Multiple Sequence Viewer

Schrödinger Software Release 2015-2

Schrödinger Press

May 2015

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].

Multiple Sequence Viewer

The Multiple Sequence Viewer panel is an alignment, visualization, and manipulation toolkit for multiple sequences, which was developed in collaboration with Dr. Jano Jusuf & Dr. Stanley Krystek from Bristol-Myers Squibb.

The multiple sequence viewer (MSV) has its own projects, which contain all the sequences in the project along with associated data. The project is stored in a single file with a `.msv` extension, and by default is stored inside the Maestro project. You can save it externally if you wish.

The project can include sequences imported directly into the project, and sequences that are displayed in the Workspace. Directly imported sequences remain in the project unless explicitly deleted. Sequences in the Workspace are transient: when structures are included in or excluded from the Workspace, the sequence is added to or removed from the project.

Although the data is stored separately from the Maestro project data, there are interactions between the Workspace and the MSV, which depend on settings that you make in the MSV:

- Changes that are made in the Workspace can be propagated to the MSV. These changes cover inclusion and exclusion of entries; deletion, mutation, or insertion of residues; and selection of residues.
- Changes to the sequences in the MSV are not propagated back to the Maestro structure.
- Changes to the residue selection in the MSV can be propagated back to the Workspace.
- Color schemes can be transferred between the MSV and the Workspace.

To open the Multiple Sequence Viewer panel, choose Tools → Multiple Sequence Viewer.

The Multiple Sequence Viewer panel has a menu bar, a set of toolbars, the sequence display area, and a status area. These features are described in the sections below.

The speed of the sequence viewer depends on the number and length of sequences loaded into the viewer. Typically, alignments of 20 sequences of 300 residues or fewer can be interactively viewed and edited. Below are a few speed optimization tips.

- Hide the ruler (Settings → Display Ruler off)
- Use unwrapped mode (Settings → Wrap Sequences off)
- Disable annotation grouping (Settings → Group Annotations off)
- Collapse all sequences (Sequences → Collapse All Sequences)
- Hide annotations (Annotations → Clear All Annotations)

- Reduce the size of the sequence area: move the sequence area splitter to the right, shrink the entire window

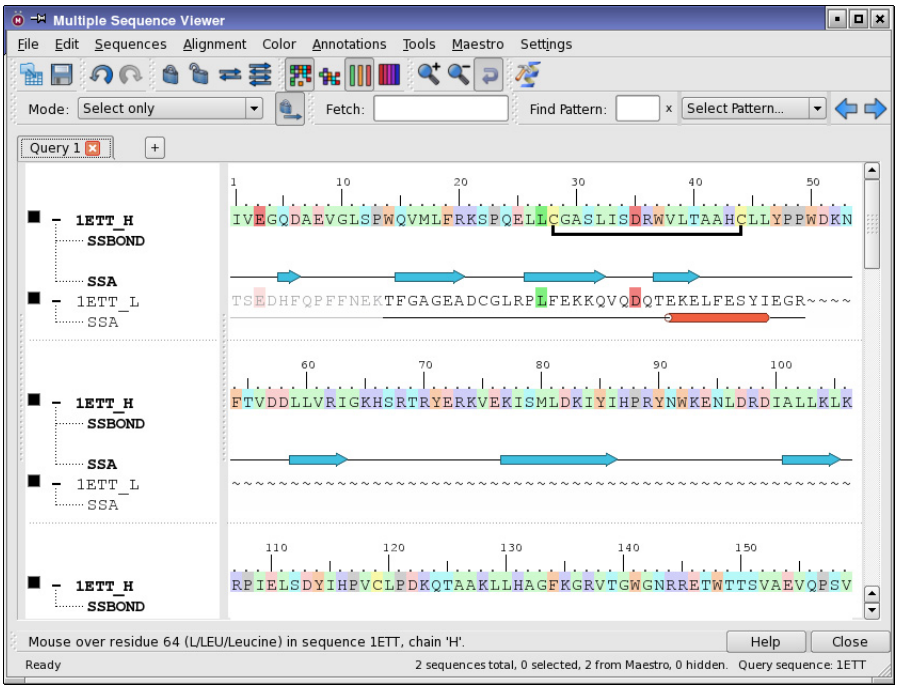


Figure 1. The Multiple Sequence Viewer panel.

1 The Menu Bar

The menu bar contains eight menus: File, Edit, Sequences, Alignment, Color, Annotations, Tools, and Settings. These menus are described below.

1.1 File Menu

The File menu provides tools for creating and managing queries, opening and saving MSV projects; importing and exporting sequences; saving images, and closing the panel.

- New Query—Create a new, empty query in a new tab.
- Rename Query—Rename the current query. The name appears on the tab for the query.
- Duplicate Query—Create a copy of the selected sequences of the current query in a new tab.

- Delete Query—Delete the current query. All sequences, structures, and data associated with the query are removed (unless they are also in another query tab).
- Open—Open an existing MSV project. Opens a file selector, in which you can navigate to and select the project.
- Save—Save the current MSV project. If the project has not yet been saved, a file selector opens, in which you can navigate to a location and name the project. If the project has been saved previously, the project is simply saved.
- Save As—Save the current MSV project with a new name. Opens a file selector, in which you can navigate to a location and name the project. After saving, the current project is the one with the new name.
- Import Sequences—Import sequences into the project. Sequences can be imported from a range of file formats: FASTA, SWISSPROT, GCG, PIR, EMBL, as well as PDB and Maestro. Opens a file selector, in which you can navigate to and select the sequence file. The file selector has these options:
  - Align to query sequence—When the sequences are imported, align them to the query sequence.
  - Replace matching sequences—If an imported sequence exactly matches a sequence that is already in the MSV, replace the existing sequence with the imported sequence, while preserving the alignment.
  - Translate DNA / RNA sequences—Translate DNA and RNA sequences into the sequence for the proteins they code for, using standard genetic code. The protein sequence is imported instead of the nucleic acid sequence.
  - Incorporate PDB files into Maestro—When PDB sequences are imported, add the corresponding structures as entries to the Maestro project.

  Structural data (ATOM records), B-factors, and secondary structure assignments are also imported if the data are in the PDB file. A nonstandard version of FASTA format is accepted, in which a residue can be preceded by its residue number; numbering is otherwise sequential, starting from 1 by default.
- Export Sequences—Export sequences from the project to a file in FASTA format. Opens a file selector, in which you can navigate to a location and name the file. The file selector has these options:
  - Save annotations—Save the SSP and SSA annotations in the file.
  - Save similarity values—Save the calculated percent similarity between sequences. The values are added to the sequence name in the exported FASTA file as ID (identity), SIM (similarity), HOM (homology), e.g.

    ```
    >1abc|ID:7.14|SIM:19.64|HOM:17.86
    ```

    - Export only selected part of the alignment—Export only the residues that are selected in all sequences.

- Save Image—Save a bitmap image of the sequence viewer in PNG format. Opens a file selector, in which you can navigate to a location and name the file.
- Close—Close the current MSV project and the panel.

1.2 Edit Menu

From the Edit menu you can choose to undo or redo actions, edit sequences, and delete sequences. The sequence editor panel accepts standard editor key strokes, such as CTRL+C, CTRL+X, and CTRL+V for copying, cutting, and pasting text.

- Undo—Undo the last editing operation. The operation is appended to the menu item text, for example Undo Load File.
- Redo—Redo the last undone editing operation. The operation is appended to the menu item text, for example Redo Load File.
- New Sequence—Create a new sequence by entering the letter codes for the residues. Opens the Sequence Editor dialog box, in which you can name the sequence and type in the letter codes for the sequence.
- Edit Sequence—Edit an existing sequence as a string of letter codes. Opens the Sequence Editor dialog box, in which you can change the name of the sequence, and add or delete residues or gaps by character code. Only the 20 standard amino acid codes, `X` (unknown residue), and `-` and `~` (gap symbols) are recognized. When you have finished editing, you can either replace the existing sequence, or add it as a new sequence to the end of the list.

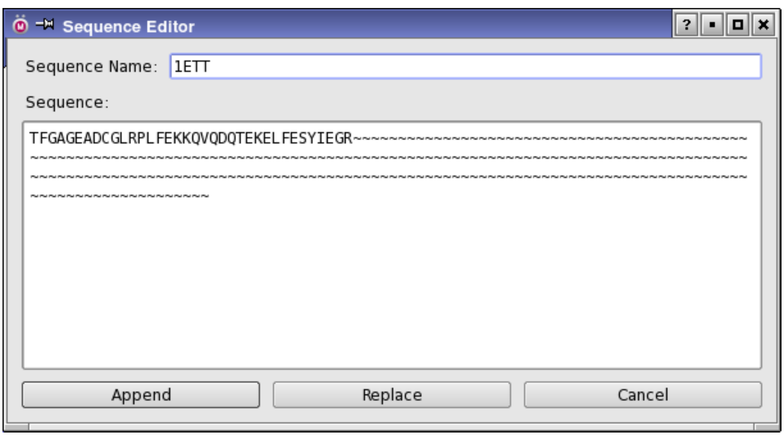


Figure 2. The Sequence Editor panel.

- Paste in FASTA Format—Insert entire sequences into the sequence viewer in FASTA format. Opens the Sequence Editor dialog box, in which you can paste the sequences and edit them. The same editing rules apply, except that lines beginning with a > character are treated as a sequence header. These lines also divide sequences, which are saved as separate sequences.

