Chimera Interface to Modeller

Chimera provides a graphical interface to running the program Modeller, either locally or via a web service hosted by the UCSF RBVI. Two types of calculations are available:

- **Comparative (homology) modeling**. Theoretical models of a protein are generated using at least one known related structure and a sequence alignment of the known and unknown structures. The protein to be modeled is the *target*, and a related known structure used for modeling is a *template*. The inputs for comparative modeling can be generated in several ways.

- **Building** parts of a protein without using a template. Missing segments can be built *de novo*, or existing segments *refined* by generating additional possible conformations. Parts that need building or refinement are often loop regions.

Modeller is developed by the Sali Lab. Use of Modeller, whether a previously downloaded copy or via web service, requires a *license key*. Academic users can register free of charge to receive a license key. (Commercial entities and government research labs, please see Modeller licensing.) Modeller users should cite:

> Comparative protein modelling by satisfaction of spatial restraints. Šali A, Blundell TL. *J Mol Biol*. 1993 Dec 5;234(3):779-815.

See also: **Build Structure**, **Rotamers**, **Multalign Viewer**, **mda**, ModBase fetch, modifying and saving data, the Comparative Modeling tutorial

← Comparative Modeling with Modeller

Comparative modeling requires a structure to serve as a template, and a target-template sequence alignment. There are several ways to obtain these inputs using Chimera. Sequence alignments are shown in **Multalign Viewer**. The Chimera interface to comparative modeling with Modeller can be started by choosing **Structure... Modeller (homology)** from the **Multalign Viewer** menu, or called with **mda**. With this interface, *only a single chain or subunit can be modeled at a time*.* Modeling a multimer or complex requires running Modeller outside of Chimera.

Choose the target (sequence to be modeled) - the name of the target sequence should be chosen from the pulldown menu of all sequences in the current alignment

Choose at least one template - at least one structure to use as a template should be chosen from the table, by clicking or dragging with the left mouse button to highlight the corresponding rows. **Ctrl**-click toggles the status of a single row. The choices are actually the sequences in the alignment, and choosing a sequence indicates using its associated structure(s). Some of the columns are blank initially, but clicking **Fetch Structures/Annotations** loads the structures into Chimera and fills in the table, where possible. The table can be sorted by the contents of any column by clicking the column header:
- **Sequence** - sequence name; choosing a sequence indicates using its associated structure as a template (the table lists all of the sequences in the alignment except the target)
- **Structure ID** - identifier for the structure associated with the sequence, if any; usually a 4-

character PDB ID, but could be a [SCOP domain ID](#) (which includes a PDB ID)
  - **%ID** - percent sequence identity of the sequence as compared to the target, computed from the alignment
  - **Title** - title of structure (from PDB entry)
  - **Organism** - source organism of structure (from PDB entry)

Clicking **Fetch Structures/Annotations**:
  1. fetches the structure for any sequence not already associated with a structure, if that structure can be deduced from the sequence name (as in the **Structure** preferences in **Multalign Viewer**)
  2. uses the structure IDs to look up additional information

Before comparative modeling with multiple templates fetched by **mda**, use the command **reset overlay** to position the template structures for best results.

Choose where to run Modeller:

  - **Run Modeller via web service**
    - **Modeller license key** - a [license key](#) is required to run the program; the [Modeller home page](#) includes links to register for a key and to download the program
  - **Run Modeller locally**
    - **Location of Modeller executable** - the location of the Modeller executable file; the [license key](#) will have been entered somewhere already during local installation
    - **Modeller script file (optional, overrides dialog)** - use the specified Modeller script to control the calculation; this will override the settings in the dialog. The script corresponding to the current dialog settings can be viewed in **IDLE** by clicking **Get Current Modeller Script**, saved to a file using the IDLE menu, and edited by hand as desired. For more details on scripting Modeller, consult the [Modeller manual](#).

Advanced Options:

  - **Number of output models [*N*] (max 1000)** (default **5**)
  - **Include non-water HETATM residues from template** (off by default) - whether to include HETATM residues other than water (ligands, ions, detergent molecules, *etc*.) from the template in the output models. This option will propagate all qualifying residues, even from multiple templates; those not desired in the output should be [deleted](#) from the template(s) beforehand.
  - **Include water molecules from template** (off by default) - whether to include water residues from the template in the output models. This option will propagate all qualifying residues, even from multiple templates; those not desired in the output should be [deleted](#) from the template(s) beforehand.
  - **Build models with hydrogens (warning: slow)** (off by default)
  - **Use fast/approximate mode (produces only one model)** (off by default) - use fast/approximate mode (~3 times faster) to get a rough idea of model appearance or to confirm that the alignment is reasonable. This mode does not randomize the starting structure (generates only a single model) and performs very little optimization of the target function.
  - **Use thorough optimization (recommended with MDA)** - optimize more thoroughly than the default, as recommended for modeling large targets with **M**ulti**D**omain **A**ssembler (the **mda** command)
  - **Temporary folder location (optional)** - use the specified location for temporary files; otherwise, a location will be generated automatically

- **Distance restraints file (optional)** - Specify an input file containing distance restraints (see example file distres.txt). Each line in the file should be of the format:

  res1 res2 dist stdev

  where *res1* and *res2* are residue numbers or ranges of residue numbers in the target sequence, *dist* is the distance in Å, and *stdev* is the standard deviation. If a single residue is specified, its Cα will be used to anchor the restraint. If a residue range (*e.g.* 233-275) is specified, the range's center of mass will be used to anchor the restraint.

