Ligand and Structure-Based Descriptors

Schrödinger Suite 2012 Update 2

Schrödinger Press

Revision A, September 2012

Document Conventions

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

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

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

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

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

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

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

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

Ligand and Structure-Based Descriptors

The Ligand and Structure-Based Descriptors panel in Maestro provides a convenient interface to several Schrödinger programs, which are used to generate descriptors for a QSAR model. The focus is on generating descriptors for a set of ligands that are docked to a receptor. Three of the programs, Liaison, Prime MM-GBSA, and the Embrace module of MacroModel, operate on the ligand and the receptor. The other two, QikProp and Ligparse, operate on the ligand only. The descriptors extracted from Liaison, Prime MM-GBSA, and Embrace are energetic properties related to ligand binding. QikProp generates ADME properties, and Ligparse generates structure-based properties such as functional group counts. These descriptors can be used as input to the model-generation facility in Strike.

To run any of the descriptor generation tasks, you must have an installed and licensed version of the appropriate software: Liaison 5.8, Prime 3.1, MacroModel 9.9 (for Embrace), and QikProp 3.5.

The results are collected as a set of descriptors in a comma-separated value (`.csv`) file, and are added to the structures in the output file, *jobname*`-out.maegz` file, which are copies of the input structures. If you chose entries from the Project Table for input, the descriptors are imported into the project when the calculations finish. You can then select the relevant entries in the Project Table, and use Strike to build a QSAR model.

In addition, the output structures from Prime MM-GBSA, Liaison, and Embrace are written to output files *jobname*`-`*product*`-out.maegz`

Note: The receptor and the ligands must be properly prepared beforehand. See the *LigPrep User Manual* for information on ligand preparation, and the *Protein Preparation Guide* for information on protein preparation.

1 Running Schrödinger Software

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

Linux:

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

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

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

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

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

You can start the Maestro interface with the following command:

```
$SCHRODINGER/maestro &
```

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

Windows:

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

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

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

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

Mac:

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

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

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

2 The Ligand and Structure-Based Descriptors Panel

The Ligand and Structure-Based Descriptors panel in Maestro is used to generate ligand and structure-based descriptors for a structure-based QSAR model of ligand binding to a receptor. The panel is divided into three sections, which are described below.

To open the Ligand and Structure-Based Descriptors panel, choose Ligand and Structure-Based Descriptors from the Workflows menu in the main window.

2.1 Selecting the Ligands and the Receptor

To set up a calculation, you must select a receptor and a source of ligands. The receptor and ligands can come from a pose viewer file, or from separate files, or from the current project. The selections are made in the Source of Ligands and Receptor section.

Maestro Pose Viewer file options

If you want to use a pose viewer file for both the receptor and the ligands, select this option. To specify the file, enter the file name and path in the text box, or click Browse and navigate to the file. This option is useful if you have run Glide and want to obtain descriptors for a selected set of poses.

Separate options

If you choose to obtain the receptor and the ligands from separate sources, you can read them from files, or obtain them from the project.

To use the selected entries in the Project Table as the source of ligands, select Selected entries in Project Table, and select the ligands. When the job starts, the ligands are written to a Maestro

file. To read the ligands from a file, select File and enter the file name and path in the text box, or click Browse and navigate to the file.

If the receptor is in the current project, you can display it in the Workspace and select Workspace entry. The receptor structure is written to a file when the job is launched. Otherwise you can read the receptor from a file. To do so, select File and enter the file name and path in the text box, or click Browse and navigate to a file.

2.2 Selecting the Descriptors To Generate

After choosing a receptor and ligands, you can choose the descriptors you want to generate. For each of the five descriptor sets, there is an option in the Descriptors to generate section of the panel. You can select more than one option; the jobs will be run simultaneously.

The programs that generate the descriptors are run with default options that were chosen to produce reasonable descriptors. To change any of these options or change some of the job options, click Advanced Options, and make your choices in the Ligand and Structure-Based Descriptors - Advanced Options dialog box. The settings in this dialog box are a limited set of the full range of settings, designed to provide useful descriptors for a QSAR model. For more information on the advanced options, see page 12. The descriptor types are described below.

