QikProp 3.5

User Manual

Schrödinger Press

QikProp User Manual Copyright © 2012 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.

BioLuminate, 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 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, see the Legal Notices, or use your browser to open $SCHRODINGER/docs/html/third_party_legal.html (Linux and Mac OS) or %SCHRODINGER%\docs\html\third_party_legal.html (Windows OS).

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.

Revision A, September 2012

Contents

Document Conventions v

Chapter 1: Introduction 1

1.1 QikProp Overview 1

1.2 Citing QikProp in Publications 6

Chapter 2: QikProp Tutorial 7

2.1 Preparing for the Exercises 7

2.2 Importing Structures 8

2.3 Running QikProp in Normal Processing Mode 9

2.4 Running QikProp in Fast Processing Mode 11

2.5 Sorting Project Table Entries 11

Chapter 3: Running QikProp 13

3.1 QikProp Input 13

3.2 Running QikProp From Maestro 14

3.2.1 Setting QikProp Options 14
3.2.2 Setting Job Options 15
3.2.3 Fixing Noncompliant Structures 16

3.3 Running QikProp From the Command Line 16

3.4 Monitoring QikProp Jobs 18

3.5 QikProp Output Files 18

3.6 Customizing QikProp with the QPlimits File 20

3.7 QikProp Methodology 22

3.8 Use of Similarity 22

3.8.1 Improvement of logS and logP 22
3.8.2 Retrieval of the Most Similar Molecules 23

Chapter 4: Quality of QikProp Results 25

4.1 Test Set Results 25

4.2 Accuracy Issues 29

4.3 Conformation Dependence of Results 30

4.4 Data Sources 32

Appendix A: Descriptor Information 33

References 35

Getting Help 37

Index 41

Document Conventions

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

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

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

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

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

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

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

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

Introduction

1.1 QikProp Overview

QikProp is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program designed by Professor William L. Jorgensen. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches.

In addition to predicting molecular properties, QikProp provides ranges for comparing a particular molecule's properties with those of 95% of known drugs. QikProp also flags 30 types of reactive functional groups that may cause false positives in high-throughput screening (HTS) assays. The range of values that cause a molecule to be flagged as dissimilar to other known drugs can be modified in the `QPlimits` file—see Section 3.6 on page 20 for details.

QikProp can be run either from the Maestro GUI or from the command line. QikProp has two modes: normal mode, and fast mode. In fast mode, certain time-consuming calculations are omitted, some properties are not predicted, and some have different values. For more information, see Chapter 3.

QikProp properties and descriptors can be used as input to Strike, which is a collection of chemically-aware statistical tools for examining correlations within data. It can develop and employ QSAR/QSPR models using partial least squares, principal component analysis, and multiple linear regression; generate univariate and bivariate statistics; and perform similarity/diversity analysis in descriptor and 2D-structure space. Strike can be run from Maestro. For further information about Strike, see the *Strike User Manual* or contact your Schrödinger sales representative.

For each successfully processed molecule, QikProp produces the following descriptors and properties. Those that are not predicted in fast mode are marked with a dagger (†). Those whose values differ between fast and normal mode are marked with a double dagger (‡).

Table 1.1. QikProp properties and descriptors.

Property or Descriptor	Description	Range[a] or recommended values
molecule name	Molecule name taken from the title line in the input structure file. If the title line is blank, the input file name is used.	
#stars	Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. Outlying descriptors and predicted properties are denoted with asterisks (*) in the `.out` file. A large number of stars suggests that a molecule is less drug-like than molecules with few stars. The following properties and descriptors are included in the determination of #stars: MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, accptHB, glob, QPpolrz, QPlogPC16, QPlogPoct, QPlogPw, QPlogPo/w, logS, QPLogKhsa, QPlogBB, #metabol	0 – 5
#amine	Number of non-conjugated amine groups.	0 – 1
#amidine	Number of amidine and guanidine groups.	0
#acid	Number of carboxylic acid groups.	0 – 1
#amide	Number of non-conjugated amide groups.	0 – 1
#rotor	Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds.	0 – 15
#rtvFG	Number of reactive functional groups; the specific groups are listed in the *jobname* `.out` file. The presence of these groups can lead to false positives in HTS assays and to decomposition, reactivity, or toxicity problems *in vivo*. See Appendix A for a complete list.	0 – 2
CNS	Predicted central nervous system activity on a –2 (inactive) to +2 (active) scale.	–2 (inactive), +2 (active)
mol_MW	Molecular weight of the molecule.	130.0 – 725.0
dipole†	Computed dipole moment of the molecule.	1.0 – 12.5
SASA	Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius.	300.0 – 1000.0
FOSA	Hydrophobic component of the SASA (saturated carbon and attached hydrogen).	0.0 – 750.0
FISA	Hydrophilic component of the SASA (SASA on N, O, and H on heteroatoms).	7.0 – 330.0
PISA	π (carbon and attached hydrogen) component of the SASA.	0.0 – 450.0

Table 1.1. QikProp properties and descriptors. (Continued)

Property or Descriptor	Description	Range[a] or recommended values
WPSA	Weakly polar component of the SASA (halogens, P, and S).	0.0 – 175.0
volume	Total solvent-accessible volume in cubic angstroms using a probe with a 1.4 Å radius.	500.0 – 2000.0
donorHB	Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.	0.0 – 6.0
accptHB	Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.	2.0 – 20.0
dip^2/V†	Square of the dipole moment divided by the molecular volume. This is the key term in the Kirkwood-Onsager equation for the free energy of solvation of a dipole with volume V.	0.0 – 0.13
ACxDN^.5/SA	Index of cohesive interaction in solids. This term represents the relationship $(accptHB(\sqrt{donorHB}))/(SA)$; see Ref. [2].	0.0 – 0.05
glob	Globularity descriptor, $(4\pi r^2)/(SASA)$, where r is the radius of a sphere with a volume equal to the molecular volume. Globularity is 1.0 for a spherical molecule.	0.75 – 0.95
QPpolrz	Predicted polarizability in cubic angstroms.	13.0 – 70.0
QPlogPC16	Predicted hexadecane/gas partition coefficient.	4.0 – 18.0
QPlogPoct‡	Predicted octanol/gas partition coefficient.	8.0 – 35.0
QPlogPw	Predicted water/gas partition coefficient.	4.0 – 45.0
QPlogPo/w	Predicted octanol/water partition coefficient.	–2.0 – 6.5
QPlogS	Predicted aqueous solubility, log S. S in mol dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.	–6.5 – 0.5
CIQPlogS	Conformation-independent predicted aqueous solubility, log S. S in mol dm^{-3} is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.	–6.5 – 0.5
QPlogHERG	Predicted IC_{50} value for blockage of HERG K^+ channels.	concern below –5

Table 1.1. QikProp properties and descriptors. (Continued)

Property or Descriptor	Description	Range[a] or recommended values
QPPCaco	Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells are a model for the gut-blood barrier. QikProp predictions are for non-active transport.	<25 poor, >500 great
QPlogBB	Predicted brain/blood partition coefficient. Note: QikProp predictions are for orally delivered drugs so, for example, dopamine and serotonin are CNS negative because they are too polar to cross the blood-brain barrier	–3.0 – 1.2
QPPMDCK	Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood-brain barrier. QikProp predictions are for non-active transport.	<25 poor, >500 great
QPlogKp	Predicted skin permeability, log K_p.	–8.0 – –1.0
IP(ev)†	PM3 calculated ionization potential.	7.9 – 10.5
EA(eV)†	PM3 calculated electron affinity.	–0.9 – 1.7
#metab‡	Number of likely metabolic reactions. See Appendix A for a complete list of reactions.	1 – 8
QPlogKhsa	Prediction of binding to human serum albumin.	–1.5 – 1.5
HumanOralAbsorption	Predicted qualitative human oral absorption: 1, 2, or 3 for low, medium, or high. The text version is reported in the output. The assessment uses a knowledge-based set of rules, including checking for suitable values of PercentHumanOralAbsorption, number of metabolites, number of rotatable bonds, logP, solubility and cell permeability.	
PercentHuman-OralAbsorption	Predicted human oral absorption on 0 to 100% scale. The prediction is based on a quantitative multiple linear regression model. This property usually correlates well with HumanOral-Absorption, as both measure the same property.	>80% is high <25% is poor
SAFluorine	Solvent-accessible surface area of fluorine atoms.	0.0 – 100.0
SAamideO	Solvent-accessible surface area of amide oxygen atoms.	0.0 – 35.0
PSA	Van der Waals surface area of polar nitrogen and oxygen atoms.	7.0 – 200.0
#NandO	Number of nitrogen and oxygen atoms.	2 – 15

