available steps include those that have already passed through in the current run, the present step, and (once step-specific criteria have been met) the next step beyond the current location. When the next step is available, clicking on it behaves the same as clicking the Next button; likewise clicking on the previous step behaves the same as clicking the Back button.

When you return to a previous step and change data, then move forward in the workflow to the following step, you are given a choice of what to do with the existing data in the following step. You can choose **Overwrite**, to discard the existing data and use the new data, **Ignore Changes**, to discard the new data and use the existing data, or **Cancel**, to remain in the current step.

**Note:** You cannot launch searches and jobs by clicking the steps in the Guide. To launch actions pertaining to a step, click the appropriate buttons within the step panel.

Below the Guide is the status area, which shows text related to the current action or the action performed by a control in the panel, and the help button.

## 2.3 The Menu Bar

The menu bar consists of four menus: File, Edit, Display, and Step. Some of the options in these menus are also available as toolbar buttons.

### 2.3.1 The File Menu

The File menu contains commands for managing runs, exporting sequences and setting job options. The commands are described below.

New

Asks for a new run name (all run names must be unique), and starts a new run at the Input Sequence step.

#### Save As

Copies the current run under a new run name and switches to the new run. The original run is saved under the original name. If you open a new project and begin a workflow without explicitly giving the run a name, it will be given the default name of run1.

#### Rename

Asks for a new name for the current run. All run names must be unique.

#### Delete

Deletes the current run, on confirmation.

#### Open

Displays options for switching to another run. The list includes the four most recently used runs. If there are more than four runs in the current project, click **More** to select a run by name.

#### Export Sequences

Exports the sequences in the Prime sequence viewer to a file.

#### Job Options

Opens the Structure Prediction - Job Options dialog box so that you can set options for running jobs for each stage of the workflow. See [Section 2.6 on page 20](#) for information on job options.

### 2.3.2 The Edit Menu

The Edit menu contains commands for editing sequences and secondary structure predictions. These editing tools are also available from the Prime toolbar, and are described with the toolbar in [Section 2.4 on page 16](#).

### 2.3.3 The Display Menu

The Display menu includes the following commands for controlling the appearance of the sequence viewer and structures in the Maestro Workspace.

#### Color Scheme

This command opens a menu of coloring options. In Prime–SP, the coloring schemes act as modes; that is, only one color scheme can be used at any given time. Each step in Prime has a default color scheme. You can use this menu to choose a different one. When amino acid sequences are added to the sequence viewer, they are colored according to the scheme that is currently selected.

When the structure associated with a sequence is displayed in the Workspace, it is generally colored using the same scheme as the sequence. However, protein structures in the Workspace are sometimes colored using other color schemes:

- When a PDB structure is imported, it is displayed in the Workspace using an error reporting color scheme in which complete recognized residues are gray, incomplete residues are red, unknown or nonstandard residues are orange, residues with multiple location indicators are green, and recognized residues with some unrecognized atoms are blue. For more information about conversion of PDB structures, see [Section 3.1.7](#) of the *Maestro User Manual*.
- When the Build Structure step is complete, a color scheme is applied to the structure in which blue atoms are those derived directly from the template, cyan atoms have backbone atoms from the template, and predicted side chains, and red atoms have been predicted or modeled.



The Prime color schemes are:

- **None:** All sequences are colored standard gray.
- **Color by Sequence:** The query sequence is colored standard gray. Each of the other sequences uses a different color.
- **Residue Type:** Each amino acid sequence is colored according to the array of residue type colors.
- **Residue Property:** Each amino acid sequence is colored according to the array of residue property colors (fewer colors than residue type).
- **Residue Matching:** The query and templates are colored in a pairwise manner. Matching residues are colored according to their residue type, and nonmatching residues are colored standard gray. If more than one template is chosen, the query will have more colored residues because the effect is cumulative. If only the query sequence is shown, it is colored standard gray for all residues.
- **Residue Homology:** Aligned query and template residues that are identical (highly conserved) are colored red. Aligned query and template residues that are similar according to the BLOSUM62 scoring matrix (conserved) are colored yellow. Gray is used for residues with no homology.
- **Secondary Structure:** The query sequence is colored standard gray. All other AA sequences should have SSAs, which are colored according to the SSA colors.



- **Template ID:** Only available for the Comparative Modeling path. The query sequence is colored standard gray. The predicted structure is colored according to template ID for each region (using the cycle colors). Each template sequence is colored with its color, but only in the region that was used (the rest of the template is colored dark gray to indicate it is not used).
- **Proximity:** At any step in which the sequence viewer is displaying sequences that have associated structure (all steps except Input Sequence), color by proximity mode is available. When you select this color scheme, the current sequence colors do not change immediately. Instead, when the mouse is over a sequence that has structure, the cursor changes to indicate that proximity can be calculated. You can then select one or more contiguous residues in the sequence, and the system will color that sequence based on the proximity to the selected residues. No other sequences will have modified colors (even if they have structure too and are in proximity of the selected sequence).

Proximity is calculated based on the shortest distance between any two atoms in two residues, not including the bonded C and N atoms in the residues next to the selected residues.

To change the proximity cutoff, choose Preferences from the Display menu and change the number in the Proximity cutoff text box.

### Font Size

The font size of the text in the sequence viewer can be changed using this option. The menu offers a selection of point sizes for the text. The default is 10 points.

### View Template Structure

View the template structures in the Maestro Workspace.

### View SSA

This option controls whether secondary structure assignments are shown in the sequence viewer. The default is not to show secondary structure assignments.

### View Full Pfam Annotation

Select this option to view the full Pfam annotation. By default, only the highly conserved residues are shown in the annotation.

### Legend

Opens the Legend panel, which includes the assignment of each color and symbol used in each color scheme and sequence representation. Click the Colors tab and select a Color scheme

from the list, or click the Symbols tab and select a type of Sequence data, such as Pfam/HMMER or SSP. This option is also available when you right-click on a sequence in the sequence viewer.

### Preferences

Opens the Preferences dialog box, which has a single item:

- Proximity cutoff—The value entered in this text box defines proximity for the Proximity color scheme. The default is 4.00 Å.

### 2.3.4 The Step Menu

The Step menu provides another way to navigate between steps in Prime. All the steps are listed on the menu. The current step is selected. To go to a different step, select that step on the menu. Steps that are not available are dimmed.

You can also show or hide the Guide, by selecting Guide on this menu. When the check box is selected, the Prime Guide is displayed. Deselect the check box to hide the Guide. The Guide is hidden by default.

## 2.4 The Prime Toolbar

The Prime toolbar is the row of buttons just under the Prime menu bar. The Prime toolbar provides convenient access to commonly used Prime tasks. Pause the pointer over a button to see its name in a tool tip and its description in the status bar. These tasks are also available from the Edit and Display menus. The toolbar buttons are shown below as they appear on the toolbar, with their names, the corresponding menu, and their functions.



Select (Edit menu): Select a region of the sequence to see the corresponding residues highlighted in the 3D Workspace.



Crop (Edit menu): The cropping tool, which can be used in the Build Structure step to select ends of the query sequence that are not to be used for building.



Edit SSP (Edit menu): Edit secondary structure predictions. Highlight a region of the SSP, and then type either H for  $\alpha$ -helix, E for  $\beta$ -strand, or - for loop.



Edit Query Sequence (Edit menu): Edit the query sequence, for example to remove extra terminal residues, remove insertions, or mutate residues. Opens a dialog box in which you can edit the sequence as a text string.



**Slide Freely (Edit menu):** Use this button to edit the alignment between query and template. Selecting a residue and sliding to the right creates N-terminal gaps. Selecting a residue and sliding to the left creates C-terminal gaps.



**Slide As Block (Edit menu):** Use this button to edit the alignment between query and template. Selecting a residue and sliding to the right creates N-terminal gaps. Selecting a residue and sliding to the left slides all residues as a block to the left.



**Add Anchors (Edit menu):** Set alignment constraints by setting anchors at residue positions. To remove anchors, click on the same residue again.



**Lock Gaps (Edit menu):** To prevent gaps from being collapsed during manual editing, lock the gaps using this button. Locked gaps are indicated with a - symbol.



**Unlock Gaps (Edit menu)** To allow gaps to collapse during manual editing, unlock the gaps using this button. Unlocked gaps are indicated with a ~ symbol.



**Color Scheme (Display menu):** Select a color scheme for coloring sequences in the sequence viewer. If there is a structure in the 3D Workspace associated with the sequence, it will also be colored.



**View Template Structure (Display menu):** Use this button to display and undisplay template structures in the Workspace.



**View SSA (Display menu):** Use this button to turn the secondary structure assignment (SSA) in the sequence viewer on and off.



**Wrap Sequences (Display menu):** Wrap sequences that are longer than the display area of the sequence viewer.

## 2.5 The Prime Sequence Viewer

The sequence viewer displays the sequences for the query and templates, and in some steps it shows secondary structure assignments (SSAs) and secondary structure predictions (SSPs). It is divided into two sections. The name of the sequence, SSA, or SSP is given in the left section, and the sequence, SSP, or SSA itself is displayed in the right section. The sequence viewer includes a shortcut menu (right click) that provides access to some functions.

To resize the sequence viewer, drag the dotted line below the sequence viewer. If you cannot enlarge it any further, drag the bottom border of the panel to make the panel larger, and then you will be able to enlarge the sequence viewer some more. You can also drag the dotted line between the right and left sections of the sequence viewer to resize the sections.

The sequence viewer can be configured to display a single line for each sequence that can be scrolled horizontally, or the sequences can be wrapped so that they scroll vertically. The configuration is controlled by the **Wrap Sequences** option on the sequence viewer shortcut menu.

The sequence viewer also includes a “ruler”, which marks the residue position relative to the leftmost residue displayed in the sequence viewer. This residue position is displayed in parentheses in the tool tip for each residue. The positions are arbitrary, and can change when you display residues in the viewer. They are intended to help you keep track of the sequence as you scroll the sequence viewer.

You can select a single residue in a sequence by clicking on it, and you can use control-click to select or deselect individual residues. To select a range of residues, you can click the first and shift-click the last, or drag over the residues. The sequence in which the residues are selected is determined by where you start to drag. Likewise, you can invert the selection of multiple residues by holding down CTRL when you drag.

The commands that are available on the shortcut menu are:

### Show All/Hide All

This command is present when you right-click on a sequence or an SSP. It can be used to show all the family and SSP information that is associated with a sequence. It is equivalent to clicking the + or – sign to the left of the sequence name.

### Run SSP

Run a secondary structure prediction for a sequence. This command is only present for sequences, in the Find Homologs, Fold Recognition, and Edit Alignment steps.

### View SSA

Show or hide the secondary structure assignment. This command is only present when you right-click on an SSA or a sequence that has an SSA.

### View Full Pfam Annotation

Show the full Pfam annotation. By default, only the highly conserved residues are shown in the annotation. This command is only present when you right-click on a Pfam annotation.

### Revert

Revert the SSP to that generated by the SSP program. This command is useful to restore the original SSP after making changes. It is only present for SSPs, and is only available in the Edit Alignment step.

### Delete

Delete the SSP that you right-clicked on. This command is only present for SSPs, and is only available in the Edit Alignment step.

### Export Single Sequence

Export an SSP to a file. A file selector opens, in which you can navigate to a location and name the file. This command is only present for SSPs.

### Save Image

Save an image of the sequence viewer. A file selector opens, in which you can select a format, from PNG, TIFF, and JPEG, and save the image. This command is always present on the shortcut menu.

### Export Sequences

Export the sequences to a single FASTA file. A file selector opens, in which you can navigate to and name the file. You can also export sequences by choosing File → Export Sequences. This command is always present on the shortcut menu.

### View Template Structure

Display the structure for the sequence in the Workspace. This command is only present when there is a structure for the sequence. The query sequence does not have a structure associated with it until a structure has been built. The template sequences do have structures, which can be displayed.

### Wrap Sequences

When this option is selected, the sequences in the sequence viewer are wrapped, so that you can scroll vertically through the sequences. When it is deselected, the sequences are displayed on a single line, which you can scroll horizontally. This command is always present on the shortcut menu.

### Legend

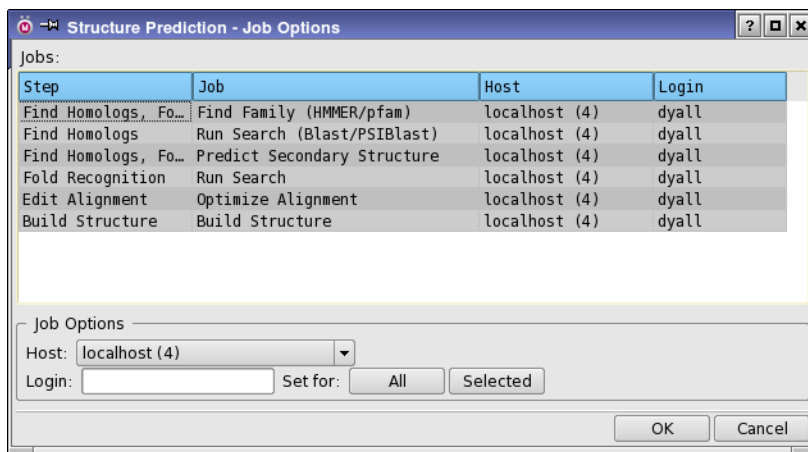
Open the Sequence Viewer Legend panel, which explains the color schemes and the symbols used in the sequence viewer. This command is always present on the shortcut menu.

## 2.6 Prime Job Options and Control

Before launching a Prime calculation from a step panel, you may want to choose which machine host the job will run on and check or modify job settings. When a run is complete, you can incorporate the results into the Maestro Project Table and save them as a project.

### 2.6.1 The Job Options Dialog Box

The Structure Prediction–Job Options dialog box, shown in [Figure 2.2](#), allows you to specify which machine each type of Prime job will run on, and under which user name. To open the dialog box, choose Job Options from the File menu on the Prime menu bar. The Job Options dialog box lists 6 types of jobs, from Find Family (run from the Find Homologs step) to Build Structure (run from the Build Structure step). Specify the host machine and the user name for each type of job. The default host is localhost. Make sure your `schrodinger.hosts` file contains a list of available machines to run on. See the [Installation Guide](#) for information on the `schrodinger.hosts` file.



**Figure 2.2.** The Structure Prediction–Job Options dialog box.

### 2.6.2 Setting Up and Launching Jobs

Though you specify the machines used to run jobs in the Job Options panel, you select Prime job settings and launch jobs from the Prime step panels. This can be as simple as clicking an action button, such as Search. In other steps, you select the type of job from a Task menu and then launch the job by clicking Run.

You can modify settings for step actions or tasks using the Options button. For more information on setting up and launching jobs, see the discussions of specific steps in the chapters that follow.

### **2.6.3 Monitoring Jobs**

After you launch a job, the job status icon in the upper right corner of the step panel turns green and spins.

To monitor a job using the Monitor panel, click the job status icon. You can use this panel to monitor jobs from the current session or your previous sessions.

To kill a job, select the job name and then click Kill. If the job still appears to be running, you may have to go to the Project directory and remove the job file.

For more information about the Monitor panel and monitoring jobs, see [Section 3.2](#) of the *Job Control Guide* and the online help.

### **2.6.4 Adding Structures to the Project Table**

In the Build Structure step, predicted structures can be added to the Maestro Project Table as new entries. This is useful for comparing predicted structures produced from different runs, and for refining structures.





---

# Structure Prediction: Initial Steps

This chapter outlines the initial steps in the Prime–SP workflow. In the Input Sequence step, a query sequence is imported. In the Find Homologs step, potential templates are identified. Later chapters describe the steps in which the templates are aligned and a model is built.

## 3.1 Input Sequence Step

This step is used to read in an amino acid sequence, edit the sequence if desired, and select the query sequence.

### To read a sequence from a file:

- Click File and specify a file containing the desired sequence.

Several file formats are supported, including FASTA, EMBL, GENBANK, PDB, and PIR. For a complete list, see [Appendix C](#). File formats are auto-detected by the program. Sequences must be in standard uppercase one-letter code format.

### To read a sequence from the PDB:

- Click PDB ID and specify the PDB ID and optionally the chain name.

The PDB file is downloaded and the sequences are read in.

### To enter a sequence as text:

- Click Text and type in the sequence or paste it from the clipboard.

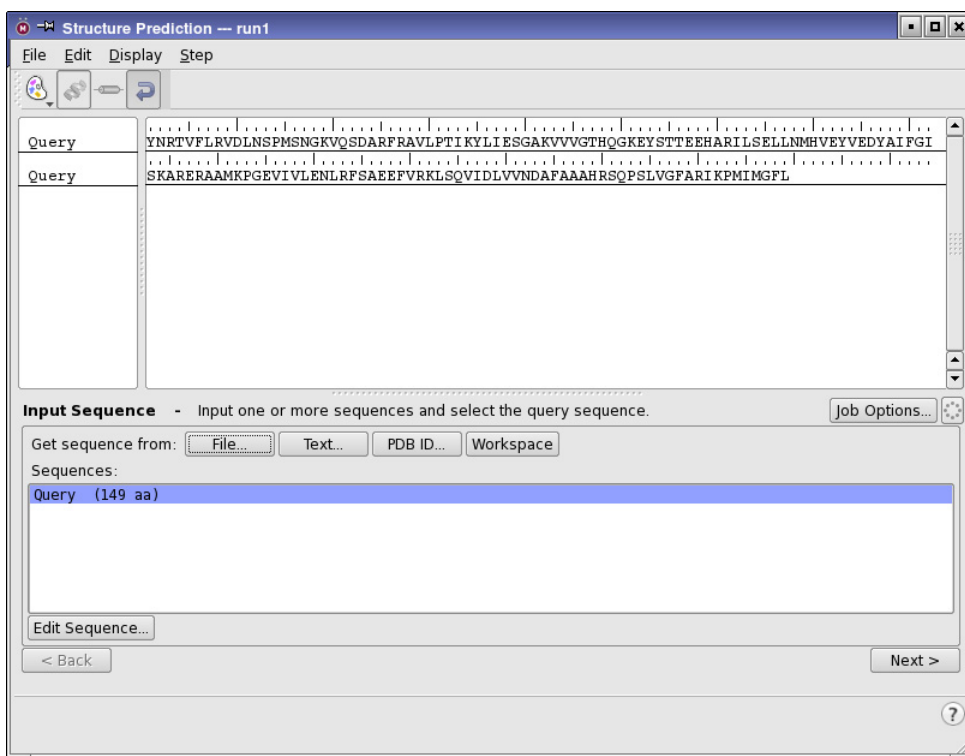
If there are multiple chains with the same sequence, the sequence will be read in only once.

### To read a sequence from the Workspace:

- Click Workspace.

All sequences represented in the Workspace are added to the Sequences table. If there are multiple chains with the same sequence, the sequence is read in only once.

After input, the Sequences table in the Step View area shows all available unique sequences with their corresponding chain lengths in the order in which they are read. By default, the first sequence read in is displayed in the sequence viewer in the upper section of the panel. If multiple sequences have been read in, only one can be brought forward to the next step in a given Prime run. If you want to build a multimer, you can do so by combining information from separate runs for each chain in the Build Structure step.



**Figure 3.1.** The Input Sequence step, after input.

#### To select a query from the Sequences table:

- Click on a sequence to select it for this run.

This brings the selected sequence into the sequence viewer. The sequences are named by extracting the name from either the PDB header or the FASTA file header. Non-unique sequences are ignored. Non-unique names are modified to make them unique.

If you want to modify the sequence before proceeding, for example, to remove extra residues at the termini, delete an insertion, or mutate residues, you can edit the sequence.

**To edit a sequence:**

1. Select the sequence in the Sequences table.
2. Click Edit Sequence.
3. Edit the sequence in the dialog box that opens, and click OK.

**Note:** Sequences longer than 1,000 residues generally include more than one domain. You should consider dividing these sequences into probable domains and treat each domain as an individual query.

**To continue:**

- Click Next to go to the Find Homologs step.

If an error message appears when you click Next, you cannot proceed until the problem is resolved. Error messages are listed in [Appendix D](#). See [Section D.1 on page 139](#) for help with Input Sequence error messages.

## 3.2 Find Homologs Step

This step is used to search for homologs of the query sequence. These homologs can then be used as templates to build a model structure.

### 3.2.1 Importing Homologs

To import homologs, use the Import buttons. You might want to do this for homologs that are not in the NCBI sequence database, or homologs that have been rotated into a common reference frame when multiple templates have been selected. You can import from two sources:

- From File—You can import homologs from a PDB file or a Maestro file.
- From Project Table—Select the desired homologs in the Project Table and click this button. This allows you to import a homolog in any format recognized by Maestro.

Imported homologs are aligned quickly using ClustalW.

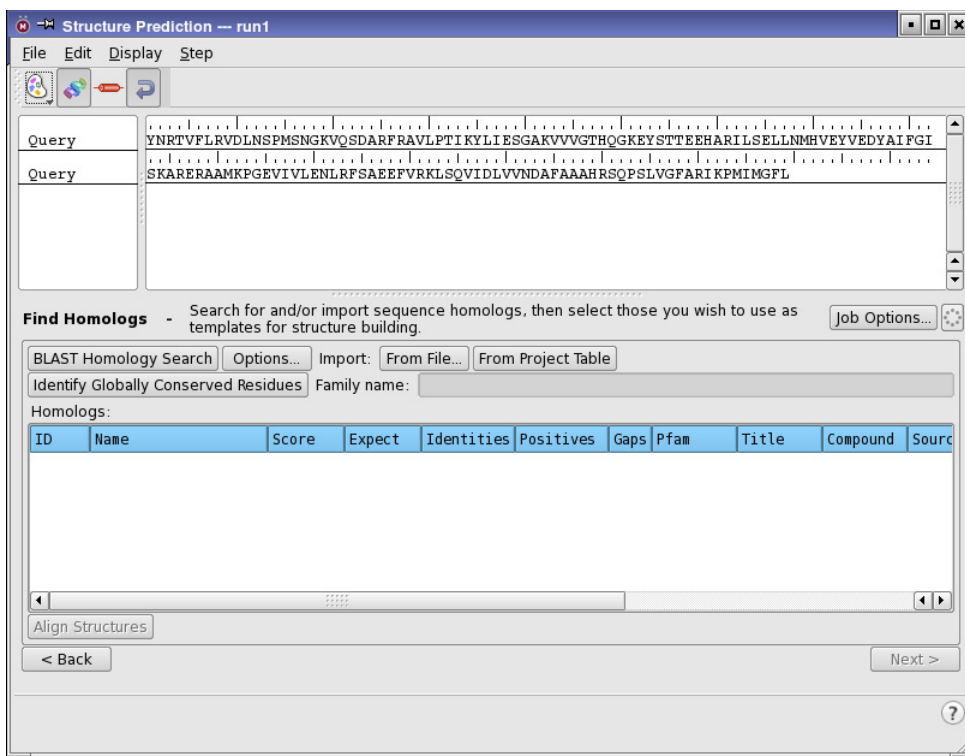


Figure 3.2. The Find Homologs step, initial view.

### 3.2.2 Searching for Homologs

To launch a search for homologs:

- Click BLAST Homology Search (near the middle of the panel).

Depending on what is selected in the Find Homologs–Options dialog box, BLAST or PSI-BLAST is used to search for templates. The default is to use BLAST to search the non-redundant PDB database that has been installed with Prime. If the database is not accessible locally, the BLAST web site can be used to search for templates. To use the web site rather than a local installation, you must allow web searches in the Find Homologs–Options dialog box. For search options, see [Section 3.2.3](#).

**To identify globally conserved residues for the query (optional):**

- Click Identify Globally Conserved Residues to run HMMER on the Pfam database.

A Hidden Markov Model (HMM) is generated from a multiple sequence alignment and used to identify the query family and provide information about which residues are conserved in the consensus sequence. The job should take only a few minutes.

The query family name found appears in the Family name text box with an E-value in parentheses. If the E-value of the family is too large, it is not meaningful to consider the family.

The sequence viewer shows the match between the query and the consensus sequence of the family. If this sequence, labeled *NAME*\_pfam, is not visible, check for a plus sign to the left of the query name: *+NAME*. Click the plus sign to show any hidden sequences. The match sequence displays the residues that are conserved in the multiple sequence alignment used to generate the Hidden Markov Model (HMM). The conserved residues that are present in both the query and the template are colored with the template color. Conserved residues that are present in only one sequence or in neither are not colored. This scheme allows you to easily identify residues in the template or in the query that are not conserved according to the model.

If you want to see the full Pfam annotation, right-click on the annotation and choose View Full Pfam Annotation. The full annotation displays more information about which residues are conserved in the multiple sequence alignment used to generate the HMM. Capital letters indicate highly conserved residues, lowercase letters indicate a match to the HMM, + means the match is conservative, and a blank indicates that the residue does not match the HMM.

### 3.2.3 Find Homologs Search Options

**To change search options:**

- Click Options to open the Find Homologs–Options dialog box.

You can use this dialog box to specify a PSI-BLAST or BLAST search, change the settings for either type of search, and choose whether to run the search on the BLAST web site or use a local installation of the BLAST database. For more information about BLAST, see the documentation on the NIH web site, at <http://www.ncbi.nlm.nih.gov/books/NBK21097/>.

The default database selection is NCBI PDB (all). Other selections include:

- NCBI PDB (non-redundant)—Does not include PDB files with identical sequences.
- NCBI NR—Recommended for PSI-BLAST runs. With this selection, only the sequences with structures available in the PDB are displayed in the Homologs table.

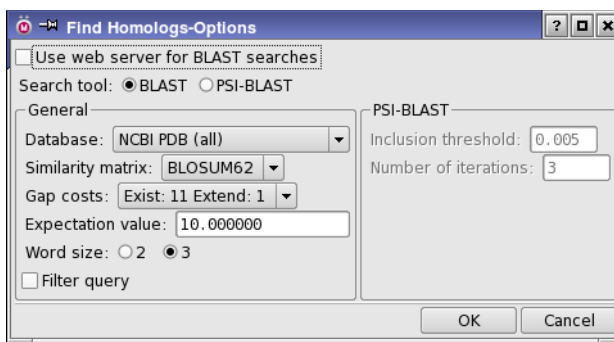


Figure 3.3. The Find Homologs–Options dialog box.

### 3.2.4 Search Results

The output from the homolog search appears in the table labeled Homologs: ( $N$  found,  $n$  selected). For each potential template, the row contains the ID, the name, score, expectation value, % identities, % positives, % gaps, and header information (if taken from a PDB file). All percentages are rounded to the nearest whole number, which may be zero. A description of the table columns is given in Table 3.1.

Table 3.1. Homologs table data

Column Header	Description
ID	The PDB ID code, including chain name. If a chain name is specified, then the chain, not the entire protein, is the potential template.
Name	Sequence name provided by BLAST.
Score	BLAST bit score.
Expect	BLAST expectation value.
Identities (%)	Percentage of residues that are identical between the sequences.
Positives (%)	Percentage of residues that are positive matches according to the similarity matrix selected in the Options dialog (default matrix is BLOSUM62).
Gaps (%)	Percentage of gaps in both query and homolog as returned by BLAST.
Title	From the PDB TITLE field.
Compound	From the PDB COMPND field.
Source	From the PDB SOURCE field.
Experiment	From the PDB EXPDTA field. This gives the experimental technique used to obtain the structure, e.g., X-RAY DIFFRACTION or NMR.



If no coordinates are available for a template, that row is dimmed and cannot be selected.

If a chain ID is given for a homolog (e.g., Chain A in 1EMS\_A), then that chain, not the entire protein, is the potential template. When aligning multiple templates with multiple chains, you may need to extract single chains using the `getpdb` utility, as described in Section 2.3 of the *General Utilities* manual.

Once homologs are found, click on a row to view the structure of that template in the Workspace. Selecting a template brings the sequence alignment of the template and the query sequence into the sequence viewer. Click on another template to deselect the first one and display the selection in the Workspace.

After you review the alignments, select the template that you want to build on by clicking on its row in the table. To select multiple templates, see Section 3.2.5.