  This feature is useful for copying an alignment from a web site and adding it directly to the MSV without having to save a file. You can also add sequences by importing a file.

- Duplicate Selected Sequences—Duplicate the selected sequences, and place each duplicate immediately below its parent.

- Find Pattern—Find a pattern in the sequences. Displays the Find toolbar if it is not displayed, and puts focus in the text box so you can start typing the pattern.

- Renumber Residues—Renumber the residues in one or more sequences. Opens the Renumber Residues dialog box, which offers two renumbering methods.

  The first is a simple sequential renumbering scheme, in which you can specify the first residue number and the increment. Insertion codes can be kept, but the residue numbers are not kept the same for residues with insertion codes, they are incremented as for the rest of the sequence. You can renumber multiple sequences with this method, with the result that the residues in all the sequences have the same number at the same position in the sequence viewer. This is useful for identifying particular regions in a set of aligned sequences.

  The second method is to renumber the sequence with the residue numbers of a template, which you import, and which is aligned to the sequence. Insertions in the sequence with respect to the template are not renumbered, so this method is useful mainly for highly similar sequences.

- Delete Selection—Delete the sequences that are selected in the viewer.

- Remove Redundant Sequences—Remove sequences whose sequence identity is greater than a given threshold. Opens a dialog box in which you can set the threshold, then click Remove to perform the action. The default threshold is 100%. For each pair of sequences that are considered identical, the shorter of the two is deleted; if they are of the same length, the second (lower down in the sequence viewer) is deleted. In the latter case, you can change which sequence is discarded by reordering the sequences. This task operates on all sequences. If you click Cancel in the dialog box, the redundant sequences become the selection, so you can perform operations on them.

1.3 Sequences Menu

From the Sequences menu you can control which sequences are selected, which sequences are shown, and the order in which they are listed; and you can manage the query tabs.

- Hide Selected—Hide the selected sequences
- Show All—Display all sequences
- Clear All—Delete all the sequences. This is equivalent to creating a new project (which can be saved with File → Save As).
- Select All—Select all sequences
- Deselect All—Deselect all sequences
- Invert Selection—Invert the selection of the sequences: select the unselected sequences, and deselect the selected sequences.
- Expand All—Expand all sequences so that the associated data, such as secondary structure assignment, is displayed.
- Collapse All—Hide the associated data for all sequences.
- Set Color of Sequence Name—Set the color used for the name of the selected sequences. Opens a color selector, in which you can select a color. This feature allows you to color-code the names of the sequences.
- Sort by Tree Order—Order sequences by the phylogeny tree generated by ClustalW. The tree is displayed in the leftmost part of the display area, which is normally hidden. When this option is selected, the other sorting items are not available. The nodes of the tree have a shortcut menu, with items Swap Branches, to swap the order of the branches originating at the node, Select Sequences, for selecting sequences from the branches originating at the node, and Hide Branch, for hiding the sequences for the branch.
- Sort Ascending—Sort the sequences by the value of a given property, in ascending order. The properties that can be used are Name, Chain ID, Length, Number of Gaps, Sequence Identity, Sequence Similarity, Sequence Homology, and Sequence Score. Homology is calculated as the percentage of residues with identical side-chain chemical properties (as defined for the Side-Chain Chemistry color scheme).
- Sort Descending—Sort the sequences by the value of a given property, in descending order. The properties that can be used are the same as for Sort Ascending.
- Move Up—Move the selected sequences up one position in the list.
- Move Down—Move the selected sequences down one position in the list.

- Move to Top—Move the selected sequences to the top of the list.
- Move to Bottom—Move the selected sequences to the bottom of the list.
- Get PDB Structures—Download structural information (secondary structures, B-factors, coordinates) for sequences that came from the PDB, such as in a Blast search. The information is obtained from a local copy of the PDB or from the RCSB web site. If sequences are selected, information is obtained for these sequences, otherwise it is obtained for all sequences. The PDB sequence replaces the corresponding sequence in the MSV.

1.4 Alignment Menu

The Alignment menu provides tools for automatic sequence alignment, residue selection and manual sequence alignment tasks. Some of the features are also available on the toolbar. Residues are not renumbered when the gaps are added or removed.

- Multiple Alignment—Align the selected sequences simultaneously using ClustalW. If there are columns (residues) selected, the alignment is performed only on the selected residues. You can run an alignment on several discontinuous selected regions at the same time.
- Pairwise Alignment—Align the selected sequences pairwise using a Smith-Waterman algorithm, with the settings from the Alignment Settings dialog box.
- Align and Merge—Align new sequences with a query sequence without changing the existing alignment. Gaps are inserted into the existing alignment to preserve the residue matching.
- Align by Residue Numbers—Align sequences so that residues with identical residue numbers (and insertion codes) are aligned. This is useful for families of proteins that share common numbering schemes, such as antibodies.

  You can also renumber aligned sequences so that the aligned residues have the same residue numbers, with Edit → Renumber Residues.
- Use Constraints— Apply constraints on pairwise alignments, so that the constrained residues are in the same position (same column) after the alignment.

  When you select this option, a constraint row is displayed between the query sequence and the other sequences. To add a constraint, click on a residue in the query sequence and then on a position in one of the other sequences. The constraints are displayed as blue lines connecting the constrained residue pair. To remove a constraint, click on the constrained residue pair again. To remove all constraints, choose Clear Constraints from the Alignment menu or from the sequence shortcut menu.

You can also allow or disallow gaps in secondary structure elements, in the Pairwise Alignment Settings dialog box.

- Clear Constraints—Remove all constraints on the alignment.
- Alignment Settings—Opens the Pairwise Alignment Settings dialog box, in which you can choose the similarity matrix type, set the gap opening penalty and the gap extension penalty, choose whether to allow gaps in secondary structure elements, and create a substitution matrix from an existing alignment. This substitution matrix can then be selected as Custom from the matrix type option menu for subsequent alignments.
- Lock Gaps—Lock gaps in the alignment so that they are not filled when performing manual alignment. If you have a residue selection, the gaps are only locked in the selected region. If you insert a gap after locking the gaps, the new gap is not automatically locked. Locked gaps are shown by a dash (-); unlocked gaps are shown by a tilde (~).
- Unlock Gaps—Unlock previously locked gaps so that they can be filled.
- Select Identities—Select residues that are identical in all sequences. Gaps are ignored.
- Select Aligned Blocks—Select blocks of residues for which there are no gaps in any of the sequences.
- Select Columns with Structure—Select columns in which at least one of the residues has 3D structure (atom coordinates) associated with it. This is useful for multiple aligned structures with SEQRES regions that don't have crystal coordinates in parts of their sequences.
- Expand Selection—Expand the residue selection in the selected sequences to include the corresponding residues in all of the selected sequences. If no sequences are selected, the residue selection is expanded to cover all sequences.
- Expand Selection from Query—Expand the residue selection in the query sequence to include the corresponding residues in the selected sequences, or in all sequences if no sequences are selected.
- Hide Selected Columns—Hide the columns in which residues in all sequences are selected.
- Hide Unselected Columns—Hide columns that contain one or more unselected residues.
- Show All Columns—Show all columns for all sequences.
- Select All—Select all residues.
- Invert Selection—Invert the selection of the residues: select the unselected residues, and deselect the selected residues.

- Deselect All—Deselect all residues.
- Delete—Delete the selected residues. You can also use the keyboard shortcut, SHIFT+BACKSPACE (⇧DELETE).
- Crop—Delete the unselected residues.
- Remove Empty Columns—Remove columns that consist entirely of gaps.
- Fill with Gaps—Replace the selected residues in the selected sequences with gaps.
- Remove Gaps—Remove gaps from the residue selection in the selected sequences by shifting residues to the left. If there is no residue selection, all gaps are removed, including gaps at the beginning of the sequences.
- Track Changes—Track the regions of the sequence that have been edited, by showing an annotation that marks the edited regions. The latest edit is shown in black, and earlier edits in progressively lighter shades, up to the fifth last edit.
- Reset History—Clear the history of changes stored when changes are tracked.

1.5 Color Menu

From the Color menu, you can apply a color scheme to the sequences. The color schemes are:

- Residue Type—Color the residues by residue type. The colors used for each residue are:

ACFILMPVW	blue	(hydrophobic)
DE	red	(acidic)
HKR	green-yellow	(basic)
GNQSTY	orange	(other)

- Residue Similarity—Color residues by similarity. Identical residues are red, similar residues (positive BLOSUM62 pairwise score) are orange, other residues are white.
- Hydrophobicity (Kyte-Doolittle)—Color residues by Kyte-Doolittle hydrophobicity. Hydrophilic residues are blue, hydrophobic residues are red, residues with zero hydrophobicity are white.
- Hydrophilicity (Hopp-Woods)—Color residues by Hopp-Woods hydrophilicity. Hydrophilic residues are red, hydrophobic residues are blue, residues with zero hydrophobicity are white.
- Taylor Scheme—Color residues with the Aminochromography color scheme developed by William Taylor (*Protein Engineering* **1997**, *10*, 743). In this scheme, well conserved parts of the alignment exhibit bright, clear colors, while parts that are not well conserved have brownish, dull colors.