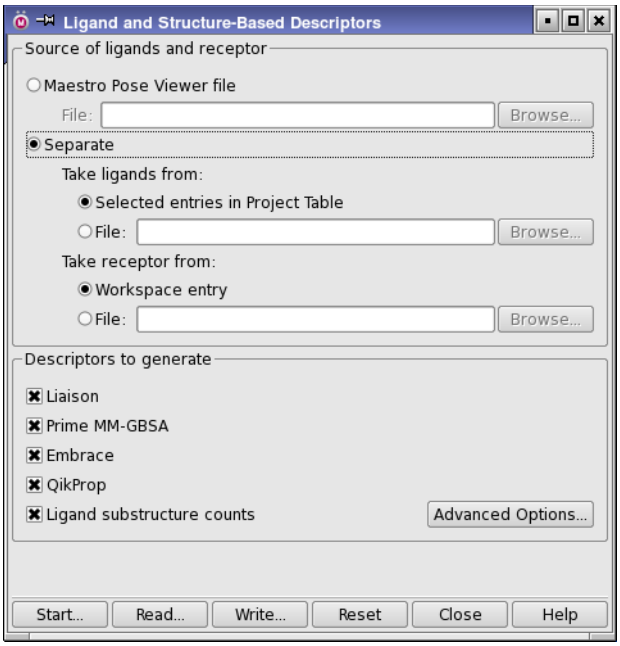

Figure 1. The Ligand and Structure-Based Descriptors panel.

2.2.1 Liaison Descriptors

Liaison calculates ligand-receptor binding affinities using a linear interaction approximation. It runs molecular mechanics (MM) simulations of the ligand-receptor complex and of the free ligand and free receptor, with an SGB continuum solvation model. Three simulation methods are available: energy minimization (the default), molecular dynamics simulation, or hybrid Monte Carlo simulation. The latter explicitly include temperature effects. For more information on the Liaison model and methods, see Chapter 1 of the *Liaison User Manual*.

The Liaison descriptors include the five terms that contribute to the linear interaction model energy.

2.2.2 Prime MM-GBSA Descriptors

The Prime MM-GBSA approach is used to predict the free energy of binding for a receptor and a set of ligands. MM-GBSA is an acronym for a method that combines OPLS molecular mechanics energies (E_{MM}), an SGB solvation model for polar solvation (G_{SGB}), and a nonpolar solvation term (G_{NP}) composed of the nonpolar solvent accessible surface area and van der Waals interactions. The total free energy of binding is then expressed as:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

where

$$G = E_{MM} + G_{SGB} + G_{NP}$$

The ligand in the unbound state is minimized in SGB solvent but is not otherwise sampled. In the calculation of the complex, the ligand is minimized in the context of the receptor. The protein is currently held fixed in all calculations. The following descriptors generated by the Prime MM-GBSA approach:

MM-GBSA_DG_bind	Ligand binding energy, ΔG_{bind}
MM-GBSA_E_complex	Energy of the complex, $G_{complex}$
MM-GBSA_E_protein	Energy of the receptor without the ligand, $G_{protein}$
MM-GBSA_E_ligand	Energy of the unbound ligand, G_{ligand}

2.2.3 Embrace Descriptors

Embrace calculates ligand-receptor binding energies by molecular mechanics energy minimization of the complex and the separated receptor and ligand, with or without continuum solvation. The Embrace calculation is run in energy difference mode. The following descriptors are generated from the calculation:

Embrace_Total_Energy_without_constraints	Ligand binding energy
Embrace_Valence_Energy	Valence energy difference
Embrace_vdW_Energy	van der Waals energy difference
Embrace_Electrostatic_Energy	Coulomb energy difference
Embrace_Solvation_Energy	Solvation energy difference
Embrace_Constraint_Energy	Constraint energy difference

For more information on Embrace, see Chapter 14 of the *MacroModel User Manual*.