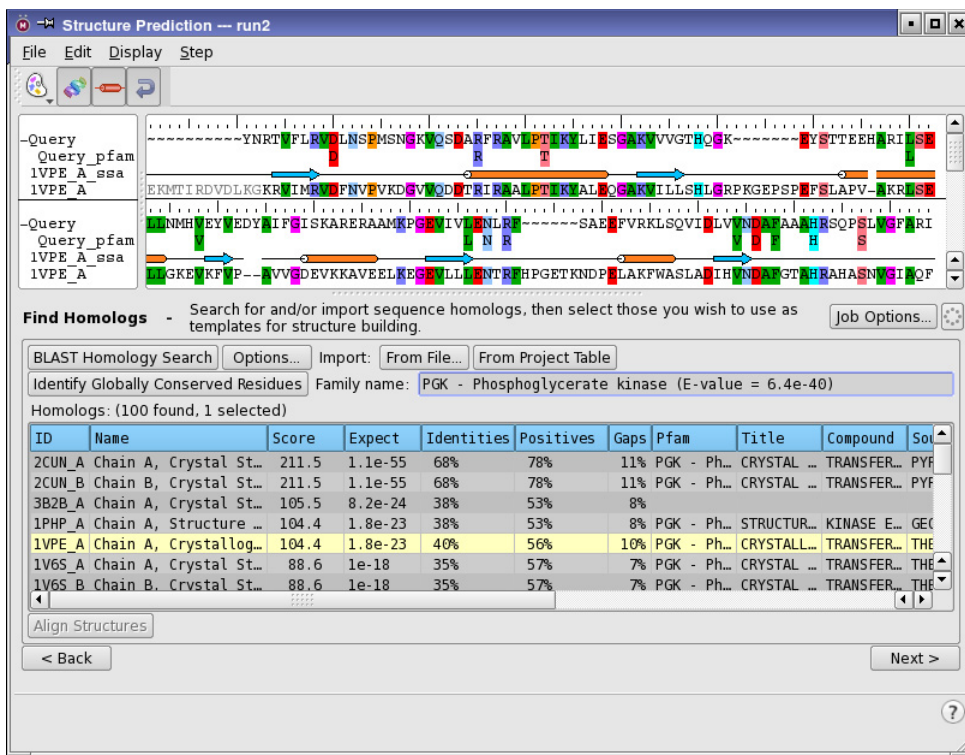


Figure 3.5. The Find Homologs step with Blast and HMMER/Pfam search results.



### 3.2.5 Multiple Templates and Structure Alignment

Multiple templates can be used in several ways in Prime to build a model. For a single query chain, you can build a composite model, where each template contributes to a different part of the structure, or you can build a consensus model, where each template contributes to all parts of the structure. If you want to build a homomultimer, you will need templates for each chain (even though the sequences are identical). For heteromultimers, you will usually only need a single template for each query sequence in a given run. You can select more than one template for building with shift-click and control-click.

If you are building a composite model with more than one template, they will need to be *structurally aligned*, that is, rotated to bring them into a common reference frame, before they can be used in the next step. However, if you are building a multimer of any sort, you should *not* align the templates, because they must already be in the correct spatial relation to each other to construct the multimer from them.

- If you have already created a file with structurally aligned multiple templates, you can import it using the Import button, as described in [Section 3.2.1](#).
- If you have not yet performed structural alignment on the selected templates for a composite model, click Align Structures.

If structure alignment fails or does not give the intended result, the problem may be one of the following:

- The selected templates are structurally too dissimilar for a meaningful alignment.
- One or more of the template structures includes multiple chains, and either
  - Two or more templates are chains from one .pdb file (the structure alignment facility will not align chains from the same file), or
  - The chains that were aligned were not the chains you intended to align.

If the problem involves a template with multiple chains, you can use the `getpdb` utility (see [Section 2.3](#) of the *General Utilities* manual) to extract each chain into a separate .pdb file, then run structure alignment again.

For information on running structure alignment from the command line, see [Section 11.1](#).

### **3.2.6 Continuing to the Next Step**

When you are satisfied with the template (or structurally aligned multiple templates) you have selected, proceed to the Edit Alignment step.

If there is an error, a message appears when you click Next to go to the next step. See [Section D.2 on page 139](#) for a list of error messages for Find Homologs.

### **3.2.7 Saving Runs With Different Templates**

If you decide to try different templates, you can avoid overwriting your existing workflow by starting a new one before you navigate back to the Find Homologs step to choose a different homolog. Choose Save As from the File menu. This creates a copy of the current run, allowing you to select a new template and build on it. You can then compare results from each run.

If you are building a heteromultimer, you must model each unique chain of the multimer, and save a run for each chain. The runs are combined to build the final model.

---

# Structure Prediction—Edit Alignment

## 4.1 Overview of the Edit Alignment Step

In order to build a model structure of the query, the Build Structure step requires a satisfactory alignment between templates and query. The alignment of templates to query generated in the Find Homologs step is based only on sequence. This alignment usually needs some improvement. The Edit Alignment step allows you to improve alignments using various tools before building a model in the Build Structure step.

There are several ways to produce an alignment for use in Build Structure:

- Accept the BLAST/PSI-BLAST alignment generated in the Find Homologs step and currently displayed in the sequence viewer.
- Import an alignment.
- Generate or import secondary structure predictions, then use the Align program to create a new alignment.
- Edit any of these alignments manually.

If you want to identify globally conserved residues to assist in the alignment process, but did not do so in the Find Homologs step, click Identify Globally Conserved Residues to run HMMER/Pfam. For more information on this action, see [Section 3.2.2 on page 26](#).

When you are satisfied with the alignment, click the Next button to proceed to the Build Structure step.

## 4.2 Using Existing Alignments

When you start the Edit Alignment step, the BLAST/PSI-BLAST alignment for the template you selected in the Find Homologs step is displayed in the sequence viewer—see [Figure 4.1](#).

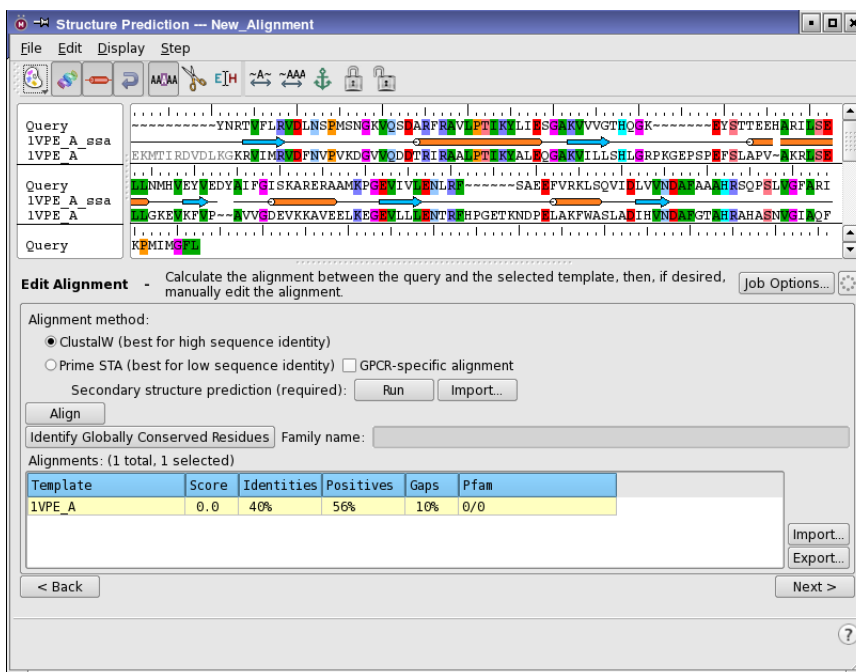


Figure 4.1. The Edit Alignment panel, initial view.

- If you chose a single template and this is the template and the alignment you want to use, click Next to go to the Build Structure step.
- If you chose multiple templates, select the alignment you want to use in the alignments table, and click Next.
- If you want to use an alignment generated in another program (or exported previously from Prime), click Import (to the right of the Alignments table), and navigate to and select the alignment in the Import Alignments file selector.

The sequences in the alignment file must be exactly the same as those used by Prime, and so must the sequence and structure (PDB) names. For details of the allowed formats, see [Appendix C](#).

**Note:** The template PDB ID must be uppercase in the alignment file.

You may want to save the existing alignment for later use before making changes. To do so, click Export, and navigate to the desired location in the Export Alignments file selector and name the file.

## 4.3 Generating an Alignment

If the existing alignments are not satisfactory, you can generate an alignment using one of two programs, listed under Alignment method:

- **ClustalW**—Running this program generates an alignment quickly, from sequence information only. It is best used when there is high sequence identity between the query and the template.
- **Prime STA**—This program generates an alignment based on secondary structure predictions (SSPs) for the query as well as sequence alignments. Once you have these predictions, you can set up and run the alignment program. This program is best used for cases of medium sequence identity (20-50%), though it can of course be used for high sequence identity as well.

### 4.3.1 Generating SSPs

To run all available SSP programs, click Run. One secondary structure prediction program (SSpro) is bundled with Prime. However, the optional third-party program PSIPRED is highly recommended for optimal results, especially for GPCRs. PSIPRED is not available on Windows, but you can use the PSIPRED web site to generate the SSP and then download and import the results. See [Appendix A](#) for more information on this program. If there is a problem running the structure prediction programs, an error message is displayed.

When the SSP run is finished, the secondary structure predictions are displayed in the sequence viewer below the query sequence. If you want to hide the SSPs, right-click on an SSP in the sequence viewer and choose Hide All.

It can be useful to generate or import SSPs even if you perform an alignment with ClustalW, because you might want to use these SSPs to guide a manual adjustment of the alignment.

### 4.3.2 Importing and Exporting SSPs

To import a secondary structure prediction for the query, click the Import button for Secondary structure prediction and select the SSP file, which may be in FASTA, PSIPRED, or Maestro format. If necessary, use the `seqconvert` utility to change the file format—see [Section 11.2](#) for more information. If the file contains bad data or the sequence doesn't match the query sequence, an error message is displayed.

If you want to save an SSP for later use, right-click on it and choose Export Single Sequence. By default, the SSP is saved in native Schrödinger format. If there is a problem exporting a file, an error message is displayed.

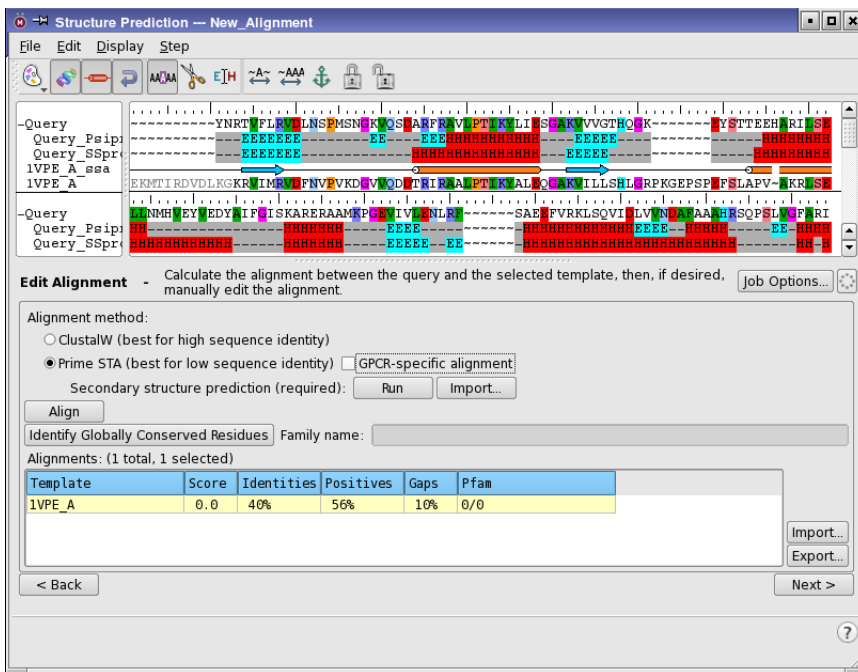


Figure 4.2. The Edit Alignment step, after running SSPs.

### 4.3.3 Adjusting SSPs

If you believe an SSP is incorrect, you can delete it entirely or edit the incorrect portions. You can revert to the original SSP if you are not satisfied with your adjustments.

#### To delete an SSP:

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

#### To edit an SSP:

1. Click the Edit SSP toolbar button:



2. Highlight the part of the SSP you want to change.
3. Type the character e, h, or -.

E represents strand, H represents helix, and – represents loop.

The highlighted region of the prediction is changed to your choice.

**To retrieve the original SSP after editing:**

- Right-click on the SSP and choose Revert from the shortcut menu.

### 4.3.4 Running the Alignment

Before running an alignment, if you want to take advantage of the special capabilities for aligning GPCRs, select GPCR-specific alignment. These capabilities include fingerprint matching, a customized GPCR sequence database, and identification of transmembrane helices. For GPCR alignment, a set of seven transmembrane helices in the SSP is needed. In general, PSIPRED is more reliable than SSPro in producing a satisfactory SSP. If the SSPs for the GPCR are not satisfactory, you can adjust them manually, by using the instructions in the previous section.

To take advantage of your knowledge of the system, one or more pairs of query/template residues that you know should be aligned can be constrained to remain so, by using alignment constraints.

**To constrain a pair of residues to remain aligned:**

1. Click the Add Anchors toolbar button.



2. Select one of the residues in the pair that you want to remain aligned during the Align calculation.

An anchor symbol is displayed in the ruler at the corresponding residue position.

When you are satisfied with the secondary structure predictions for the query, click Align to start the alignment job.

After the job finishes, the Alignments table is updated with the new alignment score, the percent identity, the percent similarity, and the percent gaps, all rounded to the nearest integer. The changed alignment is also reflected in the Maestro Workspace display of the structure.

If you are satisfied with the new alignment, you can continue to the next step by clicking Next. Otherwise, you can edit it manually as described in the next section.

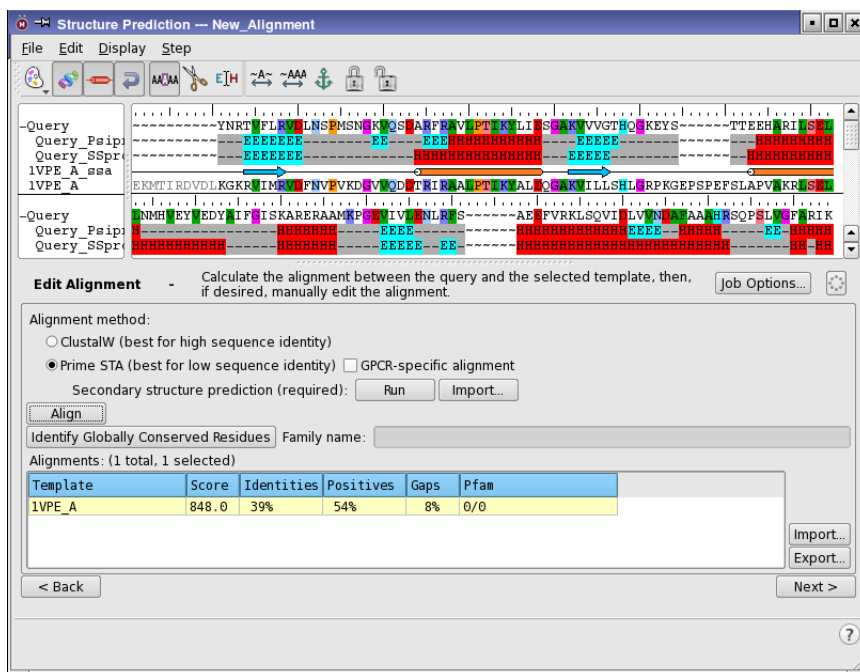


Figure 4.3. The Edit Alignment panel after alignment.

### 4.3.5 Align Program Technical Details

The goal of the program launched from the Edit Alignment step, *sta* (Single Template Alignment), is to generate an accurate alignment between proteins with medium to high sequence identity (20-90%). Features of the program include:

- A position-specific substitutional matrix (PSSM) for the query sequence, derived from PSI-BLAST, is used to match the template sequence.
- In order to minimize the inaccuracy in a single secondary structure prediction, a novel and robust algorithm has been designed to derive a composite secondary structure for the query sequence from multiple predictions. The composite SSP is aligned to the SSA of the template.

See [Section 10.5 on page 120](#) for more information on this program, including instructions for running it from the command line.



## 4.4 Manually Editing an Alignment

You can edit any alignment, however obtained, using combinations of these Prime toolbar buttons: Slide Freely, Slide as Block, Add/(Remove) Anchors, Lock Gaps, and Unlock Gaps.



**Slide Freely (Edit menu):** Edit the alignment between query and template. Selecting a residue and sliding to the right creates N-terminal gaps. Selecting a residue and sliding to the left creates C-terminal gaps.



**Slide As Block (Edit menu):** Edit the alignment between query and template. Selecting a residue and sliding to the right slides all residues as a block to the right. Selecting a residue and sliding to the left slides all residues as a block to the left.



**Add Anchors (Edit menu):** Set alignment constraints by setting anchors at residue positions. To remove anchors, click on the same residue again.



**Lock Gaps (Edit menu):** To prevent gaps from being collapsed during manual editing, lock the gaps using this button. Locked gaps are indicated with a - symbol.



**Unlock Gaps (Edit menu):** To allow gaps to collapse during manual editing, unlock the gaps using this button. Unlocked gaps are indicated with a ~ symbol.

The Workspace shows the template structure colored by query residue property, displayed as a ribbon and side chains. Each template residue is colored according to the property of the corresponding query residue in the alignment, revealing the location of charged, polar, and hydrophobic query residues. You can use this feature to make sure that insertions or gaps are not in the middle of regions of secondary structure such as alpha helices or beta sheets, and to check that hydrophobic residues point into hydrophobic regions and polar residues point towards solvent. As you make changes to the alignment, the Workspace is updated to show the mapping of the query's residue properties onto the template structure.

## 4.5 Error Messages

When you click Next at the end of the Edit Alignment step, the Check Alignment warning dialog box may appear. The warning lists possible alignment problems that you may want to fix before continuing to the next step—for example, gaps in secondary structure elements of the template, or gaps in the query that are aligned to secondary structure elements of the template.

If you do not want to make corrections to the alignment, click Continue to go to the next step. To make corrections based on the information in the error message, make a note of the gap locations given, click Cancel to close the alert box. The Edit Alignment step is still displayed, and you can change the alignment accordingly.



# Structure Prediction—Build Structure

## 5.1 The Build Process

The process of building a structure from one or more templates involves the following steps.

- Copying of backbone atom coordinates for aligned regions and side chains of conserved residues
- Optimization of side chains
- Minimization of non-template residues
- Building of insertions and closing of deletions in the alignment

By default, all of these steps are performed, except that insertions of more than 20 residues are omitted. Also by default, only the protein structure is built: ligands, cofactors and waters are deleted before building. However, you can choose to leave out one or more steps (except the first), and you can choose to build in ligands and cofactors. (Water molecules must be added afterwards.) These choices must be made before you build the structure.

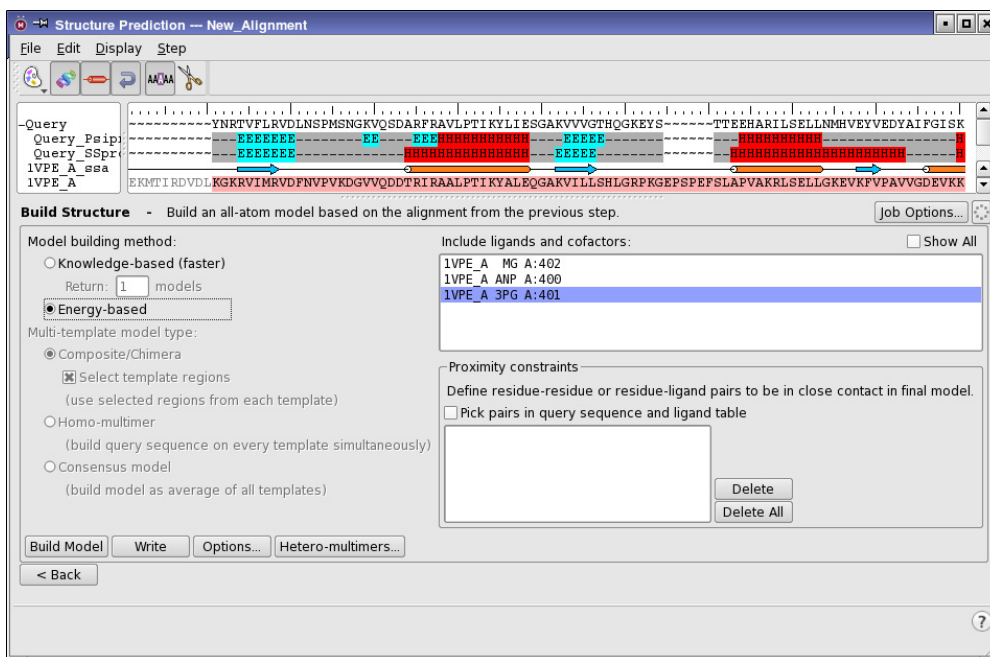


Figure 5.1. The Build Structure step for a single template.

The Build Structure step provides tools for building models of multimers as well as models of single chains, and for building models of single chains from several templates. It also allows you to set constraints and control some aspects of the model building.

The build process does not perform any refinement of the structure. Once you have built a model, you should refine it, by using the Refinement panel, described in [Chapter 6](#).

## **5.2 Choosing the Model Type**

If you are using multiple templates, you must choose the type of model you are building: composite, consensus, homo-multimer, or hetero-multimer. For composite models, you must select the regions from each template that are used to build the structure. The following sections provide more detail on the choices to be made.

If you are using a single template, these choices are not available, and you can proceed to the setting of other options.

### **5.2.1 Building a Composite Model**

If you want to build a composite model from the templates you selected in the Find Homologs step, select Composite/Chimera under Multi-template model type section. You must specify which template should be used for each region of the query, otherwise, the first template in the table is used for the entire query.

To specify the regions to be used for each template, first select Select template regions if it is not selected, then select the desired region on each template in the Sequence Viewer. When you begin to select the template regions, only the first template is in use. As you select regions in the second template, the corresponding residues in the first template are automatically deselected. If you select regions for use from a third template, the corresponding regions of the first and second templates are deselected, and so on. Thus, the selection process ensures that only one template residue is assigned to a query residue (the most recent one selected). Regions of a template that are not used to model the query are colored dark gray (regardless of color schemes.)

### **5.2.2 Building a Consensus Model**

If you want to build a consensus model from the templates, select Consensus under Multi-template model type. The model is then built from the consensus between the templates at each residue position: for example, if two templates have a loop in the same spatial region but in the third template, the loop occupies a different region, the model is built from the two that agree. The structural alignment is done as part of the model-building job, so there is no need to align the structures for the templates.

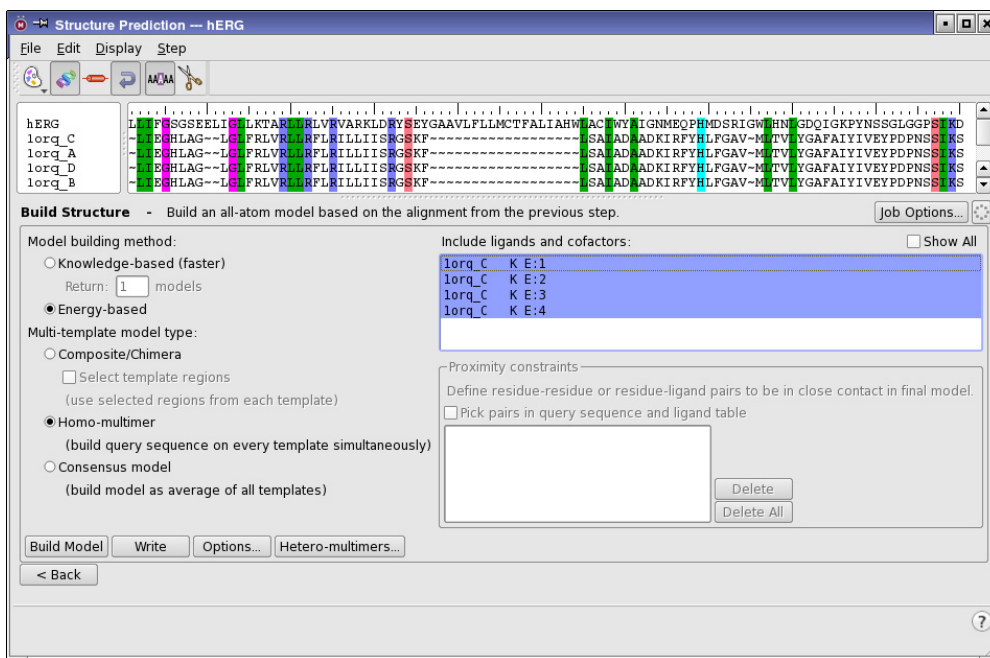


Figure 5.2. The Build Structure step with multiple templates.

### 5.2.3 Building Multimers

In addition to building models for a single chain, you can build models for multimers that are composed of copies of the same chain (homomultimers) or of different chains (heteromultimers). To build the multimer, the structures for all chains in the templates must be in the correct location (position, orientation) to build the multimer.

To build homomultimers, select Homo-multimer under Multi-template model type. All chains of the template must be aligned to the query.

To build heteromultimers, you must first model the individual units of the multimer (which can be homomultimers) in separate runs, up to this step. The templates that you choose to model the individual units must be structurally position relative to each other so that the resulting built structure has the desired configuration of the units.

To build the model, click Hetero-multimers, choose the runs in the Hetero-multimers Modeling dialog box, and click Start.

## 5.3 Preparing to Build

Before you build the structure, you can make settings that affect the build process. These settings are described in the following sections.

### 5.3.1 Including Ligands and Cofactors

Any cofactors and ligands present in the template you have chosen (except water) are listed in the Include ligands and cofactors list. To include one or more ligands or cofactors, select them in this list. You can use shift-click and control-click to select multiple rows. The selections are marked in the Workspace. DNA and RNA can be included as cofactors or ligands. By default, only the ligands and cofactors within 5 Å of the chain used for the template are included in the list. To display the full list, select Show All.

### 5.3.2 Setting Constraints

If you know that certain residues must be in close contact in the final structure, such as cysteines forming a disulfide bond, you can constrain their positions in the final model by setting proximity constraints on residue pairs. A disulfide bond is formally constructed between pairs of CYS residues. For other residues, a harmonic constraint is added for the terminal heavy atoms of the side chains. To apply these constraints, select Pick residue pairs in query sequence, and pick the pairs of residues that you want to be in contact in the final model.

You can also apply proximity constraints between a ligand and the query, by picking a query residue and a ligand in the ligands list rather than two residues. Harmonic constraints are added to both the residue and the ligand. This is useful to ensure that the ligand is kept in the active site when the model is built.

Proximity constraints can only be applied if at least one of the residues in the pair is in a loop, as helices and strands can't be adjusted to satisfy proximity constraints. The same applies to ligands: the residue to which the ligand is constrained must be in a loop.

Proximity constraints are only available for the energy-based model-building method.

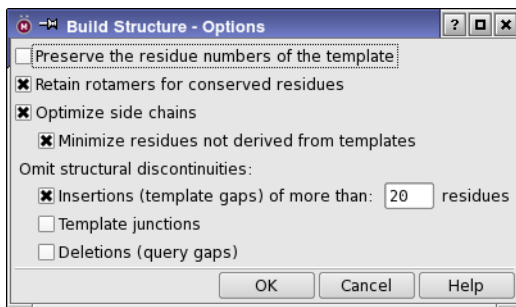
### 5.3.3 Build Options

Click Options to open the Build Structure—Options dialog box and specify which optional model building steps to perform:

- Preserve the residue numbers of the template—Number the residues in the built structure the same as in the template, as far as possible. Does not apply to consensus modeling. When using multiple templates for a single chain, the residue numbers are taken from the first template. Sequential numbering from 1 is used if the template numbering fails.

- Retain rotamers for conserved residues—retain side-chain rotamers for conserved residues.
- Optimize side chains—Optimize the side chains.
- Minimize residues not derived from templates—minimize residues that are not derived from the templates. Only available if Optimize side chains has been selected.

By default, all of these options are selected except the first.



**Figure 5.3.** The *Build Structure–Options* dialog box.

In this dialog box, you can also disable building through structural discontinuities arising from insertions, template junctions, and deletions. The discontinuities are capped with NMA and ACE, so the final structure is discontinuous. The query chain retains its name, regardless of the discontinuities. If you choose any of these options, you should check that your final structure is reasonable.

If there is an insertion in the query that requires the building of a long loop, you can select Insertions (template gaps) under Omit structural discontinuities to cut the query sequence instead, cap it with NMA and ACE, and not build the loop. You can also specify the maximum length of loops that will be built in the text box, so that only longer loops are cut rather than built. When the template that spans the insertion is cut to build the query, the residue on the high side of the cut (higher residue number) is also deleted, to avoid introducing steric clashes. Cutting could be useful where there are long insertions that are not in the region of interest. The option is selected by default, with a length of 20. If it is not selected, all loops are built.

Likewise, if there are deletions in the query (a query gap), or if there is a junction between two templates, you can choose to cap the ends of the two pieces and not try to join them when building the structure. To do this, select Deletions (query gaps) or Template junctions under Omit structural discontinuities.