- Constant Color—Color all residues with a color selected from this submenu. The submenu offers a choice of twelve colors, and a Custom item that allows you to choose a color in a color selector.
- Secondary Structure—Color the sequence by the secondary structure assignment. If no SSA is available but secondary structure predictions are available, the predictions are used to color the sequence. The colors for multiple predictions are averaged, so positions where all predictions agree have bright colors, and the positions of disagreement are more gray. If no SSP nor SSA is available, the color of the sequence is not changed.
- B Factor—Color the residues by their temperature factor (PDB B factor), on a green-white-red scale, with green for the lowest values and red for the highest.
- Residue Propensities—Color residues by their propensities for various secondary structures and their chemical properties. The schemes that are available on the submenu are described in Table 3.1.
- Mark Residues—Mark the selected residues with the color chosen from this submenu. This color overrides any other color applied. To remove this color, select the residues and choose Unmark Residues from the submenu.
- Adjust Color Range—Set the limits of sequence identity used when weighting the color density by alignment quality. Opens a dialog box, in which you can set the lower and upper threshold for the sequence identity. Residues with identity below the lower threshold are colored white; residues with identity above the upper threshold have full color, and the color density for residues with identity between the two thresholds is set using a linear scale.
- Color Sequences—Show or hide the sequence coloring.
- Adjust Text Contrast—Use white for the text on residues colored with dark colors, and black for the text on residues colored with light colors. If this option is not selected, the text is black for all residues.

Table 3.1. Color schemes for residue propensities and color blocks

Scheme	Residues	Color	Description
Helix Propensity	AMLEQK	red	helix-forming
	VIFW	magenta	weak helix-forming
	CSTNDHR	gray	ambivalent
	PGY	blue	helix-breaking
Strand Propensity	VILMTFWY	blue	strand-forming
	ACSNQHR	gray	ambivalent
	DEKGP	red	strand-breaking
Turn Propensity	GSDNP	cyan	turn-forming
	AVLIMHFWC	magenta	turn-breaking
	EQTKRY	gray	ambivalent
Helix Terminators[a]	GTMRKHF	green	helix-starting
	SNDELWP	red	helix-ending
	CQAVIY	gray	ambivalent
Exposure Tendency	RNDQEHK	blue	surface
	ACGPSTWY	gray	ambiguous
	ILMFV	orange	buried
Steric Group	GACS	red	small, noninterfering
	TVNDILPM	magenta	ambiguous
	QEKR	cyan	sticky polar
	HFYW	blue	aromatic
Side-Chain Chemistry	DE	red	acidic, hydrophilic
(default scheme)	RKH	blue	basic, hydrophilic
	GAVILM	green	neutral, hydrophobic, aliphatic
	FYW	orange	neutral, hydrophobic, aromatic
	STNQ	cyan	neutral, hydrophilic
	C	yellow	primary thiol
	P	dark gray	imino acid

a. Only available on the Color Block submenu of the Annotations menu.

1.6 Annotations Menu

The Annotations menu provides a means of representing additional information associated with the sequences. There are two classes of annotations: global annotations (consensus sequence, mean hydrophobicity) and local annotations. The global annotations are calculated for the entire set of sequences, the local annotations are computed for each sequence individually. Depending on the annotation type, they are presented as histogram plots (hydrophobicity, B-factor), color bars (Color Blocks), alphanumeric strings (consensus sequence, SSP, Pfam), graphical representations (secondary structure assignments).

- Consensus Sequence—Display the consensus sequence at the top of the panel. The consensus sequence is the sequence that is composed of the most frequently occurring residue at each position in the sequence; if there are two residues that have the same frequency of occurrence, a + symbol is used, and the residues for this position are shown in its tool tip. The sequence is annotated with a histogram of the number of sequences that are represented by each residue in the consensus, with information on the percentage consensus in the tool tip.
- Consensus Symbols—Display a row at the top of the panel that contains symbols for the degree of consensus. The symbols follow the ClustalW conventions:

*	Single, fully conserved residue.
:	One of the following "strong" groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.
.	One of the following "weaker" groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY.

- Sequence Logo—Display logo annotation. In this annotation, residue symbols at each position whose frequency of occurrence at that position is greater than a threshold are drawn in a vertical stack in order of frequency, with the height of the residue symbols proportional to the frequency of occurrence. (See Schneider T.D.; Stephens R.M. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* **1990**, *18*, 6097.)
- Mean Hydrophobicity—Display histogram of Kyte-Doolittle hydrophobicity for each residue in the alignment, averaged over all sequences. Hydrophobic residues have positive values; hydrophilic residues have negative values.
- Mean Isoelectric Point—Display histogram of isoelectric points for each residue in the alignment, averaged over all sequences.
- Add Global Annotations—Add all global annotations (listed above) to the display.

- Remove Global Annotations—Remove all global annotations from the display. Does not affect sequence-dependent annotations.
- Residue Numbers—Display residue numbers and insertion codes above the sequence. The numbers are given for residue numbers that are divisible by 5, and are left-aligned to the left edge of the residue in the sequence. This is useful for tracking sequence changes after residue deletion, for example. The ruler only gives absolute alignments.
- Secondary Structure Assignment—Display secondary structure assignment for the sequence.
- B-Factor—Display histogram of temperature factors for each residue in the sequence.
- Disulfide Bonds—Display disulfide bonds as lines connecting cysteine residues, colored from black (strongest prediction) to light gray (weakest prediction).
- Hydrophobicity—Display histogram of Kyte-Doolittle hydrophobicity for each residue in the sequence. Hydrophobic residues have positive values; hydrophilic residues have negative values.
- Isoelectric Point—Display histogram of isoelectric points for each residue in the alignment.
- Ligand Contacts—Display a row in which residue positions are colored by the shortest distance between any ligand heavy atom and any heavy atom in the residue at that position. Red is used if the distance is less than 4 Å, orange is used if the distance is less than 6 Å, and gray is used otherwise. This annotation takes a little time to generate, and requires a structure for the sequence.
- Antibody CDRs—Display assignment of the three VL and VH regions. The residues are colored red, and a red line is displayed in the annotation for each region, labeled L*n* and H*n*, for *n*=1, 2, and 3.
- Select Antibody CDRs—If the sequence is an antibody, select the antibody CDRs in the sequence.
- Antibody Numbering Scheme—Choose the numbering scheme to use for the antibody residues. The choices are Chothia (*J. Mol. Biol.* **1987**, *196*, 901), Enhanced Chothia (http://www.bioinf.org.uk/abs/#martinnum), Kabat (*J. Exp. Med.* **1970**, *132*, 211), IMGT (*Nucl. Acids Res.* **1999**, *27*, 209), and AHo (*J. Mol. Biol.* **2001**, *309*, 657). This numbering scheme is only in effect when the Antibody CDRs annotation is turned on. When you export the antibody to a file, it is exported with the selected numbering scheme.
- Color Blocks—Display a single row of color blocks that are colored according to one of the residue properties described in Table 3.1. The row is presented like a sequence, but without letter codes, only the colors for the property. This submenu includes all the items

that are on the corresponding submenu of the Color menu, and also has an item to display annotations for all properties on the submenu.

- User Annotations—Add or remove customizable annotations. There are three types of user annotation offered:
  - Mark Alignment Region—Draw a rectangle over the selected part of the entire alignment (all columns). The rectangle position is fixed at the initial ruler position: it doesn't change if you change the alignment.
  - Mark Rectangular Region—Draw a rectangle over the currently selected block of residues. The rectangle is anchored to the top left residue and its position moves with this residue if you edit the alignment.
  - Add Custom Annotation—Add an annotation that you can edit to mark residues with whatever text symbols you choose. You can edit the annotation in Edit mode, and change its name from the default "Custom annotation" by right-clicking on the annotation and choosing Rename Sequence.
- Clear Annotations—Remove all annotations, leaving just the sequences. This action removes any secondary structure assignments and predictions, as well as the annotations added from this menu.

1.7 Tools Menu

The Tools menu provides access to programs for finding homologs, finding families, predicting secondary structure, and building a 3D model.

- Find Homologs (BLAST)—Run a BLAST search to find homologs for the first sequence. Opens the Blast Search Settings panel, in which you can make choices for the search and start the job.
- Show Results of BLAST Search—Show the output of the latest BLAST search, in the BLAST Search Results dialog box. The dialog box allows you to select homologs, sort homologs by one of a number of properties, download PDB structures for the homologs, and incorporate selected homologs into the project.
- Find Family (Pfam)—Run a Pfam search to find families for the selected sequences, or for all sequences if no sequences are selected.
- Predict—Run prediction programs. This submenu has the following items:
  - Run All Predictions—Run all the predictions listed below.
  - Secondary Structure—Run the secondary structure programs to obtain a prediction of the secondary structure of the selected sequences, or of all sequences if no sequences are selected. A dialog box opens that shows job progress.