OK starts the calculation and dismisses the panel, while **Apply** starts the calculation without dismissing the panel. **Close** dismisses the panel without performing any calculation. **Help** brings up this manual page in a browser window.

Running Modeller is a background task. Clicking the information icon 🛈 in the Chimera status line will bring up the **Task Panel**, in which the job can be canceled if desired.

After the calculation has finished, the comparative models are opened in Chimera and can be saved in the usual ways. The models are automatically superimposed onto the template (or the lowest-numbered of multiple templates) using **matchmaker** defaults, and the view is focused on that template. The models are associated with the target sequence, and the **RMSD** header displayed in the sequence window. Model scores are shown in a **Model List**, the same dialog used for comparative models fetched from ModBase. Loops or other parts of a model can be subjected to further refinement.

Running Modeller with identical inputs on different machines may give different (but equally valid) results, due to small numerical differences that can lead to finding different local optima of the modeling objective function.

← Building/Refinement with Modeller (Model/Refine Loops ![Model/Refine Loops]))

The only required input for Modeller building or refinement is a protein structure. Missing segments can be built *de novo*, or existing segments *refined* by generating additional possible conformations. Building and refinement can be applied to protein structures regardless of whether they were modeled or determined experimentally.

The Chimera interface to Modeller for building or refinement can be accessed by starting **Model/Refine Loops**, a tool in the **Structure Editing** category, or by choosing **Structure... Modeller (loops/refinement)** from the **Multalign Viewer** (**Sequence** tool) menu. Using **Model/Refine Loops** is equivalent to using the **Sequence** tool to show the sequence of a chain, then using its menu to show the interface. If the structure chain is associated with a sequence in an alignment in **Multalign Viewer**, however, the alignment sequence can be used instead of the individual sequence.

   The output models will include any water molecules and other HETATM residues from the input model; those not desired in the output should be deleted from the input beforehand.

- **Model/remodel** - which parts of the input model to build or refine; they will be considered collectively even if composed of multiple noncontiguous sequence segments

- **active region** - residues in the [active region](#) in **Multalign Viewer** (the region most recently drawn or clicked); this [region](#) must contain only residues from one sequence, and the model [associated](#) with that sequence will be used as the input
- **Chimera selection region** - the currently [selected](#) amino acid residues, as represented by the **Chimera selection** region in **Multalign Viewer**; this [region](#) must contain only residues from one sequence, and the model [associated](#) with that sequence will be used as the input
- **non-terminal missing structure** - *missing structure* refers to residues that appear in PDB SEQRES records but are missing from the coordinates section of a PDB file. The model of interest should be chosen from the pulldown menu on the right. By default, missing structure is indicated in the sequence with red outline boxes, corresponding to the **missing structure...** [region](#) for the model. The ***non-terminal*** missing structure consists of only the missing segments that are constrained at both ends by existing structure.

  ** Note: if the input model contains a non-terminal missing segment that the user opts not to build, inevitably Modeller will bring the existing structure on either side together to "close the gap" in the output models. **

- **all missing structure** - all missing segments, including N-terminal and C-terminal parts only constrained at one end by existing structure; all residues in the **missing structure...** [region](#) for the model chosen from the pulldown menu on the right

- **Allow this many residues adjacent to missing regions to move** (default **1**) - if one of the **missing structure** options is used, how many existing residues at each missing/existing junction to allow to move relative to the input structure
- **Number of models to generate** (default **5**)
- **Loop modeling protocol**
  - **standard** (default)
  - **DOPE** - Discrete Optimized Protein Energy score (see [Shen and Sali](#), *Protein Sci* **15**:2507 (2006)) with Lennard-Jones potential and GB/SA implicit solvent interaction; generally higher quality but more computationally expensive than the standard method, and more prone to calculation failure (resulting in fewer models than requested)
  - **DOPE-HR** - same as DOPE, except higher precision
- **Run modeller using**
  - **web service** (default)
  - **local installation**
  If web service:
  - **Modeller license key** - a [license key](#) is required to run the program; the [Modeller home page](#) includes links to register for a key and to download the program
  If local installation:
  - **Location of Modeller executable** - the location of the Modeller executable file; the [license key](#) will have been entered somewhere already during local installation
  - **Custom Modeller script file (optional, overrides dialog)** - use the specified Modeller script to control the calculation; this will override the settings in the dialog. The script corresponding to the current dialog settings can be viewed in **IDLE** by clicking **Get Current Modeller Script**, saved to a file using the IDLE menu, and edited by hand as desired. For more details on scripting Modeller, consult the [Modeller manual](#).

- **Temporary folder location (optional)** - use the specified location for temporary files; otherwise, a location will be generated automatically

OK starts the calculation and dismisses the panel, while **Apply** starts the calculation without dismissing the panel. **Close** dismisses the panel without performing any calculation. **Help** brings up this manual page in a browser window.

Running Modeller is a background task. Clicking the information icon in the Chimera status line will bring up the **Task Panel**, in which the job can be canceled if desired.

After the calculation has finished, the models (each including the unchanged parts of the protein in addition to what was built or refined) are opened in Chimera and can be saved in the usual ways. The models are associated with the sequence, and the **RMSD** header displayed in the sequence window. Model scores are shown in a **Model List**, the same dialog used for comparative models fetched from ModBase.