Table 1.1. QikProp properties and descriptors. (Continued)

Property or Descriptor	Description	Range[a] or recommended values
RuleOfFive	Number of violations of Lipinski's rule of five [3]. The rules are: mol_MW < 500, QPlogPo/w < 5, donorHB ≤ 5, accptHB ≤ 10. Compounds that satisfy these rules are considered drug-like. (The "five" refers to the limits, which are multiples of 5.)	maximum is 4
RuleOfThree	Number of violations of Jorgensen's rule of three. The three rules are: QPlogS > -5.7, QP PCaco > 22 nm/s, # Primary Metabolites < 7. Compounds with fewer (and preferably no) violations of these rules are more likely to be orally available.	maximum is 3
#ringatoms	Number of atoms in a ring	
#in34	Number of atoms in 3- or 4-membered rings	
#in56	Number of atoms in 5- or 6-membered rings	
#noncon	number of ring atoms not able to form conjugated aromatic systems (e.g. sp^3 C).	
#nonHatm	Number of heavy atoms (nonhydrogen atoms)	
Jm	Predicted maximum transdermal transport rate, $K_p \times MW \times S$ ($\mu g\ cm^{-2}\ hr^{-1}$). K_p and S are obtained from the aqueous solubility and skin permeability, QPlogKp and QPlogS. This property is only written to the output file: it is not used in any other calculations.	

[a] For 95% of known drugs.