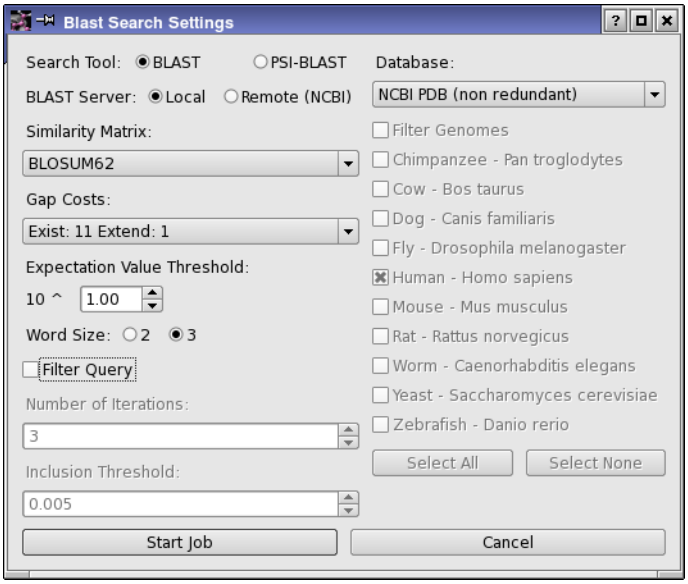
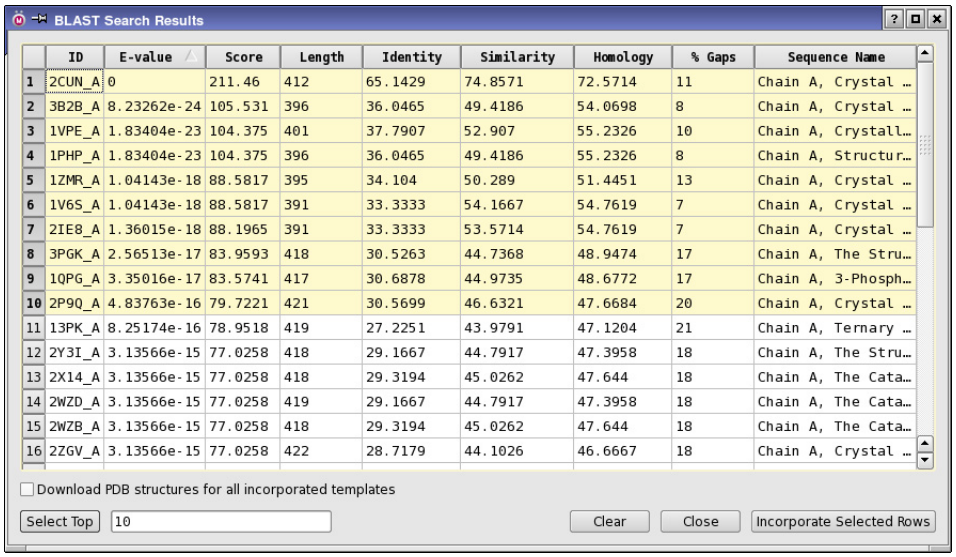


Figure 3. The Blast Search Settings panel.



	ID	E-value	Score	Length	Identity	Similarity	Homology	% Gaps	Sequence Name
1	2CUN_A	0	211.46	412	65.1429	74.8571	72.5714	11	Chain A, Crystal …
2	3B2B_A	8.23262e-24	105.531	396	36.0465	49.4186	54.0698	8	Chain A, Crystal …
3	1VPE_A	1.83404e-23	104.375	401	37.7907	52.907	55.2326	10	Chain A, Crystall…
4	1PHP_A	1.83404e-23	104.375	396	36.0465	49.4186	55.2326	8	Chain A, Structur…
5	1ZMR_A	1.04143e-18	88.5817	395	34.104	50.289	51.4451	13	Chain A, Crystal …
6	1V6S_A	1.04143e-18	88.5817	391	33.3333	54.1667	54.7619	7	Chain A, Crystal …
7	2IE8_A	1.36015e-18	88.1965	391	33.3333	53.5714	54.7619	7	Chain A, Crystal …
8	3PGK_A	2.56513e-17	83.9593	418	30.5263	44.7368	48.9474	17	Chain A, The Stru…
9	1QPG_A	3.35016e-17	83.5741	417	30.6878	44.9735	48.6772	17	Chain A, 3-Phosph…
10	2P9Q_A	4.83763e-16	79.7221	421	30.5699	46.6321	47.6684	20	Chain A, Crystal …
11	13PK_A	8.25174e-16	78.9518	419	27.2251	43.9791	47.1204	21	Chain A, Ternary …
12	2Y3I_A	3.13566e-15	77.0258	418	29.1667	44.7917	47.3958	18	Chain A, The Stru…
13	2X14_A	3.13566e-15	77.0258	418	29.3194	45.0262	47.644	18	Chain A, The Cata…
14	2WZD_A	3.13566e-15	77.0258	419	29.1667	44.7917	47.3958	18	Chain A, The Cata…
15	2WZB_A	3.13566e-15	77.0258	418	29.3194	45.0262	47.644	18	Chain A, The Cata…
16	2ZGV_A	3.13566e-15	77.0258	422	28.7179	44.1026	46.6667	18	Chain A, Crystal …

Figure 4. The Blast Search Results panel.

  - Solvent Accessibility—Predict accessibility of each residue to solvent. If more than 25% of total residue surface area is predicted to be exposed to the solvent, the residue is marked “e” (exposed, colored blue), otherwise it is marked “b” (buried, colored yellow).
  - Domain Arrangement—Predict the arrangement of domains. Residues marked gray are likely to form a domain. Residues marked red are likely to be in linker (interdomain) regions.
  - Disordered Regions—Calculate a disorder score and classify residues by this score The score is normalized to a 0 to 1 range. If a residue has a disorder score less than 0.5, it is marked light gray. If the score is between 0.5 and 0.9, the residue is marked orange. If the score is greater than 0.9, the residue is marked red.
  - Disulfide Bridges—Predict disulfide bridges between cysteines. Predicted bonds are drawn as lines connecting cysteine residues, colored from black (strongest prediction) to light gray (weakest prediction).
  - Beta Strand Contacts—Predict the contacts between beta sheets.

- Clear Predictions—Remove the secondary structure predictions.
- Build Homology Model—Build a 3D structure for the selected sequence, using the Prime Build Structure tools. The structure is built for the first sequence, and templates must be selected from among the sequences that have PDB structures. The Build Homology Model panel opens, in which you can select options for use of templates, for multimer models, and for building the structure; and then run the job. The structure is incorporated into the Maestro project when the job finishes. See Section 5 on page 35 for details.
- Analyze Binding Site—Calculate sequence identity in multiple alignment columns that are within a certain spatial distance from ligand residues. You must have a set of aligned sequences, for example from a BLAST search (Tools → Find Homologs) followed by an alignment (Alignment → Multiple Alignment or toolbar button). Choosing this menu item opens the Analyze Binding Site dialog box.

  When you open the dialog box, the query sequence is analyzed and annotated with ligand contacts, and the columns in contact with the ligand (or first ligand) are selected. Sequences that have 100% identity with the query residues in contact with the ligand (or the first ligand) are also selected.

  You can choose any sequence from the alignment and the ligand for that sequence (if any) from the option menus at the top of the panel. If no structure was imported for the sequence, then there is no ligand contact information, the Ligand menu will be empty, and the analysis cannot be done.

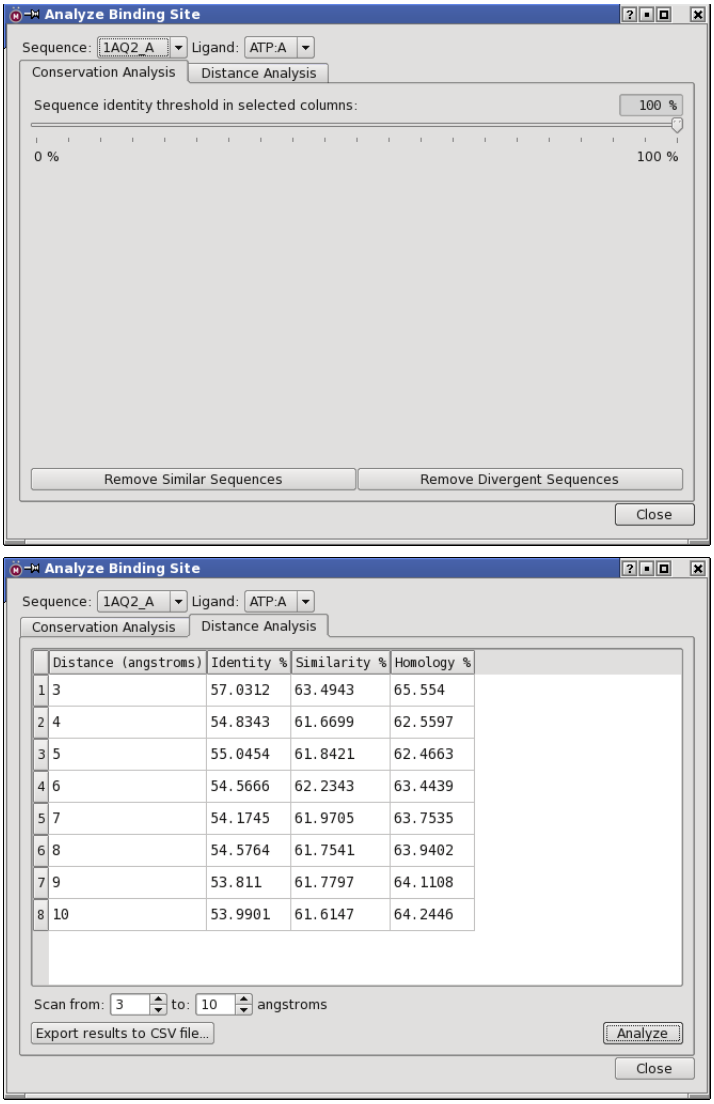


Figure 5. The Analyze Binding Site panel, Conservation Analysis tab (above), Distance Analysis tab (below.).

