Quick-start guide to homology- or loop- modeling using Rosetta 3.1

This is one of a series of tutorials designed to get you started with using Rosetta 3.1. It was produced to accompany Kaufmann et. al. (2010) Biochemistry, and the latest version can be found at http://meilerlab.org/. The algorithm is described in Mandell et. al. Nat. Methods 6(8):551-2. It is assumed that you have installed the Rosetta suite from rosettacommons.org, and that you are comfortable working in Linux. You will also want to become familiar with the documentation that can be found in the /manual/ and /demos/ subdirectories, as well as the online Rosetta 3 User Manual, FAQ, and forums at http://www.rosettacommons.org/tiki/.

Comparative (or homology) modeling is the construction of a 3D model for a protein of unknown structure (the target sequence), based on a related protein of known structure (the template structure). The process can be thought of in three steps: 1) template selection (not covered here), 2) copying the backbone coordinates of corresponding amino acids from the template structure to the target sequence, and 3) dealing with the differences between the target and template by replacing side-chains and accounting for insertions and deletions. Because these typically occur in regions lacking defined secondary structure, this process is usually referred to as loop modeling.

Preparing the input files:

The loop modeling application requires as input: a pdbfile, a loops file, and a set of fragment files. There are sample files for modeling bovine pancreatic phospholipase A2 on the structure of a cobra phospholipase in the /loop_modeling/ directory included with this tutorial.

The input pdbfile

1. The input pdbfile contains the backbone structure of the portions of the target protein that can be copied directly from the template structure. This typically means elements of secondary structure- the conserved helices and strands. For historical reasons, this file is sometimes confusingly referred to as the "template" pdb. This file can be generated by a script to copy the backbone coordinates from the template to the "template" while substituting the residue identities from the target sequence. See the createTemplate.pl script in the BioTools directory of the Rosetta distribution.

The loops file

2. The loops file indicates to Rosetta which stretches of the protein are to be sampled by loop modeling. This usually comprises any regions which contain insertions deletions in the sequence alignment between template and target. The choice of start and endpoints of each loop can significantly impact the results; if a short loop is not being closed successfully, try including an additional residue or two in the sampling. Reasonable loop lengths are 4-14 residues. The loops file is a simple text file whose format is documented in the user manual.

The fragment files

3. The loops are rebuilt using a cyclic coordinate descent algorithm combined with fragment assembly. You can generate fragment files most easily using the Robetta server at http://robetta.bakerlab.org/. Making the fragments locally can also be done, by following the instructions found here:

http://www.rosettacommons.org/manuals/rosetta3_user_guide/file_fragments.html
(Note that the new Kinematic Loop Closure (q.v.) algorithm does not use fragment files.)

4. Create a flags file to specify the parameters Rosetta will use during this run, including the location and format of input and output files, and options that specify details of the algorithm. There is a sample flags file in the /loop_modeling/ directory included with this tutorial. It is worth reading the file in a text editor becuase it is well-commented and contains pointers to more documentation.
5. Conduct the folding run:

```
loopmodel.linuxgccrelease @flags > loops.log
```

Analysis and notes

1. Examine the output files generated during this loop modeling run: the logfile, the output pdb structure files, and the scorefile (loops_output.sc). These are plain text files, so you can read them with a pager or text editor. Explanations of the file formats can be found in the manual. The output structures can be compared using the Rosetta energy score, where the lowest energy indicates the best structure. Per-residue scores can be found at the end of each output structure (or decoy) in the pdbfiles, and total scores are summarized in the scorefile. Load the decoys into your favorite molecular visualization program (Chimera, PyMOL, etc) and compare them visually.