Schrödinger applications can be started 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 &
```

Windows:

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:

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

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

1.2 Citing QikProp in Publications

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

QikProp, version 3.5, Schrödinger, LLC, New York, NY, 2012.

QikProp Tutorial

This chapter is designed to assist you to rapidly become familiar with the capabilities of QikProp 3.5. Once you have worked through these exercises, you will have an understanding of the basic QikProp features.

This chapter assumes that you have already installed the Schrödinger software, or that you have access to an installation of the Schrödinger software. If you do not have access to the software, see the *Installation Guide* for installation instructions.

2.1 Preparing for the Exercises

To run the exercises, you need a working directory in which to store the input and output, and you need to copy the input files from the installation into your working directory. This is done automatically in the Tutorials panel, as described below. To copy the input files manually, just unzip the `qikprop` zip file from the `tutorials` directory of your installation into your working directory.

On Linux, you should first set the `SCHRODINGER` environment variable to the Schrödinger software installation directory, if it is not already set:

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

If Maestro is not running, start it as follows:

- **Linux:** Enter the following command:

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

- **Windows:** Double-click the Maestro icon on the desktop.

  You can also use Start → All Programs → Schrodinger-2012 → Maestro.

- **Mac:** Click the Maestro icon on the dock.

  If it is not on the dock, drag it there from the SchrodingerSuite2012 folder in your Applications folder, or start Maestro from that folder.

Now that Maestro is running, you can start the setup.

1. Choose Help → Tutorials.

   The Tutorials panel opens.

2. Ensure that the Show tutorials by option menu is set to Product, and the option menu below is labeled Product and set to All.
3. Select QikProp Tutorial in the table.
4. Enter the directory that you want to use for the tutorial in the Copy to text box, or click Browse and navigate to the directory.

   If the directory does not exist, it will be created for you, on confirmation. The default is your current working directory.

5. Click Copy.

   The tutorial files are copied to the specified directory, and a progress dialog box is displayed briefly.

If you used the default directory, the files are now in your current working directory, and you can skip the next two steps. Otherwise, you should set the working directory to the place that your tutorial files were copied to.

6. Choose Project → Change Directory.
7. Navigate to the directory you specified for the tutorial files, and click OK.

You can close the Tutorials panel now, and proceed with the exercises.

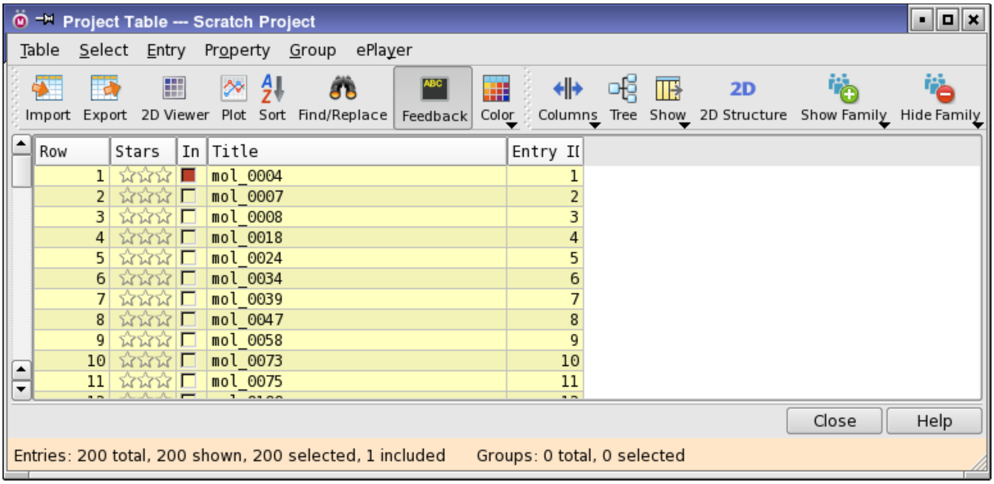
2.2 Importing Structures

When running QikProp from Maestro, you can use as input for a job either the structures in the Workspace or entries that are selected in the Project Table panel. In this exercise, you will import structures into Maestro to use as QikProp input.

1. Click the Import button on the Project toolbar.

   The Import panel is displayed.

2. If the import options are not displayed, click Options.
3. From the Create groups for option menu, choose None.


Project Table — Scratch Project
Table Select Entry Property Group ePlayer
Import Export 2D Viewer Plot Sort Find/Replace Feedback Color Columns Tree Show 2D Structure Show Family Hide Family
Close Help

Row	Stars	In	Title	Entry ID
1	☆☆☆	■	mol_0004	1
2	☆☆☆	☐	mol_0007	2
3	☆☆☆	☐	mol_0008	3
4	☆☆☆	☐	mol_0018	4
5	☆☆☆	☐	mol_0024	5
6	☆☆☆	☐	mol_0034	6
7	☆☆☆	☐	mol_0039	7
8	☆☆☆	☐	mol_0047	8
9	☆☆☆	☐	mol_0058	9
10	☆☆☆	☐	mol_0073	10
11	☆☆☆	☐	mol_0075	11

Entries: 200 total, 200 shown, 200 selected, 1 included   Groups: 0 total, 0 selected

Figure 2.1. The Project Table panel with all entries selected.

4. Select `ligs.mae`, and click Open.

   The structures are imported as individual entries, and the first structure is displayed in the Workspace.

5. Click the Table button on the Project toolbar.

   The Project Table panel opens. The contents of the `ligs.mae` file are listed, and all 200 entries are selected.

2.3 Running QikProp in Normal Processing Mode

In this exercise, you will run QikProp in normal processing mode on the imported structures.

1. Choose Applications > QikProp.

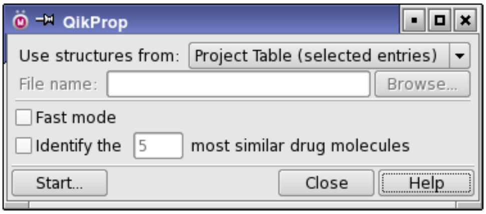
   The QikProp panel opens.

2. Choose Project Table (selected entries) from the Use structures from option menu.

3. Click Start.

   The Start dialog box is displayed. In this dialog box you can make settings for the output incorporation and for running the job.

4. From the Incorporate option menu, choose Replace existing entries.



QikProp

Use structures from: Project Table (selected entries)

File name:

☐ Fast mode

☐ Identify the 5 most similar drug molecules

Figure 2.2. The QikProp panel.

5. In the Name text box, type `ligands`.
6. Click Start.

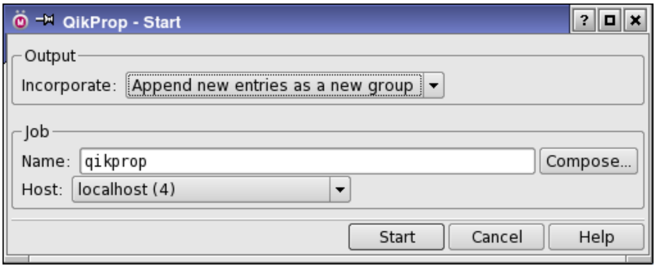
   The job begins running. When the job finishes, the following files should be in your working directory:

   `ligands-out.mae`
   `ligands.CSV`
   `ligands.qpsa`
   `ligands.out`
   `ligands.mae`
   `ligands.log`

   The file `ligands-out.mae` is a Maestro-formatted file that contains property data and structural information. The `ligands.mae` file is written by QikProp at the start of the job and contains only structural information for the selected entries. The `ligands.out` file lists properties predicted and the limits specified in `QPlimits`. It also includes a section that shows how some of the properties were formed by linear combinations of more basic molecular properties. The `ligands.qpsa` file provides a detailed surface area analysis, and `ligands.CSV` contains all the predicted properties for each structure in comma-separated value (CSV) format.
7. Choose Project → Save As.

   The Save Project dialog box opens.
8. Name the project `ligs.prj` and click Save.

   The project is saved as `ligs.prj`.



QikProp - Start

Output

Incorporate: Append new entries as a new group

Job

Name: qikprop

Host: localhost (4)

Figure 2.3. The Start dialog box.

2.4 Running QikProp in Fast Processing Mode

In this exercise, you will run QikProp in fast mode. You will use the same set of structures you processed in normal mode in the previous exercise so that you can compare the output files produced by each mode.

1. Ensure that all entries in the Project Table panel are selected.

   This should be the case if you are proceeding from the last exercise.

2. If the QikProp panel is not open, open it from the Applications menu in the main window.
3. Ensure that Project Table (selected entries) is chosen from the Use structures from option menu.
4. Select Fast Mode.
5. Click Start.
6. From the Incorporate option menu, choose Append new entries individually.
7. In the Name text box, type `QPfast`.
8. Click Start.

   The job runs in a few seconds. When it finishes, the Project Table has 400 entries.

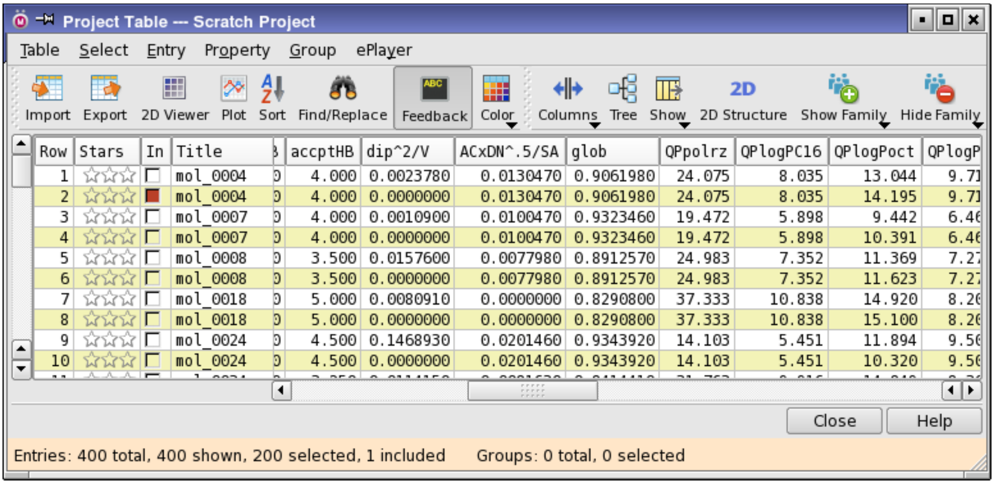
2.5 Sorting Project Table Entries

In this exercise, you will sort the entries in the Project Table to compare the results from fast mode to the results from normal mode.

1. Right-click on the Title heading in the Project Table and choose Sort → All Rows.

   The row for each fast mode entry is now directly underneath the row for the corresponding normal mode entry.

2. In the Project Table panel, choose Fit To Header+Data from the Columns button menu on the toolbar.
3. Scroll through the Project Table and compare the fast mode results with the normal mode results.


Project Table — Scratch Project
Table
Select
Entry
Property
Group
ePlayer
Import
Export
2D Viewer
Plot
Sort
Find/Replace
Feedback
Color
Columns
Tree
Show
2D Structure
Show Family
Hide Family
Close
Help

Row	Stars	In	Title	accptHB	dip^2/V	ACxDN^.5/SA	glob	QPpolrz	QPlogPC16	QPlogPoct	QPlogP
1	☆☆☆	☐	mol_0004	4.000	0.0023780	0.0130470	0.9061980	24.075	8.035	13.044	9.71
2	☆☆☆	🟥	mol_0004	4.000	0.0000000	0.0130470	0.9061980	24.075	8.035	14.195	9.71
3	☆☆☆	☐	mol_0007	4.000	0.0010900	0.0100470	0.9323460	19.472	5.898	9.442	6.4
4	☆☆☆	☐	mol_0007	4.000	0.0000000	0.0100470	0.9323460	19.472	5.898	10.391	6.4
5	☆☆☆	☐	mol_0008	3.500	0.0157600	0.0077980	0.8912570	24.983	7.352	11.369	7.2
6	☆☆☆	☐	mol_0008	3.500	0.0000000	0.0077980	0.8912570	24.983	7.352	11.623	7.2
7	☆☆☆	☐	mol_0018	5.000	0.0080910	0.0000000	0.8290800	37.333	10.838	14.920	8.2
8	☆☆☆	☐	mol_0018	5.000	0.0000000	0.0000000	0.8290800	37.333	10.838	15.100	8.2
9	☆☆☆	☐	mol_0024	4.500	0.1468930	0.0201460	0.9343920	14.103	5.451	11.894	9.5
10	☆☆☆	☐	mol_0024	4.500	0.0000000	0.0201460	0.9343920	14.103	5.451	10.320	9.5

Entries: 400 total, 400 shown, 200 selected, 1 included   Groups: 0 total, 0 selected

Figure 2.4. The Project Table with fast and normal mode properties sorted.