The dialog box has two tabs: Conservation Analysis and Distance Analysis. The Conservation Analysis tab is displayed by default. In this tab you can adjust the threshold for selecting sequences by the sequence identity in the selected columns (residues in contact with the ligand), using the slider. You can then choose to remove sequences from the alignment (and the panel) that are more similar to the query than the threshold (the

selected sequences) with the Remove Similar Sequences button, or remove sequences that differ from the query by more than the threshold (the unselected sequences) with the Remove Divergent Sequences button. These buttons allows you to prune the alignment. For example, you could choose templates for homology modeling based on binding site conservation by removing divergent sequences.

In the Distance Analysis tab, you can analyze the similarity as a function of distance. Clicking Analyze analyzes the alignment and displays the results in a table. Each row represents a neighborhood of the selected residues. The first column is the distance cutoff for that neighborhood: any residue in the query that has a heavy atom within that distance of a heavy atom in the ligand is considered to be in that neighborhood. The remaining columns contain the average identity, similarity and homology values of the aligned sequences to the selected sequence for the residues in each neighborhood. The results can be exported to a CSV file.

- Compare Sequences—Compare all sequences or the selected sequences, by identity, similarity, or homology. The percentage is displayed in a table, like a heat map, with cells color-coded by the percentage value. Opens the Compare Sequences panel, in which you can choose the comparison measure to display, switch between all or selected sequences, and refresh the display after changing the alignment.

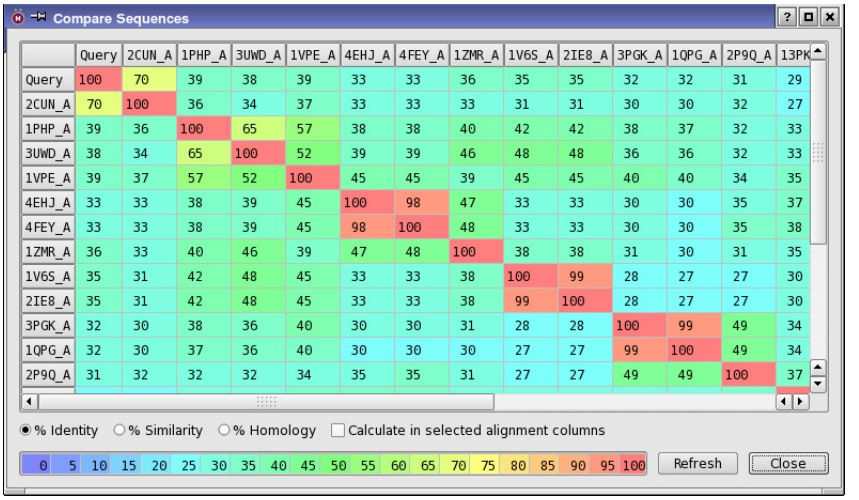


Figure 6. The Compare Sequences panel.

- Show Job Settings—Opens the Multiple Sequence Viewer Job Settings dialog box, so that you can make settings for any job that is run from the MSV.
- Show Job Log—Show the log file for the most recent job in a dialog box.

1.8 Maestro Menu

This menu provides options and actions for the interaction with the sequences and their structures as stored in the Maestro project and displayed in the Workspace.

- Incorporate Entries from Workspace—Incorporate the sequences and the corresponding structures from the entries that are in the Workspace into the MSV.
- Incorporate Selected Entries from Project Table—Incorporate the sequences and the corresponding structures from the entries that are selected in the Project Table into the MSV.
- Include Incorporated Entries in Workspace—When structures are imported into the MSV and incorporated into the Maestro project, include the structures in the Workspace.
- Associate Maestro Entries—Associate Maestro entries with sequences in the MSV without adding new sequences, modifying the alignment,or importing structures into the MSV. Opens the Associate Maestro Entries with MSV Sequences dialog box, in which you can choose an entry chain from the Workspace (list on the left) and an MSV sequence (list on the right), and click Associate Selected Pair to associate the sequence with the entry. The sequence identity for the selected pair is shown below the lists, and colored green if the identity exceeds 95% or red if it does not. The residues that do not match the entry sequence are marked as structureless in the MSV by using a less intense color. This facility is useful if the sequences are imported into the MSV and the structures are independently imported into Maestro.
- Superimpose Structures According to Sequence Alignment—Superimpose structures according to their sequence alignment. Opens the Superposition panel, with sequence identities selected as the atoms for superposition.
- Protein Structure Alignment—Run the Prime Protein Structure Alignment program on the selected (or all) protein structures and return the alignments. The sequences you select must have structures associated with them. See Section 10.3.4 of the *Maestro User Manual* for more information.
- Get Colors from Maestro Workspace—Color sequences with the colors that they have in the Maestro Workspace. The color of the alpha carbon is used to color the residues.
- Apply Colors to Workspace—Apply the colors from the MSV to the sequences and structures in Maestro.
- Color Entry Surface—Color the molecular surface in the Workspace using the colors from the corresponding sequence in the MSV. If sequences are selected, only the colors of the selected sequences are applied.
- Update from Maestro—Update the sequences in the MSV that originated from Maestro with any changes made in Maestro.

- Update Maestro Workspace Selection from MSV—Update the atom selection in the Maestro Workspace from the residue selection in the MSV.
- Update Automatically from Maestro—Automatically propagate changes made to sequences in the Maestro project to the MSV. If not selected, you can choose Synchronize with Maestro to apply changes made in Maestro to the MSV.
- Allow Structural Changes—When in Edit mode, allow mutation operations to change the structure. This choice has no effect for deletions, which are never made in the structure.

1.9 Settings Menu

The Settings menu provides access to a range of settings that affect the display of the sequences and operations on the sequences.

- Wrap Sequences—When selected, wrap the sequences so that the display consists of multiple rows of sequences, and can be scrolled vertically. When unselected, display the sequences in a single row that scrolls horizontally. Operations on unwrapped sequences are generally faster, especially when there are many sequences.
- Group Annotations by Type—When selected, group the annotations of the same type and display the groups as separate rows, below the sequences. When unselected, display the annotations for each sequence directly below the sequence.
- Font Size—Change the font size for the text in the sequence viewer. This submenu offers a selection of point sizes for the font.
- Display Ruler—Display a ruler that marks the residue positions in the sequence viewer. These positions are not the same as the residue numbers in each sequence, which can be offset from the origin and have gaps.
- Display Tooltips—Show information about the panel and its contents in tool tips (text displayed when the pointer pauses over the relevant part of the panel).
- Display Header Row—Display a heading row above the sequences with labels for each section of the viewer.
- Replace Identities with Dots—Replace residue symbols with dots for all residues that are identical to those in the query sequence. This feature makes it easy to find the mutations from the query sequence in the other sequences.
- Pad Alignment with Gaps—Add gaps to the end of each sequence so that the sequences are the same length.
- Display Sequence Boundaries—Display the residue number of the first and last visible residue in each row for each sequence. The numbers are displayed to the left and right of the sequence.

- Display Percentage Identity—Display the percentage identity with the query sequence to the right of the sequence. By default, the query sequence is the consensus sequence. You can set the query sequence by right-clicking on a sequence and choosing Set as Query from the shortcut menu. The name of the query sequence is displayed in the status area.
- Display Percentage Similarity—Display the percentage similarity to the query sequence to the right of the sequence. The similarity value is calculated by dividing the number of similar positions by the number of aligned positions. Two residues are similar if their similarity matrix value (BLOSUM62 by default) is positive. By default, the query sequence is the consensus sequence. You can set the query sequence by right-clicking on a sequence and choosing Set as Query from the shortcut menu. The name of the query sequence is displayed in the status area.
- Display Percentage Homology—Display the percentage homology to the query sequence to the right of the sequence. Homology is calculated as the percentage of residues with identical side-chain chemical properties (as defined for the Side-Chain Chemistry color scheme). By default, the query sequence is the consensus sequence. You can set the query sequence by right-clicking on a sequence and choosing Set as Query from the shortcut menu. The name of the query sequence is displayed in the status area.
- Display Score—Display the BLOSUM62 similarity score to the right of the sequence. The score is calculated relative to the query sequence.
- Include Gaps in Sequence Identity Calculations—When calculating sequence identity, count gaps as though they were residues, rather than ignoring them. For example, a column consisting of 2 different residues and 8 gaps would have a sequence identity of 20% if gaps are included, but 50% if gaps are ignored.
- Calculate Sequence Identity Only in Selected Columns—When calculating sequence identity, restrict the calculation to the columns (residues) that are selected. This allows you to calculate the identity of regions of a sequence rather than the whole sequence.
- Update Sequence Profile—Update the internal profile that is used for sequence identity and consensus calculations, and sequence coloring. If Automatically Update Sequence Profile is off, use this command to manually update the profile before doing calculations or applying coloring.
- Automatically Update Sequence Profile—Update the internal profile that is used for sequence identity and consensus calculations, and sequence coloring automatically when changes are made. Deselect this option to improve performance when you have a large number of sequences.
- Ask Before Accessing a Remote Server—Display a dialog box requesting confirmation of the action before retrieving information from a web server.

- Return to Default Settings—Reset all settings on the Settings menu to the default values.

2 The Toolbars

There are two rows of tools on the toolbars. The first row contains toolbars that only have buttons. The second row contains toolbars that have other tools besides buttons. You can show or hide any of the toolbars from the shortcut menu. The buttons are as follows:

Import Sequences
Import sequences into the MSV project from a FASTA or PDB file. Opens a file selector, in which you can navigate to and select the file. Same as File > Import Sequences.

Export Sequences
Export sequences to a FASTA file. Opens a file selector, in which you can navigate to a location and name the file. Same as File > Export Sequences.