When the structure is built with discontinuities, the results appear to be misaligned in the sequence viewer. To display the correct alignment, right-click in the sequence viewer and choose Align by Residue Number.

## 5.4 Building the Structure

There are two choices of method for building the structure:

- Knowledge-based—Construct insertions and close gaps using segments from known structures. With this option, you can choose to return multiple models of the structure. By default, only one is returned. The models are approximately ordered by the sequence identity and the incidence of clashes in the structure.
- Energy-based—Construct and refine residues that do not come from the template based on the energy, including building through structural discontinuities and optimizing side chains.

When the model is built, the structure is automatically incorporated into the project as a project entry, and the model is displayed in the Workspace. For composite and multimer models, the residues are colored according to how the template was used:

- blue—all coordinates were taken from the template
- cyan—the backbone coordinates were taken from the template
- red—the residue was rebuilt

This information is stored in the Homology Status property (`i_osp_Homology_Status`), with values of 0, 1, and 2 respectively, but the color scheme is not stored.

The sequence in the Structure Prediction panel remains as it was before building: it reflects only the state of the sequences brought into the Build Structure step. The Workspace sequence viewer displays the full sequence of the built model.

You can also run the Build Structure job from the command line. To do so, click Write to write the input file. The file is written to the current working directory, and is named according to the run and project. A dialog box informs you of the file name and the command to use, which is:

```
$SCHRODINGER/prime input-file
```

For more information on this command, see [Section 10.2 on page 99](#).



## 5.5 Technical Notes for the Build Structure Step

The energy-based model building method in the Build Structure step uses the following resources and methods:

- The OPLS\_2005 all-atom force field for energy scoring of proteins as well as for ligands and other non-amino-acid residues.
- A Surface Generalized Born (SGB) continuum solvation model for treating solvation energies and effects.
- Residue-specific side-chain rotamer and backbone dihedral libraries, derived from the non-redundant data sets extracted from the PDB, and representing values most commonly observed in protein crystal structures. The libraries allow for both the evaluation of existing rotamer/dihedral values and the systematic prediction of new ones.

Build Structure can treat the 20 standard amino acids as well as modified residues and ligands.

Model building uses the following procedure:

1. Non-conserved residues are mutated to the desired identity. Side chains are added by finding the first rotamer in the library that does not produce a clash.
2. Insertions, deletions, and template transitions are built. These are cases in which the backbone itself needs to be reconstructed, either due to gaps in the alignment or template transitions that produce gaps in the 3D structure. These gaps are closed by reconstruction of the affected region *ab initio*, using a backbone dihedral library. If no reasonable conformation can be generated that closes the loop structure, the region being reconstructed is expanded until closure is possible.
3. Gap reconstruction is done by finding any single loop conformation that closes the structure and is physically reasonable. If multiple conformations are found, the one that deviates from the template structure as little as possible is chosen. No attempt is made to perform an exhaustive optimization of the region.
4. Side-chain prediction (optional). Side chains may be optimized. The two choices are either: (a) optimize all side chains from scratch, or (b) optimize only those that are not in the template.
5. Minimization of non-template regions (optional). Any portions of the backbone that were not derived directly from one of the templates (and were thus reconstructed *ab initio*) are subject to a local minimization as described in [Section 6.5 on page 57](#).



# Refining Protein Structures

The Refinement module in Prime is used to refine protein structures. It can be used for loop refinement, side-chain prediction, minimization, rigid-body minimization, hybrid Monte Carlo conformational searching, binding site refinement, and energy analysis at the current geometry of the protein structure. The refinement can be performed in implicit solvent, an implicit membrane, or both.

Each task from this module is presented in Maestro as a separate panel, which you can open as follows:

## Maestro:

- Applications → Prime → Protein Refinement → *task*
- Tasks → Protein Refinement → *task*

## BioLuminate:

- Tasks → Loop + Sidechain Prediction → *task*
- Tasks → Implicit Solvent Refinement + Analysis → *task*

The refinement is performed on the structure that is displayed in the Workspace.

Model structures built in the Build Structure step of the Structure Prediction module usually require further refinement. If there are insertions or deletions in the alignment between query and template, it is recommended that you perform a loop refinement, since insertions and deletions are most frequently found in loop regions. As another example, if you have added a water molecule to the binding site of the protein, you can run a refinement on the composite entry.

Refinement can be done on structures with covalently bound ligands. The ligands can be multiply connected, covalently bound to each other and to the protein. This capability allows the refinement of proteins with phosphorylated residues or attached sugars, for example.

Maestro provides several tools to analyze protein structures, which can give information on the parts of the structure that may need refinement. See [Section 10.6](#) of the *Maestro User Manual* for more information on these tools.

This chapter describes how to run Prime refinements from Maestro. Each task is described in a separate section. You can also run refinements from the command line, with the `prime` command. See [Section 10.2 on page 99](#) for details of the command syntax.

## 6.1 Preparing Structures for Refinement

The refinement module can handle many types of structures, and can fix many structural problems. For standard residues, Refinement fixes formal charges and bond orders, and corrects disparities between the sequence and the structure. For example, if a residue has the coordinates of ALA but is called SER, the 3HB is ignored and the OG and HG added during refinement. The standard residues are the 20 canonical amino acids; ACE and NMA; HOH; CYX (disulfide); and the acid/base variants ASH/AS1 (ASP), GLH/GL1 (GLU), ARN (ARG), LYN (LYS), HIE/HIP/HID (HIS), CYT (CYS), SRO (SER), TYO (TYR).

The input structures must meet certain conditions for a refinement job to be run successfully. Some of these conditions only apply to covalently bound ligands.

For all kinds of structures, the following condition must be met:

- The structure must be an **all-atom structure** (i.e. with explicit hydrogens).

This condition applies to the protein and any ligand, waters or cofactors present in the structure: all hydrogens must be present.

You can add hydrogens to a structure by displaying the structure in the Workspace and double-clicking the Add hydrogens button in the toolbar.



However, for structures containing nonstandard residues, you must correct bond orders and formal charges first.

For structures with nonstandard residues, the following conditions must be met:

- No residue can consist of disconnected pieces (“fragmented residues”) without formal bonds. All atoms within a residue must be connected via bonds, including zero-order bonds in the case of metals.
- Bond orders and formal charges must be corrected.

For nonstandard residues the supplied structure will be used, and you must check for correct bond orders and formal charges. Metals and their ligating groups must be connected by zero-order bonds.

When you rename residues, you should be aware that OPLS\_2005 parameters are used for standard residues and nonstandard residues.

Proteins can also be used as ligands, and the same conditions apply. No special treatment is needed for two distinct chains connected by a disulfide or similar side-chain linkage. If the

“ligand” chain has nonstandard terminal groups or is connected to the main chain by its backbone, two steps are required to make sure that the “ligand” chain is treated correctly:

1. Rename standard residues appearing in the ligand to nonstandard names.
2. Rename the atoms in inter-residue C–N bonds between renamed (ligand) residues. For example, change “\_C\_” to “\_C\*\_”, where the underscores represent spaces. Other atoms can keep their PDB names; only one of C or N needs to be changed.

If your structure does not meet one or more of the conditions outlined above, you must fix it before you can successfully run a refinement. Some of the problems only become apparent when you have run a refinement job, so it can be a useful diagnostic procedure to run a Prime energy calculation first. This job is quick, and produces warnings in the log file, which you can check in the Monitor panel.

Many of the preparation tasks are performed automatically or interactively with the Protein Preparation Wizard. The Protein Preparation Wizard panel can be opened from the Applications menu or the Project toolbar. When the preparation has finished, you should check that all changes made are correct. You can fix any remaining errors with the procedures below. For more information on the Protein Preparation Wizard, see the *Protein Preparation Guide*.

Procedures for fixing bond orders, formal charges, and atom names are given below. These procedures use the Build toolbar and the Build panel. To display the Build toolbar, click Build on the manager toolbar or choose Windows → Toolbars → Build. You can open the Build panel with Edit → Build → *tab*.

#### To assign bond orders:

1. Choose Tools → Assign Bond Orders.
2. Inspect the residues in the structure for any remaining bond orders that are incorrect.
3. If there are still incorrect bond orders, click the Increment bond order or Decrement bond order button on the Build toolbar.



4. Click the bonds in the Workspace structure that needs correction.

You can also right-click the bonds in the Workspace structure, and choose the correct order from the Order submenu of the shortcut menu.

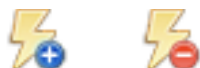
**To view and change formal charges:**

1. From the Label atoms button menu on the Labels toolbar, choose Formal Charge.



Charges are displayed for atoms that are charged.

2. Click the Increment formal charge or Decrement formal charge button on the Build toolbar.



3. Click on an incorrectly charged atom in the Workspace until its charge is correct.  
The formal charge is incremented or decremented by one unit for each click.

**To change PDB atom names:**

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



2. In the Atom Properties tab of the Build panel, choose PDB Atom Name from the Property option menu.
3. Ensure that Pick is selected in the Apply PDB atom name section, and that Atoms is chosen from the Pick option menu.
4. Zoom in on one of the residues (middle+right mouse buttons or mouse wheel).
5. For each atom that needs renaming, enter the desired PDB atom name in the PDB atom name text box, then click on the atom.
6. Repeat [Step 4](#) and [Step 5](#) for each residue whose atoms need renaming.

**To change residue names:**

1. From the Label atoms button menu on the main toolbar, choose Residue information.



2. In the Residue Properties tab of the Build panel, choose Residue Name from the Property option menu.

3. Ensure that Pick is selected in the Apply residue PDB name section, and that Residues is chosen from the Pick option menu.
4. For each residue that needs renaming, enter the desired PDB name in the Residue PDB name text box, then click on the residue.

You might need to zoom in to select the residues (middle+right mouse buttons or mouse wheel).

## 6.2 The Refinement Panels

The refinement task panels have a number of features in common, which are described below. The task-specific features are described in the following sections.

When selecting regions for refinement, you may want to view information about possible structural problems. This information is available in the Protein Reports panel, which can be opened from the Tools menu. For more information on this panel, see [Section 10.6.2](#) of the *Maestro User Manual*.

**Force field:** The OPLS\_2005 force field is used for all refinement tasks.

**Structure source:** Most of the panels allow you to choose the source of the structure, from the Use structures from option menu. This menu has three choices, Workspace, Project Table, and File. Multiple structures can be refined (sequentially) in a single job. The exception is the Refine Loops panel, for which only one structure can be refined in a single job, and the structure must be displayed in the Workspace.

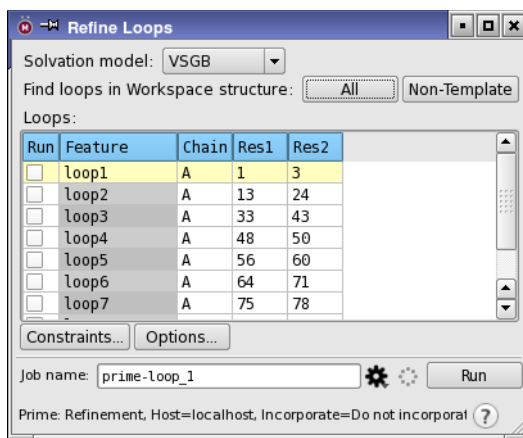
**Solvation:** The Solvation model option menu allow you to choose the model used to solvate the protein. Three models are provided: VSGB, which is an aqueous model using the VSGB method; vacuum, which turns off solvation; and chloroform, for which the SGB method is used.

Each panel has an Options button and a Constraints button, which open dialog boxes. Both dialog boxes have common features. The options are described in [Section 6.10 on page 62](#), and setting constraints is described in [Section 6.12 on page 73](#).

Jobs are started with the Job toolbar, which is described in [Section 1.4 on page 6](#).

## 6.3 Refining Loops

The loop refinement task is capable of refining loop structures of various lengths, and provides algorithms for different loop lengths. In addition, loops whose structure affects other loops can be cooperatively refined in pairs.



**Figure 6.1. The Refine Loops panel.**

**To refine one or more loops serially:**

1. Click one of the Find loops in Workspace structure buttons.

If the structure in the Workspace came from a homology model (the structure has a Homology Status atom property), you can click Non-Template to load only the loops that did not come from the templates. Otherwise, click All to load all loops.

The loop features of the Workspace structure appear in the Loops table. For each loop, the table includes a Run check box, a feature name such as loop1, the chain name, and the numbers of the first (Res1) and last (Res2) residues in the loop. Note that loops of one or two residues are not listed, as they are unsuitable for the refinement protocol.

2. Select a loop by clicking its name in the Feature column.

The loop residues are highlighted in the Workspace. If you are refining a homology model, you should consider residues with a Homology Status of 2 for loop refinement (non-template loops).

3. If necessary, shorten the loop region to be refined by changing the first and last residue numbers (Res1 and Res2).

The time required to refine a loop scales approximately linearly with the length of the loop, and the sampling options recommended for loops longer than five residues are also more time-intensive.

4. (Optional) Click Options to select a sampling method and make settings for this loop in the Refine Loops - Options dialog box.



If the loop is longer than five residues, you should select a sampling method other than the default. See [Section 6.10.2](#) for information on the loop refinement options. The approximate number of subjobs that will be run is indicated along with the sampling method.

5. To designate this loop for refinement, select the check box in the Run column.
6. Repeat [Step 2](#) through [Step 5](#) for each loop you want to refine.
7. Click the **Settings** button to open the Job Settings dialog box.



It is usually necessary to make job settings the first time you run a job. Thereafter, you can click **Run** to run the job with the most recent settings made for this type of job.

8. Make any job settings, then click **Run**.

If you are running refinements on longer loops, you should choose a host with multiple processors or cores.

When you choose several loops for refinement, they are refined in series. The first selected loop is refined, the best scoring structure for that loop is determined, that structure is used in refining the next loop, and so on. When all loops have been refined, one final structure is returned. If only one loop is refined, the four highest scoring structures are returned by default.

If you want to refine two loops cooperatively, you should replace [Step 2](#) through [Step 6](#) above with the following steps:

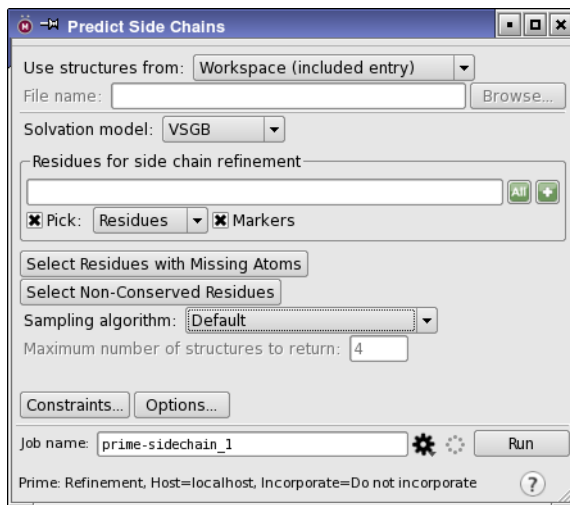
1. Click the **Run** column for the two loops you want to refine cooperatively.
2. If necessary, shorten the loop region to be refined by changing the first and last residue numbers (Res1 and Res2).
3. Click **Options** to open the Refine Loops - Options dialog box.
4. Select **Cooperative** loop sampling.
5. Set any other options, then click **OK**.

For technical details about loop refinement, see [Section 6.13.1](#).

When a loop refinement begins, a validation program checks the loop features of the structure. Apparent errors are reported in a warning dialog box. You may choose **Run All Features**, **Run Only Valid Features**, or **Cancel**. It is recommended that you make a note of the invalid features, click **Cancel**, and correct the structure if appropriate.

## 6.4 Predicting Side Chains

The Predict Side Chains task predicts the conformation of side chains in one or more proteins by sampling orientations to minimize the energy. The sampling can include the orientation with respect to the backbone, and sampling of the backbone as well.



**Figure 6.2.** The Predict Side Chains panel.

You can select residues for side-chain prediction with the Residues for side chain refinement atom selection tools. You might want to use the Atom Selection dialog box, which provides proximity options that allow you to select all residues within a certain distance of a ligand, for example.

If you imported a PDB structure that has residues with missing atoms, you can select those residues by clicking **Select Residues with Missing Atoms**. The selection is based on the information generated when the structure was imported, and does not use the current structural information. The selection therefore includes any residues that were fixed since import.

If you are refining a homology model, you should consider refining residues whose side-chain coordinates did not come from the templates. These are residues for which the Homology Status is equal to 1 and 2. The backbone coordinates for residues with a Homology Status of 1 were taken from the templates, but not the side chain coordinates. For residues with a Homology Status of 2 (none of the coordinates came from the templates), you should consider doing a loop refinement first. You can select all residues with non-template side chains by clicking **Select Non-Conserved Residues**.

The Sampling algorithm option menu allows you to choose a sampling algorithm for the prediction of side chains. These are primarily aimed at cases where some degree of backbone sampling is likely to be required, such as in homology modeling or cross-docking, where the backbone conformation of the initial structure might not be entirely correct. These methods allow for progressively more backbone movement during the side-chain prediction calculations. There are three choices on the Sampling algorithm option menu:

- **Default**—No backbone sampling or reorientation of the CA – CB bond is performed.
- **Monte Carlo**—Perform Monte-Carlo sampling of side-chain conformations. When this algorithm is selected, you can select **Maximum number of structures to return** to return more than one side-chain conformation, and enter the desired number in the text box.
- **With CA-CB vector sampling**—Vary the orientation of the CA – CB bond by up to 30° from the initial direction.
- **With backbone sampling**—Sample the backbone by running a loop prediction on a set of 3 residues centered on the residue for which the side chain is being refined.

While these algorithms can be run on any arbitrary set of residues in a given structure, the general intent is that they should be run on a relatively small subset of the structure, such as a binding site or other region of interest. Running these calculations on an entire structure is likely to result in more structural changes than were intended. For example, since the CA-CB sampling method involves a minimization of all residues being predicted, using this method on the entire structure results in minimization of the entire structure in the final stage, which is generally undesirable.

If heavy atoms are added to a residue because they were missing, a boolean property `side chain added (b_psp_side_chain_added)` is added to all atoms in the residue. You can use this property to select these side chains for other tasks, such as minimization.

For technical details about side-chain prediction, see [Section 6.13.3](#).

## 6.5 Minimizing Structures

The Minimize refinement task minimizes the energy of selected parts of proteins or entire proteins with respect to coordinate changes.

You can specify a set of atoms or residues for minimization with the **Atoms for minimization** picking tools, or click **Select Non-Conserved Residues** to minimize atoms in a homology model whose coordinates did not come from the templates. The atoms in these residues have a **Homology Status** property of 1 or 2.

For more information on atom selection, see [Chapter 6](#) of the *Maestro User Manual*.

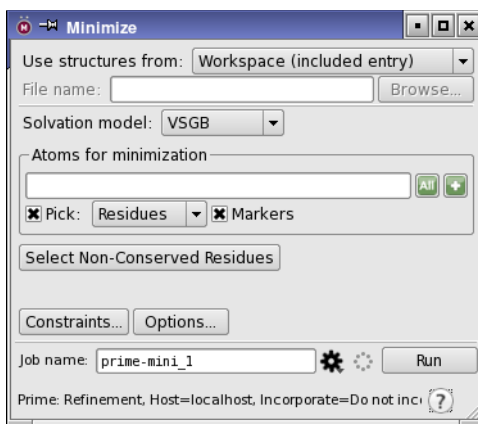


Figure 6.3. The Minimize panel.

## 6.6 Minimizing Parts of Structures as Rigid Bodies

The Minimize as Rigid Body task allows you to treat part of a system as a rigid body or a set of rigid bodies, with optional freely moving atoms, and freeze the rest of the system. The panel offers two standard selections of the rigid body, Protein and Ligand, which are detected automatically, and a Custom feature that allows you to choose the rigid body or bodies.

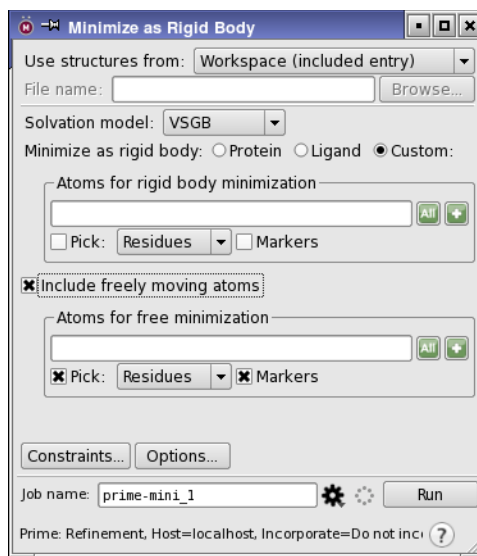
Rigid-body minimization is a lot faster than a full minimization, so it is useful for obtaining results where the full optimization is too time-consuming, or to get a good starting position for a fuller refinement. It is often used in the early stages of X-ray refinement of a protein structure, for example. Below are some scenarios where rigid-body minimization might be useful.

- Move a ligand into position in a receptor site, with the possibility of treating the side chains of the receptor flexibly.
- Minimize the orientation of all side chains in a protein simultaneously.
- Move a hinged protein into the open or closed conformation rigidly.
- Refine the structure of a dimer resulting from protein-protein docking, with constraints to keep the protein in place, and flexible residues at the interface.

### To set up a rigid-body minimization:

1. Select an option for the part of the structure to minimize as a rigid body.

The options are Protein, Ligand, and Custom. The protein and the ligand are automatically detected. If you select Custom, the Atoms for rigid-body minimization picking tools are displayed, so you can specify the rigid parts of the structure.



**Figure 6.4. The Minimize as Rigid Body panel.**

2. If you selected Custom, pick the atoms to use for the rigid body using the Atoms for rigid-body minimization standard picking tools.

The atoms that you choose do not have to be connected: if they are not, each connected piece is treated as a rigid body. For example, if you choose Protein Side Chains from the Atom Selections button menu, each side chain is treated as a rigid body.

3. Select Include freely moving atoms if you want to add atoms that are minimized without being constrained to a rigid body.

When you select this option, the Atoms for free minimization picking tools are displayed below the option.

4. If you selected the option to include freely moving atoms, select the desired atoms using the Atoms for free minimization standard picking tools.

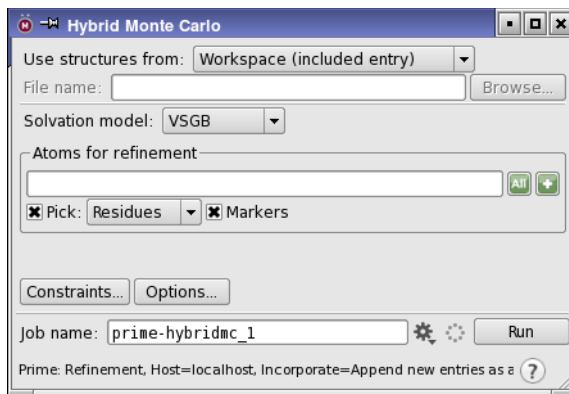
These atoms are minimized individually along with the rigid body. You might, for example, pick atoms on the boundary between the rigid body and the rest of the structure for free minimization. The atom selection should not contain any atoms that are part of the rigid body region: if it does, they are removed from the freely moving selection.

5. (Optional) Click Constraints to set up any additional constraints on the structures.

You might, for example, apply a distance constraint between two rigid proteins, to prevent them from drifting apart rather than moving closer together.

## 6.7 Hybrid Monte Carlo Conformational Sampling

The Hybrid Monte Carlo task uses high-temperature molecular dynamics simulations followed by minimization to explore conformational space. Each MD/minimization stage produces a minimized structure. The high temperature MD simulation allows the structure to move between minima on the potential energy surface. New conformations are accepted or rejected based on standard Monte Carlo criteria.



**Figure 6.5.** The Hybrid Monte Carlo panel.

You can specify the part of the structure that you want to refine by using the standard picking tools in the Atoms for refinement section.

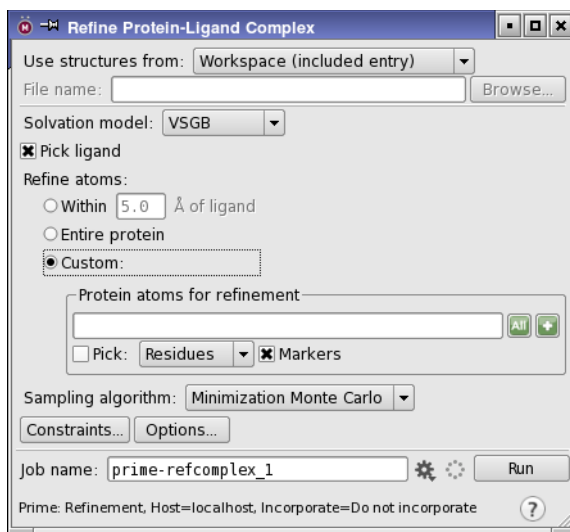
By default, only the lowest energy structure is returned, so this method can be used to minimize the energy of a structure to a point other than the nearest minimum. You can choose to return more than one structure (by setting an option), which allows you to examine the conformational space explored by the search.

## 6.8 Refining Protein-Ligand Complexes

The Refine Protein-Ligand Complex task refines the structure of protein-ligand complexes. The ligand can be covalently bound or bound by noncovalent interactions. The refinement can be restricted to the region around the ligand, and can sample positions and conformations of the ligand.

### To set up a protein-ligand refinement job:

1. Pick a ligand atom in the Workspace to define the ligand.



**Figure 6.6.** The Refine Protein-Ligand Complex panel.

2. Select an option for the protein atoms to refine along with the ligand:
  - Within  $N \text{ \AA}$  of ligand—Refine the protein residues that have atoms within the specified distance of the ligand.
  - Entire protein—Refine all residues in the protein. This is likely to take some time.
  - Custom—Select the protein atoms to refine. When you select this option, the Protein atoms for refinement picking tools are displayed, so you can pick the atoms.
3. Choose the algorithm used for sampling the part of the structure that is being refined, from the Sampling algorithm option menu:
  - Local optimization—Predict side chains and minimize the selected protein residues, then minimize the ligand and the protein residues together.
  - Covalent ligand sampling—Perform a complete sampling of the rotatable bonds of the ligand, then cluster the resultant conformations, and choose a representative from each cluster. Then for each representative, perform side-chain prediction of the protein residues, followed by a minimization of the ligand and all residues selected for refinement. This procedure is similar to the loop refinement procedure.
  - Minimization Monte Carlo—Perform a Monte Carlo sampling of the conformations of the refinement region. The sampling includes inner steps, in which only the short-range forces are updated, and outer steps, in which the long-range forces and solvent interaction is updated.

- Hierarchical optimization—Systematically sample ligand positions, orientations, and conformations along with receptor residues.

4. (Optional) Set constraints or options.

## 6.9 Calculating the Energy

The Calculate Energy task performs a single-point molecular mechanics energy calculation which includes a breakdown of the energy into various terms. The calculation uses the OPLS\_2005 all-atom force field for protein residues as well as for ligands and cofactors. The output includes a breakdown of the energy into covalent, Coulombic, van der Waals and solvation energy contributions. Each of these is added as a Maestro property to the output structure file, named Prime Covalent, Prime Coulomb, Prime vdW, and Prime Solvation. The output also includes a breakdown of the energy on a per-residue basis and residue-residue interactions. The energy breakdown can be visualized on the structure in the Workspace—see [Chapter 9](#) for details.

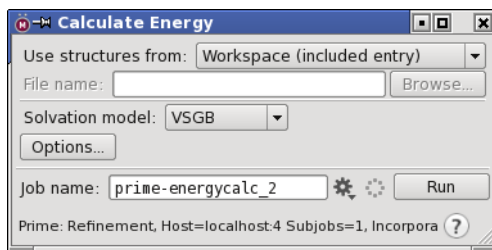


Figure 6.7. The Calculate Energy panel.

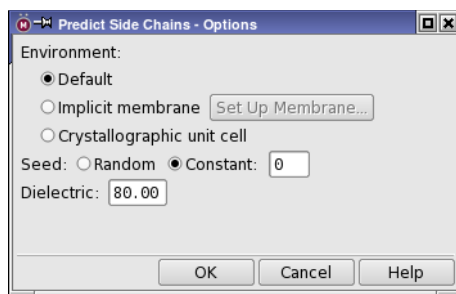
## 6.10 Setting Refinement Options

The Options dialog boxes for the refinement tasks have several common options at the top, followed by options for the specific task. The Calculate Energy and Predict Side Chains tasks only have the common options. Click Options in any of the refinement panels to view or change settings.

### 6.10.1 Common Options

These options apply to all types of refinement or to more than one type of refinement. The Seed option is not available for minimization tasks, which do not use a random seed. The Environment options allow you to set up a special environment for the protein. You can choose one of the three options.