Notice that, in fast mode, the dipole moment (“dipole”), dipole^2/volume (“d2ov”), the ionization potential (“ip”), and the electron affinity (“ea”) are not predicted. Instead 0.0 is entered for these values. This is because, in the interest of saving time, the PM3 single point calculation is not performed. Notice also that the prediction for the free energy of solvation in octanol (“log poct”) differs between fast and normal processing modes because, in fast mode, the QSPR equations are refit without the dipole terms. The `#metab` descriptor also differs between the two modes.

Running QikProp

QikProp can be run either from the Maestro GUI or from the command line. Using Maestro to run QikProp enables you to build structures or import structures into Maestro, use them as input for a QikProp job, and then use Maestro's analysis tools and Schrödinger's other computational programs to further investigate lead compounds. You might want to run QikProp from the command line if you are processing a large number of structures, or do not immediately want to use Maestro for analysis of the results.

QikProp has two processing modes. By default, QikProp runs in *normal* mode, predicting 44 properties for as many as 10,000 molecules in an hour. Alternatively, QikProp can run in *fast* mode, which predicts 40 properties rather than the default 44. In fast mode, QikProp is able to evaluate approximately 300,000 compounds per hour on a 2 GHz Pentium processor. Fast mode skips the PM3 calculation, which produces the dipole moment, ionization potential and electron affinity. As a result, the octanol/gas partition coefficient is different. Fast mode also bypasses some of the tests performed in normal mode, such as checking for a single molecule, neutral functional groups and valid atom types. The properties generated in fast mode for molecules with these violations are not valid, though the descriptors are still valid.

3.1 QikProp Input

QikProp input structures and files must meet the following criteria:

- Structures must be three-dimensional, not two-dimensional.
- Hydrogen atoms must be explicit. You cannot, therefore, use structures with united-atom types from molecular mechanics programs. If your input contains such structures, Maestro prompts you to correct the hydrogen treatment. There is no check for missing hydrogens when you run QikProp from the command line.
- With the exception of $$$$-delimited MDL SD mol files and Maestro-formatted files, each structure file must contain only one molecule. You can concatenate MDL SD mol (*jobname*`.mol`) files into a single file for input to QikProp. For an example, see the file `$SCHRODINGER/qikprop-v`*version*`/molfiles/many.mol`.
- The allowed elements are H, C, N, O, F, Al, Si, P, S, Cl, Br, and I. In fast processing mode all atom types are permitted, though some results might not be valid if atoms other than the above PM3 atoms are used.

- The number of atoms must not exceed 1000 atoms in the fast mode or 100 non-hydrogen and 150 hydrogen atoms in normal processing mode.

Charged input structures do not have to be neutralized prior to processing, because QikProp neutralizes them internally, and copies the QikProp properties for the neutral species to the original (charged) input structure for output to file. In normal mode, if QikProp is unable to neutralize a structure, such as in the case of a quarternary ammonium compound, that ligand is skipped and no properties are generated.

If your structures do not meet any of the above criteria, you can use LigPrep to preprocess the structures so that they are ready for input to QikProp. LigPrep converts structures from 2D to 3D, ensures that explicit hydrogen atoms are used, and performs a molecular mechanics minimization. See the *LigPrep User Manual* for more information on using this product.

QikProp results are not very sensitive to the conformation of the input structure. Where there is variation of the results with conformation, extended structures generally give results that are closer to the experimental data than compact structures. QikProp results are also not very sensitive to the details of structure optimization, though some form of optimized structure should be used for the input. See Section 4.3 on page 30 for more information.

QikProp processes some additional entries in standard MDL SD mol files, as follows. If there is an entry called `<MOLNAME>`, QikProp reads the following line for the molecule name that is written to all files. If there is an entry called `<MDLNUMBER>` or `<MOLCODE>`, this entry is written after the molecule name in the *jobname*`.out` file.

3.2 Running QikProp From Maestro

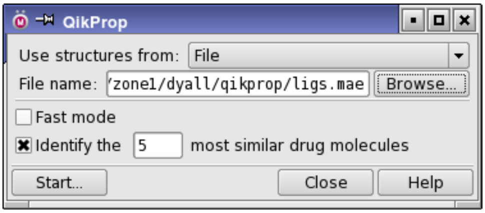
To run QikProp from Maestro, choose Applications → QikProp or Tasks → ADME. The QikProp panel opens.

To launch a job from the QikProp panel, specify an input source and select options. After you make your selections, click the Start button. The Start dialog box is displayed, in which you can specify an incorporating mechanism, a job name, and a host.

3.2.1 Setting QikProp Options

Specifying an input source

You can use Workspace structures, selected entries in the Project Table panel, or structures from a Maestro file as input for a QikProp job. The default setting is Selected Entries. To evaluate structures from a file, choose File from the Use structures from option menu, and enter the file name in the File name text box, or click Browse and navigate to the file in the file selector that is displayed.



Use structures from: File

File name: 'zone1/dyall/qikprop/ligs.mae

☐ Fast mode

☒ Identify the 5 most similar drug molecules

Figure 3.1. The QikProp panel.

Selecting a mode

To run in fast mode, select Fast Mode. To run in normal mode, deselect Fast Mode.

Identifying similar drug molecules

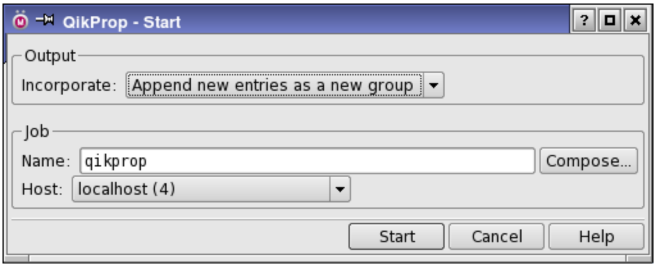
If you want to identify the drug molecules that are most similar to the input molecules, select Identify the *N* most similar drug molecules, and enter the number of similar drug molecules to return for each input molecule in the text box.

3.2.2 Setting Job Options

The Start dialog box provides a standard set of controls for incorporation of job output, naming the job, and selection of the execution host. These controls are described in detail in Section 2.2 of the *Job Control Guide*.

If you selected project entries rather than a file for input, a file named *jobname*`.mae` containing these structures is written to the working directory when you start the job.

Generally the output structure does not differ from the structure used as input. However, if QikProp determines that a structure is missing hydrogens, QikProp modifies the input structure and replaces it with the corrected structure. If you do not want to replace the input structures, you should choose one of the two Append options for incorporation of results, or choose not to incorporate the results.

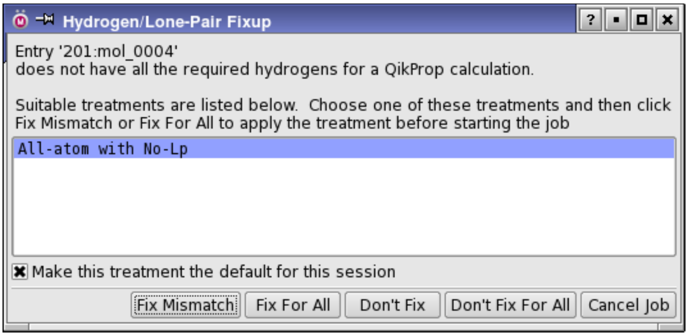


Incorporate: Append new entries as a new group

Name: qikprop

Host: localhost (4)

Figure 3.2. The Start dialog box.



Entry '201:mol_0004'
does not have all the required hydrogens for a QikProp calculation.

Suitable treatments are listed below. Choose one of these treatments and then click Fix Mismatch or Fix For All to apply the treatment before starting the job

All-atom with No-Lp

☒ Make this treatment the default for this session

Figure 3.3. The Hydrogen/Lone-Pair Fixup dialog box.

Before you can run a QikProp job on a remote host, you must configure the remote host and set up the `schrodinger.hosts` file on the local and remote hosts. Information on these tasks can be found in the *Installation Guide* and the *Job Control Guide*. To run QikProp on a remote host, select a host from the Host option menu. All hosts that have been configured in the `schrodinger.hosts` file appear on this menu.

3.2.3 Fixing Noncompliant Structures

If your input structures are missing hydrogen atoms or other structural information needed by QikProp, the Hydrogen/Lone-Pair Fixup dialog box (Figure 3.3) is displayed when you click Start in the Start dialog box. This dialog box presents a list of valid treatments to fix up the structures, and buttons to select the action to take.

3.3 Running QikProp From the Command Line

To run QikProp from the command line on Linux, enter the command:

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

QikProp accepts the standard Job Control options, and the common options `-LOCAL`, `-WAIT`, and `-HELP`. These options are described in Section 2.3 of the *Job Control Guide*. The QikProp command line options are described in Table 3.1.

To run QikProp from the command line on Windows, open a Schrödinger Command Prompt window from the Start menu, enter the command `sh`, and use the command above.

Table 3.1. Options for the qikprop command.

Option	Description
`-fast`	Turn on fast mode.
`-nofast`	Turn off fast mode.
`-neut`	Neutralize molecules in Maestro-formatted files before processing, by adding or removing protons (default).
`-noneut`	Do not neutralize molecules in Maestro files before processing.
`-sim`	Generate a list of known drugs most similar to each processed molecule.
`-nosim`	Do not generate of a list of known drugs most similar to each processed molecule (default).
`-simf`	Generate a file of the *n* most similar drug molecules to the last molecule processed. The number of drug molecules is specified by `-nsim`.
`-nsim` *int*	Number of the most similar drug molecules to report. Default: 5.
`-nosa`	Do not write the *jobname*`.qpsa` file.