2.2.4 QikProp Descriptors

QikProp produces a list of 44 descriptors related to absorption, distribution, metabolism and excretion. These descriptors include properties like skin permeability and octanol/water partition coefficients, and counts of important functional groups. For a complete list of descriptors, see Chapter 1 of the *QikProp User Manual*

2.2.5 Ligand Substructure Counts

Counts of various substructures are generated from `ligparse`, and include counts of a wide range of functional groups (defined by single SMARTS pattern), counts of composite groups (defined by multiple SMARTS patterns), and some other counts. Over 100 functional groups are identified by `ligparse`. The following composite group counts are reported:

Num acceptor groups	Num acidic hydrogens
Num amide hydrogens	Num charged acceptor groups
Num charged donor groups	Num divalent oxygen atoms
Num donor groups	Num neutral acceptor groups
Num neutral amines	Num neutral donor groups
Num reactive groups	

and the remaining counts are of the following:

Num rings	Num heteroaromatic rings
Num aromatic rings	Num aliphatic rings
Num rotatable bonds	Num atoms
Molecular weight	Num chiral centers

2.3 Running the Jobs

When you have made all your selections, click Start. The Start dialog box opens, in which you can set a job name and select a host for running the jobs.

The choice you make from the Host option menu is the host on which the computationally intensive jobs will be run. The master job and the ligand substructure counts job always run on the local host. The other jobs are run concurrently on the chosen host if the host allows it, and jobs for individual programs can be distributed over multiple processors if the program allows it.

If you choose a remote host or batch queue as the host for the job, you must ensure that the current working directory is mounted on the remote host. This requirement is equivalent to running the computational jobs from the Linux command line with the `-LOCAL` option. All driver jobs are run locally.

When you click Start in the Start dialog box, the input files are written, directories are set up for each kind of job, and the jobs are started. The calculations are run under Job Control, and can be monitored in the Monitor panel.

If you want to postpone running the jobs, save settings for use with other systems, or edit the input files to change options that are not available from the panel, you can click Write to write the input files without running the jobs. The main input file is written to *jobname*`.inp`, and is used to set up input files for the various programs. Job settings made in the Advanced Options panel override those made in the main panel, and are stored in the input file.

To retrieve the settings from an input file, either to modify them or to run a job, click Read. A dialog box opens, in which you can navigate to and choose an input file, which has a file extension of `.inp`.

If you edit the input files, you must then run the jobs for each program from the Linux command line with the following command:

`$SCHRODINGER/lsbd` [*options*] *input-file* [*output-dir*]

The standard Job Control options, described in Section 2.3 of the *Job Control Guide*, are supported. `$SCHRODINGER/lsbd -h` displays a usage message that describes other options.

If you want to clear all custom settings and return to the default settings, click Reset. The calculation settings for all programs are returned to their defaults, which are as follows:

Liaison: Truncated Newton minimization with OPLS_2005 force field, 1000 steps, 15 Å residue-based cutoff, medium constraints option.

Embrace: OPLS_2005 force field, no solvent, constant dielectric (1.0), normal cutoffs, PRCG minimization method with 5000 iterations, tiny constraints option.

QikProp: Normal mode.

3 Advanced Options for Descriptor Generation

You can set nondefault options for the various programs that generate descriptors in the Ligand and Structure-Based Descriptors - Advanced Options panel. To open this panel, click Advanced Options. There are no options for ligand substructure counts.

3.1 Liaison Options

For Liaison, the options related to the calculation itself are the same as in the Parameters tab of the Liaison panel. For details, see Section 4.2.2 of the *Liaison User Manual*. These options provide most of the flexibility that is available for Liaison calculations.