**Figure 6.8. The Predict Side Chains - Options dialog box.**

#### Default

Use the environment specified in the main panel, which could be a vacuum or an implicit solvent.

#### Implicit membrane

This option embeds the protein in an implicit membrane, which must be defined. It does not add a membrane. To create a membrane, you must either add it in the Prime Membrane Setup panel (see below), or select an entry that already has a membrane.

#### Set Up Membrane

Click this button to open the Prime Membrane Setup panel, which is described in detail in [Section 6.11 on page 70](#).

#### Crystallographic unit cell

If crystal symmetry is known for this protein, apply periodic boundary conditions so that the crystal symmetry is satisfied. The protein is refined with its crystallographic neighbors in place.

#### Seed

Specify the seed for the random-number generator used in side-chain predictions and Monte Carlo sampling. The seed is used for side-chain prediction in both the loop refinement and the side-chain prediction tasks. The options are:

- Random: use a random seed.
- Constant: use the integer given in the text box as the seed. The default is 0.

#### Dielectric

Set the dielectric constant for the (continuum) solvent. Note that the solvent models also use a solvent radius parameter, so changing the dielectric constant only is not sufficient to change to another solvent, unless the radius is very similar to that of the chosen solvent model.

## 6.10.2 Loop Refinement Options

The loop refinement options include selecting the sampling method, subjob options, output options, and some general constraints on the optimization.

Prime can perform cooperative loop sampling as well as serial loop sampling. Both the size of the loop and the sampling options chosen affect the time needed for loop refinement. The options are:

### Method

- Cooperative loop sampling—Select this option to refine two loops together. Each loop is refined in turn until the refinement has converged for both loops. You must have only two loops selected to use this option.
- Serial loop sampling—Select this option to refine each loop independently. The option menu provides three levels of sampling accuracy: Default (up to 5 residues), Extended (6 to 11 residues), and Ultra Extended (10 or more residues). The text below the option and menu displays the recommended loop length for the selected accuracy. The menu options indicate how many subjobs will be used for the loop refinement, so you can choose an appropriate number of processors for the job.

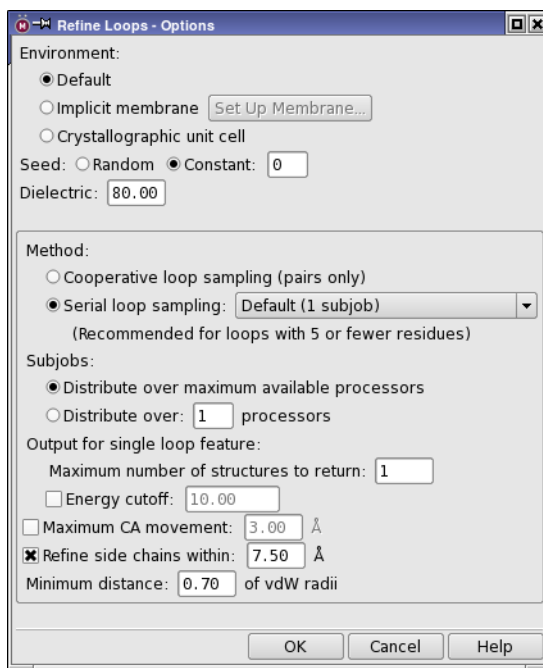


Figure 6.9. The Refine Loops - Options dialog box.

Extended sampling samples loop conformations more thoroughly, using a series of loop constraint settings. Up to eight processors can be used simultaneously to perform loop refinement. If you select Ultra Extended, you can only refine one loop at a time.

### Subjobs

These options allow you to set a limit on the number of processors over which to distribute the loop sampling subjobs. Each sampling method runs in several stages, and each stage can use a different number of processors. The maximum number of processors that can be used by the different sampling methods are:

Default	1
Extended	4
Ultra Extended	8*L (L is the loop length)

- To make use of as many processors as are available, select Distribute subjobs over maximum available processors.
- To set the maximum number of processors, select Distribute subjobs over and enter a value in the text box.

### Output for single loop feature

These options determine how many structures are returned:

- Maximum number of structures to return: Select this option to return more than one structure for each loop specified for refinement. The default when this option is selected is to return 4 structures for each loop.
- Energy cutoff: Select this option to prevent higher-energy structures from being returned. Structures whose energy is higher than that of the lowest-energy structure by more than the specified amount are not returned; the default cutoff is 10 kcal/mol.

The following settings are applied to the loops you have selected for refinement.

- Maximum CA movement: To limit the extent to which alpha carbon atoms can move from their original positions, select this option, and enter a limit in the text box. The default distance is 3.0 Å. The extended sampling procedure tests multiple values for Maximum CA movement, so this option is unavailable when extended sampling is selected. This option applies to all alpha carbon atoms. For more control over which alpha carbons are restrained and by how much, use spatial constraints, which are described in [Section 6.12 on page 73](#).
- Refine side chains within: Select this option to refine side chains within a given distance of the loops, and specify the distance in the text box (default 7.5 Å). Side chains are

included in the refinement if the distance between any heavy atom in the side chain and any CB atom in the initial set of loop structures is less than the specified distance.

- Minimum distance:  $N$  of vdW radii: Specify the minimum distance between atoms in terms of the fraction of their 'ideal' van der Waals radii. The default for this value is 0.70, meaning that the distance between atoms must be at least 70% of their 'ideal' van der Waals separation. The extended sampling procedure tests multiple values for the minimum overlap, and therefore this option is unavailable when extended sampling has been selected.

### 6.10.3 Minimization Options

The minimization options allow you to specify the minimization method and set options for convergence. These options are the same for the Minimize and Minimize as Rigid Body tasks.

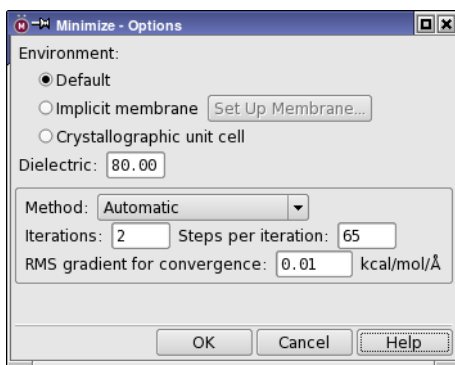


Figure 6.10. The Minimize - Options dialog box.

#### Method

Choose the minimization method from this option menu. The choices are Quasi-Newton (LFBGS), Truncated Newton, Conjugate gradient, and Automatic. The Automatic method uses conjugate gradient minimization when the gradients are large, and switches to the truncated Newton method when they are sufficiently small.

#### Iterations

Specify the number of iterations to be performed for the minimization in this text box. At the beginning of each iteration, various updates are performed, such as on solvent parameters and long-range pairs. Iterating reduces the accumulation of errors in the minimization. The minimization stops if the specified number of iterations is completed.

### Steps per iteration

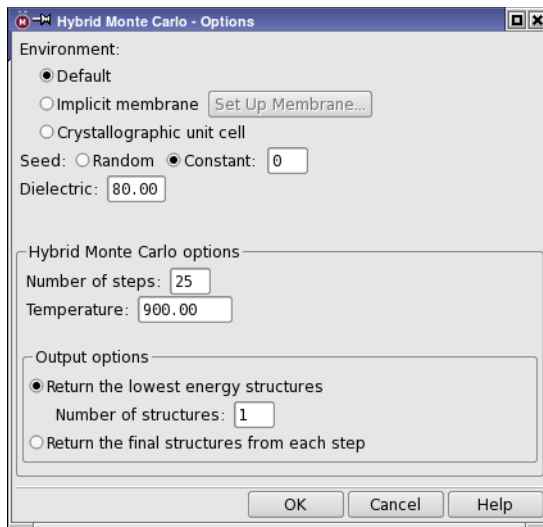
Specify the maximum number of steps to take in each minimization iteration. Each step involves one application of the chosen algorithm to obtain a new geometry. The iteration cycle finishes when the maximum number of steps is completed.

### RMS gradient for convergence

Specify the RMS gradient threshold for convergence of the minimization, in  $\text{kcal mol}^{-1} \text{\AA}^{-1}$ . The minimization finishes if the RMS gradient is smaller than the specified value.

## 6.10.4 Hybrid Monte Carlo Options

The options presented in the Hybrid Monte Carlo - Options dialog box allow you to set the temperature and length of the sampling, and decide which structures to return.



**Figure 6.11.** The Hybrid Monte Carlo - Options dialog box.

### Number of steps

Specify the number of steps to take in the sampling procedure. Each step consists of a short, high-temperature MD simulation followed by a minimization.

### Temperature

Set the temperature for the MD simulations. The temperature should be high enough to allow exploration of new conformations, but not so high as to allow larger changes such as interconversion of chiral centers.

### Output options

Choose an option for the selection of structure to return.

- Return the lowest energy structures—Return the structures that have the lowest energy, selected from the results of each minimization. The number of structures is set in the Number of structures text box. This is the default option, and the default number of structures is 1.
- Return the final structures from each step—Return the final structure from each step in the sampling, which represents a trajectory for the Monte Carlo sampling. This is a way of checking the course of the sampling for convergence, for example.

## **6.10.5 Protein-Ligand Complex Refinement Options**

The options presented in the Refine Protein-Ligand Complex - Options dialog box depend on the sampling algorithm chosen in the Refine Protein-Ligand Complex panel. The options are described in subsections for each algorithm.

### **6.10.5.1 Local Optimization Options**

Set the number of passes through the prediction of side chains and minimization of the protein residues, in the Number of passes text box. Each pass consists of a side-chain prediction, followed by a minimization of the protein residues. The ligand is minimized with the rest of the protein after the specified number of passes is complete.

### **6.10.5.2 Covalent Ligand Sampling Options**

Set the maximum number of structures to return from the sampling, in the Maximum number of structures to return text box. The number of structures is limited by the number of clusters formed when sampling the ligand dihedrals, so it may be less than the maximum set in this section.

### **6.10.5.3 Minimization Monte Carlo Options**

Set parameters for the Monte Carlo sampling of the refinement region.

#### Number of steps

Specify the number of steps to take in the minimization. These steps are the “outer” steps, in which the long-range forces and the interaction with the solvent are updated. The number of “inner” steps (moves), in which short-range forces are updated, is set to an optimal value. The random selection of the type of move is made for each outer step, so the inner steps are all moves of the same type.

### Step probabilities

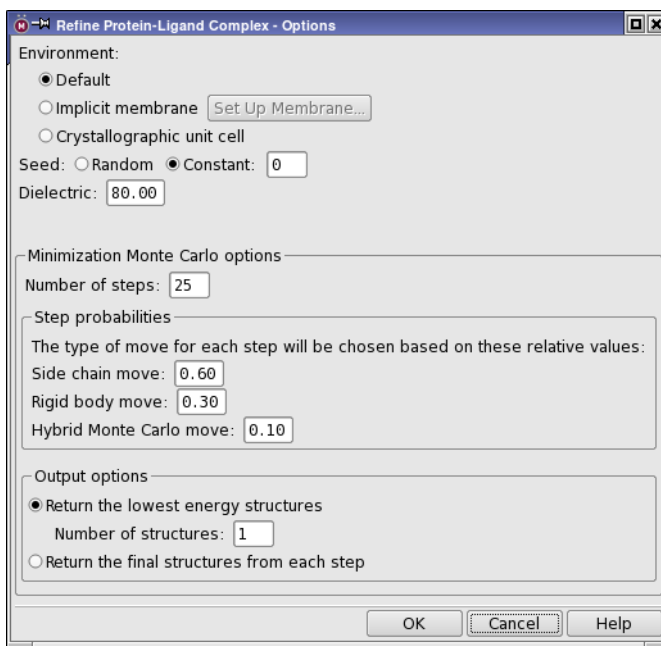
Specify the probabilities of making each of three kinds of moves in these text boxes. These probabilities can be used to guide the minimization. The numbers you enter are relative frequencies of each move, so they do not have to sum to 1.

- Side chain move—Sample the torsions of the side chains.
- Rigid body move—Move molecules (protein chains, ligands, waters, cofactors) as rigid bodies.
- Hybrid Monte Carlo move—Use the final frame of a very short MD simulation as the move.

### Output options

Choose an option for the structures to return.

- Return the lowest energy structures—Return the structures that have the lowest energy. The number of structures is set in the Number of structures text box.
- Return the final structures from each step—Return the final structure from each step in the sampling, which represents a trajectory for the Monte Carlo sampling. This is a way of checking the course of the sampling, for example for convergence.



**Figure 6.12.** The Refine Protein-Ligand Complex - Options dialog box, showing Minimization Monte Carlo options.

#### 6.10.5.4 Hierarchical Optimization Options

Set options for hierarchical optimization.

Number of clusters

Specify the number of clusters to use when clustering the conformations of the complex. A k-means algorithm is used for the clustering.

Maximum number of structures to return

By default, only one structure is returned. If you want to examine multiple possible structures, specify the maximum number to return in this text box. The number of structures is limited also by the number of clusters.

### 6.11 Setting Up an Implicit Membrane

The Prime implicit membrane model is intended for proteins embedded in a membrane. The implicit membrane is a low-dielectric slab-shaped region, which is treated in the same way as the high-dielectric implicit solvent region. Hydrophobic groups, which normally pay a solvation penalty for creating their hydrophobic pocket in the high dielectric region, do not have to pay that penalty while in the membrane slab. Conversely, hydrophilic groups lose any short-ranged solvation energy from the high dielectric region when moving into the low dielectric region.

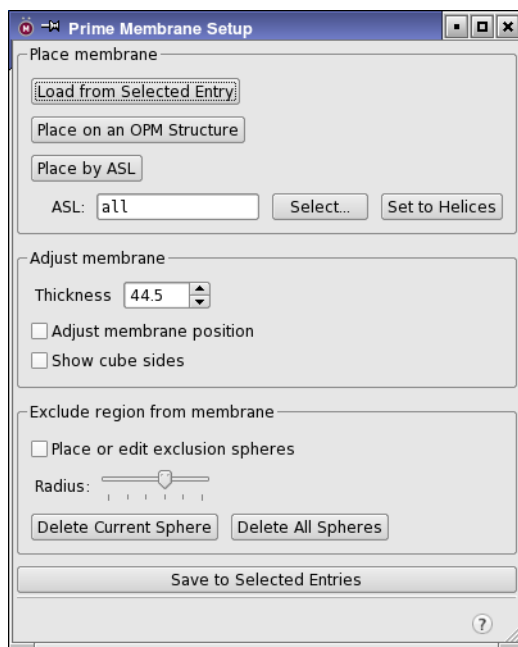
The implicit membrane model is intended for use with proteins that span the membrane. The slab is trimmed for efficiency, so that any structure that is too small or too far away from the membrane is likely to result in too much trimming and might not produce useful results.

The membrane is marked in the Workspace by a semitransparent cube with a white line that is perpendicular to the surfaces of the membrane (which are planar). The membrane surfaces are colored red, and the other faces of the cube (which are interior to the membrane) are colored green. You can control whether the sides of the cube are displayed with the **Show cube sides** option. The direction of the line shows the orientation of the membrane. You can adjust both the thickness and the orientation of the membrane.

Setting up a membrane is done in the Prime Membrane Setup panel, which you open by clicking **Set Up Membrane**.

There are three sections in the panel, which correspond to three tasks: initial placement, adjustment, and excluding regions from the membrane.





**Figure 6.13.** The Prime Membrane Setup panel.

In the Place membrane section, there are three choices for placement.

- **Load from Selected Entry**—Load a membrane from the entry that is selected in the Project Table. The membrane is placed at the same location in space as in the source entry, so you should ensure that the protein in the source entry and the current entry are aligned.
- **Place on an OPM Structure**—If you are refining a structure from the OPM database (<http://opm.phar.umich.edu>), click this button to place the membrane according to the information stored with the structure.
- **Place by ASL**—Place the membrane so that the atoms given by the expression in the ASL text box are inside the membrane. The ASL expression can be entered in the text box, or you can click **Select** to specify the atoms in the Atom Selection dialog box.

To place the membrane on a set of helices, click **Set to Helices**. The secondary structure assignment is then used to select the helices for placement inside the membrane. The protein must have a secondary structure assignment before you use this button. To do the assignment, choose **Tools** → **Assign Secondary Structure** in the main window.

Once you have made an initial placement, you can adjust the position of the membrane. If you started with an ASL expression, you can click **Select**, and add to the ASL expression used in

the Atom Selection dialog box. The Previous button should recall the last ASL expression used. You can adjust the membrane position manually, in the Adjust membrane section, as follows:

1. Select Adjust membrane position.
2. Use the middle mouse button to rotate the membrane and the right mouse button to translate the membrane.
3. Adjust the thickness by entering a value in the text box or using the arrow buttons to adjust the thickness up or down by 0.1 Å at a time.
4. Deselect Adjust membrane position.

This is required to exit membrane adjustment and return to normal Workspace operations.

If the protein has pockets that would be occupied by solvent (water), you can select these regions to exclude from the implicit membrane using the tools in the Exclude region from membrane section. The solvent dielectric constant is then used for these excluded regions.

Excluded regions are defined by a set of spheres that are placed on atoms adjacent to the region. The spheres can overlap, and should cover the entire excluded region. It does not matter that they overlap the protein, because the protein dielectric is not affected.

- To define a region, select Place or edit exclusion spheres and pick atoms in the Workspace that are adjacent to the region. For each pick a sphere is placed on the atom.
- To adjust the volume of one or more spheres, pick the spheres in the Workspace, and use the Radius slider. This slider only uses integer values, because the exact size of the sphere isn't critical: the spheres only have to cover the excluded region.
- To delete a sphere, pick the sphere and click Delete Current Sphere.

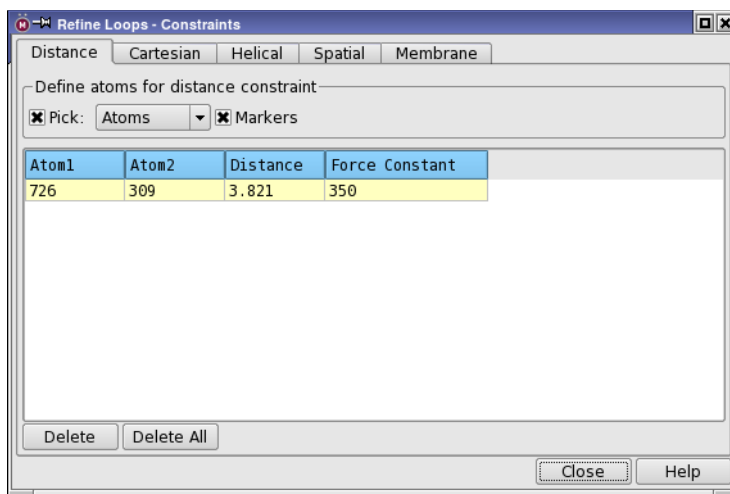
When you have finished, deselect Place or edit exclusion spheres.

If you want to add the membrane to the entries that are selected in the Project Table, click Save to Selected Entries. The membrane is stored as a set of coordinates for each end of the white line. The entries that you select should therefore contain proteins that occupy approximately the same region of space as the protein used to define the membrane. Otherwise, you will have to adjust the membrane location for these entries.

## 6.12 Applying Constraints

Prime provides five kinds of constraints that you can apply in a refinement: distance constraints (also known as pairwise constraints), cartesian constraints, helical constraints, spatial constraints, and membrane constraints. You can apply the constraints in the Constraints dialog box for the refinement task, which you open by clicking Constraints in any of the refinement task panels. The various kinds of constraints are set in separate tabs. Distance and Cartesian constraints are available for all types of refinements; helical, spatial, and membrane constraints are only available for loop refinement.

Distance constraints can be applied to any pair of atoms, and are applied as harmonic constraints. Normally, these constraints are made between nonbonded atoms, but you can pick bonded atoms as well. To set up a constraint, select Pick in the Define atoms for distance constraint section, and pick two atoms in the Workspace. The atom indexes are recorded in the table, with the current distance and a default strength (“force constant”, actually twice the force constant) of  $350 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . You can edit the distance to force the atoms to a new distance, and you can edit the force constant. To delete a constraint, select it in the table and click Delete; to delete all constraints, click Delete All. These constraints set `PAIR_CONSTRAINT_i` keywords in the input file.



**Figure 6.14.** The Distance tab in the Constraints panels.

Cartesian constraints restrain atoms to their initial positions with a flat-bottomed harmonic potential. To set up a constraint, ensure that Pick is selected in the Restrain atoms to their initial coordinates section, and pick atoms in the Workspace. The ASL expression for the picked atoms is shown in the text box. You can set the force constant and the radius of the flat bottom for the potential in the Force constant and Distance text boxes.

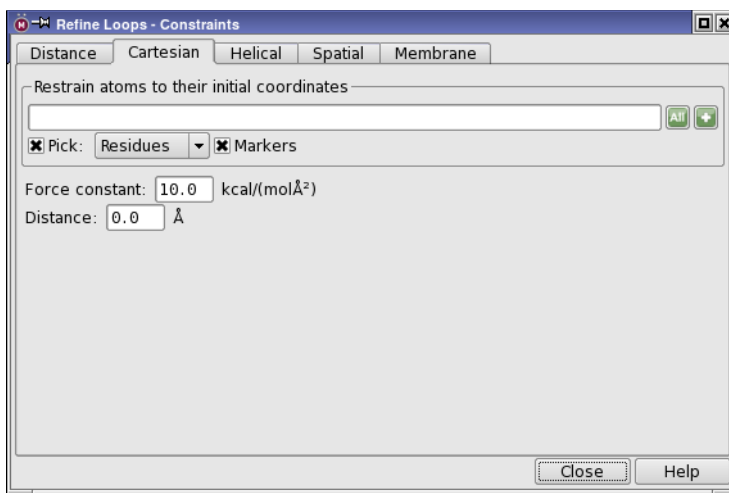


Figure 6.15. The Cartesian tab in the Constraints panels.

Helical constraints can only be used in loop refinement. The constraint forces a subset of the residues in the loop to adopt an alpha helical conformation. You can specify the first and last residues in the helix in the Helix Start and Helix End boxes. By default, these are set 4 residues from the end of the loop. If you do not leave enough residues outside the helical region, the loop prediction might not produce good conformations, because it is too constrained. To turn on the constraint, select Constrain loop subset to helical conformation. This constraint sets a `HELIX_BUILD` keyword in the input file.

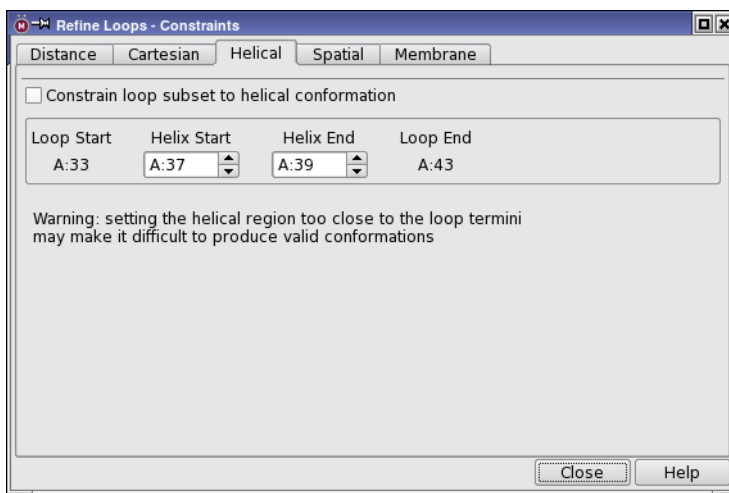
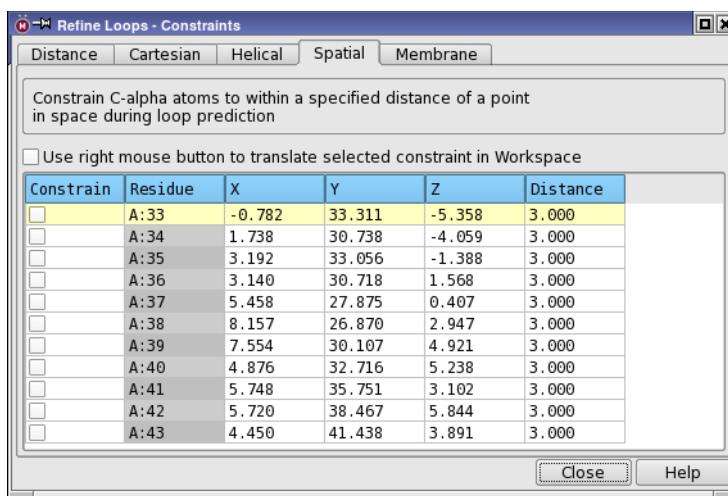


Figure 6.16. The Helical tab in the Refine Loops - Constraints panel.

Spatial constraints are also only available for loop refinement. They specify the maximum distance that any alpha carbon atom in the loop should be allowed to move during refinement. The C-alpha atoms are listed in the table. To apply the constraint, check the check box for the residue in the Constrain column. The constraint is represented as a sphere in the Workspace. The current position of the C-alpha atom and the maximum distance are listed in the table.

You can edit the coordinates to constrain the C-alpha to a different position, and you can edit the maximum distance. You can also select a constraint in the table, select Translate constraint, and drag the sphere in the Workspace to a new location to change the position for the constraint. You might want to do this, for example, to force a loop into a different region of space. These constraints set `MAX_CA_MOVEMENT` keywords in the input file.



**Figure 6.17.** The Spatial tab in the Refine Loops - Constraints panel.

Membrane constraints restrict the loop conformation placement with respect to the implicit membrane. You can restrict the conformation to be outside the membrane (Outside membrane) or inside the membrane (Inside membrane), or you can allow the loop to be placed anywhere (No restriction).

The constraints are applied to the C $\alpha$  atoms only: other atoms of the residues are not restricted. Side chains can cross the boundary, for example. The constraints are applied only in the building stage of loop refinement, and subsequent minimization can move the residues across the membrane boundary.

There is no requirement that the implicit membrane model, with the adjusted dielectric constant, is actually used in the loop refinement. You can use the membrane to define a spatial exclusion zone, and then the constraints are just spatial constraints.

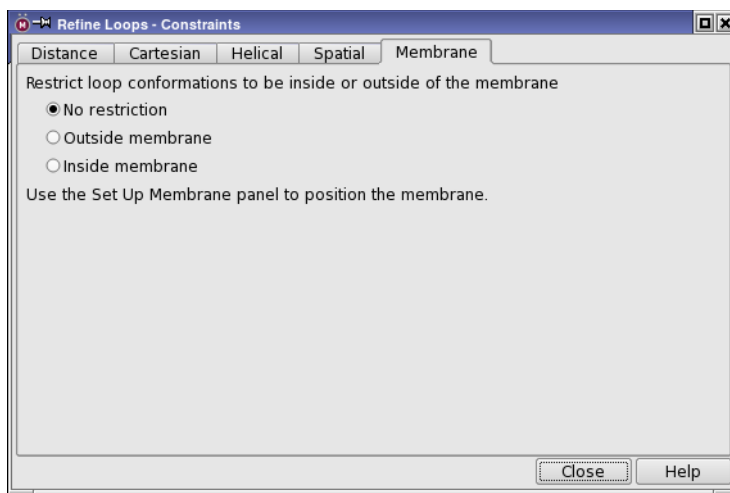


Figure 6.18. The Membrane tab in the Refine Loops - Constraints panel.

## 6.13 Technical Details for Refinement

### 6.13.1 Loop Refinement

Loop refinement using the Default option proceeds as follows:

1. The loop is reconstructed using the backbone dihedral library, by building up half from each direction. The resolution starts off very coarsely, becoming finer until it manages to produce a required number of physically realistic loop conformations (based on the length of the loop).
2. The large number of loops generated in the first step are clustered, and representatives of each cluster are selected.
3. These cluster representatives are then scored as follows:
  - a. Side chains are re-added to the loop residues, as well as any residues within a specified cutoff, according to the side-chain optimization procedure.
  - b. The loop and contact side chains are minimized.
  - c. The energy is calculated.
4. The best scoring loop structures are returned.

Loop prediction with extended sampling is specifically designed to overcome sampling problems with long loops, in which the number of residues results in an unwieldy number of possible loop conformations.

Extended sampling uses the following procedure:

1. Two initial predictions are run on the loop. These are intended to coarsely sample conformational space and are carried out with two slightly different sets of prediction parameters. Each initial prediction returns up to four top-scoring conformations, resulting in up to eight structures. These eight structures are intended as initial points for finer sampling of what appears to be favorable regions (basins) of conformational space.
2. Each of the eight structures generated in [Step 1](#) is run through another stage of refinement, with a 4 Å restriction on C $\alpha$  movement. This is intended to sample the basins more finely.
3. The eight refined structures from [Step 2](#) are combined with the original structures from [Step 1](#), and this set of 16 structures is ranked by energy. The five top-scoring structures are then subjected to a final stage of refinement, with a 2 Å restriction on C $\alpha$  movement. This is intended to perform fine-grained sampling of the best basins.