`-n` *lo*`:`*hi*	Process the specified range of molecules from the input file. The lower and upper limits on the molecule number, *lo* and *hi*, must be separated by a colon, with no spaces.
`-outcomp`	Compress the output structure file. This option has no effect if the input file is compressed. The extension for compressed output does not have a period, i.e `.maegz` is used for Maestro files and `.sdfgz` is used for MDL files.
`-outname` *name*	Specify the base name for the output files. This name takes the place of *jobname* in the output file names—see Section 3.5 on page 18. The output structure file format is the same as the input file format, including the compression level (unless `-outcomp` is used).
`-logpdb` *filename*	Specify the database (full path) for the use of experimental logP values in prediction of logP. See Section 3.8 on page 22 for more information.
`-logsdb` *filename*	Specify the database (full path) for the use of experimental logS values in prediction of logS. See Section 3.8 on page 22 for more information.
`-simdb` *filename*	Specify the database (full path) for the retrieval of the most similar molecules. See Section 3.8 on page 22 for more information.

QikProp input must be a file containing the 3D structure (x, y, and z coordinates and atomic numbers) of one or more molecules. Five file formats are accepted:

- Maestro files, uncompressed or compressed (`.mae`, `.maegz`, `.mae.gz`)
- MDL mol files, also known as 3D SD files, uncompressed or compressed (`.mol`, `.molgz`, `.sd`, `.sd.gz`, `.sdgz`, `.sdf`, `.sdf.gz`, `.sdfgz`)

- MOL2 files
- PDB files (`.pdb`)
- BOSS/MCPRO Z-matrix files (`.z`)

An example of each file type is provided in the directory `$SCHRODINGER/qikprop-v`*version*`/molfiles`. In addition, this subdirectory contains the mol files for 30 common drugs.

To run a QikProp job on a remote host, use the `-HOST` jobcontrol option.

3.4 Monitoring QikProp Jobs

QikProp jobs, whether they are run from Maestro or from the command line, are managed by the Schrödinger Job Control Facility. All jobs, whether run from the command line or from Maestro, can be monitored from the Monitor panel. The contents of the *jobname*`.out` file are displayed in the File tab as the file is written, and the progress of the job is displayed as a progress bar in the File tab.

For information on the Job Control Facility and how to use this facility to monitor QikProp jobs, see the *Job Control Guide*.

When you start a QikProp job from Maestro, the Monitor panel is displayed. The name of the job is listed in the Job tab of the Monitor panel, and the job is selected in the jobs table.

When there are no jobs currently running, only the Monitor and Delete buttons are active. While a job is running, the Pause, Stop, Kill, and Update buttons are active, but only the Kill button has any effect for QikProp jobs. To terminate a QikProp job, select the desired job in the database list, then click Kill.

3.5 QikProp Output Files

For every completed QikProp job, the following files are created:

- *jobname*`.CSV`
- *jobname*`.qpsa`
- *jobname*`.out`
- *jobname*`.log`

Here, *jobname* is the stem of the input file name. If you started the job from Maestro, *jobname* is the name you entered in the Name text box. In addition to these standard files, a *jobname*`.warning` file is written if QikProp has difficulty processing a structure.

A structure output file that contains the QikProp properties is written if you ran the job from Maestro, or if the input file was in Maestro or MDL mol or SD format. The structure file is named *jobname*`-out.`*ext*, where *ext* is the same as the input file extension, unless you force compression of the file with the `-outcomp` option.

QikProp places any generated output files in the current working directory. If you run QikProp from Maestro, this directory is the file i/o directory specified in the Directory tab of the Preferences panel.

The default file i/o directory is Maestro's current working directory. To change the file i/o directory preference, open the Preferences panel from the Maestro menu in the main window and select a directory option. The default file i/o directory, as well as most of the other directory choices, is relative. For instance, if you select Maestro's current working directory, Maestro writes output files to the directory from which you launched Maestro, unless you have since entered the `cd` or `chdir` command in the Command Input Area.

The Output Structure File

For structures that do not meet the input criteria, no properties are generated, but the structures are still written to the output file. Any preexisting QikProp properties are cleared, so that the "failed" structures do not have any QikProp properties in the output file. The names of the properties in a Maestro output file are formed by prepending `i_qp_` or `r_qp_` to the names given in Table 1.1, depending on whether the property can take real or integer values.

The *jobname*`.qpsa` File

The *jobname*`.qpsa` file contains additional information for each molecule, including the covalent neighbors list, atom-type assignments, associated atom-type force field parameters, and a solvent accessible surface area (SASA) analysis. At the end of the file, there is a summary giving the number of molecules that were successfully processed and the number of molecules that were not successfully processed, and a report of the processor time used.

The *jobname*`.out` File

The *jobname*`.out` file lists the output descriptors, predicted properties, and results of comparisons between the compound's properties and the properties of typical drugs. Each property with a value outside the 95% range for known drugs is marked with an asterisk (*).

The *jobname*`.out` file also contains a "QP Breakdown" section for each structure. The four quantities listed in this section, log Po/w, –log S, log BB and log PMDCK, are computed as a linear combination of more basic molecular properties. QikProp lists the descriptor values multiplied by their coefficients for each derived property. The predicted property value, listed under `Total`, is the sum of these. Descriptor values that are above the maximum observed in the training set used to develop QikProp are marked with an angle bracket (<).

The *jobname*`.CSV` File

The *jobname*`.CSV` file is a comma-separated-value (CSV) file that lists the 44 QikProp descriptors and predicted properties on one comma-separated line per molecule. The stem of the *jobname*`.CSV` file is to the job name entered in the Start panel. When QikProp writes this file, it converts any commas appearing in molecule names to semicolons, and it removes any blanks from molecule names.

You can import the *jobname*`.CSV` file into spreadsheet programs such as Microsoft Excel, OpenOffice and Gnumeric, or analysis programs such as JMP. You might want to do this to create a database for logP or logS predictions, for example—see Section 3.8 on page 22.

QikProp makes entries in the *jobname*`.CSV` file for any molecules with fatal errors. Check the file for these messages after multiple-structure runs.

The *jobname*`.warning` File

Warnings about potential problems with the input structures are written to the *jobname*`.warning` file. For instance, warnings are issued when QikProp encounters empty input files, structures with multiple fragments, and structures with missing hydrogens.

In addition to statements placed in the *jobname*`.warning` file, if multiple structures are being processed in a single run, any structures with fatal errors (e.g., disallowed atom types) are skipped, and entries are made in the *jobname*`.CSV` file stating that the processing of the molecules failed. After executing QikProp for multiple structures, you should display the contents of the *jobname*`.warning` file, if present, and search for any occurrences of `empty` or `failed` in the *jobname*`.CSV` file. If QikProp encounters no problems with the processed structures, a *jobname*`.warning` file is not generated.

The *jobname*`.log` File

This file contains a brief summary of the options used to run the job.

3.6 Customizing QikProp with the QPlimits File

The `QPlimits` file can be used to set the defaults for some QikProp settings or to customize the ranges of values that determine which molecules are flagged as being dissimilar to other known drugs. This file can be copied from the `$SCHRODINGER/qikprop-v`*version*`/data` directory to the current working directory or to `$HOME/.schrodinger/qikprop`.

By default the molecules are flagged as being dissimilar to other known drugs based on comparison to a set of test structures studied by Professor William L. Jorgensen. If you want to use limits other than the default limits, you can edit `QPlimits`. The current limits for each property are given beneath the name of the property. These limits can be deleted and replaced by user-specified minimum and maximum values. Structures that have properties that do not lie within the limits are flagged with an asterisk in the output files.

You can suppress the *jobname*`.qpsa` file (called `QPSA.out` in the example), and choose fast or normal processing modes by changing the settings at the top of the `QPlimits` file. Changes to the names of the output and CSV files are ignored. Command-line options and Maestro settings supersede any settings made in the `QPlimits` file. This file is reproduced below.