Undo
Undo the last action. Same as Undo on the Edit menu.

Redo
Redo the last action. Same as Redo on the Edit menu.

Lock Gaps
Lock gaps in the sequences so that they are not filled when performing manual alignment. Same as Alignment > Lock Gaps.

Unlock Gaps
Unlock gaps in the sequences after they have been locked. Same as Alignment > Unlock Gaps.

Pairwise Sequence Alignment
Align multiple sequences pairwise using ClustalW. Same as Alignment > Pairwise Alignment.

Multiple Sequence Alignment
Align multiple sequences simultaneously using ClustalW. Same as Alignment > Multiple Alignment.

Color Matching Residues Only
Apply the current color scheme only to residues that are identical in all sequences (gaps are ignored

Weight Colors by Alignment Quality
Set the color density according to the sequence identity.

Average Colors in Columns
Average the colors in each column and color all residues in the column by the average color.

Zoom in
Increase the width of each residue so that the horizontal scale is expanded.

Zoom out
Decrease the width of each residue so that the horizontal scale is contracted. When the residues are narrower than the text, the text is no longer displayed. The residues can be identified by their tooltips.

Wrap Sequences
Wrap the sequences so that the display consists of multiple rows of sequences, and can be scrolled vertically. When unselected, display the sequences in a single row that scrolls horizontally. Same as Settings > Wrap Sequences.

Build Homology Model
Build a homology model of the sequence using Prime. Same as Tools > Build Homology Model.

2.1 Sequence Editing Modes

The Mode option menu allows you to select one of four sequence editing modes so that you can edit the alignment, and in some cases, edit the sequence itself.

You can lock the sequence downstream (to the right) of the residue or block that you are moving, so that the downstream part of the sequence moves as a block, without creating or removing gaps. To lock or unlock the sequence downstream, click the Lock Sequence Downstream button to the right of the Mode option menu. Locking is on by default.

You can use the Lock Gaps and Unlock Gaps toolbar buttons to prevent gaps from collapsing while editing the alignment. You can set anchors on the residue to the left and the residue to the right of the selection, so that only residues in the selection can move, by making a selection and choosing Anchor Residues Outside Selection from the alignment shortcut menu. To remove the anchors, you can click in a blank area, or choose Clear Restricted Regions from the alignment shortcut menu.

The four sequence editing modes are described below.

- Select and Slide—Use this mode to select multiple residues and slide them.
  - To select residues, drag over the residues, or shift-click the first and last residues in the range. You can drag across multiple sequences to select residues, and you can drag in the ruler to select residues in all sequences.
  - To deselect selected residues, control-click the residues.
  - To slide the selected residues, drag them to their new location.

- Grab and drag—Use this mode to drag residues to a new location. To select residues for setting anchors, use control-drag, or drag in the ruler.
- Edit—Edit the sequence or the SSPs by typing. You can mutate residues by typing in the replacement residue code. If the sequence is a Maestro sequence, this operation also mutates the residues in the structure, but you must have Allow Structural Changes selected on the Maestro menu to perform the mutations. You can also delete residues from the sequence, but the deletion is not done on the structure in Maestro, regardless of the Allow Structural Changes setting. You can change an SSP by typing in the replacement code (E, H, or –). The allowed key strokes are listed below (Mac keys are in parentheses where they differ from PC usage).
  - SPACE—insert a single gap under the cursor and move the downstream sequence one position to the right.
  - BACKSPACE (DELETE)—delete a single residue or gap to the left of the cursor and move the downstream sequence one position to the left. The cursor stays on the current residue, which moves to the left. (This is the normal text-editing behavior.)
  - SHIFT+BACKSPACE (⇧DELETE)—delete the selected residues.
  - DELETE (FN+DELETE)—delete the single residue or gap under the cursor and move the downstream sequence one position to the left. The cursor stays at the same ruler position. (This is the normal text-editing behavior.)
  - Arrow keys—move the cursor around the alignment. Using CTRL with the left and right arrow keys moves the cursor by 10 residues.
  - HOME—move the cursor to the leftmost position that is displayed in the current sequence (the left side of the sequence display). Pressing HOME again moves the cursor to the first residue in unwrapped mode, scrolling the display.
  - END—move the cursor to the rightmost position that is displayed in the current sequence (the right side of the sequence display). Pressing END again moves the cursor to the last position in unwrapped mode, scrolling the display. The cursor is not necessarily over a residue.

  You can also edit sequences as text by using the tools on the Edit menu.
- Insert and remove gaps—Insert single gaps by clicking with the left mouse button. Delete single gaps by clicking with the right mouse button. If multiple sequences are selected and you click in one of the selected sequences, the gaps are inserted or deleted in all selected sequences. If you click in a single selected sequence or in a sequence that is not selected, the gaps are inserted or deleted only in the sequence you clicked on.

2.2 Fetching Sequences

The Fetch tool allows you to fetch sequences from the Protein Data Bank or the Entrez Protein Database, from the appropriate web sites.

To fetch sequences from the PDB, type the four-character code into the Fetch text box and press ENTER. The sequence is retrieved from the RCSB web site and added to the project.

To fetch sequences from Entrez, type the access code into the Fetch text box and press ENTER. The access code format is *database*|*code*, where *database* is the code for the database (`gi` for GeneBank, `pdb` for the Protein Data Bank, `emb` for the EMBL Sequence Database, and so on), and *code* is the sequence code for the database. Examples: `gi|12345`, `pdb|2aba`, `emb|CAA44029.1`. The sequence is retrieved from the web site and added to the project.

2.3 Finding Patterns

The Find Pattern toolbar provides a tool for finding a pattern in the sequences. The pattern is merely located and highlighted, but not selected. The pattern used in the search is an extended PROSITE pattern, which has the following syntax:

sequence	Find occurrences of the specified sequence. The sequence must be given as upper case letters. For example, `AIL` finds contiguous occurrences of A, I, and L in the order given.
[*list*]	Find occurrences of any of the residues listed. For example `[AIL]` finds occurrences of A, I, or L.
{*list*}	Exclude all occurrences of any of the residues listed. For example {ED} ensures that occurrences of E and D are not found.
(m) (*m*,*n*)	Find the specified number of occurrences. (2) means find two occurrences, (2,5) means find two to five occurrences.
`a`	Find acidic residues (D and E).
`b`	Find basic residues (K and R).
`e`	Find residues in an extended region.
`f`	Find residues in a flexible region.
`h`	Find residues in a helical region.
`o`	Find hydrophobic residues (A, C, F, I, L, P, V, W, and Y).
`p`	Find aromatic residues (F, W, and Y).
`s`	Find solvent-exposed residues.

`x . ?`	Find any residue. Any of these three characters can be used.
`@`*number*	Find the residue with the specified PDB residue number (not ruler position). Insertion codes are not recognized, so all residues with a given insertion code are found.

If you want to find a simple sequence, just type in the sequence. For more complex search patterns, you can combine these elements to create patterns. There is an implied AND between each consecutive element in a sequence that is used to match a single residue. Each such sequence of elements must be separated from the next by a `-` character. The search takes place when you press ENTER. The patterns that are found are highlighted, and all other residues are colored white.

Below are some examples of valid and invalid patterns in a complex search.

`N-{P}-[ST]`	Asn followed by any residue but Pro, then Ser or Thr
`N[sf]-{P}[sf]-[ST][sf]`	as above, but all residues flexible OR solvent exposed
`Nsf-{P}sf-[ST]sf`	as above, but all residues flexible AND solvent exposed
`Ns{f}`	Asn solvent exposed AND not in flexible region
`N[s{f}]`	Asn solvent exposed OR not in flexible region
`[ab]{K}{s}f`	Acidic OR basic, with exception of Lys, flexible AND not solvent exposed
`Ahe`	Ala helical AND extended - no match possible
`A[he]`	Ala helical OR extended
`A{he}`	Ala not helical nor extended
`[ST]`	Ser OR Thr
`ST`	Ser AND Thr - no match possible

As well as typing in patterns, you can store and retrieve your own patterns from the Select Pattern option menu. The patterns are listed at the top of this option menu, with several default items, Deamidation Site, Glycosylation Site, Proteolysis Site, and Oxidation Site. The last item is Edit Patterns, which opens the Edit Patterns dialog box, in which you can change patterns, including these default patterns, and add and delete patterns. You can edit the table cells to change the pattern name, the definition, and the “hotspot”. The Hotspot column contains the residue index in the pattern that should be selected when the pattern is found. This feature is ignored by the MSV but is used in other applications that rely on the MSV.