This procedure defines the “Extended High” sampling procedure. The “Extended Medium” and “Extended Low” carry four and two structures, respectively, into [Step 2](#) and [Step 3](#).

In ultra-extended sampling, the three steps of extended sampling are followed by the steps below.

4. The top four structures are selected, and two predictions run on each, with part of the loop frozen. Runs are done with “moving windows” with varying numbers of contiguous residues: all  $n$  residues, the two possible contiguous sets of  $n-1$  residues, the three sets of  $n-2$  residues, and so on. The structures are sorted after each cycle of a given number of residues.
5. [Step 2](#) is repeated with the top eight structures cumulatively generated, then [Step 3](#) is repeated with the top eight structures cumulatively generated.

### 6.13.2 Cooperative Loop Refinement

Cooperative loop refinement proceeds as follows:

1. A list of side chains that will be optimized is generated. By default this list includes any residue with an atom within 7.5 Å of either loop 1 or loop 2.
2. Loop 1 is sampled, allowing side chains in the list from [Step 1](#) to move, holding the backbone of loop 2 fixed at its original conformation. Eight loop conformations are generated.

3. Loop 2 is sampled, allowing side chains in the list from [Step 1](#) to move, holding the backbone of loop 1 fixed at its original conformation. Eight loop conformations are generated. This step runs concurrently with [Step 2](#).
4. Conformations of loop 1 from [Step 2](#) with loop 2 from [Step 3](#) are merged into a set of structures.
5. All side chains in the list determined in [Step 1](#) are reoptimized, keeping only those structures whose energy is below the energy cutoff.
6. All residues in both loops are then minimized, including the backbone, keeping only those structures whose energy is below the energy cutoff.
7. The 4 best new structures are taken as input for resampling, starting from [Step 2](#). The process is repeated until convergence is obtained.

### 6.13.3 Side-Chain Prediction

The structure refinement program uses the following procedure to re-predict conformations for the side chain set that you select. The default algorithm is as follows:

1. Side-chain rotamers are randomized for nonconserved residues (the default) or for all residues.
2. Beginning with the first residue to be predicted, the side-chain rotamer library is used to find the rotamer with the lowest energy while keeping all other side chains fixed. Once the process is complete for the first residue, the next residue is treated, and so forth until all have been done once (a single pass has been completed).
3. Once the pass is complete, [Step 2](#) is repeated from the beginning, and then repeated again until the side-chain rotamers appear to be converged (no more changes are occurring).
4. Minimization is run on all of the side-chain atoms (but not backbone atoms) of the residues being treated.

Two methods are available for situations where some degree of backbone movement is likely to be required to get the correct side-chain conformation. This typically happens when working with homology models or when performing cross-docking studies, where the backbone conformation of the initial structure might not be entirely correct. These two methods allow for progressively more backbone movement during the side chain prediction calculations, and are described below. They are intended for use on a small set of residues, not the entire structure.



### With CA-CB vector sampling

Even minor discrepancies in backbone conformation can result in a slight shift in the location of the CB atom. This slight shift can often cause subsequent side-chain predictions to fail in placing the side chain optimally, particularly on large rigid aromatic residues. Allowing for sampling of the CA-CB vector in conjunction with the side-chain prediction can alleviate this problem. The method is turned on in the input file with the keyword `SAMPLE_CBETA`. The algorithm is as follows:

1. CB positions in a conical region are systematically sampled around the initial position with a maximum displacement in the CA-CB vector of 30.0 degrees (settable with input file keyword `MAXCONEANG`).
2. At each position, the optimal side chain rotamer is selected.
3. The best overall combination of CB position and rotamer for the residue is selected from the results for all positions.
4. The sampling iterates over all requested residues until convergence is achieved in the same way as with conventional side chain prediction (see above).
5. Once all side-chain sampling has completed, the side chains and backbone atoms of the predicted residues are minimized.

### With backbone sampling

This method is intended for situations where even larger backbone movements may be required, but where a full loop prediction on the entire region may produce more movement than desired. It is turned on in the input file with the keyword `SAMPLE_BACKBONE`.

For each residue for which side chains are predicted, a backbone loop segment of 3 residues (settable with the input file keyword `BACKBONE_LEN`), centered about that residue is reconstructed using the normal loop prediction algorithm. Because this loop prediction includes the additional residues on either side of each requested residue, you should expect to see the conformations of these residues optimized as well (i.e. the region that is optimized extends somewhat beyond the residues actually selected for side-chain optimization).

Side-chain prediction is not performed for GLY and ALA residues with any of these methods.

## 6.14 Refinement with Nonstandard Residues

If you want to refine a structure with nonstandard residues, you can make changes to the residues in the structures before refinement. Several nonstandard residues are supported in the Build panel: the neutralized forms of ASP, LYS, ARG, and GLU, which are named ASH, LYN, ARN, and GLH; and the tautomers of HIS (HID=HIS and HIE) and its protonated form, HIP.

Ionization (deprotonation) of four other residues are supported, CYS, SER, THR, and TYR. The names of the ionized residues are CYT (thiolate), SRO (alkoxide), THO (alkoxide), and TYO (phenoxide). In addition, the disulfide-bridged CYS is supported as CYX. To change one of these residues, you can either change the name, or change the structure. For example, you can generate an ionized SER by either of the following methods:

- Changing the residue name to SRO. The HG is deleted and a formal negative charge is added to the OG.
- Deleting the HG (if present) and adding a formal charge to the OG. The residue is renamed to SRO.

## **6.15 Output of Refinement Jobs**

By default, the refined structures are appended to the Project Table for the project currently open. If you would prefer the output structure to replace the input structure, or would rather not incorporate the output structure into the project at all, select the appropriate option from the Incorporate option menu in the Job Settings dialog box:

- Append new entries as a new group
- Append new entries individually
- Replace existing entries
- Do not incorporate

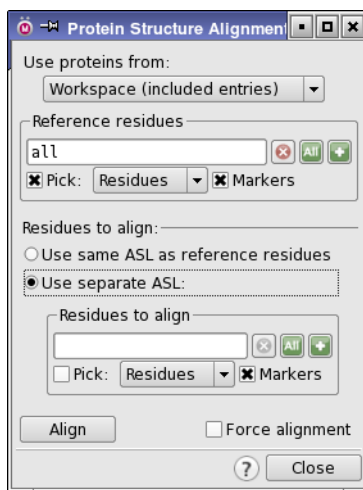
In addition to returning the structures, the output of a refinement job include several Maestro properties. The main properties are Prime Energy and Prime Energy Difference, which reports the difference in energy between the input and output structures. The energy difference is cumulative: if the property already exists, the difference is incremented in the current job. For loop refinement and side-chain prediction jobs, the lipophilic term is reported separately as Prime Lipo; the sum of the Prime energy and the lipophilic term is used to rank loop conformations.

# Protein Structure Alignment

It is sometimes useful to be able to align proteins outside the usual Prime workflow. Maestro provides access to the structure alignment facilities of Prime so that you can perform such alignments. Two panels are available from the tools menu: the Protein Structure Alignment panel, and the Align Binding Sites panel.

## 7.1 The Protein Structure Alignment Panel

To open the Protein Structure Alignment panel, choose Tools → Protein Structure Alignment.



**Figure 7.1. The Protein Structure Alignment panel.**

In the Protein Structure Alignment panel, structure alignment is performed on Project Table entries that have been included in the Workspace or on the selected entries. You can choose which to use from the Use proteins from option menu. If you choose Workspace (included entries) the reference structure, whose frame of reference is used for the alignment, must be the first included entry (the entry with the lowest row number). To make another protein the reference, move that protein above all the other included entries in the Project Table. If you choose Project Table (selected entries), the reference structure must be the only structure included in the Workspace.

By default, the alignment includes all residues. However, it is often useful to align a subset of residues that have greater similarity than the structures as a whole, or are more informative to compare. Because this program uses matching of secondary structure elements, the subset should have enough contiguous residues to include at least one helix or strand. Very small numbers of residues do not produce useful alignments.

When the alignment calculation is complete, the structures are placed in the same frame of reference in the Workspace, and the results of the calculation are listed in the Protein Structure Alignment - Results panel. The aligned residues are listed and an RMSD and Alignment Score are reported. The RMSD and alignment score are also added to the structures as Maestro properties, `StructAlign RMSD (r_esp_StructAlign_RMSD)` and `StructAlign Score (r_esp_StructAlign_Score)`. The RMSD is calculated based only on those residues that are considered to have been successfully aligned, and therefore does not represent an overall comparison of the structures. It is computed from the C-alpha atoms of the aligned residues. An Alignment Score lower than 0.6–0.7 indicates a good alignment. Alignment Scores greater than about 0.7–0.8, or a failure of the structural alignment calculation, indicates there is insufficient structural similarity for a meaningful alignment. For details on the algorithm, see the paper by Honig and Yang (*J. Mol. Biol.* **2000**, 301, 665).

To run a default alignment job, include two or more protein structures in the Workspace, open the Protein Structure Alignment panel, and click **Align** to start the alignment calculation.

The Protein Structure Alignment panel provides some tools for controlling which parts of the protein are aligned and how they are aligned. These tools are described below.

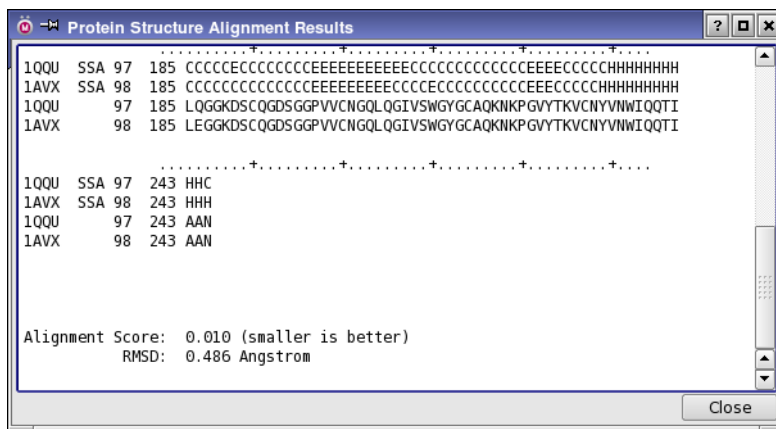
If you do not want to use all residues in the reference protein for the alignment, you can select the residues you want to use in the Reference residues section, by using the picking tools. The remaining residues are ignored in the alignment.

Alignment can be performed for all residues or for a subset of residues. There are two options:

- **Use same ASL as reference residues**—If you restricted the residues in the reference protein, select this option to use the same expression for the proteins to be aligned as for the reference. This does not necessarily mean the exact same residues (although it can), because the ASL expression is evaluated independently for each protein being aligned.
- **Use separate ASL**—Select this option to specify the subset with a different residue selection (ASL expression) from the reference, and use the picking tools to make the selection. You might need to try different subsets to obtain the most useful alignment. Subsets should include at least one secondary structure element (helix or strand). Clicking the **Common objects** button then choosing **Select** opens the **Atom Selection** dialog box to the **Residue** tab. Here you can select a range of residue numbers, specify portions of the residue sequence, or click **Secondary Structure** in the list of options on the left to choose res-

idues based on their location in Helix, Strand, or Loop structures. Subsets can be combined, intersected, or otherwise modified in the Atom Selection dialog box.

When the job starts, the Protein Structure Alignment Results panel opens, displaying the text “Calculating alignment”. When the job finishes, the results are displayed and the structures are aligned.



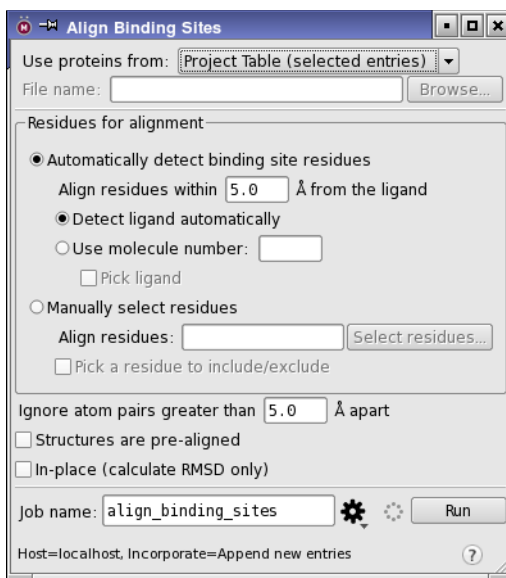
**Figure 7.2.** The Protein Structure Alignment Results panel.

## 7.2 The Align Binding Sites Panel

The Align Binding Sites panel allows you to align the binding sites (or other structural features) of members of a family of proteins. The Align Binding Sites job first runs a Protein Structure Alignment to obtain the global alignment and then automatically generates the list of atoms to use in a pairwise alignment of the C-alpha atoms from the residues selected.

To open the Align Binding Sites panel, choose Tools > Align Binding Sites in the main window.

The Use proteins from option menu provides choices of the source of protein structures. If you choose Project Table (selected entries), the entries that are selected in the Project Table are used. If you choose Workspace (included entries), the entries that are include in the Workspace are used. If you choose File, the Input file text box and Browse button become available, and you can browse for the input file. The first entry, either from the Project Table (lowest row index) or from the file, is the reference structure to which the other proteins are aligned.



**Figure 7.3. The Align Binding Sites panel.**

In the Residues for alignment section you select the residues that are to be used to align the proteins. There are two options, described below.

- **Automatically detect binding site residues**—Align residues within a specified distance of the ligand. The distance is specified in the Align residues within  $N$  Å from the ligand text box. Two options are available for specifying the ligand:
  - **Detect ligand automatically**—Detect the ligand by selecting the molecule that is most ligand-like (number of atoms).
  - **Use molecule number**—Specify the ligand by molecule number. You can enter a number in the text box, or select Pick ligand and pick a ligand atom in the Workspace.
- **Manually select residues**—Specify the residues to align by using any of the following methods:
  - Enter the residue specification in the Align residues text box, in the format *chain: number*.
  - Click Select residues to open the Atom Selection dialog box and define the residues.
  - Select Pick a residue to include/exclude and pick residues in the Workspace. Picking an already included residue excludes it.

The residues are listed in the Align residues text box, which you can edit to remove or add residues.

The alignment is performed on atom pairs. You can ignore atom pairs in the alignment process if they are further apart than a specified distance, by entering the distance in the Ignore pairs greater than  $N \text{ \AA}$  apart text box.

If you have proteins that are prealigned, you can select Structures are prealigned to suppress the global alignment of the protein structures, and align only by using the selected residues. If you want to calculate the RMSD between prealigned protein structures, select In place (calculate RMSD only).

When you have finished making settings, click Start to start the job.

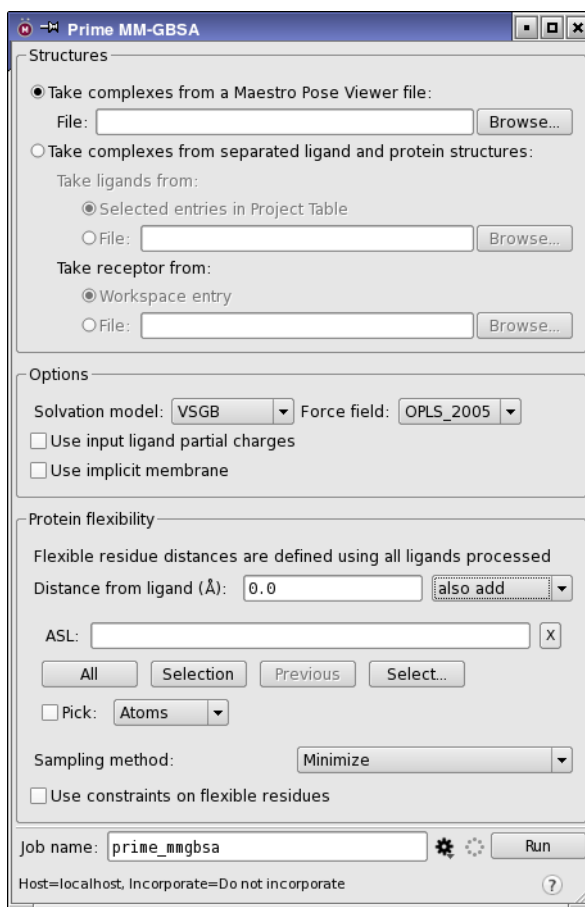
You can also run a job to align binding sites from the command line—see [Section 11.3 on page 129](#) for more information.





# Prime MM-GBSA

The Prime MM-GBSA panel can be used to calculate ligand binding energies and ligand strain energies for a set of ligands and a single receptor, using the MM-GBSA technology available with Prime. You can also run the job from the command line, which offers more flexibility than the panel.



**Figure 8.1. The Prime MM-GBSA panel.**

The ligands and the receptor must be properly prepared beforehand, for example, by using LigPrep and the Protein Preparation Wizard. Waters and ions must be removed from the protein, as an implicit solvent model is used. The ligands must be pre-positioned with respect to the receptor, and the receptor must be prepared as for a Prime refinement calculation.

## 8.1 Running Prime MM-GBSA from Maestro

To open the Prime MM-GBSA panel, choose Applications → Prime → MM-GBSA.

After preparing your structures, you can set up the calculation as follows:

1. Specify the source of the structures.

You can take structures from a Pose Viewer file (Glide output), or from separated ligand and protein structures. If you choose the latter option, you must ensure that the ligands and the protein are properly prepared and aligned.

2. (optional) Choose calculation settings.

To use partial charges from the input ligands rather than from the force field, select Use input ligand partial charges.

If you want to use the Prime implicit membrane model, select Use implicit membrane. The receptor you choose must have been already set up with the membrane, as described in [Section 6.11 on page 70](#).

3. (optional) Define a region within a certain distance of the ligand for which the protein structure will be relaxed in the calculation, in the Protein flexibility section.

Residues that have any atoms within the specified distance of the first ligand processed are included in the flexible region. By default, all protein atoms are frozen, and only the ligand structure is relaxed. The larger the flexible region, the longer the calculation takes.

You can adjust this region by choosing also add, exclude, or intersect with from the option menu to the right of the text box, and specify the residues to add (logical OR), exclude (logical AND NOT), or intersect with (logical AND) using the picking tools below.

You can choose the method used to treat the flexible region from the Sampling method option menu, from Minimize, Minimize side chains only, and Minimize polar hydrogens only.

The movement of the flexible residues can be constrained with a harmonic potential (default strength  $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ), so that they do not deviate too much from their initial position. To apply the constraints, select Use constraints on flexible residues.

4. Click Run to start the job with the current job name, host, and job settings, or click the Settings button to open the Job Settings dialog box to choose host and job settings, then click Run to start the job.

In the Job Settings dialog box, you can choose to incorporate the results into the Maestro project, set the job name, select the host, and set the number of CPUs and number of subjobs, if you are running the job on a multiprocessor host or submitting it to a queuing system. You can divide the ligand set between a specified number of subjobs, to achieve load balancing. The number of subjobs must be no smaller than the number of CPUs, and for optimal load balancing, should be several times the number of CPUs.

Instead of starting the job, you can choose Write from the Settings button menu to write the input files (structure input *jobname.maegz* and command input *jobname.inp*) and run the job from the command line.

To clear all the settings in the panel, choose Reset Panel from the Settings button menu.

## 8.2 Running Prime MM-GBSA from the Command Line

You can run Prime MM-GBSA from the command line, with the input file written from the panel. The syntax is as follows:

```
$SCHRODINGER/prime_mmgsa [options] input-file
```

The input file must be a Maestro file. By default, the receptor must be the first entry, followed by one or more ligands (i.e. Glide pose viewer file format). The file can be compressed (.mae.gz, .maegz) or uncompressed (.mae). If the `-ligand` option is specified, the file must contain ligand-receptor complexes.

To restart a failed job, run it again with the `-RESTART` option.

If you have a command input file, *jobname.inp*, as written from the panel, for example, you can replace the input structure file and the command options with this file. The command input file contains a keyword for the structure input, so it need not be specified. You can also specify command options as well as a command input file, to override values specified in the file.

The standard Job Control options are accepted—see [Section 2.3](#) of the *Job Control Guide*—except that `-HOST` and `-SUBHOST` are used for distributing jobs, not `-DRIVERHOST`. For a description of the program options, run the command with the `-h` option.

The MM-GBSA program was substantially rewritten in Suite 2012, with new output. If you are relying on output in the earlier format, you can run the old version with the command

```
$SCHRODINGER/run -FROM psp prime_mmgsa2011.py
```

## 8.2.1 Command Input File

If you want to use a command input file to store options, you can include any option other than the Job Control options. The input structure should be specified on a line with the following syntax:

```
STRUCT_FILE filename
```

All other options should be placed on separate lines with the preceding dash removed. For example, the following two calculations are equivalent:

```
prime_mmgsa file_pv.mae -job_type ENERGY -lcons SMARTS.C
```

and

```
prime_mmgsa input.inp
```

where the contents of `input.inp` are:

```
STRUCT_FILE file_pv.mae
JOB_TYPE      ENERGY
LCONS         SMARTS.C
```

The standard Job Control options must be set on the command line, because these are parsed by Job Control, not the MM-GBSA program. Other command options can also be specified on the command line, and override those in the file.

## 8.2.2 ASL Expressions

All atoms in the ligand have the `i_psp_Prime_MMGBSA_Ligand` property set to 1, so that ASL expressions can then be used for specifying parts of the structure relative to the ligand. For example,

```
fillres within 5 (atom.i_psp_Prime_MMGBSA_Ligand 1)
```

selects all residues within 5 Å of the ligand.

## 8.3 Prime MM-GBSA Output

The output Maestro structure file, which is compressed by default, includes a number of properties. Properties are calculated for the complex, the ligand, and the receptor, and strain energies can be calculated for the ligand and the receptor. For each of these properties, individual contributions of various types are calculated, as described in [Table 8.1](#). These properties are Prime properties, for which further information can be found in the following paper:

Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A., “The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling,” *Proteins*, **2011**, *79*, 2794.

The default file type is a pose-viewer file, except if the receptor is treated flexibly in a run started from Maestro, when the file contains complexes rather than a receptor and ligands.

The binding energy is calculated according to the equation:

$$DG_{\text{bind}} = E_{\text{complex}}(\text{minimized}) - E_{\text{ligand}}(\text{minimized}) - E_{\text{receptor}}(\text{minimized})$$

*Table 8.1. Output property types from a Prime MM-GBSA calculation.*

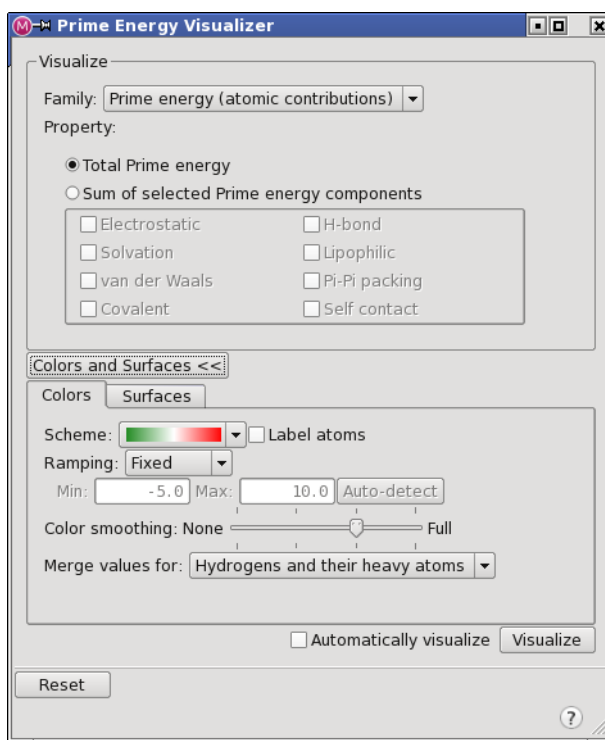
<b>Property</b>	<b>Description</b>
Coulomb	Coulomb energy
Covalent	Covalent binding energy
vdW	Van der Waals energy
Lipo	Lipophilic energy
Solv GB	Generalized Born electrostatic solvation energy
Energy	Prime energy
Hbond	Hydrogen-bonding energy
Packing	Pi-pi packing energy
SelfCont	Self-contact correction



# Visualizing Prime Energy Terms

The Prime energy and its components from an Analyze Energy job or an MM-GBSA job can be visualized in the Workspace, by using the Prime Energy Visualizer panel. This panel also supports visualization of other properties, such as the partial charge, ligand strain, solvent accessibility, residue interactions, and membrane permeability. The visualization colors the structure and an optional molecular surface by the values of a chosen property or sum of properties.

To open the Prime Energy Visualizer panel, choose Applications → Prime → Energy Visualizer or Tasks → Protein Analysis → Energy Visualization.



**Figure 9.1.** The Prime Energy Visualizer panel.

The structure whose properties you want to visualize must be displayed in the Workspace. If the Workspace is empty, you are prompted to display a structure.

### To visualize a property and its components:

1. Choose the property family to display from the Family option menu.

The properties listed under Property are updated to reflect your choice. The option menu shows all possible choices, but there might not be values for the properties for all of these choices, depending on the calculations you performed for the Workspace structure.

If the Workspace structure does not have the required properties, the area under Property is replaced with a warning message indicating what type of job to run to generate the data, and a button is displayed to open the panel for setting up and running the job. For properties from a Prime energy calculation, a Calculate button is shown, which runs the energy calculation with the default settings, and Maestro is unavailable while the job runs. For large structures, the job may take several minutes, so it may be better to run the job from the Calculate Energy panel. The panel also provides access to the other controls that are available in the Calculate Energy panel, if you want to use nondefault settings.

2. Select the property that you want to display.

You can choose to show the total property value or the components. When you select Sum of *property* Components, the check boxes for the individual components are made available. The values for the components are weighted according to their contribution to the total property value. Selecting multiple components displays the sum of the contributions from those components.

If you select Prime energy (residue interactions), you must pick a residue in the Workspace to show the interactions for the selected energy components. The residue you pick is displayed in the Residue for interaction display text box. The Pick a residue option is selected by default, for picking the desired residue.

3. Click Visualize to color the structure according to the values of the selected property.

If you want to update the Workspace automatically for every change that you make, select Automatically visualize. Residue interactions are automatically visualized, so these controls are unavailable if you select Prime energy (residue interactions).

If you want to modify the color scheme settings, or create surfaces, click Colors and Surfaces to display the controls for doing so. You can hide the controls by clicking the button again.

Four color schemes are available from the Scheme option menu: red-white-green, red-white-blue, red-blue, and rainbow. The blue or green values represent low energies or favorable values, the red values represent high energies or unfavorable values. In addition, you can choose None to color the atoms by element. This can be useful when labeling atoms by the property values rather than coloring them.



If you want to label the atoms with the property values, select Label atoms. The labels are colored by the atom color.

The color schemes are actually color ramps, for which the numerical values for the currently selected properties are mapped to colors. The ramp is simply defined by a minimum and maximum value: anything below the minimum is colored the same as the minimum, and anything above the maximum is colored the same as the maximum. Values in between are scaled linearly along the currently selected color ramp.

You can choose the following settings for the ranges represented by the ramp, from the Ramping option menu:

- **Fixed**—Use hard-wired minimum and maximum values that were chosen to give the most broadly useful range of colors. Each property has its own minimum and maximum values that are set according to the usual range of property values. If multiple properties are shown, the minimum and maximum values are set to the smallest of the minimum values for the selected properties and the largest of the maximum values.

Many of the terms have different inherent ranges. For example, the range for Pi-Pi packing is [-0.08,0.08], whereas the range for Electrostatic is [-20.0,20.0]. If you visualize Pi-Pi packing by itself, the overall ramp will be [-0.08,0.08], and the various pi-pi interactions will show up as strong blue. However, visualizing Pi-Pi packing in combination with Electrostatic means the overall ramp will be [-20.0,20.0], and the pi-pi contributions will be very subtle in comparison to the electrostatic contributions.

- **Adaptive**—Use the maximum and minimum values over the currently selected properties as the end points for the color. This choice can be useful for amplifying the differences in the current structure, to better differentiate the variations. However, if the properties only have favorable values, some of these values are colored red, which otherwise indicates that the values are unfavorable. In this context it means that they are the least favorable.
- **Custom**—Use the values given in the Min and Max text boxes as the ends of the ramp. These values are applied to all properties, which places them on the same scale. You can use the Auto-detect button to set the values in the Min and Max text boxes to the minimum and maximum property values found in the Workspace structure.

The changes in color from one part of the structure to another can be smoothed out to some extent by using the Color smoothing slider. Increasing the smoothing spreads the color over neighboring atoms by performing some averaging over the property values.