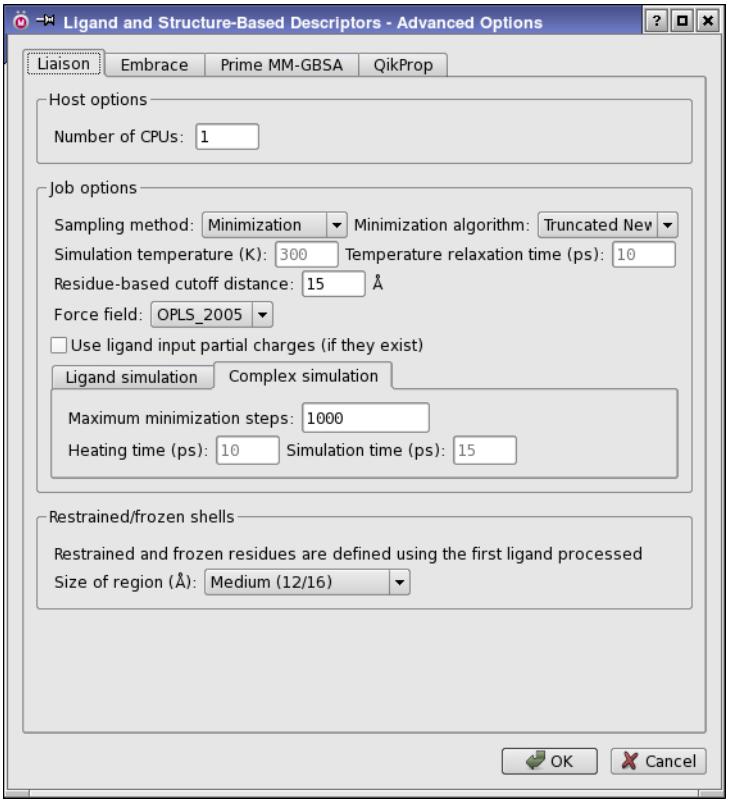

Figure 2. The Liaison tab of the Advanced Options panel.

The Host options section provides a choice of the number of processors over which to distribute the Liaison calculation.

3.2 Embrace Options

The Host options section allows you to select the number of CPUs to use for the job and the number of subjobs to split the calculation into.

The options for the potential and minimization are those from the Potential tab and Mini tab of the MacroModel panels. For more information on these options, see Section 4.2 and Section 6.2 of the *MacroModel User Manual*.

Constraints can be set by choosing from a menu that supplies a set of predetermined distances for restrained, frozen, and ignored shells. A Maestro file named *jobname*`-emb_cons.mae` is written and used to specify constraints relative to a single structure, so the calculations will be repeatable. From this file, the substructure `.sbc` file is written for each complex.

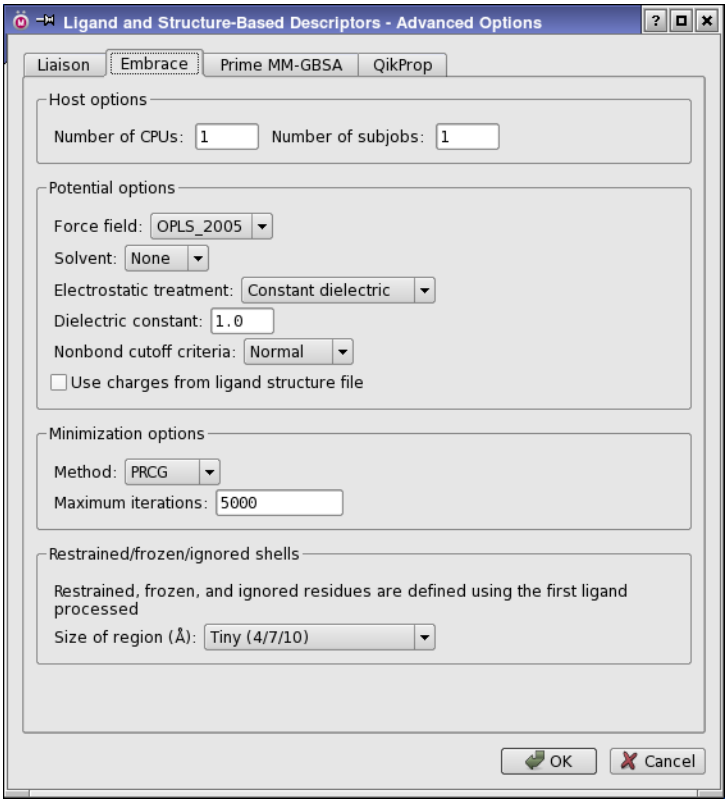

Figure 3. The Embrace tab of the Advanced Options panel.

3.3 Prime MM-GBSA Options

The Host options section allows you to select the number of CPUs to use for the job and the number of subjobs to split the calculation into.

The Settings section provides two options, for calculating ligand strain energies, and for using partial charges read in as properties with the input ligand (rather than those calculated by the force field). For the strain energies, calculations are done on the free ligand and on the ligand in the geometry it adopts in the complex, both with implicit solvent. The strain energy is the difference between the two energies.

The Protein Flexibility section contains an option menu from which you can choose various ranges relative to the first input ligand for which protein atoms are allowed to move.