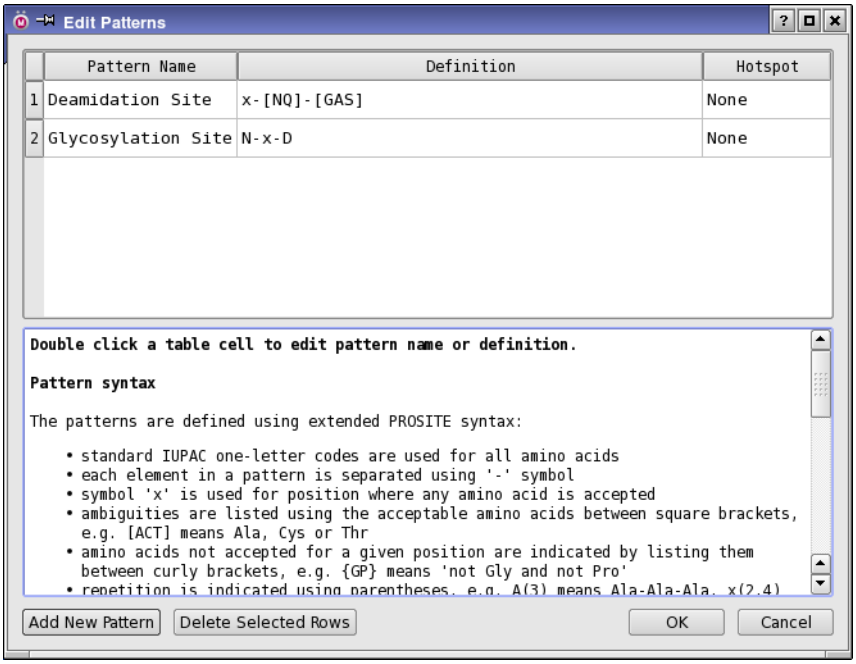


Figure 7. The Edit Patterns panel.

3 The Sequence Display Area

The sequence display area is divided into three sections.

The first section is empty until you do an alignment with ClustalW. When the alignment is done, this area contains a phylogeny tree diagram.

The second section displays the sequence names and the names of the various annotations. Selected sequences are colored slate blue. The annotations can be expanded or collapsed individually, by clicking the “tree node” (box) immediately to the left of the sequence name. They can also be expanded or collapsed globally with CTRL+DOWN ARROW and CTRL+UP ARROW. Information on the sequence is displayed in a tool tip when you pause the pointer over the sequence name.

The third section displays the sequences and their annotations, global annotations, and a ruler.

Sequences are represented by the standard residue letter symbols. Residues for which atom coordinates are missing are colored in a paler shade than residues for which atom coordinates are available.

Secondary structure predictions are represented by the characters H (helix), E (extended), and - (everything else). Secondary structure assignments are indicated by tubes for helices and arrows for extended structures.

To select sequences, click on them in the sequence name section. Use shift-click to select a contiguous range of sequences, and control-click to select or deselect individual sequences.

Selection of residues depends on the sequence editing mode—see Section 2.1 on page 27.

There are two shortcut menus. The Sequence shortcut menu opens when you right-click in the sequence name section. The Alignment shortcut menu opens when you right-click on a sequence. The tab shortcut menu opens when you right-click on the tab name.These shortcut menus are described in the following sections.

3.1 Sequence Shortcut Menu

This shortcut menu contains items from several of the main menus, and some additional items. If you right-click on a sequence when there is no selection, the sequence you click on is selected. Otherwise, right-clicking does not change the selection. The menu items are listed below, with descriptions or links to their descriptions in the main menus.

- Set As Query—Make the selected sequence the query sequence. This item is only present on the menu if there is a single sequence selected. To make the consensus sequence the query, first display it with Annotations > Consensus Sequence, then use this command to make it the query.
- Select Ligand Contacts—Select the residues that are in contact with the ligand. This is a useful way of selecting residues in a binding site. This item is present when you right-click on a Ligand Contacts annotation.
- Rename Sequence—Rename the sequence. Opens a dialog box in which you can enter a new name.
- Translate DNA / RNA sequence—Translate a DNA or RNA sequence into the sequence for the protein it codes for, using standard genetic code.
- Annotations—This submenu is a copy of the Annotations menu (Section 1.6 on page 16).
- Clear Annotations—Clear (remove) the annotations for the selected sequences.

Items that are on the Tools menu (Section 1.7 on page 18):

- Find Homologs (BLAST Search)
- Find Family (Pfam Search)
- Predict

Items that are on the Sequences menu (Section 1.3 on page 10):

- Get PDB Structures
- Select All
- Deselect All
- Hide Selection
- Show All
- Invert Selection
- Move Up
- Move Down
- Move to Top
- Move to Bottom

Items that are on the Edit menu:

- Delete Selection

3.2 Alignment Shortcut Menu

This shortcut menu contains several items from the Alignment menu, and two additional items:

- Anchor Residues Outside Selection—Set anchors on the residues outside the selection so that they do not move at all during alignment.This action prevents new gaps from being created in a sequence, but you can slide residues into existing gaps. The residues outside the selection are grayed out. To remove the anchors, click in an area outside the sequences.
- Clear Restricted Region—Remove the anchors that were set by Anchor Residues Outside Selection.

Items that are on the Alignment menu (Section 1.4 on page 11):

- Select Identities
- Select All
- Invert Selection
- Deselect All
- Delete
- Multiple Sequence Alignment

3.3 Query Shortcut Menu

This shortcut menu has two items, which are also on the Sequences menu.

- Rename Query—Rename the current query. The name appears on the tab for the query.
- Duplicate Query—Create a copy of the selected sequences of the current query in a new tab.

4 The Status Area

The status area at the foot of the panel displays information on the current task, on the sequences in the MSV project, and on the query sequence.

5 Building a Homology Model

You can build a homology model for your query, using the technology in Prime. The query can be a monomer, a homomultimer, or a heteromultimer. You can set options for building the model, preview the model, and start the job to build the model in the Build Homology Model panel, which you open by choosing Tools → Build Homology Model, or clicking the Build Homology Model toolbar button.

This feature requires a Prime license.

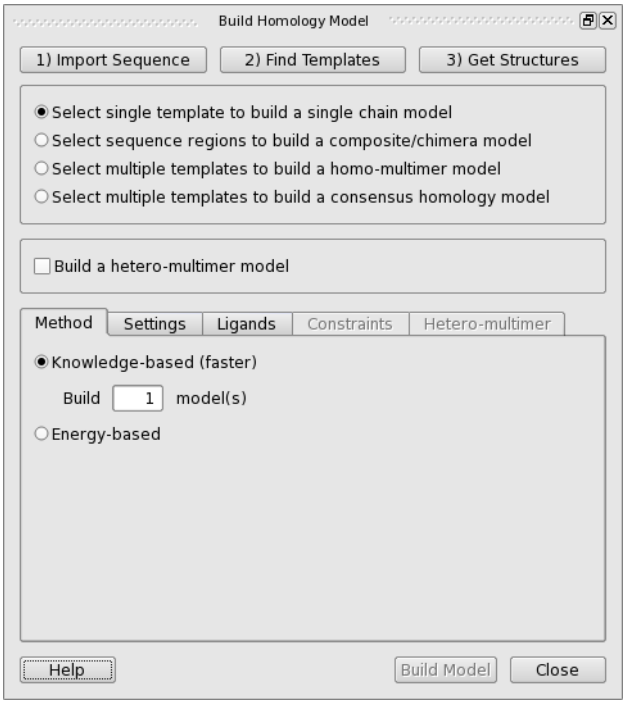


Figure 8. The Build Homology Model panel, showing the Method tab.

If you don't already have a query sequence, an alignment, and structures for the templates, you can use the three buttons at the top of the panel, Import Sequence, Find Templates, and Get Structures, to perform these tasks.

5.1 Choosing a Model Type

The first task in building the model is to choose the kind of model that is built: a single chain model from a single template, from selected regions of multiple templates, or from a consensus of multiple templates; a homo-multimer model; or a hetero-multimer model. The templates must have PDB structures associated with them before you build the model. If any of the templates do not have PDB structures, you must import the structures. You choose the model by selecting one of the following options:

- Select single template to build a single chain model—Select this option to build a model for a single chain from a single template. The template must be selected from the available sequences in the MSV.
- Select sequence regions to build a composite/chimera model—To build a model for a single chain that makes use of several templates, each in a different region of the query, select this option. To select the regions from each template, drag over the residues that you want to use in the template.

  When you select this option, the sequence section is outlined in red and the color scheme is removed. The residues used for each region are marked with a different color for each template. The templates that are used are the ones in which you select residues. By default, all residues in the first template are selected, but you can select all residues in another template by clicking the template name.

  To select the template regions, drag over the desired residues in each template that you want to use. These residues are colored to indicate that they are selected. Dragging over residue positions in a sequence that are already selected in another sequence clears the selected residues: that is, the last selection of residue positions takes precedence.
- Select multiple templates to build a homo-multimer model—Select this option to build a homo-multimer. Each template chain must be aligned to the query, and you must select all of the chains that you want to use to build the model. The structures for each template chain must be in the correct relative orientation and location to build the multimer. This usually means that the template will itself be a homo-multimer, but it could also be a hetero-multimer with a high sequence identity between the chains. You can align the chains of the template using multiple sequence alignment, or separately with Align and Merge. If the template is a heteromultimer, you could also build the homomultimer as if it was a heteromultimer.
- Select multiple templates to build a consensus model—Select this option to build a model that is a consensus of several templates. Each alpha carbon is placed at the position where the largest number of templates places that carbon. The template structures must be pre-aligned.

- Build a hetero-multimer model—Select this option if you want to build a hetero-multimer model. The Hetero-multimer tab is displayed, so you can select the queries and templates for the multimer. You must do the alignment for each query before you select the queries as templates for the multimer.