It can be useful to display values for different structural units, such as residues. The values are summed over all atoms in the structural unit. You can merge values for hydrogens into the attached heavy atoms, or merge values for entire residues, by choosing from the Merge values for option menu.

If you want to view the values on a surface, rather than the structure itself, click **Create Surface** in the **Surfaces** tab. A molecular surface is created and colored by the color of the nearest atom. The color is updated as you change the property that is used to color the atoms, so it is only necessary to create one surface. The resolution of the grid used to create the surface, and hence the quality of the surface, can be selected from the **Resolution** option menu. Once you have created a surface, you can show or hide it with the **Hide surface** option, which allows you to switch between viewing properties on the surface and on the structure.

You can choose a surface style, from **Solid**, **Dot**, or **Mesh**, and you can set the transparency of the surface (for solid surfaces) using the **Transparency** sliders. See [Chapter 12](#) of the *Maestro User Manual* for more information on surfaces.

# Command Syntax

This chapter summarizes the command syntax for various programs that are part of Prime. All the programs are executed by driver scripts, which run under the Schrödinger Job Control facility. For more information on this facility, see the *Job Control Guide*. The conventions used in describing the command syntax are outlined in the *Document Conventions* section at the front of this manual. For information on command options, run the command with the `-h` option.

Residue specifications are used with a number of the input file keywords for several of the scripts. These have the format `C:N`, where `C` is a one-character chain ID (‘\_’ if the chain ID is blank) and `N` is the residue number and insertion code (if any). Atom specifications similarly have the format `C:N:A`, where `A` is the 4-character PDB atom name with spaces replaced by underscores, e.g. `A:185:_SG_`.

## 10.1 blast—Run Blast Searches

The `blast` script provides an interface to the Blast and PSIBlast programs from NCBI (see <http://www.ncbi.nlm.nih.gov/books/NBK21097/> for information), while also supporting standard Schrodinger Job Control options. All options are passed on to the script via an input file.

### Syntax

```
$SCHRODINGER/blast jobname [options]
```

Input is read from the file `jobname.inp`. Output is written by default to `jobname.out`.

### Input File

Keywords control the operation of the `blast` script, and are listed in [Table 10.1](#). Most keywords are simply a translation of options available from the `blastall` or `blastpgp` executables distributed by NCBI. For optional keywords, the Blast default values are used. Since these defaults can differ between different version of Blast they are not listed here.

Table 10.1. Keywords for the blast script

Keyword	Description
QUERY_FILE <i>filename</i>	Required. Name of the FASTA file containing the input/query sequence
DATABASE {nr   pdb}	Database to be used in the search. In a default installation the two allowed values are nr and pdb.
ALIGNMENT_FORMAT <i>value</i>	Alignment view options (Blast -m option). Allowed values are: 0 pairwise 1 query-anchored showing identities 2 query-anchored no identities 3 flat query-anchored, show identities 4 flat query-anchored, no identities 5 query-anchored no identities and blunt ends 6 flat query-anchored, no identities and blunt ends 7 XML Blast output 8 tabular 9 tabular with comment lines m2io Maestro format (default)
EXPECT <i>value</i>	Expectation value (E) (Blast -e option).
FILTER {true   false}	Filter query sequence: DUST with blastn, SEG with others. (Blast -F option).
GAP_OPEN_COST <i>value</i>	Cost to open a gap (Blast -G option)
GAP_EXTEND_COST <i>value</i>	Cost to extend a gap (Blast -E option)
MATRIX <i>type</i>	Sequence similarity matrix to use for the alignments. Allowed values of <i>type</i> are BLOSUM45, BLOSUM62, BLOSUM80, PAM30, PAM70
MAX_ROUNDS <i>n</i>	Maximum number of passes to use in multipass version (PSI-Blast -j option); only applies to runs using PROGRAM blastpgp
E_VALUE_THRESHOLD <i>value</i>	e-value threshold for inclusion in multipass model (PSIBlast -h option).
OUTPUT_FILE <i>filename</i>	Name of the output file. Default: <i>jobname</i> .out.
PROGNAME <i>name</i>	Type of blast search to run. The allowed values are blastp, blastpgp, and b12seq (for pairwise alignments). Default: blastp.
TEMPLATE_FILE <i>filename</i>	Name of the FASTA file that contains the template sequence (b12seq -j option). Only applies to pairwise alignments using b12seq

Table 10.1. Keywords for the blast script (Continued)

Keyword	Description
PSSM_ASCII_OUTFILE <i>filename</i>	ASCII output file for PSI-BLAST matrix (PSIBlast -Q option).
EXPAND_HITS {true   false}	Expand redundant hits, based on information in the NCBI define for a particular hit. Only available when using Maestro format.
SHOW {all   pdb}	Allows the user to restrict the types of hits that are included in the output. all includes all hits; pdb only includes hits from the PDB. Only available when using Maestro format.

### Return Value and Errors

blast returns 0 if successful and a non-zero value in all other cases.

### Examples

The following is an example input file; query.aa is the query sequence in FASTA format.

```
QUERY_FILE      query.aa
PROGNAME        blastp
DATABASE        pdb
```

### Environment Variables

The following environment variables are supported by the blast script:

```
PSP_BLASTDB    directory containing the customized Blast databases
PSP_BLAST_DIR  directory containing a custom Blast installation
```

## 10.2 prime—Build and Refine Protein Structures

The prime program can be used to build initial (unrefined) protein models from an alignment to one or more templates, and to refine protein structures, using the Protein Local Optimization Program (PLOP). Structure refinement includes energy minimization, side-chain prediction, helix prediction, and loop prediction, with protocols for extended loop sampling, ultra-extended loop sampling, and cooperative loop pair sampling.

Multithreaded execution (with Open MP) can be enabled by setting the environment variable OMP\_NUM\_THREADS (see [Section 10.2.4 on page 111](#)).

The command syntax is

```
$SCHRODINGER/prime [options] jobname
```

The input is read from the file *jobname.inp* and the structural output is written to the file *jobname-out.maegz*. If the input structure file for a job has alternate coordinates, a refinement is performed on the first set of alternate coordinates. When building models, you can specify multiple input files to build a multimer model. The file *jobname.log* contains text output, and *jobname.err* contains any error messages.

The `-examples` option displays a set of example input files, which are included in [Section 10.2.5 on page 112](#).

## 10.2.1 Command Input File

The command input file contains a list of keywords that control the execution of the program, one per line. The keywords are listed in the tables below. Generally valid keywords are listed in [Table 10.4](#) and protocol-specific keywords are listed in [Table 10.5](#), under the protocol to which they apply. The keywords and values are case-insensitive, except where the case is important, such as for file names.

*Table 10.2. General keywords*

Keyword syntax	Description
JOB_TYPE <i>job</i>	Type of job to run. The available options are: MODEL           Build a structure for a homology model. REFINE           Refine a structure.
REDUN_CUTOFF <i>tol</i>	RMSD cutoff for eliminating redundant conformers, where relevant.
REMOVE_REDUNDANT {true false}	Where multiple structures can be generated, remove redundant structures, as determined by REDUN_CUTOFF.

Table 10.3. Keywords for homology model building jobs

Keyword syntax	Description
BUILD_DELETIONS {true   false}	Required. Turn on or off closure of the chain breaks near deletions in the alignment. If not closed, they will remain as chain breaks in the output structure.
BUILD_TRANSITIONS {true   false}	Required. Turn on or off closure of the junctions between templates in multi-template homology modelling jobs. If not closed, they will remain as chain breaks in the output structure.
COMPOSITE_ARRAY <i>string</i>	A string of integers indicating from which template (or, more specifically, from which alignment) coordinates for a given residue should be obtained. Thus, this string of integers must be equal in length to the query sequence being used in the alignment files. COMPOSITE_ARRAY is required for multiple template jobs. It is not required for single template jobs, but if it is present, it is ignored.
FAST_SIDE_OPT {true   false}	Turn on or off fast side chain optimization, in which a set of side chains is selected that has no steric clashes but no energy minimization is performed. Requires SIDE_OPT. Default: false.
KEEP_ROTAMERS {true   false}	Required. Specifies which side chains to optimize. If true, the rotamers for conserved side chains are preserved, and only non-conserved side chains are predicted. If false, all side chains are predicted. This keyword is ignored if SIDE_OPT is false.
KNOWLEDGE_BASED {true   false}	Specify the method used to build the model: knowledge-based (true), which is faster and does not do any minimization, or energy-based (false), which uses the full model-building capabilities. Default: false.
MAX_INSERTION_SIZE <i>n</i>	Specifies the longest insertion in the alignment that will be built. Insertions longer than this will be omitted, and not appear in the output structure. Residue numbering will reflect the full query sequence however. Default: 1000.
MINIMIZE {true   false}	Turn on or off minimization of regions that were not directly copied from the template structure. These regions primarily include portions of the structure involved in closing gaps or building insertions, but also include any side chains optimized due to the SIDE_OPT keyword. Default: true.
MIN_OVERLAP <i>value</i>	Fraction of van der Waals distance used to define a clash. Smaller values allow structures with worse potential atomic overlaps to be generated. Default: 0.70
NUM_OUTPUT_STRUCT <i>n</i>	Number of output model structures to return. Only valid with KNOWLEDGE_BASED true.

Table 10.3. Keywords for homology model building jobs (Continued)

Keyword syntax	Description
SIDE_OPT {true   false}	Turn on or off a side chain optimization stage. Default: true.
<i>template_ALIGN_FILE</i> <i>filename</i>	Required. The name of the file containing the alignment between this template and the query. Details on the format are provided in the Examples section below.
<i>template_HETERO_i</i> <i>ligand-spec</i>	Required if a ligand is present. Specifies a ligand to include from this template. The format of the specification is AAA C:### where AAA is the ligand's three letter code, C is its chain ID, ### is its residue number. Multiple ligands from the template are numbered using <i>i</i> as an index, starting from zero.
TEMPLATE_NAME <i>template</i>	Required. A label used to identify a given template in subsequent data fields. For example, if TEMPLATE_NAME is set to 1BPJ_A, other template-specific fields are given as 1BPJ_A_STRUCT_FILE, 1BPJ_A_ALIGN_FILE, and so on. The chain to be used is specified by an underscore and letter suffix: for example, for 1BPJ_A, chain A will be used. The chain suffix must be included in the label.
<i>template_NUMBER n</i>	Required. Used to identify a given template in the COMPOSITE_ARRAY (see below). Numbering goes from 1 to 9.
<i>template_STRUCT_FILE</i> <i>filename</i>	Required. The PDB file to be used for this particular template.
TEMPLATE_RESIDUE_NUMBERS {true   false}	Number the residues in the built structure the same as in the template, as far as possible. Does not apply to consensus modeling. When using multiple templates for a single chain, the residue numbers are taken from the first template. Sequential numbering from 1 is used if the attempt to use template numbering fails.
USE_SYMMETRY {true   false}	Turn on or off use of the crystallographic symmetry of the template file when constructing the homology model. Default: false.



Table 10.4. Keywords for all refinement jobs

Keyword syntax	Description
ADD_MISSING_SIDE_CHAINS {yes   no}	Add missing side chains in a default extended conformation. No sampling or refinement of the added side chains is performed.
CONSTRAINT <i>i constraint</i>	Define a constraint on atom positions, distances, angles or dihedral angles. See text for syntax details.
ECUTOFF <i>value</i>	Energy cutoff for return of conformations. Returns all conformations within <i>value</i> kcal/mol of the lowest-energy structure. It can be used in conjunction with NUM_OUTPUT_STRUCT, to impose a maximum on the number of such conformations to return. As with NUM_OUTPUT_STRUCT, it only applies to single loop refinements.
ENTRY <i>list</i>	Comma-separated list of entries in the input structure file to refine. Default: refine all structures.
EXT_DIEL <i>value</i>	Dielectric constant of the continuum solvation model. Default: 80.0. Used with SGB_MOD <i>sgbnp</i> .
INT_DIEL <i>value</i>	Dielectric constant used within the radius of an atom. Default: 1.0.
MIN_OVERLAP <i>value</i>	Fraction of van der Waals distance used to define a clash. Smaller values allow structures with worse potential atomic overlaps to be generated. Valid for LOOP_BLD and HELIX_BLD jobs. Default: 0.70
NUM_OUTPUT_STRUCT <i>n</i>	Number of conformations to return for single loop refinements or cooperative side-chain refinements. This has no effect on helix refinements or jobs composed of more than one refinement (e.g. 2 loops, 1 helix + 1 loop, etc.). All structures are returned in a single structure file in Maestro format. Default: 1.
PAIR_CONSTRAINT <i>i atom1, atom2, dist [, strength]</i>	Specify a constraint on a pair of atoms. The format of the atom specifications is given at the beginning of this chapter. The value of <i>dist</i> is the target distance between the atoms in angstroms. The value of <i>strength</i> is the coefficient of the harmonic constraint potential in kcal mol <sup>-1</sup> Å <sup>-2</sup> , which has a default value of 350.
PLANARITY_RESTRAINT <i>value</i>	Specify the factor used to multiply the potential for improper torsions, to reduce the amount by which rings can deviate from planarity.

Table 10.4. Keywords for all refinement jobs (Continued)

Keyword syntax	Description
PRIME_TYPE <i>type</i>	Type of calculation to perform when running a refinement job (JOB_TYPE REFIN). The available options are: SIDE_PRED Predict side chains (prefix SC) SIDE_COMBI Predict side chains cooperatively and exhaustively LOOP_BLD Predict loops and helices EXTENDED Predict extended loops (6-11 residues) LONG_LOOP_2 Predict ultra-extended loops (10 or more residues) LOOP_PAIR Cooperative refinement of two loops HELIX_BLD Rigid-body refinement of helices REAL_MIN Minimize energy (prefix MINI) SITE_OPT Active site minimization (including ligand) ENERGY Energy calculation
RESIDUE_ <i>i spec</i>	Specify a residue to refine. The numbering <i>i</i> begins at 0, and increases incrementally. The format of the residue specification <i>spec</i> is given at the beginning of this chapter.
RETAIN_CORRECTIONS {yes   no}	Retain corrections made when duplicate residue numbers are found. The duplications detected are between non-protein residues or between protein and non-protein residues. The non-protein residue is assigned a new, unique chain name temporarily. If you set this keyword to yes, the new chain name is written to the output file. Default: no.
SEED <i>n</i>	The integer to use as the seed for the random number generator if USE_RANDOM_SEED is no. If it is yes, SEED is ignored. Default: -1, to generate a random seed rather than use the supplied seed.
SELECT <i>string</i>	Specify the manner for specifying which residues to refine. Available options are: all Refine all residues pick Refine the specific residues, indicated by included RESIDUE_ <i>i</i> parameters (see above). file Read list of residues to refine from a file. The file consists of a series of lines with a residue specification on each line.
SGB_MOD {sgbnp   vdgbnp   vsgb1.0   vsgb2.0}	Specify the implicit solvation model to use. The choices are: sgbnp—use the standard generalized Born model. vdgbnp, vsgb1.0—use version 1 of the the variable-dielectric generalized Born model, which adapts the dielectric constant for residue type. vsgb2.0—use version 2 of the variable-dielectric generalized Born model, which incorporates a wider range of residue-dependent effects. Default: vsgb2.0.

Table 10.4. Keywords for all refinement jobs (Continued)

Keyword syntax	Description
STRUCT_FILE <i>filename</i>	Input structure file in Maestro format, compressed (.mae.gz, .maegz) or uncompressed (.mae). If the file contains multiple structures, each structure is refined, or the structures specified by ENTRY.
USE_CRYSTAL_SYMMETRY {yes   no}	Set to true if the input structure contains unit cell information and crystal symmetric atoms are to be included in calculation. Default: no.
USE_RANDOM_SEED {yes   no}	Indicates whether to use a random seed for the random number generator. This keyword has no effect on Minimization tasks. Default: no.
USE_MAE_CHARGES {yes   no}	Indicates whether to use atomic partial charges from the Maestro input file for untemplated residues (ligands, cofactors). Default: no.
USE_MEMBRANE {yes   no}	Indicates whether to use the implicit membrane model. The model must be set up from the Setup Membrane panel in Maestro. Default: no.
USE_SGB {yes   no}	Use the SGB implicit solvation model. If the model is not used, calculations are run in vacuum with a dielectric of 6.0. Default: no.

Table 10.5. Keywords for loop refinement jobs

Keyword syntax	Description
BURIED <i>i spec</i>	Constrain the specified residue to be buried (no more than 20% of surface area exposed to solvent).
CA_CONSTRAINT <i>i spec, [x, y, z,] value</i>	Specify a constraint on the position of a specific alpha carbon atom. The format of the residue specification <i>spec</i> is given at the beginning of this chapter. The optional coordinates specify the target position; if omitted, the target position is the initial position. The value is the maximum distance that the C-alpha atom can move from the target position, in angstroms. Specifying a target position allows you to move a loop to a desired location.
EXPOSED <i>i spec</i>	Constrain the specified residue to be exposed (more than 40% of surface area exposed to solvent).
LOOP <i>i_RES_j spec</i>	Specify residue <i>j</i> for defining loop <i>i</i> . Loops are numbered sequentially, beginning at 0. Two occurrences of this keyword are required, with <i>j</i> =0 (the beginning of the loop) and <i>j</i> =1 (the end of the loop). <i>spec</i> is a residue specification as defined above.
HELIX_BUILD <i>spec1/spec2</i>	Build the specified sequence as an alpha helix. The format of the residue specifications <i>spec1</i> and <i>spec2</i> is given at the beginning of this chapter. You should include enough residues on either side of the helical region in the loop prediction to ensure that sufficient flexibility is available to build the loop.
MAX_CA_MOVEMENT <i>value</i>	Specify the maximum distance that any alpha carbon atom in the loop should be allowed to move during refinement. Omitting this parameter indicates that no restriction should be applied.
MEMBRANE <i>inside outside</i>	Place the C-alpha atoms of the loop either inside or outside the membrane in the building stage. Refinement can move the residues across the membrane boundary.
RES_SPHERE <i>value</i>	All side chains that lie within this distance of the loop will also be optimized during refinement. This allows these nearby side chains to “react” to the loop being predicted. Default 0.0.
LOOP_NCLUST <i>number</i>	Specify the number of clusters to score for loop prediction. More clusters should result in more accurate predictions albeit with greater computational cost.

Table 10.6. Additional keywords for extended loop and multiple loop refinement jobs.

Keyword syntax	Description
MAX_CA_REF1 <i>value</i>	Specify the maximum distance that any alpha carbon atom in the loop should be allowed to move during refinement stage 1 (Ref1). Omitting this parameter indicates that no restriction should be applied.
MAX_CA_REF2 <i>value</i>	Specify the maximum distance that any alpha carbon atom in the loop should be allowed to move during refinement stage 2 (Ref2). Omitting this parameter indicates that no restriction should be applied.
MAX_JOBS <i>n</i>	Maximum number of subjobs that can run at one time. This can be set to a value less than the value of <code>THREADS</code> if sufficient resources are not available. Not available with <code>refinestruct</code> .
THREADS <i>n</i>	Number of simultaneous jobs to be run during selected refinement stages. Determines the sampling level for some protocols. Note: the jobs specified by <code>THREADS</code> do not have to actually run in parallel—see <code>MAX_JOBS</code> .
NUM_FIXED_STAGE <i>n</i>	Specify the number of stages in which backbones of residues at the beginning or end of the loop are fixed. These stages are inserted between the CA refinement stages. Each stage fixes progressively more residues, starting with 1. The fixed residues are distributed in all combinations between the ends of the loop, so, for example, stage 4 has 5 combinations and therefore 5 refinement jobs. Default: 5.
OFAC_INIT_STAGE <i>factors</i>	Run several initial stages with the specified minimum overlap factors (see <a href="#">MIN_OVERLAP</a> ). The factors are specified as a list separated by slashes, with no spaces. One stage is run for each factor. Default: 0.65/0.70/0.75.
PRIME_TYPE <i>n</i>	Type of refinement to be carried out. Can be specified as either a number or text string. Different from <code>refinestruct</code> . Available choices are: 0 or <code>default</code> : Equivalent to running a single loop refinement using <code>refinestruct</code> . Not particularly useful except for completeness and testing purposes. Does not recognize <code>THREADS</code> keyword. 1 or <code>extended</code> : Used for Extended Sampling from the GUI. Does not recognize <code>MIN_OVERLAP</code> and <code>MAX_CA_MOVEMENT</code> keywords, as these are used internally by the protocol. Accepts multiple loops as input which are run sequentially. 2 or <code>long_loop</code> : No longer used. 3 or <code>loop_pair</code> : Used for Cooperative Loop Sampling from the GUI. Exactly two loops must be specified. 4 or <code>long_loop_2</code> : Used for Ultra-extended Sampling from the GUI. Does not recognize the <code>MAX_CA_MOVEMENT</code> keyword, as this is used internally by the protocol. Only one loop can be specified.
STRUCT_FILE	Required. The input structure file in Maestro format, compressed or uncompressed. Same as for <code>refinestruct</code> .

Table 10.6. Additional keywords for extended loop and multiple loop refinement jobs.

Keyword syntax	Description
SEG_LIST <i>filename</i>	Name of a file containing the list of loops to be refined. For the example given for the keyword LOOP_i_RES_j, the file would have a single line: loop A:10 A:20 Not available with refinestruct.

Table 10.7. Keywords for active site optimization (SITE\_OPT) jobs

Keyword syntax	Description
BACKBONE_LEN <i>n</i>	Number of residues to include in the loop defined for backbone sampling in side-chain prediction. Only used with SAMPLE_BACKBONE. Default: 3.
LIGAND <i>string</i>	Residue specification for the ligand.
MAXCONEANG <i>value</i>	Maximum angular displacement of the CA-CB vector from the initial position when sampling CB positions. Only used with SAMPLE_CBETA. Default: 30°.
NITER_SIDE <i>n</i>	Number of iterations of side-chain prediction to perform. Larger numbers generally give better results. If set to zero, only the initial selection of rotamers without clashes is performed. Default: 1.
NPASSES <i>n</i>	Number of passes through side chain optimization of the protein residues followed by minimization of the protein residues (including the backbone). These two steps are repeated the specified number of times before proceeding to optimization of the protein and the ligand. Default: 2. Files written from Maestro have NPASSES set to 1.
SAMPLE_BACKBONE {yes   no}	Sample the backbone by running a loop prediction on a set of residues centered on the residue for which the side chain is being refined. The number of residues in the loop is defined by the BACKBONE_LEN keyword. Default: no.
SAMPLE_CBETA {yes   no}	Sample CB positions in a conical region around the initial position. The maximum displacement of the CA-CB vector in this region is set by the MAXCONEANG keyword. Default: no.

Table 10.8. Keywords for helix rigid refinement (*HELIX\_BLD*) jobs

Keyword syntax	Description
<i>HELIX_0_RES_i spec</i>	Specify helix and region containing helix. Values of <i>i</i> are: 0 start of region containing helix 1 start of helix 2 end of helix 3 end of region containing helix A keyword with each value of <i>i</i> must be included in the input.
<i>MAX_MOVEMENT value</i>	Maximum amount of movement allowed in the C and the N terminus of the helix. Overrides <i>MAX_C_MOVEMENT</i> and <i>MAX_N_MOVEMENT</i> . Default: 5 Å.
<i>MAX_C_MOVEMENT value</i>	Maximum amount of movement allowed in the C terminus of the helix. Default: 5 Å.
<i>MAX_N_MOVEMENT value</i>	Maximum amount of movement allowed in the N terminus of the helix. Default: 5 Å.
<i>ORIGIN_FOR_ROTATION value</i>	Fraction of the distance from the start to the end of the helix that is used as the origin for rotation. Default: 0.5.
<i>ROLL_HI value</i>	Minimum amount of roll sampled, in degrees. Default: 0.
<i>ROLL_LO value</i>	Maximum amount of roll sampled, in degrees. Default: 0.
<i>ROLL_STEP value</i>	Increment in roll between <i>ROLL_LO</i> and <i>ROLL_HI</i> , in degrees.
<i>SIDE_OPT_DISTANCE_CUTOFF</i>	Distance in angstroms to optimize side chains around the helix region.

Table 10.9. Keywords for side-chain prediction (*SIDE\_PRED*) jobs

Keyword syntax	Description
<i>NITER_SIDE n</i>	Number of iterations of side-chain prediction to perform. Larger numbers generally give better results. If set to zero, only the initial selection of rotamers without clashes is performed. Default: 1.

Table 10.10. Keywords for minimization (*REAL\_MIN*) jobs

Keyword syntax	Description
<i>MINIM_NITER n</i>	Maximum number of minimization cycles for the solvent treatment. Each cycle consists of a minimization to the specified RMSD with a particular set of solvent parameters, after which the solvent parameters are updated. The iterative procedure finishes if the energy changes by less than 1 kcal/mol or the number of cycles is exceeded. Default: 2.
<i>MINIM_RMSG value</i>	RMSD gradient convergence threshold. Default 0.01 kcal mol <sup>-1</sup> Å <sup>-1</sup>

## 10.2.2 Constraint Syntax

The syntax of a general constraint (CONSTRAINT\_*i*) is:

*atom-spec1*; *atom-spec2*; ... ; *strength*; *equilibrium-value*

From one to five atom specifications can be given, in the format described at the beginning of this chapter. Positional and distance constraints have other specification options, as described below. A positional constraint requires one atom (or atom group) specification, a distance requires two atom (or atom group) specifications, an angle requires three atom specifications, and a dihedral requires four atom specifications.

*strength* is the constraint strength, the coefficient of the harmonic potential. It is given in kcal mol<sup>-1</sup> Å<sup>-2</sup> for distances and kcal mol<sup>-1</sup> radian<sup>-2</sup> for angles.

*equilibrium-value* is the equilibrium value of the coordinate, and can be a single value or a range specified by *a+-b* (meaning from *a-b* to *a+b*), *a+* (greater than *a*) or *a-* (less than *a*). The constraint potential is zero inside the range (flat-bottomed potential) and harmonic outside. The format *a+* is useful for keeping atoms separated; and *a-* is useful for constraining atoms near a starting point.

**Positional constraints:** specify one or more atoms with an atom or residue specification, or an ASL expression in the form *asl=(expression)*. The strength must be given; the equilibrium value is optional. If it is omitted, the atoms are constrained to their starting positions. If it is given, the syntax *a-* specifies the maximum distance from the starting position.

**Distance constraints:** specify two sets of atoms or a set of atoms and a point in space. The center of mass of the sets of atoms is constrained. The atoms can be specified with a atom or residue specification, or an ASL expression in the form *asl=(expression)*. The second specification can also be a set of Cartesian coordinates *x, y, z* or *current*, meaning the current location. For example *A:12;0.0,0.0,1.0;6.0;0.5-* constrains the center of mass of residue *A:12* to a position within 0.5 Å of the position at 1.0 Å in the *z* direction from the origin.

**Angle and dihedral constraints:** specify three atoms for an angle constraint and four atoms for a dihedral constraint. The strength and equilibrium distance are required.

To specify degrees for any quantity instead of radians, prepend the letter *D*. For example, in the constraint specification

```
CONSTRAINT_0 _:900:_C12;_:900:_C11;_:900:_N1_;_:900:_C19;D300;D60+-10
```

the force constant is specified in degrees as *D300*, and the dihedral value and range in degrees as *D60+-10*.



### 10.2.3 Restarting an Incomplete Loop Refinement Job

If a loop refinement job fails (one with protocol `EXTENDED`, `LOOP_PAIR`, or `LONG_LOOP_2`), you can restart it by submitting it again from the same directory. The `jobname.restart` file and the archives of subjobs must be present in the directory, and the directory must be accessible from the host on which the driver job runs. Also, the job records for the failed jobs must not be purged from the job database, so that the restart mechanism has access to information on the failed subjobs. If the job records are missing, the subjobs are repeated.

Failed jobs that were started from Maestro must be restarted from the command line. Use `-PROJ` to assign the job to the Maestro project and `-DISP` to set the incorporation mode.

Once the job finishes successfully, the subjob archives and the restart file are removed.

It is not necessary to use `-LOCAL` or `-SAVE` to ensure that a job can be restarted, but there must be sufficient space in the working directory for the subjob archives.