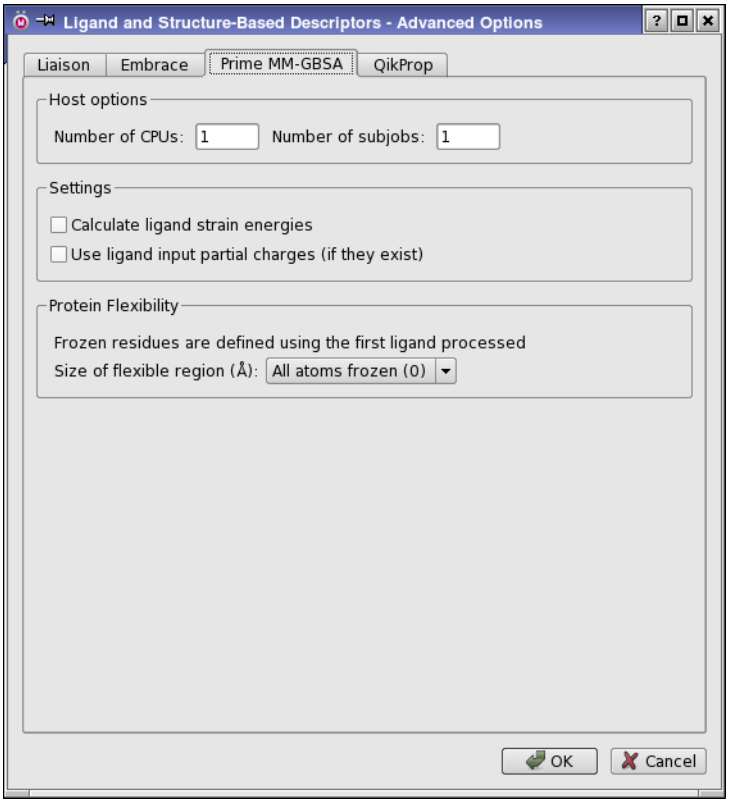

Figure 4. The Prime MM-GBSA tab of the Advanced Options panel.

3.4 QikProp Options

For QikProp, the only program setting is whether or not to use fast mode. For a description of fast mode, see Chapter 1 of the *QikProp User Manual*.

4 Citing Ligand and Structure-Based Descriptors in Publications

Schrödinger Suite 2012 Ligand and Structure-Based Descriptors protocol; Liaison version 5.8, Schrödinger, LLC, New York, NY, 2012; MacroModel version 9.9, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012; QikProp version 3.5, Schrödinger, LLC, New York, NY, 2012.

Getting Help

Information about Schrödinger software is available in two main places:

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

Finding Information in Maestro

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

To get information:

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

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

For information on:

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

Contacting Technical Support

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

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

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

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

Gathering Information for Technical Support

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

For general enquiries or problems:

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

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

If your job failed:

1. Open the Monitor panel in Maestro.

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

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

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

If Maestro failed:

1. Open the Diagnostics panel.

   - **Windows:** Start → All Programs → Schrodinger-2012 → Diagnostics
   - **Mac:** Applications → Schrodinger2012 → Diagnostics
   - **Linux/command line:** `$SCHRODINGER/diagnostics`

2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.

3. Attach the file specified in the dialog box to your e-mail message.

4. Attach the file `maestro_error.txt` to your e-mail message.

   This file should be in the following location:

   - **Windows:** `%LOCALAPPDATA%\Schrodinger\appcrash`
     (Choose Start → Run and paste this location into the Open text box.)
   - **Mac:** `Documents/Schrodinger`
   - **Linux:** Maestro's working directory specified in the dialog box (the location is given in the terminal window).

5. On Windows, also attach the file `maestro.EXE.dmp`, which is in the same location as `maestro_error.txt`.

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

101 SW Main Street
Suite 1300
Portland, OR 97204

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

8910 University Center Lane
Suite 270
San Diego, CA 92122

155 Gibbs St
Suite 430
Rockville, MD 20850-0353

Dynamostraße 13
D-68165 Mannheim
Germany

Zeppelinstraße 73
D-81669 München
Germany

Potsdamer Platz 11
D-10785 Berlin
Germany

Quatro House
Frimley Road
Camberley GU16 7ER
United Kingdom

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

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

SCHRÖDINGER®