```
This file lets the user change the names of the output QP.out and QP.CSV files,
suppress the QPSA.out file, choose the fast (no QM) or slower processing modes, and
change the default ranges for 95% of Drugs that are used to determine the "star" count.
Enter name of QP.out file (20 characters max):
QP.out
Enter name of QP.CSV file (20 characters max):
QP.CSV
Suppress output of QPSA.out file? (y or n)
n
Processing Mode (FAST or NORMAL; FAST does not output dipole moment, IP and EA)
normal
Molecular Weight                     standard is:  ( 130.0 / 725.0)
 130.0 / 725.0
Dipole Moment (D)                    standard is:  (   1.0 /  12.5)
   1.0 /  12.5
Total        SASA                    standard is:  ( 300.0 /1000.0)
 300.0 /1000.0
Hydrophobic  SASA                    standard is:  (   0.0 / 750.0)
   0.0 / 750.0
Hydrophilic  SASA                    standard is:  (   7.0 / 330.0)
   7.0 / 330.0
Carbon Pi    SASA                    standard is:  (   0.0 / 400.0)
   0.0 / 400.0
Weakly Polar SASA                    standard is:  (   0.0 / 150.0)
   0.0 / 150.0
Molecular Volume (A^3)               standard is:  ( 500.0 /2000.0)
 500.0 /2000.0
No. of Rotatable Bonds               standard is:  (   0.0 /  15.0)
   0.0 /  15.0
Solute as Donor -    Hydrogen Bonds   standard is:  (   0.0 /   6.0)
   0.0 /   6.0
Solute as Acceptor - Hydrogen Bonds   standard is:  (   2.0 /  20.0)
   2.0 /  20.0
Solute Globularity   (Sphere = 1)     standard is:  (  0.75 /  0.95)
  0.75 /  0.95
Solute Ionization Potential (eV)      standard is:  (   7.9 /  10.5)
   7.9 /  10.5
Solute Electron Affinity    (eV)      standard is:  (  -0.7 /   1.7)
  -0.7 /   1.7
QP Polarizability (Angstroms^3)       standard is:  (  13.0 /  70.0)
  13.0 /  70.0
QP log P  for     hexadecane/gas      standard is:  (   4.0 /  18.0)
   4.0 /  18.0
QP log P  for     octanol/gas         standard is:  (   8.0 /  43.0)
   8.0 /  43.0
QP log P  for     water/gas           standard is:  (   5.0 /  48.0)
   5.0 /  48.0
QP log P  for     octanol/water       standard is:  (  -2.0 /   6.0)
  -2.0 /   6.0
QP log S  for   aqueous solubility    standard is:  (  -6.0 /   0.5)
  -6.0 /   0.5
QP log K hsa Serum Protein Binding    standard is:  (  -1.5 /   1.2)
  -1.5 /   1.2
```