5.2 Choosing a Method, Ligands, and Constraints

Once you have selected the type of model and the templates, you can perform these tasks:

- Choose the method for building the model in the Method tab. There are two methods:
  - Knowledge-based—Construct insertions and close gaps using segments from known structures. With this option, you can choose to return multiple models of the structure. By default, only one is returned. The models are approximately ordered by the sequence identity and the incidence of clashes in the structure.
  - Energy-based—Construct and refine residues that do not come from the template based on the energy, including building through structural discontinuities and optimizing side chains.
- Select ligands for the final model in the Ligands tab, by choosing from a list of ligands. By default, only the ligands and cofactors within 5 Å of the chain used for the template are included in the list. To display the full list, select Show All.
- Set proximity constraints on the protein. Click the Constraints tab, then select pairs of query residues that you want to be in contact. The constraints are marked in an annotation row with a red curve between the residues. A harmonic constraint is added for the terminal heavy atoms of the side chains. If you specify a pair of CYS residues, a disulfide bond is formed.

  One of the constrained residues must be in a loop, since helices and strands do not have the flexibility to adjust to the constraint.

  To remove a constraint, select the two residues that are constrained. To remove all constraints, click Clear All Constraints in the Constraints tab.
- Set proximity constraints on the ligand. Click the Ligands tab and select the ligand, then click in the empty annotation row for the ligand, which is below the query, at the position for the residue you want to constrain to be in proximity to the ligand. (This annotation is only displayed when the Ligands tab is displayed: it is not displayed at the same time as the protein-protein proximity constraints annotation.) It might be useful to show ligand contacts for the template, to provide some guidance on where to place the ligand constraints on the query.

5.3 Making Settings for Building the Model

You can make settings that control how the model is built, in the Settings tab. The options that you can set are:

- Preserve the residue numbers of the template—Number the residues in the built structure the same as in the template, as far as possible. Does no apply to consensus modeling. When using multiple templates for a single chain, the residue numbers are taken from the first template. Sequential numbering from 1 is used if the attempt to use template numbering fails.
- Retain rotamers for conserved residues—Retain the rotamers for the side chains of conserved residues. These side chains will not be optimized. This option is selected by default.
- Optimize side chains—Optimize side chains, except for those attached to conserved residues if Retain rotamers for conserved residues is selected. This option is selected by default.

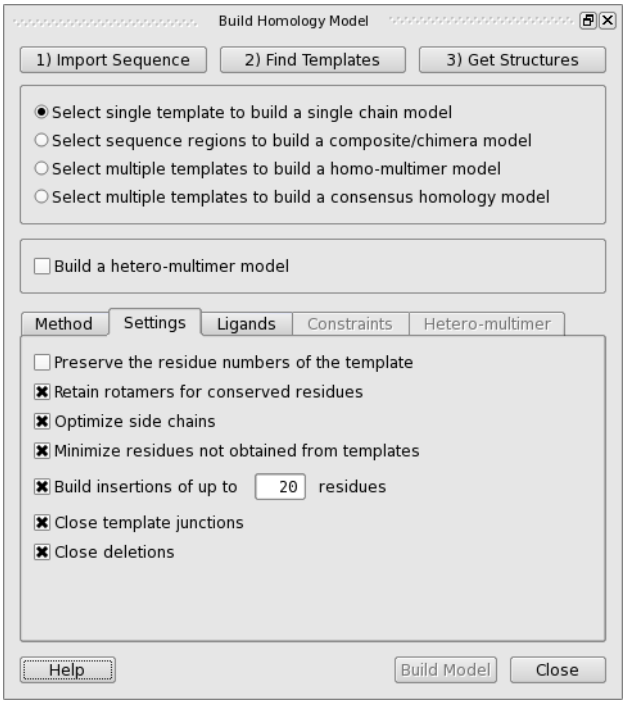


Figure 9. The Build Homology Model panel, showing the Settings tab.

- Minimize residues not obtained from templates—If side chains are optimized, also optimize the residues whose structure is not derived from any of the templates. This option is selected by default.
- Build insertions of up to *N* residues—If there is an insertion in the query (a template gap) that requires the building of a loop, build the loop if it doesn't have more than the specified number of residues, otherwise cut the query sequence and cap it with NMA and ACE. The query chain retains its name, regardless of the cuts.When the template that spans the insertion is cut to build the query, the residue on the high side of the cut (higher residue number) is also deleted, to avoid introducing steric clashes. Cutting could be useful where there are long insertions that are not in the region of interest. This option is selected by default, with a length of 20. If this option is not selected, all loops are built.
- Close template junctions—Build the structure through junctions between templates. If this option is not selected, the junctions are capped with NMA and ACE. This option is selected by default.
- Close deletions—Build the structure through deletions in the query (a query gap). If this option is not selected, the ends of the deletion are capped with NMA and ACE. This option is selected by default.

When discontinuities are capped, the final structure is discontinuous, and you should check that your final structure is reasonable.

5.4 Building the Model

When you have made all the desired settings, click Build Model to build the model. When the job finishes, the model is imported into Maestro and the sequence is included in the MSV and aligned with the query (or queries, for a hetero-multimer).

5.5 Building a Heteromultimer

If you want to build a heteromultimer, you must have queries for each chain in the multimer, and each query must be aligned to one or more appropriate templates. The template structures must also be properly positioned with respect to each other to build the structure—no translation or rotation of the individual units is performed when assembling the multimer. Usually the templates themselves are chains of a similar hetero-multimer. The process is as follows.

For each chain of the heteromultimer, do the following:

1. In the Multiple Sequence Viewer panel, import the chain as a new query.

   If some of the chains form a homomultimer, you can import only one of these chains as a query, and skip the rest.

2. Find templates for the query sequence.

   If you already know which template you want to use, you can import the template with its structure. Only the chain of the template that is used for the query sequence is needed. If you don't have a template, you can do a search to find a template.

   If the query is one chain of a homomultimeric component of the heteromultimer, you must import all the chains of the homomultimer as templates. Usually this means importing chains from a homomultimer template (or the homomultimer in a heteromultimer template).

3. Align the template (or templates) to the query.

4. Open the Build Homology Model panel and make settings as if you were building this query independently, but do not build the model.

   Making the settings is necessary so that each part of the heteromultimer is built correctly, and the model is included in the list of available queries for the heteromultimer. You should ensure that the options in the Method and Settings tabs are the same for all queries that you use.

5. Close the Build Homology Model panel.

   This is necessary so that you can add a new query in the Multiple Sequence Viewer panel.

The template structures for all queries must be properly aligned structurally to each other to build the heteromultimer. This will be the case if you use a heteromultimer as the template for your query; otherwise you must translate and rotate the template structures so that they are in the correct position.

When all the queries have been set up, you can build the heteromultimer:

1. Open the Build Homology Model panel.
2. Select Build a hetero-multimer model.
3. In the Hetero-multimer tab, select the queries used for each unit of the heteromultimer. The queries are listed by the tab name.
4. Build the model.

Getting Help

Information about Schrödinger software is available in two main places:

- The `docs` folder (directory) of your software installation, which contains HTML and PDF documentation. Index pages are available in this folder.
- The Schrödinger web site, http://www.schrodinger.com/, In particular, you can use the Knowledge Base, http://www.schrodinger.com/kb, to find current information on a range of topics, and the Known Issues page, http://www.schrodinger.com/knownissues, to find information on software issues.

Finding Information in Maestro

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

To get information:

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

  If the tooltip does not appear within a second, check that Show tooltips is selected under General → Appearance in the Preferences panel, which you can open with CTRL+, (⌘,). Not all features have tooltips.

- Click the Help button in the lower right corner of a panel or press F1, for information about a panel or the tab that is displayed in a panel. The help topic is displayed in the Help panel. The button may have text or an icon:

- Choose Help → Online Help or press CTRL+H (⌘H) to open the default help topic.
- When help is displayed in the Help panel, use the navigation links in the help topic or search the help.
- Choose Help → Documentation Index, to open a page that has links to all the documents. Click a link to open the document.

- Choose Help → Search Manuals to search the manuals. The search tab in Adobe Reader opens, and you can search across all the PDF documents. You must have Adobe Reader installed to use this feature.

For information on:

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

Contacting Technical Support

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

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

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

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

Gathering Information for Technical Support

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

For general enquiries or problems:

1. Open the Diagnostics panel.
   - **Maestro:** Help → Diagnostics
   - **Windows:** Start → All Programs → Schrodinger-2015-2 → Diagnostics
   - **Mac:** Applications → Schrodinger2015-2 → Diagnostics
   - **Command line:** `$SCHRODINGER/diagnostics`
2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.
3. Upload the file specified in the dialog box to the support web form.

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

If your job failed:

1. Open the Monitor panel, using the instructions for your profile as given below:
   - **Maestro/Jaguar/Elements:** Tasks → Monitor Jobs
   - **BioLuminate/MaterialsScience:** Tasks → Job Monitor
2. Select the failed job in the table, and click Postmortem.

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

   An archive file is created, and an information dialog box with the name and location of the file opens. You can highlight and copy the name of the file.
5. Upload the file specified in the dialog box to the support web form.

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

6. Copy and paste any log messages from the window used to start the interface or the job into the web form (or an e-mail message), or attach them as a file.
   - **Windows:** Right-click in the window and choose Select All, then press ENTER to copy the text.
   - **Mac:** Start the Console application (Applications → Utilities), filter on the application that you used to start the job (Maestro, BioLuminate, Elements), copy the text.

If Maestro failed:

1. Open the Diagnostics panel.
   - **Windows:** Start → All Programs → Schrodinger-2015-2 → Diagnostics
   - **Mac:** Applications → SchrodingerSuite2015-2 → Diagnostics
   - **Linux/command line:** `$SCHRODINGER/diagnostics`
2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.
3. Upload the file specified in the dialog box to the support web form.

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

   The files should be in the following location:

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

If a Maestro panel failed to open:

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

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®