### 10.2.4 Environment Variables

In addition to those used for all Schrödinger software, `prime` uses the environment variables listed in [Table 10.11](#):

*Table 10.11. Environment variables used by the prime script.*

Variable	Description
<code>OMP_NUM_THREADS</code>	Set the number of threads to use in multithreaded execution (with Open MP). It is recommended to set this environment variable in the <code>schrodinger.hosts</code> file for hosts on which multithreading is supported—see <a href="#">Section 7.1</a> of the <i>Installation Guide</i> .
<code>PSP_OPLS_VERSION</code>	OPLS force field version. The default is 2005. It can be set to 2001 to use the earlier set of force field parameters for ligands and non-standard residues.
<code>PSP_DEBUG</code>	If set, debugging is turned on and the subdirectory containing intermediate files is kept. Equivalent to using the command-line option <code>-DEBUG</code> .
<code>PSP_LOCAL</code>	If set, all calculations run in the local launch directory. Equivalent to using the command-line option <code>-LOCAL</code> .
<code>PSB_SGB_MOD</code>	Specify the solvent model. The value should be set to the same as the <code>SGB_MOD</code> keyword. This environment variable is overridden by a value in the input file.



### 10.2.5.3 Side-Chain Prediction

The following example demonstrates a side chain prediction of residues A:100, A:101, A:102, and A:103.

```
STRUCT_FILE input_structure.maegz
PRIME_TYPE SIDE_PRED
SELECT pick
RESIDUE_0 A:100
RESIDUE_1 A:101
RESIDUE_2 A:102
RESIDUE_3 A:103
USE_CRYSTAL_SYMMETRY no
USE_RANDOM_SEED no
SEED 0
EXT_DIEL 80.00
USE_MEMBRANE no
```

### 10.2.5.4 Minimization

The following example demonstrates the minimization of all hydrogens.

```
STRUCT_FILE input_structure.maegz
PRIME_TYPE REAL_MIN
SELECT as1 = (atom.ele H)
USE_CRYSTAL_SYMMETRY no
USE_RANDOM_SEED no
SEED 0
EXT_DIEL 80.00
USE_MEMBRANE no
```

### 10.2.5.5 Loop Prediction

The following example demonstrates loop prediction of loop A:102 to A:110.

```
STRUCT_FILE input_structure.maegz
PRIME_TYPE LOOP_BLD
LOOP_0_RES_0 A:102
LOOP_0_RES_1 A:110
RES_SPHERE 7.50
MIN_OVERLAP 0.70
USE_CRYSTAL_SYMMETRY no
USE_RANDOM_SEED no
SEED 0
EXT_DIEL 80.00
USE_MEMBRANE no
```

## 10.3 ska—Protein Structure Alignment

The SKA program carries out structural alignment of proteins by aligning protein backbones at different levels of detail.

The `ska` script provides an interface to the SKA structural alignment program. As with other computational programs in Prime its operation is controlled through simple input files. The allowed keywords and their values are discussed in more detail below. The `ska` script allows you to run the SKA program in three different modes under Job Control:

- **align mode**—SKA aligns two or more protein structures to one another.
- **dbcreate mode**—SKA creates databases for use with scan mode. You should split PDB files with multiple chains into single-chain files before adding them to the database, as SKA has difficulty aligning to multiple chains.
- **scan mode**—SKA searches a database of protein structures and identifies potential structural analogues. By default a non-redundant database of protein chains from the PDB is searched.

### Syntax

```
$SCHRODINGER/ska [options] jobname
```

By default the input is read from the file `jobname.inp` and the output is written to `jobname.out`.

`ska` does not generate output other than the explicitly defined database and a log file in `dbcreate` mode—see below for details.

### Options

The only options apart from Job Control options accepted by `ska` are `-DEBUG`, which turns on debugging mode for both Job Control and the SKA program, and `-NOJOBID`, which runs the program outside Job Control.

### Command Input File

The input file consists of lines containing keyword-value pairs, one per line. The keywords are described in [Table 10.12](#). The set of allowed and required keywords depends on the mode, which is determined by the `MODE` keyword.

### Return Value and Errors

The script returns 0 for success and 1 for failure.

Table 10.12. Keywords syntax for *ska* input file

Keyword syntax	Mode	Description
MODE <i>mode</i>	all	The SKA operating mode; allowed values are: <i>*align*</i> : structurally align two or more protein structures to on another. This is the default mode <i>*scan*</i> : given a query structure scan a database of protein structures and identify potential structural homologues <i>*dbcreate*</i> : create a database that can be search in scan mode from a list of PDB files
QUERY_FILE <i>filename</i> [ <i>name</i> ]	<i>*align*</i> <i>*scan*</i>	Required. The input query structure. In <i>*align*</i> mode, the structure against which other proteins are aligned. In <i>*scan*</i> mode, the structure used for scanning the database. The optional <i>name</i> ( <i>*align*</i> mode only) sets the name of this structure to a value other than the filename (the default). The name is displayed in the output file (see <a href="#">Examples</a> , below), <i>name</i> cannot contain the file separator character ('/' on UNIX).
TEMPLATE <i>filename</i> [ <i>name</i> ]	<i>*align*</i>	Required. PDB structures that are to be aligned to the query structure, set by QUERY_FILE. The TEMPLATE keyword can appear multiple times in order to be able to carry out multiple structure alignments (aligning more than one template structure to the query structure). The optional <i>name</i> sets the name of this entry to a value other than the filename (the default). The name is displayed in the output file (see <a href="#">Examples</a> , below), <i>name</i> cannot contain the file separator character ('/' on UNIX).
DATABASE <i>filename</i>	<i>*scan*</i>	The absolute path to the database to scan with the query structure. The database can be generated in <i>dbcreate</i> mode. Default: database supplied with Prime.
DBNAME <i>filename</i>	<i>*dbcreate*</i>	The name of the file in which to store the database create by <i>ska</i> . <i>filename</i> must be the file name without any path specification.
FILELIST <i>filename</i>	<i>*dbcreate*</i>	A file containing a list of absolute filenames for structures (in PDB format) that are to be included in a database of templates that can be search with <i>ska</i> in scan mode. Absolute filenames are required to avoid having to copy large numbers of PDB files to the temporary directory in which the job is run.

Table 10.12. Keywords syntax for *ska* input file (Continued)

Keyword syntax	Mode	Description
GAP_OPEN <i>value</i>	*align* *scan*	Gap initiation penalty.
GAP_DEL <i>value</i>	*align*	Deletion penalty.
OUTPUT_FILE <i>filename</i>	*align* *scan*	File in which to store the structural alignment. Default is <i>jobname.out</i> .
ALISTRUCS_OUTFILE <i>filename</i>	*align*	File in which to store the 3D coordinates of the aligned structures.
PSD_THRESHOLD <i>value</i>	*align* *scan*	PSD threshold for inclusion in the MSA. Overridden by RECKLESS.
SSE_MINSIM <i>value</i>	*align* *scan*	Minimum Secondary Structural Element (SSE) similarity for the alignment.
SSE_MINLEN <i>value</i>	*align* *scan*	Minimum SSE length for the alignment.
SSE_WINLEN <i>value</i>	*align* *scan*	Align only SSE windows of length <i>value</i> .
SWITCH [true false]	*align* *scan*	Carry out global alignment first and then align sub-windows if no similarity was found.
ORDER [seed psd]	*align*	psd: order the alignment by PSD to the seed structure.
CLEAN	*align*	Exclude structures with a PSD < PSD_THRESHOLD from the alignment.
RECKLESS [true false]	*align* *scan*	Force the alignment of structures even if no similarity is found.
USE_AUTOMATIC_SETTINGS	*align*	Automatically run scanning alignment with lengths 8, 9, and 13, and choose the best alignment.

In case of failure, check the log file for details on what might have caused the program to fail. To turn on debugging output use the `-DEBUG` command line option as follows:

```
$SCHRODINGER/ska -DEBUG jobname
```

### Environment Variables

The *ska* program handles all required environment variables internally and requires only that the environment variable `SCHRODINGER` be set. Should the environment variable `TROLLTOP` exist in the current environment it will be overwritten.

## Examples

The following minimal input file (assumed to be called `align.inp`) structurally aligns two protein structures to one another:

```
QUERY_FILE query.pdb
TEMPLATE template.pdb
```

where `query.pdb` is the name of the PDB file containing the query structure (the structure to which we want to align) and `template.pdb` contains the template structure (the structure we want to align to the query).

When run using `$SCHRODINGER/ska align` it should produce the following output:

- `align.out`—a file containing the structural alignment
- `align.log`—a log file that contains log information and diagnostics on the run

Note that the actual 3D coordinates of the superposed structures are not written out by default. You must use `ALISTRUCS_OUTFILE` to write out the structures.

The following input file (`align.inp`):

```
QUERY_FILE query.pdb
TEMPLATE template.pdb
ALISTRUCS_OUTFILE superposition.pdb
```

saves the 3D coordinates of the aligned structures in the file `superposition.pdb`.

The following is a sample input file for database creation in `dbcreate` mode:

```
MODE dbcreate
FILELIST filelist.dat
DBNAME mydatabase.dat
```

The file `filelist.dat` contains the file names for the templates, including the path, as follows:

```
/scr/templates/template-1.pdb
/scr/templates/template-2.pdb
/scr/templates/template-3.pdb
/scr/templates/template-4.pdb
```

A minimal input file for a scan is as follows:

```
MODE scan
QUERY_FILE query.pdb
```

Running `ska` with this input file would scan the database supplied with Prime for structural analogs. The output from the structural search contains pairwise alignments between the query

structure and every structure in the database (assuming the two structures are sufficiently similar). The results can be analyzed in a more convenient way with the graphical SKA Results Viewer utility, \$SCHRODINGER/utilities/SkaResultsViewer.

The use of the optional *name* argument is illustrated next. For the following input file:

```
QUERY_FILE query.pdb 1ctf
TEMPLATE 1dd3_A.pdb
```

the output file is as follows:

```

.....+.....+.....+.....+.....+.....
1ctf      1                                     ceeeeecc
1dd3_A    1  cchhhhhhhhhhhcchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhccceeeeecc
1ctf      1  -----EFDVILKAA
1dd3_A    1  MTIDEIIEAIEKLTVSELAELVKKLEDKFGVTAAPVAVAAAPVAGAAAGAAQEEKTEFDVVLKSF

```

```

.....+.....+.....+.....+.....+.....+.....
1ctf     62  ccchhhhhhhhhhhcchhhhhhhhhhhc c eeeeeccchhhhhhhhhhhhhhhcceeeee
1dd3_A   67  ccchhhhhhhhhhhcchhhhhhhhhhhcccccceccchhhhhhhhhhhhhhhcceeeee
1ctf     62  GANKVAVIKAVRGATGLGLKEAKDLVESA-P--AALKEGVSKDDAEALKKALEEAGAEVEVK
1dd3_A   67  QONKIQVIKVVREITGLGLKEAKDLVEKAGSPDAVIKSGVSKEEABEIKKKLEEAGAEVELK

```

```
RMSD: 0.763733
PSD: 0.024524
```

Without the *name* argument for the QUERY\_FILE keyword, the output would be

```

.....+.....+.....+.....+.....+.....
query     1                                     ceeeeecc
1dd3_A    1  cchhhhhhhhhhhcchhhhhhhhhhhhhccchhhhhhhhhhhhhhhhhhhhhccceeeeecc
query     1  -----EFDVILKAA
1dd3_A    1  MTIDEIIEAIEKLTVSELAELVKKLEDKFGVTAAPVAVAAAPVAGAAAGAAQEEKTEFDVVLKSF

```

```

.....+.....+.....+.....+.....+.....+.....
query     62  ccchhhhhhhhhhhcchhhhhhhhhhhc c eeeeeccchhhhhhhhhhhhhhhcceeeee
1dd3_A   67  ccchhhhhhhhhhhcchhhhhhhhhhhcccccceccchhhhhhhhhhhhhhhcceeeee
query     62  GANKVAVIKAVRGATGLGLKEAKDLVESA-P--AALKEGVSKDDAEALKKALEEAGAEVEVK
1dd3_A   67  QONKIQVIKVVREITGLGLKEAKDLVEKAGSPDAVIKSGVSKEEABEIKKKLEEAGAEVELK

```

```
RMSD: 0.763733
PSD: 0.024524
```



## 10.4 ssp—Secondary Structure Prediction

The `ssp` script provides an interface to the secondary structure prediction programs SSPro, which is bundled with Prime, as well as PSIPRED, which is available from a third party.

### Syntax

```
$SCHRODINGER/ssp [options] jobname
```

By default the input is read from the file `jobname.inp` and the output is written to `jobname.out`.

### Command Input File

The command input file consists of lines containing keyword-value pairs, one per line. The keywords are described in [Table 10.13](#).

Table 10.13. Keywords for the `ssp` input file.

Keyword syntax	Description
QUERY_FILE <i>filename</i>	Required. Name of the FASTA file containing the query sequence
METHOD <i>program</i>	Required. Name of the secondary structure prediction that is to be run. Allowed values are <code>sspro</code> , <code>psipred</code> , <code>profking</code> , and <code>all</code> . <code>all</code> means run all available programs. If the input file contains multiple METHOD keywords, only the last one is used.
OUTPUT_FILE <i>filename</i>	Optional. Name of the output file. Default is <code>jobname.out</code> .

### Return Value and Errors

`ssp` returns 0 if successful and a non-zero value in all other cases.

### Environment Variables

The environment variables that can be used to control the secondary structure prediction programs are listed in [Table 10.14](#).

Table 10.14. Environment variables for secondary structure prediction programs.

Environment Variable	Description
PSP_SSPRO_DB	BLAST database to be searched when assembling the multiple sequence alignment used in the SSPro prediction
PSP_PSIPRED_DB	BLAST database to be searched when assembling the sequence profile used in the PSIPRED prediction

## Examples

The following minimal input file, called `sspro.inp`, generates an SSPro secondary structure prediction for the sequence stored in file `query.fasta`:

```
QUERY_FILE query.fasta
METHOD      sspro
```

The prediction produced by the back-end is written to the file `sspro.out`.

The following input file, assumed to be called `all-methods.inp`, instructs the `ssp` script to produce secondary structure prediction for all methods available—SSPro and, depending on the local installation, PSIPRED.

```
QUERY_FILE query.fasta
METHOD all
```

## 10.5 sta—Single Template Alignment

The Single Template Alignment (STA) program in PRIME is designed for protein sequences with medium to high sequence identity (>20%). STA differs from standard sequence alignment programs such as BLAST by taking into account secondary structure matching as well as profile-sequence matching. As a result, STA often generates better alignment than BLAST in the regions where sequence conservation is relatively weak. STA can also take in constraints, such as user-selected residue pairs to be aligned together or residues in either query or template that are not to be aligned (gaps).

STA can use templates directly from the PDB, and can also use templates from the Prime fold library.

### Syntax

```
$SCHRODINGER/sta [options] jobname
```

The input is read from the file `jobname.inp` and the output is written to `jobname.out`.

### Command Input File

The input file consists of lines containing keyword-value pairs, one per line. The keywords are described in [Table 10.15](#).

Table 10.15. Keywords for the *sta* program.

Keyword syntax	Description
QUERY_NAME <i>name</i>	Name of the query sequence.
QUERY_FILE <i>filename</i>	File name of the query sequence in FASTA format.
FORMAT <i>string</i>	Alignment file format. Default is Maestro. Set to dev for plain text format—see <a href="#">page 122</a> for an example.
BGNRES <i>n</i>	The first residue of the query that is used in alignment. Default: 1.
ENDRES <i>n</i>	The last residue of the query sequence that is used. Default: the last residue of the original query sequence.
PREDSS[ <i>i</i> ] <i>filename</i>	Secondary structure prediction file of the query sequence, in CASP format. One occurrence of this keyword is required for each file, and there must be at least one SSP file specified. If only one prediction is used, the keyword is PREDSS; if more predictions are used, the keyword is suffixed with an index <i>i</i> , starting from zero: PREDSS_0, PREDSS_1, and so on.
TEMPLATE_PDB <i>filename</i>	Filename of the template structure, in PDB format.
MODE <i>string</i>	Optional. Needed in the input file only when there are user-specified constraints. In that case, the value is set to <i>user</i> .
PAIR <i>pair-list</i>	Optional. User-specified residue pairing, used only when MODE is set to <i>user</i> . All the residue pairs can be added after the keyword in one or multiple lines. For example, the following lines pair residue 12, 20, 60 from the query sequence with residue 15, 24, 65 from the template. <pre>MODE user PAIR 12 15 20 24 PAIR 60 65</pre>
GAP <i>gap-list</i>	Optional. User-specified gaps, used only when MODE is set to <i>user</i> . Can be used with or without PAIR. Gaps are specified by the residues in either query (P) or template (T) that are aligned with them. For example, the following opens gaps against query residue 10, 11, template residue 34 and 35. <pre>MODE user GAP P10 P11 T34 T35</pre>

## Files

- Input file—named *jobname.inp*. Contains keywords for running the program.
- Query sequence file—File containing the complete sequence of the query, in FASTA format. Specified in the input file.
- Secondary structure prediction files in CASP format for the query sequence.
- Template PDB file—Template structure file, in PDB format. The sequence used for the template is taken from SEQRES lines in the PDB file. If SEQRES is not found in the PDB file, the template sequence is then taken from ATOM records.
- Output alignment file—File named *jobname.out* containing pairwise alignment between the query and the template. By default, the alignment is in Maestro format. To generate plain text format, include `FORMAT dev` in the command input file. In plain text format, aligned residues from the query and the template are placed on top of each other—see below for an example.
- Log file—File named *jobname.log* containing progress of the `sta` program, including warnings and error messages.

## Return Value and Errors

`sta` returns 0 for success and non-zero for failure. In case of failure, check the log file for details, and search for `ERROR`. To further analyze the failure, use the `-DEBUG` option when you run `sta`.

## Examples

Below is an example input file, where `query_seq.fasta` stores the query sequence in FASTA format.

```
QUERY_NAME      query
QUERY_FILE      query_seq.fasta
PREDSS_0        query-ssp1.casp
PREDSS_1        query-ssp2.casp
PREDSS_3        query-ssp3.casp
BGNRES          1
ENDRES          149
TEMPLATE_PDB    pdb1vpe.ent
```

Below is an example of the plain text format for output alignment files. This format is generated when you include `ALIFORMAT dev` in the command input file.

```
probe length=70=====tmplt_T0281_d1dq3a2 length=79 TotalScore= 42 SEQ= -4 SSTR= 102
  SEQID= 0.061538 nali= 65
probe bgnRes=1 endRes=70      template bgnRes=3 endRes=75
```

```
ProbeAA: ..MWMPPRPEEVARKLRRLLGFVERMAKGGHRLYTHPDGRIVVVPFH...SGELP...KG
ProbeSS: LLLLLLHHHHHHHHHLLLEEEEEELLLLLLEEEELLLLLEEEeLLL  LLLLL  HH
Fold AA: GNFGPLPLNFNAFKEWASEYGVFVKTNGSQTIAIIND...ERISLGQWHTRNRVSKAVLVK
Fold SS: LLEELLLLHHHHHHHHHLLLEEEEEELLLLLLEEEELLLL  EEEELLLLHHHHLLLEHHHHHHH
```

```
ProbeAA: TFKRILRDAGLTEEEFHNL...
ProbeSS: HHHHHHHHhLLLHHHHHLL
Fold AA: MLRKLYEATK.DEEVKRMLHLIE
Fold SS: HHHHHHHHHL LHHHHHHHHHHL
```

The file contains header information at the top, then in the alignment section, the query sequence (ProbeAA), query composite secondary structure (ProbeSS), template sequence (Fold AA) and template secondary structure (Fold SS) are given. Gaps are denoted by periods in the sequences and by spaces in the secondary structures.

## 10.6 tfm—Tertiary Folding Module

The `tfm` program generates one or more complete backbone conformations starting from a primary sequence and secondary structure, along with optional domain boundaries and sheet topologies. In addition, `tfm` can accept several different types of constraint including multiple coordinate templates, multiple types of distance constraints, and dihedral angle constraints. A large number of run-time parameters allow the code to be used in a variety of ways, from minimizing a homology model to full ab initio folding.

### Syntax

The `tfm` program has hundreds of user-specified options contained in over a dozen input files. A detailed description of all the options is beyond the scope of this document. The `refine` command can be used as a wrapper to prepare default input files. If suitable input has been prepared, there is also a top-level wrapper for `tfm` that can be used to run the program directly under Schrodinger Job Control. It can be invoked from the command line via:

```
$SCHRODINGER/tfm input-file [options]
```

where *input-file* is typically the job name, with or without a `.inp` extension. The input file must be either the first or last command-line argument, and everything else on the command line is passed as is to Job Control.

## Command Input File

There are only three keywords recognized in the command input file, listed in [Table 10.16](#).

Table 10.16. Keywords for *tfm* program.

Keyword	Description
INPUT_FILE <i>filename</i>	The input file passed to the executable.
OUTPUT_FILE <i>filename</i>	The output produced by the executable.
STRUCT_NAME <i>name</i>	Optional. Prefix for files in the local directory that will be automatically treated as input, such as template or constraint files.

## Files

In addition to the descriptive output in the file specified by OUTPUT\_FILE and the error messages in *name.err*, *tfm* produces the following files, where the first part of the file name is *tfm* by default but can be changed by other input options, and *N* is a serial number.

*tfm.N.pdb*      Output structures, if applicable.  
*tfm.N.spair*    Possible strand pair assignments for unpaired strands, if automatic strand-pairing is being used.

## Return Value and Errors

*tfm* returns a value of 0 on successful completion, 1 otherwise. All error messages are printed to the file *name.err*, where *name* is either the STRUCT\_NAME value or the SEQNAME parameter in the file specified by INPUT\_FILE. This file is empty otherwise, and all output from the program itself is written to the file specified by OUTPUT\_FILE. Output from Job Control goes to stdout, and any errors arising before the driver script has started will go to stdout or stderr.

## Examples

Sample input for *tfm* can be generated by running *refine* with the *-LOCAL* and *-DEBUG* options. The command

```
refine sample
```

creates a directory *run\_sample* with the input file *sample.1.in* (among other files determined by the refinement type). An input file *sample\_tfm.inp* containing the following lines can be used to run the job from the command line

```
INPUT_FILE sample.1.in
OUTPUT_FILE sample.1.out
```

with the command

```
tfm sample_tfm
```

### Environment Variables

In addition to the standard environment variables `tfm` recognizes the following environment variables:

`TFM_QUIET`      These environment variables can be set to decrease or increase the job-related output in the log file. The value to which they are set is irrelevant.

`TFM_VERBOSE`





# Command-Line Utilities

In addition to the command-line programs described in [Chapter 10](#), Prime provides a number of utilities, which are described in this chapter. These utilities are installed in the `utilities` directory in your Schrödinger software installation.

To run these utilities:

- On Linux and Mac, open a terminal window, set the `SCHRODINGER` environment variable, and prefix the command with `$SCHRODINGER/utilities/`.
- On Windows, open a command prompt window with Start → All Programs → Schrodinger-2015 → Schrodinger Command Prompt, and enter the command without any prefix.

## 11.1 Multiple Template Alignment: `structalign`

If multiple templates are selected in the Find Homologs step, they must be rotated into the same coordinate reference frame before they can be used. The `Align Structures` button is the usual means of performing structural alignment of templates, but the `structalign` utility can be run from the command line, as explained below.

You can specify the input as multiple files in PDB format, or as a Maestro file. The templates should consist of single chains. To use the output files, each must be imported into Find Homologs and selected as one of the templates, replacing the original, unaligned version.

In addition to its use in Prime Structure Prediction, this utility can also be used more generally to structurally align proteins. The syntax is:

```
structalign [options] reference-file input-files
```

The reference file and the input files can be in PDB or Maestro format. The output file name is generated by adding the prefix `rot-` to the input file name. For a description of the options, run the command with the `-h` option.

The following instructions are for use with the Prime Structure Prediction workflow.

### To obtain PDB structures of selected templates:

1. Click the sequence name of a template in the Find Homologs step, to display it in the Workspace.

2. Choose Project > Export Structures in the main window.
3. Select Workspace for the source of structures.
4. Choose PDB for the file format.
5. Name the file with the homolog ID.
6. Repeat the above steps for each template of interest.

### To export templates to a Maestro file:

1. Click the sequence name of a template of interest in the Find Homologs step, to display it in the Workspace.
2. Choose Workspace > Create Project Entry.
3. Save the entry with the ID as the title.
4. Repeat the above steps for each template of interest.
5. Select the templates in the Project Table.
6. Choose Table > Export Structures.
7. Select Selected entries for the source of structures.
8. Export to a single Maestro file.

At the command line, enter the appropriate form of the command:

```
structalign template1.pdb template2.pdb ...  
structalign templates.mae
```

This command runs the `structalign` utility and returns one aligned file for each template:

```
rot-template1.pdb  
rot-template2.pdb  
...
```

For a Maestro file, the files are named `rot-templates-N.pdb`, where *N* is the entry number in the file.

**Note:** If you encounter a problem involving templates with multiple chains, use the `getpdb` utility (see [Section 2.3](#) of the *General Utilities* manual) to extract each chain into a separate `.pdb` file, then run `structalign` again.

You can then import the templates into the Find Homologs step. Once the files are imported and the aligned templates appear in the Homologs table, select them all using shift-click or control-click. (Do not select the original structures.)

## 11.2 Format Conversion: seqconvert

This utility converts sequences between any permitted input file format and output file format. This may be useful if autoconversion does not proceed as expected.

### Syntax:

```
seqconvert [options] input-format input-filename output-format output-filename
```

One use of the `seqconvert` utility is to convert SSPs produced in Prime to FASTA format with the following command:

```
seqconvert -inative infile.seq -ofasta outfile.fasta
```

For a description of the options, run the command with the `-h` option.

## 11.3 Structure Alignment: align\_binding\_sites

This script performs a pairwise superposition of multiple structures using the C-alpha atoms of selected residues. The reference structure for the alignment is the first structure in the input file. The C-alpha atoms used for the alignment come from residues within the specified cutoff distance from the ligand in the reference structure. Alternatively, a list of residues may be given for the alignment, which must correspond to residues in the reference structure. C-alpha atoms in the structures that are aligned that are greater than the specified distance from any C-alpha in the reference are not considered in the alignment. The RMSD matrix for the alignment is written to the file `jobname-matrix.csv` in three columns: the indexes for the structures, then the RMSD value. The syntax is:

```
align_binding_sites [options] input-files
```

If you specify more than one input file, the reference structure is the first structure in the first file. For a description of the options, run the command with the `-h` option.

## 11.4 Fast Loop Building: buildloop

Build a loop quickly using the knowledge-based loop builder. The input can be given either in an input file or as a set of options, with the following syntax:

```
buildloop filename.inp  
buildloop input-file res1 res2 [options]
```

The input file contains a list of keyword-value pairs that correspond to the options on the command line. The keywords and the corresponding options are described in [Table 11.1](#).

Table 11.1. Options, arguments, and keywords for the `buildloop` command.

Option/Argument	Keyword	Description
<code>-v[ersion]</code>		Show the program version and exit.
<code>-h[elp]</code>		Show the help message and exit.
<code>-jobname <i>jobname</i></code>		Job name. The default is the stem of the input command file or the input structure file.
<i>input-file</i>	STRUCT_FILE	File containing the structure for which the loop is to be built, in Maestro format.
<i>res1</i>	STEM_RES1	Residue ID for the first residue in the loop, given as <i>chain:residue</i> .
<i>res2</i>	STEM_RES2	Residue ID for the last residue in the loop, given as <i>chain:residue</i> .
<code>-seq <i>sequence</i></code>	QUERY_SEQ	Sequence of the loop to build.
<code>-nmodel <i>int</i></code>	NUM_OUTPUT_STRUCT	Number of requested models. Default: 1.
<code>-minimize <i>yes no</i></code>	MINIMIZE	Minimize the loop after buildup. Default: <i>yes</i> .
<code>-loop_resid <i>list</i></code>	RESID_LIST	Comma-separated list of residue IDs for the new loop, given as <i>chain:residue</i> . Example: 'A:62,A:62A,A:62B,A:63,A:63A,A:64'

Below is a sample input file:

```
STRUCT_FILE input.mae
STEM_RES1 A:61
STEM_RES2 A:65
QUERY_SEQ DDTLYN
NUM_OUTPUT_STRUCT 1
MINIMIZE yes
RESID_LIST 'A:62,A:62A,A:62B,A:63,A:63A,A:64'
```

## 11.5 Database Update Utilities

Prime uses two databases, the PDB database and the BLAST database. As these databases are continually being updated, the versions distributed with the software will not have the latest contributions. Utilities have been provided for updating these databases from the web sites. You can run these utilities on Windows from a Schrödinger shell—see [Section 4.2](#) of the *Installation Guide* for details. You must have permission to write to the destination in order to run these utilities.

### 11.5.1 update\_BLASTDB

This utility updates the BLAST database from a copy on the Schrödinger web site that is synchronized nightly with the BLAST web site, or via FTP from the NCBI web site. The preformatted database files are downloaded. The syntax is:

```
$SCHRODINGER/utilities/update_BLASTDB [options]
```

For a description of the options, run the command with the `-h` option.