```
QP log BB for      brain/blood        standard is:  (  -3.0 /   1.0)
  -3.0 /   1.0
No. of Primary Metabolites            standard is:  (   1.0 /   8.0)
   1.0 /   8.0
```

3.7 QikProp Methodology

QikProp was developed with the following process. The BOSS program and the OPLS-AA force field were used to perform Monte Carlo statistical mechanics simulations on organic solutes in a periodic boxes of explicit water molecules. This process resulted in configurational averages for a number of descriptors, including hydrogen bond counts and solvent-accessible surface area (SASA). Correlations of these descriptors to experimentally determined properties were obtained, and then algorithms that mimic the full Monte Carlo simulations and produce comparable results were developed.

When performing an evaluation, QikProp rapidly analyses atom types and charges, rotor counts, and the sample molecule's volume and surface area. QikProp then uses this information, along with the physical descriptors calculated using the algorithms described above, in regression equations. The result is an accurate prediction of a molecule's pharmacologically relevant properties.

3.8 Use of Similarity

QikProp uses similarity in property space to improve predictions of key properties, and to return similar molecules from a database. The similarity calculation uses a normalized Euclidean functional form with a proprietary set of per-property weights. Similarities range from a value of 0.0 (no similarity) to 1.0 (exactly identical).

3.8.1 Improvement of logS and logP

To improve predictions for the key properties solubility (logS) and permeability (logP), similarities in molecular property space are used along with a large set of experimental measurements and the corresponding QikProp predictions. The QikProp prediction is adjusted with a contribution from experimental values if the similarity is large enough, as follows. When a molecule is processed by QikProp, its similarity in molecular property space is computed to each of the molecules in a large collection of molecules whose experimental value is known.

- If the similarity is less than 0.9, the QikProp prediction is used.
- If the similarity is greater than or equal to 0.9, the predicted property P_{pred} is adjusted according to

$$P_{\text{pred}} = SP_{\text{exp}} + (1 - S)P_{QP}$$

where S is the similarity, and P_{exp} and P_{QP} are the experimental and QikProp predictions for the most similar molecule. So at a similarity of 1.0 (molecules have identical properties) the QikProp reported prediction will be the experimental value for that molecule.

Files of QikProp predictions and experimental measurements for a large collection of molecules have been created for logP and logS. They are in the same format as an output QikProp CSV file with one additional column containing the experimental (known) value for logP or logS. These files are located in the `$SCHRODINGER/qikprop-v`*version*`/data` directory as `QPxpd2.CSV` for logS and `QPxpd1.CSV` for logP.

To make use of your own data, you can modify these files or create new files. If you modify the files in the `$SCHRODINGER/qikprop-v`*version*`/data` directory and retain their original names, QikProp calculations will automatically use the modified files. If the modified files are located elsewhere, you can use them in the QikProp calculation with the `-logsdb` and `-logpdb` command line options. In the interests of speed, only the first 10 000 entries in the files are used in the logP or logS calculations. Note that you cannot use scientific notation (“E format”) in these files.

3.8.2 Retrieval of the Most Similar Molecules

QikProp uses the same molecular similarity calculation to identify and return names of the *n* most similar compounds from a CSV file that contains a set of QikProp predictions. The default file is the file `Similar.CSV`, which is stored in `$SCHRODINGER/qikprop-v`*version*`/data`. This file was generated by running QikProp on a large set of druglike molecules. You can request to return the *n* most similar compounds to each molecule processed by QikProp either from Maestro or from the command line. By default, the five most similar compounds are reported.

As for the logP and logS files, you can modify the existing file or create your own. To create your own file, simply run QikProp on the set of molecules of interest. If you modify or replace `$SCHRODINGER/qikprop-v`*version*`/data/Similar.CSV`, then the modified or replaced file is used automatically. For any other file name or location, you can select the file with the `-simdb` command-line option. This file can have a maximum of 500 000 entries.Note that you cannot use scientific notation (“E format”) in this file.

Quality of QikProp Results

Experimental results for more than 710 compounds including about 500 drugs and related heterocycles were used in developing QikProp. The following table summarizes the fits for QikProp 2.2.

Table 4.1. Statistics on QikProp 2.2 Fits to Experimental Data

Property	N	r^2	RMS error	MW range
polarizability (Å^3)	78	0.97	1.05	20–200
log P (hexadecane/gas)	392	0.93	0.37	20–200
log P (octanol/gas)	117	0.91	0.61	20–250
log P (water/gas)	421	0.93	0.58	20–250
log P (octanol/water)	448	0.93	0.50	20–735
log S (water/solid)	389	0.91	0.63	20–823
log K'hsa (serum protein binding)	90	0.82	0.25	130–765
log IC_{50} for HERG blockage	47	0.76	0.80	275–750
log BB (brain/blood)	104	0.80	0.35	20–525
log P Caco-2	126	0.72	0.61	60–520
log P MDCK	52	0.73	0.57	130–430
log Kp (skin permeability)	58	0.78	0.68	20–600

In the QikProp output file for the log P (octanol/water), log S and log BB output, if the value for a utilized descriptor exceeds the range for the experimental training set, it is flagged. Otherwise, the molecular weight ranges above indicate the domain of validated applicability of the regression equations.

4.1 Test Set Results

The following tables illustrate results for test sets of molecules that were not in the QikProp 2.2 training set. The 3D structures were obtained directly from SciFinder, then to gauge the effect of optimization, they were optimized with the BOSS program using the OPLS-AA force field and CM1P charges. Both sets of structures were processed by QikProp 2.2

Table 4.2. Log S results for a test set of molecules.

Molecule	MW	exptl log S	QPlogS[1]	QPlogS[2]
N-methylmorpholine	101.1	1.00	1.85	1.66
2,5-dimethylpiperazine	114.2	0.49	1.47	1.36
Isoniazid	137.1	0.01	–0.79	–0.85
3,3-dimethyl-1-butanol	102.2	–0.50	–0.81	–0.88
3-methyl-3-hexanol	116.2	–1.00	–1.99	–1.53
bis-(2-chloroethyl) sulfone	191.1	–1.50	–1.48	–1.42
Minoxidil	209.3	–1.98	–2.07	–1.92
2,4-D	221.0	–2.51	–2.21	–2.78
Heptabarbital	250.3	–3.00	–2.29	–2.30
Sulfadiazine	250.3	–3.51	–2.08	–2.06
Terbutyrne	241.4	–4.00	–2.89	–3.57
1,2,4-tribromobenzene	314.8	–4.50	–4.00	–4.05
Quinonamid	318.5	–5.03	–3.34	–3.89
Benfluralin	335.3	–5.53	–4.08	–4.49
Fluoranthene	202.3	–6.00	–6.44	–6.61
o,p'-DDD	320.0	–6.51	–6.53	–6.90
7,12-dimethylbenz(a)anthracene	256.3	–7.02	–6.40	–6.90
2,2',3,4,6-PCB	326.4	–7.43	–8.04	–7.71
benzo(j)fluoranthene	252.3	–8.00	–8.24	–8.63
2,2',4,4',5,5'-PCB	360.9	–8.56	–10.75	–8.50
q^2			0.89	0.95
RMS			1.01	0.70
Average error			0.77	0.55

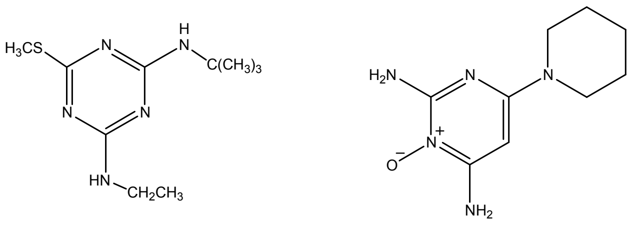
1. Using structures from SciFinder.
2. Using optimized structures.

Table 4.3. Log $P_{o/w}$ results for a test set of molecules.

Molecule	MW	exptl log $P_{o/w}$	QPlogPo/w[1]	QPlogPo/w[2]
Isoniazid	137.1	–0.70	–0.24	–0.04
Nicotinamide	122.1	–0.37	–0.13	–0.20
N-methylmorpholine	101.1	–0.33	–0.65	–0.52
Sulfadiazine	250.3	–0.09	–0.06	0.02
Pyrazole	68.1	0.26	0.21	0.20
Sulfamethiazole	270.3	0.54	–0.40	–0.39
Minoxidil	209.3	1.24	1.00	1.13
Heptabarbital	250.3	2.03	0.82	0.86
2,4-D	221.0	2.81	2.03	2.27
Ethisterone	312.5	3.11	3.78	4.35
Pericyazine	365.5	3.52	3.19	3.34
Terbutyrne	241.4	3.74	2.52	2.82
Captafol	349.1	3.80	2.83	3.13
Doxepin	279.4	4.29	3.41	4.00
Biquinoline	256.3	4.31	4.25	4.35
Prochlorperazine	373.9	4.88	3.82	3.79
Fluoranthene	202.3	5.16	5.03	5.13
Benfluralin	335.3	5.29	3.99	4.76
Amitraz	293.4	5.50	5.60	5.86
7,12-dimethylbenz(a)anthracene	256.3	5.80	5.28	5.63
q^2			0.86	0.92
Rms			0.84	0.65
ave. error			0.63	0.47

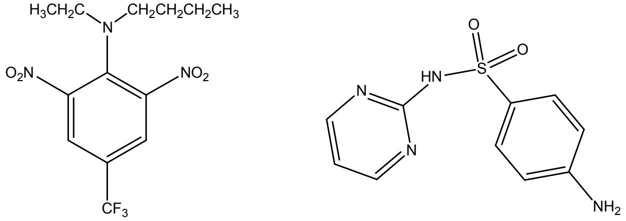
1. Using structures from SciFinder.
2. Using optimized structures.

Some of the structures from the test sets are shown below.



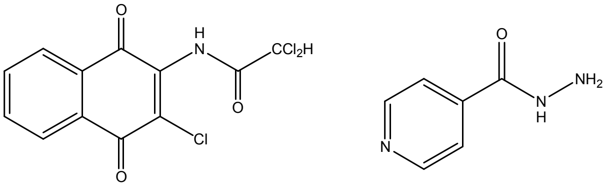
terbutyrne

minoxidil



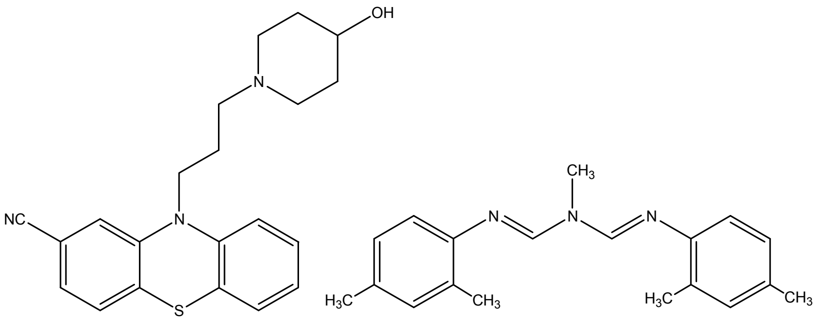
benfluralin

sulfadiazine



quinonamid

isoniazid



pericyazine

amitraz

Figure 4.1. Selected structures from the test set.

4.2 Accuracy Issues

QikProp predictions are performed for specific conditions (pH 7, for one), and comparison of the results with experiment must take into account any differences in those conditions. The following issues can affect the accuracy of the comparison:

- active vs passive transport mechanisms
- efflux pump mechanisms
- smaller number of compounds in the training sets
- different experimental conditions than that used in the regression
- noise in experimental data due to inherent complexity of chemical systems interacting with biological systems

The biggest discrepancies are for PCaco predictions, owing to these issues. Values for log BB are less problematic. The QikProp fit for log BB is very good—at the noise level of the data—but there are a few outliers such as oxazepam and tiotidine that are not fitted by any regression method.

Obtaining reliable experimental data for log S requires pure, non-hydrated, no-salt crystals in equilibrium with a saturated aqueous solution. This condition is often difficult to achieve, so it is important to check the experiments as well as the QikProp predictions.

Comparisons must take into account the protonation state of molecules at physiological pH. For example, compounds that contain a tertiary amine are protonated at that pH, and log P measurements must be performed at pH 11, and corrected to pH 7.

Experimental log P values (such as log Po/w) for amino acids, which have a saturated amine and a carboxylic acid, are affected by the protonation state. Though the Hansch * value is supposed to correspond to the un-ionized state, there is no pH at which either the amine is not protonated or the COOH deprotonated or both. The Hansch * values in these cases correspond to a pseudo log D at pH 7, at which pH the acid group is ionized, but the amino group's protonation has been corrected for. The data in Table 4.4 clearly show the problem.

*Table 4.4. Hansch * values for selected molecules*

Molecule	* Value	Molecule	* Value
ethanol HOCH2CH3	–0.31	hydroxyacetic acid HOCH2COOH	–1.11
ethylamine H2NCH2CH3	–0.13	aminoacetic acid (glycine) H2NCH2COOH	–3.21
difference	–0.18	difference	+2.10

The difference for an alcohol and an un-ionized amine should be –0.18, but for the two molecules that have an acid group, the difference is +2.10. If the acid in both cases was un-ionized, the difference should still be –0.18. In fact, the value of –1.11 is for the un-ionized acid, but the value of –3.21 is for the ionized acid. The discrepancy of 2.28 is exactly what is expected for the un-ionized acid, i.e.

pH 7.0 – pKa of RCOOH (4.8) = 2.2

The true un-ionized Hansch * value for glycine should be about –1.0; QikProp 1.5 gives –0.96.

The above issues do not apply to peptides in general, which may be capped. QikProp 1.5 agrees with Hansch * values for capped peptides, as illustrated in Table 4.5.

*Table 4.5. Comparison of Hansch * values and QikProp predictions for capped peptides.*

Peptide	Hansch * Value	QikProp 1.5
AcMetValNH2	–0.28	–0.19
Ac(Ala)3NHtBu	–0.51	–0.32
AcAlaTyrLeuNH2	–0.04	–0.37

Therefore, for QikProp versions since 1.6, the log Po/w values for amino acids have been lowered by 2.0 log units, so that QikProp reproduces the Hansch * values. The correction is noted in the output files when it is applied.

Note that this correction is not just for the standard amino acids, but for any compound with a saturated amine group and a carboxylic acid, e.g., amoxicillin.

A similar correction factor is used by Hansch and Leo to account for zwitterions. They discuss this in Ref. 16, and show that it works for molecules like amoxicillin which are zwitterionic.

4.3 Conformation Dependence of Results

The dependence of the QikProp results on molecular conformation has been studied by performing conformational searches on many molecules using BOSS with the OPLS-AA force field, followed by QikProp calculations for each conformer. The results demonstrate the following points:

1. There is generally modest difference in the QikProp predictions for different conformers. This results because the algorithms for determining the hydrogen-bond counts, which are very important descriptors, are independent of the 3D structure. The surface area and dipole moment terms are affected by conformation, but their variations are generally

small. The differences are always negligible for log BB and PCaco. The differences for log Po/w are also small, normally a few tenths of a log unit. The differences can be greater for log S, but rarely cover more than 1 log unit.

2. When there are differences, extended structures give results closer to the experimental data. The QikProp regressions have been developed using extended conformers; standard 2D–3D conversion programs usually generate extended conformers.
3. Boltzmann-weighting of the QikProp results using either the gas-phase conformer energies or the aqueous-phase conformer energies is probably not worth the effort, though it is theoretically desirable.

The results for six drugs are described below. The drugs each have multiple rotatable bonds, which were explored in the conformational searches. Clearly, negligible dependence of the results on conformation is expected for more rigid molecules.

- **acyclovir**—43 conformers covering an energy range of 9 kcal/mol were considered. The QPlogS values range from 0.3 to 0.5; the QPlogPo/w values range from –0.2 to –0.4.
- **epinephrine**—69 conformers covering an energy range of 14 kcal/mol were considered. The QPlogS values range from 0.3 to 0.6; the QPlogPo/w values range from –0.2 to –0.4.
- **haloperidol**—49 conformers covering an energy range of 18 kcal/mol were considered. The QPlogS values fall in two groups, –3 to –4 for very compact conformers with the fluorophenyl ring folded on top of the piperidine, and –4.5 to –5.0 for extended structures. The experimental log S is –4.4. The QPlogPo/w values range from 3.9 to 4.5 for all conformers.
- **linezolid**—68 conformers covering an energy range of 9 kcal/mol were considered. The QPlogS values range from –1.7 to –2.3; the QPlogPo/w values range from 0.3 to 0.9.
- **omeprazole**—42 conformers covering an energy range of 8 kcal/mol were considered. The QPlogS values range from –2.9 to –3.9 for the first 45 conformers; the QPlogPo/w values range from 1.5 to 2.2 for the first 42 conformers.
- **indinavir**—29 conformers covering an energy range of 8 kcal/mol were considered. The QPlogS values range from –2.3 to –0.2; the QP log Po/w values range from 0.8 to 1.4. This is a very flexible molecule with 16 variable dihedral angles included in the conformational search. The more extended structures give lower log S and higher log P values.

4.4 Data Sources

The sources of the experimental log P and log S values are described in Refs. 1, 2, and 4. Additional data has been kindly provided by Pharmacia Inc.

The databases of free energies of hydration and free energies of solvation in hexadecane are from Abraham et al [6].

The log BB values are mostly listed in Luco [7] and Kelder et al. [8]. The latter paper includes the experimental protocol. The values come from in vivo steady-state measurements using radiolabelled compounds with rats.The predicted CNS activities are based largely on log BB with some adjustments, and correspond well with the CNS activities reported in Table 1 of Ajay et al. [9].

The Caco-2 cell permeabilities are from Boehringer-Ingelheim (Yazdanian et al. [10] with minor additions), Affymax (Irvine et al. [11]), and AstraZeneca (Stenberg et al. [12]). The MDCK (Madin-Darby Canine Kidney) cell permeabilities are from the Affymax paper.

The experimental IC_{50} values for blockage of mammalian HERG K^+ channels are from Cavalli et al. [13], De Ponti et al. [14], and http://www.fenichel.net. There is significant scatter in the experimental data amounting to uncertainties of factors of 2 to 50. The model has been based on all available data (47 compounds); however, this is still a small data set. Drugs that have been withdrawn owing to QT-prolongation problems also exhibit a large range of IC_{50} values, e.g., cisapride (6.5 nM), sertindole (14 nM), terfenadine (56 nM), and grepafloxacin (50000 nM). For trovafloxacin, the QPlogHERG = –5, so IC_{50} = 10000 nM. Thus, the allowable limit for IC_{50} values depends on the class of compounds, dosage, and bioavailability.

The log Kp values for skin permeability are from data of Flynn noted in Potts and Guy [15].

The log K'hsa data for binding to human serum albumin are from Colmenarejo et al. [5].

Descriptor Information

The following metabolic reactions contribute to the `#metabol` descriptor.

- aromatic OH oxidation
- enol oxidation
- benzylic-like H → alcohol
- allylic H → alcohol
- secondary alcohol → ketone
- primary alcohol → acid
- tertiary alcohol E1 or SN1
- amine dealkylation
- ether dealkylation
- pyridine C2 hydroxylation
- aniline NH → NOH or NCOR
- low IP—easily oxidized
- alpha hydroxylation of cyclic ether
- sulfoxide → sulfone
- alpha hydroxylation of carbonyl
- alpha, beta dehydrogenation at carbonyl
- thiol SH → SSR, SR
- para hydroxylation of aryl
- aryl sulfide → S=O
- reduction of aryl nitro to amine
- oxidative deamination of primary amine

The following reactive functional groups contribute to the `#rctvFG` descriptor:

- acyl halide
- hetero-halogen bond
- NAS substrate
- alkyl halide
- halogen alpha to W-group
- heteroatom in 3-ring
- activated cyclopropane
- aluminum present (toxic)
- silicon present (toxic)
- hetero-hetero single bond
- azo, diazo, or azide

- acceptor carbonyl or derivative
- anhydride or analog
- unhindered ester
- sulfonate or relative
- phosphonate or relative
- acetal or analog
- carbonate
- thiol—oxidation possible
- carbonyl in 3-ring

References

1. Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000,** *122,* 2878-2888.

2. Jorgensen, W. L.; Duffy, E. M. Prediction of Drug Solubility from Monte Carlo Simulations. *Bioorg. Med. Chem. Lett.* **2000,** *10,* 1155-1158.

3. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Delivery Rev.* **2001**, *46*, 3-26.

4. Jorgensen, W. L.; Duffy, E. M. Prediction of Drug Solubility from Structure. *Adv. Drug Delivery Rev.* **2002,** *54,* 355-366.

5. Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. Cheminformatic Models To Predict Binding Affinities to Human Serum Albumin. *J. Med. Chem.* **2001,** *44,* 4370-4378.

6. Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. *J. Chem. Soc. Perkin Trans.* **1994**, *2*, 1777-1791.

7. Luco, J. M. Prediction of the Brain-Blood Distribution of a Large Set of Drugs from Structurally Derived Descriptors Using Partial Least-Squares (PLS) Modeling. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 396-404.

8. Kelder J.; Grootenhuis, P. D.;Bayada, D.M.; Delbressine, L.P., Ploemen, J.P. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* **1999**, *16*, 1514–1519.

9. Ajay; Bemis, G. W.; Murkco, M. A. Designing Libraries with CNS Activity. *J. Med. Chem.*, **1999**, *42*, 4942–4951.

10. Yazdanian M.; Glynn, S. L.; Wright, J. L.; Hawi A. Correlating partitioning and caco-2 cell permeability of structurally diverse small molecular weight compounds. *Pharm. Res.* **1998***, 15*, 1490–1494.

11. Irvine, J. D.; Takahashi, L.; Lockhart, K; Cheong, J.; Tolan, J. W.; Selick, H. E.; Grove, J. R. MDCK (Madin-Darby canine kidney) cells: a tool for membrane permeability screening. *J. Pharm. Sci.* **1999**, *88*, 28–33.

12. Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption. *J. Med. Chem.* **2001**, *44*, 1927–1937.

13. Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Toward a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of HERG K^+ Channel Blockers *J. Med. Chem.* **2002**, *45*, 3844–3853.

14. De Ponti, F.; Poluzzi, E.; Montanaro, N. Organising evidence on QT prolongation and occurrence of Torsades de Pointes with non-antiarrhythmic drugs: a call for consensus. *Eur. J. Clin. Pharmacol*. **2001**, *57*, 185–209.

15. Potts, R. O.; Guy, R. H. Predicting skin permeability. *Pharm. Res*. **1992**, *9*, 663–669; A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. *Pharm. Res*. **1995**, *12*, 1628–1633.

16. Exploring Qsar: Fundamentals and Applications in Chemistry and Biology; C Hansch, C.; Leo, A.; Hoekman, D. H., Eds.; Oxford University Press, 1995; p.156.

17. Kleinbaum, D. G.; Kupper, L. L.; Muller, K. E.; Nizam, A. Applied Regression Analysis and Other Multivariable Methods. Third Edition, Duxbury Press: Pacific Grove, 1987.

18. Miller J. C.; Miller, J. N. Statistics for Analytical Chemistry, 3rd ed.; Ellis Horwood: New York, 1993.

19. Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985,** *25,* 64-73.

20. Willet, P. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998,** *38,* 983-996.

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
- QikProp purchaser (company, research institution, or individual)
- Primary QikProp 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`.

Index

A

atoms
  allowed elements, QikProp 13
  hydrogen, requirements on 13

B

BOSS/MCPRO Z-matrix files 18

C

concatenated mol files 13
conformation dependence of QikProp results 14, 30
conventions, document v
CSV files
  logP and logS databases 23
  QikProp output 20
  similarity database 23

D

descriptors, table of 2
directory
  installation 5
  output 19

E

entries, CSV file, failed 20
entries, Project Table
  choosing for job input 9
  incorporation mode 9
  sorting 11
environment variable
  `SCHRODINGER` 5
experimental values, using for logP and logS predictions 22

F

failed entries
  QikProp output for 20
fast mode 13
  example 11
  selecting in Maestro 15
  validity of properties 13
file formats
  accepted 17
  restrictions on 13

H

host, selecting 16
hydrogen atoms, requirements on 13

L

LigPrep 14

M

mol files 17
  concatenated 13

P

PDB files 18
product installation 38
Project Table panel 9
properties
  QikProp, table of 2
  validity of QikProp 13

Q

QikProp
  command syntax 16
  conformation dependence of results 14, 30
  input files 17
  input source 14
  monitoring jobs 18
  multiple molecule processing 13
  output files 18
  panel 10, 15
QPlimits file 1, 20
`.qpsa` file, suppressing 17

R

restrictions on input structures 13

S

Schrödinger contact information 38
SD files 17
solvent accessible surface area (SASA) analysis 19

structures
  2D 13
  charged 14
  conformation sensitivity 14, 30
  disallowed, QikProp treatment of............... 20
  importing 8
  preprocessing with LigPrep........................ 14
  restrictions 13

U

united-atom types... 13

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

101 SW Main Street
Suite 1300
Portland, OR 97204

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

8910 University Center Lane
Suite 270
San Diego, CA 92122

155 Gibbs St
Suite 430
Rockville, MD 20850-0353

Dynamostraβe 13
D-68165 Mannheim
Germany

Zeppelinstraβe 73
D-81669 München
Germany

Potsdamer Platz 11
D-10785 Berlin
Germany

Quatro House
Frimley Road
Camberley GU16 7ER
United Kingdom

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

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

SCHRÖDINGER®