The BLAST database is installed in the first location found in the following list:

- The location specified by the `-f` option.
- `$PSP_BLASTDB`
- `$SCHRODINGER_THIRDPARTY/database/blast`
- `$SCHRODINGER/thirdparty/database/blast`

The environment variables are described in [Appendix B](#).

### 11.5.2 rsync\_pdb

This utility creates or updates a local mirror of the PDB. The syntax is:

```
$SCHRODINGER/utilities/rsync_pdb [-p port|-f] [-h]
```

For a description of the options, run the command with the `-h` option.

The PDB is installed in the first location found in the following list:

- `$SCHRODINGER_PDB`
- `$SCHRODINGER_THIRDPARTY/database/pdb`
- `$SCHRODINGER/thirdparty/database/pdb`

These environment variables are described in [Appendix B](#).

This utility uses an rsync server on the Schrödinger web site that enables the files to be downloaded. The files are written in the new remediated format. The Schrödinger site is updated nightly from the PDB. On Windows, the utility `wget` is used to update the local copy.



---

# Third-Party Programs

## A.1 Location of Third-Party Programs

As well as internal and bundled software and data, Prime uses certain external third-party programs and databases for sequence and homolog searches, protein family searches, and secondary structure prediction. These are described under the headings in this topic.

### A.1.1 BLAST/PSI-BLAST, Pfam, and the PDB

See the home pages for these programs and servers for more information:

- BLAST/PSI-BLAST

<http://www.ncbi.nlm.nih.gov/blast>

The BLAST version distributed with Prime is version 2.2.16. To download the BLAST database, you should use the `update_BLASTDB` utility. This utility uses `rsync` to download or update the database, in the correct format for use with Prime.

- HMMER/Pfam

<http://pfam.sanger.ac.uk>

This site has links to mirror sites that may be more convenient for downloads. The Pfam database is included in the Schrödinger software distribution, so you should not need to download it.

- PDB

<http://www.rcsb.org/pdb/>

Note that the PDB format has changed, as of August 1, 2007. You should use the `rsync_pdb` utility to update your local copy of the PDB for use with Prime. Prime uses the new format, and automatically converts files in the old format.

You can download and install the programs and databases in full, so that you have a local copy. You can also run searches of the PDB and BLAST databases on the web, which does not require a separate installation.

### **A.1.2 Secondary Structure Prediction**

Two distinct third-party secondary structure prediction programs can be used to increase the accuracy of, and get better results from, Prime. One of these programs (SSpro) is already bundled with Prime. The other (PSIPRED) is highly recommended, and may be manually installed by you, at your election, and subject to applicable license terms, if any. At the third-party web site whose address is given below, you may download the program and then install it. For instructions on how to install third-party programs, see [Section 3.6](#) of the *Installation Guide* (Linux) or [Section 4.2](#) of the *Installation Guide* (Windows), or the [Third Party Programs page](#) of our web site.

The PSIPRED home page is located at:

<http://bioinf.cs.ucl.ac.uk/psipred/>

and contains information about the program, referrals to terms of use and/or license terms (in the README file), and a link to download the program. The link to download PSIPRED is contained in the section of the page labeled “Overview of prediction methods.” PSIPRED 2.61 is required for use with Prime 4.0.

**Note:** PSIPRED is not available on Windows.

## **A.2 Third-Party Program Use Agreement**

By using the links above or the third-party programs, you acknowledge and agree to the following:

PSIPRED is a third party program (“Third Party Program”) and may therefore be subject to third party license agreements and fees. We (i.e., Schrödinger, LLC and its affiliates) have no responsibility or liability, directly or indirectly, for Third Party Programs or for any third party Web sites (“Linked Sites”) or for any damage or loss alleged to be caused by or in connection with use of or reliance thereon. Any warranties that we make regarding our own products and services do not apply to the Third Party Programs or Linked Sites, or to the interaction between, or interoperability of, our products and services and the Third Party Programs. Referrals and links to Third Party Programs and Linked Sites do not constitute an endorsement of such Third Party Programs or Linked Sites.



---

# Environment Variables

The only environment variable that *must* be set before you can run Prime is SCHRODINGER. The following list is presented for your convenience in case you prefer a different location for any Prime-related directories.

## General

SCHRODINGER\_THIRDPARTY: Directory containing third-party software and databases, which can be installed with the installers supplied by Schrödinger. The databases are installed in the database subdirectory; the software in the bin subdirectory. Each third-party product has its own subdirectory under these subdirectories. The default location if this environment variable is not set is `$$SCHRODINGER/thirdparty`.

## PDB

SCHRODINGER\_PDB: Full path to a user-supplied copy of the PDB database. The default for this directory is `$$SCHRODINGER/thirdparty/database/pdb` if only SCHRODINGER is set; if SCHRODINGER\_THIRDPARTY is set, the default is `$$SCHRODINGER_THIRDPARTY/database/pdb`. The copy of the database must retain the directory hierarchy used by the PDB and must include the following subdirectories to be usable with Prime:

```
data/structures/all/pdb
data/structures/divided/pdb
data/structures/obsolete/pdb
```

On Windows the `data/structures/all` subdirectory is absent (and is not required).

## Blast

PSP\_BLAST\_DIR: Location of Blast executables. Default is `$$SCHRODINGER_THIRDPARTY/bin/platform/blast/bin`, with the default for SCHRODINGER\_THIRDPARTY noted above.

PSP\_BLAST\_DATA: Location of Blast matrices. Default is `$$SCHRODINGER_THIRDPARTY/bin/platform/blast/data`, with the default for SCHRODINGER\_THIRDPARTY noted above.

PSP\_BLASTDB: Location of Blast databases. The BLAST databases distributed with Prime are formatted with version 2.2.16. If your BLAST database does not conform to this format, you might experience problems with Prime. Default is `$$SCHRODINGER_THIRDPARTY/database/blast`, with the default for SCHRODINGER\_THIRDPARTY noted above.

## **HMMER/Pfam**

PSP\_HMMER\_DIR: Location of HMMER distribution (the executables are assumed to be in the binaries subdirectory). Default is \$SCHRODINGER\_THIRDPARTY/bin/platform/hmmer, with the default for SCHRODINGER\_THIRDPARTY noted above.

PSP\_HMMERDB: Location of Pfam database. Default is \$SCHRODINGER\_THIRDPARTY/database/pfam, with the default for SCHRODINGER\_THIRDPARTY noted above.

## **PSIPRED**

PSP\_PSIPRED\_DIR: Location of PSIPRED installation (top level directory containing the bin and data directories). Default is \$SCHRODINGER\_THIRDPARTY/bin/platform/psipred, with the default for SCHRODINGER\_THIRDPARTY noted above.

PSP\_PSIPRED\_DB: Sequence database to use for assembling the PSI-BLAST profile. Allowed values are nr and pdb. Default: pdb.

## **SSPRO**

PSP\_SSPRO\_DB: Sequence database to use for assembling the SSPRO profile. Allowed values are nr and pdb. Default: pdb.

---

# File Formats

## C.1 Input File Formats

The input file formats are:

- native (Schrödinger) format
- FASTA format
- PDB format

Sequences are read from the PDB SEQRES records, if they exist. Otherwise they are read from the ATOM records. If SEQRES records do not exist and residues are missing from the ATOM records, the query sequence will of course contain gaps.

- NBRF-PIR format
- TEXT-PLAIN format
- GCG format
- SWISS-PROT format
- GENBANK format
- NCBI format
- PHYLIP format
- NEXUS-PAUP format
- SELEX format
- STADEN format
- PFAM-STOCKHOLM format
- CASP format

## C.2 Output File Formats

The output file formats are:

- native (Schrödinger) format
- FASTA format
- NBRF-PIR format
- TEXT-PLAIN format
- GCG format
- SWISS-PROT format
- CASP format
- GENBANK format



---

# Error Messages and Warnings

## D.1 Input Sequence

### Invalid File

This message appears if the system fails to extract a valid sequence from the file.

- Check that the sequence is in one of the supported sequence file formats.
- If FASTA format is used, make sure there are no extra spaces between the “>” and the sequence name, and that there are no extra spaces at the end of the line.

### No Sequences Found

This message appears if the system fails to extract any sequences from the workspace.

- Check the Project Table to ensure the correct entries are included in the Workspace.
- Use Open Project in the Maestro Window to open a project.

### Invalid Sequence Characters

This message appears if the sequence contains non-standard one-letter code for amino acids.

- Check that sequence uses the one-letter code and is all uppercase.
- Change/edit any non-standard characters from your sequence using an editor, and then restart job.

## D.2 Find Homologs

### D.2.1 Error Messages

#### Multiple Templates Not Aligned

- If you select multiple templates and proceed to Edit Alignment, you are reminded that the templates must be structurally aligned to one another for composite models.

If one of the following messages appears when you click **Next** to go to the Edit Alignment step, you will not be able to proceed along the Comparative Modeling path until the problem is resolved:

### No Homologs Found

- You may change your search options and continue searching, or proceed to Fold Recognition.

### Invalid File

- Check to ensure the homolog file exists in either of the following locations:  
\$SCHRODINGER/thirdparty/database/pdb/data/structures/divided/pdb/\*  
\$SCHRODINGER\_THIRDPARTY/database/pdb/data/structures/divided/pdb/\*  
and is in the right format (gzip compressed, pdb\*.ent.gz).

## D.2.2 Warnings in Homologs Table

These messages appear in the Warning column of the Homologs table for sequences that are likely to be unusable for model building:

### Many missing residues

- Too many consecutive missing residues to guarantee model building will succeed.

### !UNUSABLE: Alternate residue type

- The residue type was undetermined, causing the file to include the residue twice but with different identity (e.g. ALA20 and VAL20). This is not supported by Prime.

### !UNUSABLE: Bad SEQRES

- General problems with the SEQRES.

### !UNUSABLE: CA only

- File contains only alpha C coordinates.

### !UNUSABLE: Molecule *molname* should be labeled as HETATM

- A non-protein molecule has been listed as ATOM when it should be listed as HETATM. ATOM records are reserved for proteins only.

### !UNUSABLE: SEQRES/ATOM mismatch

- The sequence listed in the SEQRES records does not match the actual sequence in the ATOM records.

### !UNUSABLE: SEQRES is missing residues

- There are residues in the ATOM records that are not listed in the SEQRES records. PDB files require that the ATOM records be a subset of the SEQRES records.

### **!UNUSABLE: SEQRES numbering is wrong**

- The number of residues in a particular chain does not agree with the number of residues shown in columns 14 to 17 of SEQRES.

## **D.3 Edit Alignment**

### **Check Alignment**

When you click **Next** at the end of the Edit Alignment step, the Check Alignment warning may appear. The warning lists possible alignment problems that you may want to fix before continuing to the next step. For example, there may be gaps in secondary structure elements of the template, or gaps in the query that are aligned to secondary structure elements of the template.

If you do not want to make corrections to the alignment, click **Continue** to go to the next step. To make corrections based on the information in the error message, make a note of the gap locations given, click **Cancel** to close the message box while remaining in Edit Alignment, and change the alignment accordingly.

## **D.4 Build Structure**

Error messages and warnings for this generally appear in the log of a running job. The log can be viewed in the job Monitor panel until the job is completed.

## **D.5 Refinement**

Error messages for Refinement generally appear in the log of a running job. The following is an example of a warning in the log of a loop refinement.

### **Invalid Features**

When a loop refinement begins, a validation program checks the loop features of the structure. Apparent errors are reported in a warning dialog box. You may choose to **Run All Features**, **Run Only Valid Features**, or **Cancel**. It is recommended that you make a note of the invalid features, click **Cancel**, and correct the structure if appropriate.





---

# 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](mailto: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
- Prime purchaser (company, research institution, or individual)
- Primary Prime 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.

---

# Glossary

**alignment**—The optimal matching of residue positions between sequences, typically a query sequence and one or more template sequences.

**all-atom structure**—A structure in which all atoms, including hydrogen atoms, are explicitly represented. This means that the structure must have coordinates for all atoms. PDB files often do not have coordinates for hydrogen atoms, and some force fields can merge the hydrogens into the heavy atoms in a “united atom” model.

**anchor**—A constraint on alignment set at a given residue position. Alignment changes must preserve the query-template pairing at that residue until the anchor is removed.

**ASL**—Atom Specification Language.

**cofactor**—The complement of an enzyme reaction, usually a metal ion or metal complex.

**Comparative Modeling**—Protein structure modeling based on a query-template match with a substantial percentage of identical residues (usually 50% or greater sequence identity).

**constraints**—Tools to keep regions of a sequence (alignment constraints) or structure (during minimization) in a particular configuration.

**deletions**—The residues missing from a query sequence that are present in a template sequence.

**Fold Recognition**—The use of secondary structure matching and profiles generated from multiple sequence/structure alignments to find templates when sequence methods are unsuccessful.

**gaps**—The spaces in an alignment resulting from insertions and deletions.

**HETATOMs**—The atoms of residues, including amino acids, that are not one of the standard 20 amino acids. In PDB files, HETATM.

**homolog**—A sequence/structure related to the query sequence; i.e., a sequence with many of the same residues in the same patterns as the query sequence. Usually these sequences are derived from the same family and may have similar function.

**insertions**—The extra residues found in a query sequence that are not found in a template sequence.

**ligand**—A small molecule that binds to a protein; e.g., an antigen, hormone, neurotransmitter, substrate, inhibitor, and so on.

**loop**—A region of undefined secondary structure.

**project**—A collection of related data, such as structures with their associated properties. In Prime a project comprises one or more *runs* (executions of the Prime workflow). The project may include data that does not appear in the *project table*.

**Project Table**—The Maestro panel associated with a project, featuring a table with rows of entries and columns of properties. In Prime, project table entries are usually model structures.

**query sequence**—A sequence of unknown structure or fold.

**refinement**—An improvement of a model structure through energy-based optimization of selected regions.

**run**—A single execution of the Prime workflow using a particular set of choices (of templates, of paths, and of settings). Each run belongs to a *project*. Runs cannot be saved without saving the project to which they belong.

**SSA**—Secondary structure assignment.

**SSP**—Secondary structure prediction.

**template sequence**—A sequence of known structure and fold used as a basis for building a model of the query.

**Workspace**—The open area in the center of the Maestro window in which structures are displayed.

**Z-Score**—Measures the compatibility of the query sequence with the model structure, relative to the compatibility of randomly shuffled sequences of the same composition.

## Symbols

- (hyphen)
  - for locked gap ..... 17, 39
  - for loop ..... 16, 36
- ~ (tilde), for unlocked gap ..... 17, 39

## A

- Add Anchors ..... 17, 39
- alignment ..... 147
  - editing ..... 17, 39
  - exporting ..... 34
  - GPCRs ..... 37
  - importing ..... 34
  - methods ..... 35
  - with ClustalW ..... 35
  - with Prime STA ..... 35
- alignment score ..... 37
  - interpretation of values ..... 81
- alpha carbons
  - aligning structures by ..... 129
  - constraining specific ..... 106
  - maximum movement ..... 65, 106, 107
- anchors ..... 147
  - alignment constraints ..... 17, 39
  - removing ..... 17, 39
  - setting ..... 17, 39
- ASL ..... 147
- atom names, correcting ..... 52
- ATOM records ..... 137

## B

- backbone sampling, in side-chain prediction... 57, 108
- Balloon Help ..... 11
- binding sites, aligning ..... 83
- bit score ..... 28
- BLAST
  - database format ..... 135
  - expectation value ..... 28
  - percentage gaps ..... 28
  - updating database ..... 131
  - web-based searching ..... 26
- BLAST/PSI-BLAST ..... 133
- blocks, sliding ..... 17, 39
- BLOSUM62 similarity matrix ..... 28
- bond orders, assigning ..... 51

- bonds
  - disulfide ..... 44
  - to metals ..... 50
- Build Structure step ..... 41

## C

- Calculate Energy panel ..... 62
- chain lengths ..... 24
- chain name ..... 28
- chains, multiple ..... 23, 114
- charged residues ..... 39
- code, sequence ..... 23
- cofactors ..... 44, 147
- Color by Sequence ..... 14
- Color Scheme ..... 13, 17
  - Prime Display Menu ..... 14
  - Prime toolbar ..... 14
- color schemes ..... 14–15
  - built model ..... 46
- columns, sorting ..... 29
- Comparative Modeling ..... 147
- composite SSP ..... 38
- Compound (PDB COMPND) ..... 28
- consensus sequence ..... 27
- conserved residues ..... 27
- constraints ..... 147
  - alpha carbon ..... 75, 106
  - atom pair ..... 44, 73, 103
  - Cartesian ..... 73
  - helical ..... 74
  - loop placement in membrane ..... 75
  - membrane ..... 106
  - proximity ..... 44
  - solvent exposure ..... 106
  - spatial ..... 75
- continuum solvation model ..... 47, 104
- conventions, document ..... ix
- cooperative refinement of loops ..... 55
- covalently bound ligands ..... 49
- crystal symmetry ..... 63
- C-terminal gaps ..... 17, 39

## D

- deleting runs ..... 13
- deleting SSP ..... 19, 36
- deletions ..... 49, 147

- dielectric, variable..... 104
- directory  
   installation ..... 4  
   Maestro working..... 4
- Display menu (Prime)..... 13, 16
- disulfide bond..... 44
- domains ..... 25
- E**
- E (for beta strand) ..... 16, 36
- Edit Alignment step ..... 33
- Edit menu (Prime)..... 16
- Edit Query Sequence ..... 16
- Edit SSP ..... 16, 17
- editing alignment ..... 39
- editing sequences ..... 13
- EMBL ..... 23
- entries, Project Table..... 21
- environment variable  
   SCHRODINGER ..... 4  
   SCHRODINGER\_PDB..... 5, 135  
   SCHRODINGER\_THIRDPARTY..... 135
- environment variables ..... 135
- error messages..... 25, 139, 141
- Expect column ..... 28
- expectation value..... 28
- Export Sequences..... 13, 19
- Export SSP..... 19, 35
- extracting sequence names..... 24
- F**
- FASTA  
   file format ..... 23, 35  
   file header ..... 24
- file formats ..... 23
- File menu ..... 10  
   Job Options..... 20
- Find Homologs step..... 23, 25
- Fold Recognition..... 147
- Font Size, in Sequence Viewer ..... 15
- force field  
   Build Structure ..... 47  
   nonstandard residues ..... 50  
   refinement tasks..... 53  
   single-point energy ..... 62
- formal charges, correcting ..... 52
- G**
- gaps ..... 147  
   C-terminal..... 17, 39  
   locking..... 17, 39  
   N-terminal ..... 17, 39  
   percentage..... 28  
   in query ..... 28  
   in secondary structure..... 39, 141  
   in template ..... 28  
   unlocking ..... 17, 39
- GENBANK ..... 23
- getpdb utility ..... 30
- GPCRs  
   aligning..... 37  
   SSPs for ..... 35
- Guide..... 10, 11, 16
- H**
- H (for alpha helix) ..... 16, 36
- header  
   FASTA file..... 24  
   PDB ..... 24
- helices  
   building..... 74, 106  
   placing membrane on ..... 71
- HETATMs ..... 147
- HETATOMs ..... 147
- heteromultimer  
   saving runs for ..... 32  
   template alignment ..... 44
- hidden sequences ..... 27
- HMMER/Pfam..... 26, 33, 133
- homologs..... 147  
   importing ..... 25  
   searching for ..... 25
- Homologs table ..... 28
- host machine ..... 20
- Hybrid Monte Carlo - Options dialog box..... 67
- Hybrid Monte Carlo panel ..... 60
- hydrogens, adding..... 50
- hydrophobic residues ..... 39
- hyphen symbol  
   for locked gap..... 17, 39  
   for loop ..... 16, 36



**I**

ID, PDB chain .....	28
Identities column.....	28
Import SSP.....	35
importing structurally aligned templates .....	25
Include ligands and cofactors window .....	44
Input Sequence step .....	23
insertions .....	39, 49, 147

**J**

Job Monitor panel .....	21
job options.....	20
Job Options dialog box .....	13, 20
job status icon .....	11, 21
jobs	
killing.....	21
monitoring .....	21

**K**

killing jobs .....	21
--------------------	----

**L**

libraries	
backbone dihedral.....	47
side-chain rotamer .....	47
ligands.....	44, 148
covalently bound.....	49, 50
lipophilic term.....	80
locking gaps .....	17, 39
loop refinement .....	49
alpha-carbon constraints.....	75
constraining to helix .....	74
cooperative.....	55
defining region.....	54
first residue .....	54
invalid features check .....	55
last residue.....	54
minimum interatomic distance .....	66
options .....	64
order of .....	55
restriction with respect to membrane .	75, 106
solvent exposure constraints.....	106
time required.....	54
loops.....	148

**M**

maximum sequence length .....	25
membrane, implicit .....	70
adding to other structures .....	72
adjusting .....	71
constraining loops with respect to.....	75
excluding solvent pockets.....	72
from OPM database.....	71
thickness .....	72
menu bar (Prime) .....	10, 12
menus	
Display (Prime) .....	13, 16
Edit (Prime) .....	16
File (Prime) .....	10, 20
merging projects .....	9
metals, bonds to .....	50
Minimize - Options dialog box.....	66
Minimize as Rigid Body panel .....	59
Minimize panel .....	58
missing residues.....	137
model-building method.....	46
Monitor panel.....	11
monitoring jobs.....	21
mouse functions .....	20
multimer	
input for .....	24
multiple chains	
with same sequence .....	23
with SKA .....	114
multiple sequences.....	24
multiple templates.....	42

**N**

names	
chains.....	28
extracting sequence .....	24
non-unique.....	24
sequences.....	24, 28
native Schrödinger format.....	35
New command .....	12
None color scheme .....	14
N-terminal gaps .....	17, 39

**O**

OPLS_2005 force field.....	47, 50, 62
optimizing side chains .....	45

- options  
  job ..... 13, 20  
  in panel layout ..... 10
- P**
- panel layout ..... 10
- panels  
  Job Options ..... 20  
  Monitor ..... 11
- PDB ..... 133  
  ATOM records ..... 137  
  chain ID ..... 28  
  header ..... 24  
  ID code ..... 28  
  SEQRES ..... 137  
  updating database ..... 131
- PDB atom names, correcting for receptor-ligand complex ..... 52
- Pfam ..... 26, 33, 133
- PIR ..... 23
- polar residues ..... 39
- position-specific substitutional matrix (PSSM) ..... 38
- Positives column ..... 28
- Predict Side Chains - Options dialog box ..... 63
- Predict Side Chains panel ..... 56
- Preferences ..... 16  
  command ..... 15
- Prime Energy Visualizer panel ..... 93
- Prime Guide ..... 10, 16
- Prime menu bar ..... 10, 12
- Prime MM-GBSA panel ..... 88
- Prime runs ..... 9
- Prime Sequence Viewer ..... 10
- Prime toolbar ..... 10, 16
- Prime workflow ..... 11
- Prime-Refinement ..... 3
- Prime-Structure Prediction ..... 1, 10, 23
- processors, maximum number ..... 65
- product installation ..... 144
- programs, third-party ..... 133
- Project Table ..... 21, 148
- projects ..... 148
- Protein Structure Alignment panel ..... 82
- Protein Structure Alignment Results panel ..... 83
- proximity ..... 15
- Proximity color scheme ..... 15
- proximity cutoff ..... 15, 16
- PSIPRED ..... 35, 134
- Q**
- query sequence ..... 148  
  editing ..... 24
- query, maximum length ..... 25
- R**
- random seed, for side-chain prediction ..... 63
- Ranking Score ..... 148
- read sequence ..... 23
- records  
  ATOM ..... 137  
  SEQRES ..... 137
- Refine Loops - Options dialog box ..... 64
- Refine Loops panel ..... 54
- Refine Protein-Ligand Complex - Options dialog box ..... 69
- Refine Protein-Ligand Complex panel ..... 61
- Refine Structure ..... 21
- Refinement ..... 3
- refinement ..... 148  
  cooperative ..... 55
- regions of multiple templates ..... 42
- Res1 ..... 54
- Res2 ..... 54
- Residue Matching color scheme ..... 14
- residue names, correcting for receptor-ligand complex ..... 52
- Residue Property color scheme ..... 14
- Residue Type color scheme ..... 14
- residues  
  charged ..... 39  
  conserved ..... 27  
  hydrophobic ..... 39  
  missing ..... 137  
  nonstandard ..... 50  
  polar ..... 39  
  specification syntax ..... 97  
  standard, list of ..... 50
- rotamers ..... 45
- ruler ..... 18
- runs ..... 9, 148  
  deleting ..... 13  
  for heteromultimer ..... 32  
  new ..... 12

- opening ..... 13
  - renaming ..... 13
  - saving ..... 13, 32
- S**
- sampling accuracy, loop refinement ..... 64
  - saving runs ..... 13
  - Schrödinger contact information ..... 144
  - schrodinger .hosts file ..... 20
  - score
    - alignment ..... 37
    - BLAST bit ..... 28
  - scratch projects ..... 9
  - Search program
    - in Find Homologs ..... 25
  - secondary structure assignment ..... 17, 38
  - Secondary Structure color scheme ..... 14
  - secondary structure prediction programs ..... 134
    - See also* SSPs
  - Select toolbar command ..... 16
  - seqconvert utility ..... 35, 129
  - SEQRES records ..... 137
  - sequence identity ..... 28
  - sequence positives ..... 28
  - Sequence Viewer (Prime) ..... 10
  - sequences
    - coloring ..... 17
    - command menu for ..... 17
    - cropping ..... 25
    - editing ..... 13
    - editing query ..... 24
    - exporting ..... 13
    - family consensus ..... 27
    - maximum length ..... 25
    - multiple ..... 24
    - names ..... 24, 28
    - non-unique ..... 24
    - Pfam ..... 27
    - selecting ..... 24
    - showing hidden ..... 27
    - standard code ..... 23
    - unusable ..... 29
  - side chains
    - backbone sampling when predicting .. 57, 108
    - cooperative prediction ..... 103
    - including in loop refinement ..... 65
    - number of conformations returned ..... 103
    - optimizing ..... 45
    - rotamers ..... 45
  - similarity, BLOSUM62 matrix ..... 28
  - Slide As Block ..... 17, 39
  - Slide Freely ..... 17, 39
  - solvation ..... 47
  - solvent pockets
    - excluding from membrane ..... 72
  - sorting tables, by column heading ..... 29
  - SSA ..... 148
  - SSpro ..... 35, 134
  - SSPs ..... 148
    - command menu for ..... 20
    - composite ..... 38
    - deleting ..... 19, 36
    - editing ..... 16
    - exporting ..... 19, 35
    - importing ..... 35
    - retrieving unmodified ..... 37
    - in Sequence Viewer ..... 17
  - standard residues ..... 50
  - steps
    - action ..... 10
    - available and unavailable ..... 12
    - Build Structure ..... 41
    - changing data ..... 10, 12
    - Edit Alignment ..... 33
    - Find Homologs ..... 23, 25
    - Input Sequence ..... 23
      - name ..... 10
      - panel layout ..... 10
  - structalign utility ..... 127
  - structurally aligned templates ..... 31, 44
  - structure refinement, loop ..... 49
  - structures, coloring ..... 17
  - Surface Generalized Born (SGB) ..... 47
- T**
- temperature, setting for Hybrid Monte Carlo
    - refinement ..... 67
  - Template ID color scheme ..... 15
  - template sequence ..... 148
  - templates
    - homomultimer ..... 43
    - multiple ..... 42, 43
    - regions ..... 42
    - structural alignment ..... 31, 44

third-party programs ..... 133, 134  
tilde symbol, for unlocked gap..... 17, 39  
Title, PDB header..... 28  
toolbar, Prime..... 10, 16  
tooltips ..... 11, 29  
    for buttons..... 16  
    for table entries ..... 11

**U**

unlocking gaps ..... 17, 39  
unusable sequence..... 29  
utilities  
    getpdb..... 30  
    seqconvert ..... 35, 129  
    structalign..... 127

**V**

View SSA..... 17  
View Structure ..... 15, 17

**W**

warnings..... 25, 29, 139–141  
web-based BLAST searches ..... 26  
workflow ..... 11, 23  
Workspace..... 13, 17, 148

**Z**

Z-Score ..... 148



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

155 Gibbs St  
Suite 430  
Rockville, MD 20850-0353

Quatro House  
Frimley Road  
Camberley GU16 7ER  
United Kingdom

101 SW Main Street  
Suite 1300  
Portland, OR 97204

Dynamostraße 13  
D-68165 Mannheim  
Germany

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

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

Zeppelinstraße 73  
D-81669 München  
Germany

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

8910 University Center Lane  
Suite 270  
San Diego, CA 92122

Potsdamer Platz 11  
D-10785 Berlin  
Germany

**SCHRÖDINGER**