

Membrane Protein Folding

Bold text means that these files and/or this information is provided.

Italicized text means that this material will NOT be conducted during the workshop.

Fixed width text means you should type the command into your terminal.

If you want to try making files that already exist (e.g., input files), write them to a different directory!

Tutorial 1: Membrane Protein Folding with MembraneAbinitio

1. Preparation for running MembraneAbinitio

1. Prepare your working directory. You will work in this directory for the rest of the tutorial.

1. Create a directory in the protein_folding/MembraneAbinitio directory called my_files and switch to that directory.

```
cd ~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/  
mkdir my_files  
cd my_files
```

2. Save your protein sequence to a file using FASTA format.

The BRD4A.fasta file is provided for you in the

~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files directory.

1. Get the sequence in FASTA format from NCBI.

1. Go to <http://www.ncbi.nlm.nih.gov/protein/>.
2. Type in AAA72504.1 in the search bar at top.
3. Click on the “FASTA” link.
4. Open a text editor, such as gedit. Copy all the sequence information, including the line beginning with “>”, into a new file.
5. In a text editor, edit the sequence so that it corresponds to just the region we wish to model: Remove the N-term portion of the amino acid sequence up to “PIY” and the C-term portion after “GIV”.

6. Save the file as BRD4A.fasta to the MembraneAbinitio/my_files directory.

3. Generate the spanfile.

1. Go to <http://octopus.cbr.su.se/> (Sonnhammer and Krogh, Proc. Int. Conf. Intell. Syst. Mol. Biol., 1998; Viklund, et al., Bioinformatics, 2008).
2. Copy the BRD4 sequence (only the sequence!) into the provided box and click “submit.”
3. After a few minutes, it should generate a text file. Click on the OCTOPUS topology file link.

4. Copy the contents of this file into a file called BRD4A.octopus and move this file to the ~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/my_files directory.

5. Create a BRD4A.span spanfile describing the membrane-spanning regions of the four-helix bundle core of bacteriorhodopsin:

```
~/rosetta_workshop/Rosetta/main/source/src/apps/public/membrane_abinitio/octopus2span.pl\  
BRD4A.octopus > BRD4A.span
```

4. Generate the LIPS lipophilicity file <http://tanto.bioengr.uic.edu/lips/> (Adamian and Liang, BMC Struct. Biol., 2006).

- *Note: Can only generate this file if have BLAST and NR database installed. To use the LIPS server, you must have a multiple sequence alignment. We will not run this script during the workshop! See the example below:*

```
~/rosetta_workshop/Rosetta/main/source/src/apps/public/membrane_abinitio/run_lips.pl \
BRD4A.fasta BRD4A.span /blast/bin/blastpgp /nr_database \
~/rosetta_workshop/Rosetta/main/source/src/apps/public/membrane_abinitio/alignblast.pl
```

5. Prepare fragment libraries.

The BRD4 fragment files (aaBRD4A03_05.200_v1_3 and aaBRD4A09_05.200_v1_3) are already provided in the
~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files directory.

1. Make fragment libraries using Robetta (for the purposes of this workshop).
 1. If you are an academic or non-profit user of ROSETTA, make sure you're registered at <http://robetta.bakerlab.org/>.
 2. Under "Services", "Fragment Libraries" click "Submit".
 3. Type BRD4A under "Target Name".
 4. Copy/paste all the text in BRD4A.fasta into the "Paste Fasta" field.
 5. If you are benchmarking, you will want to check "Exclude Homologues".
 6. Click "Submit".
 7. You can see your position in the queue by clicking "Queue" under "Fragment Libraries". This should not take very much time unless the queue is long.
 8. When your job is finished, you should receive an email with a link called "Result Details". Click on this link. Then, click on "DOWNLOADS".
 9. Your fragment files should be called aaBRD4A_03_05.200_v1_3 and aaBRD4A_09_05.200_v1_3. However, it is wise to keep all of the files, especially the checkpoint, psidpred_ss2, and jufo_ss2 files. Right click on the files, and save the files to the data directory.
2. *Or, make fragment libraries using make_fragments.pl (not covered by this tutorial; see Appendix of the regular Abinitio tutorial).*

6. Prepare the options file.

The BRD4_mem_abrlx.options file is already provided in the
~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files
directory.

1. ROSETTA ignores lines beginning with # (these are comments.)
2. Avoid mixing tabs and spaces. Be consistent in your formatting (tab-delimited or colon-separated).

7. Prepare the rigid file.

The BRD4A_TM_rms.txt file is already provided in the
~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files directory.

1. BRD4A_TM_rms.txt is a file containing the residues over which you want to compute the CA-RMSD (the membrane-spanning regions in this case). It has the format:


```
RIGID 6 26
RIGID 31 51
RIGID 58 78
RIGID 97 117
```

8. Provide a native structure for comparison (if available).

The native file 1PY6A.pdb is already provided in the directory
~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files

For this example run, we know what the output structure should look like. In the option file we provide the native structure for comparison. Note that this doesn't change the folding protocol - the native is only used for outputting additional analysis metrics (RMSD, GDT, etc.) The native file can be omitted if one is not available.

2. Running the MembraneAbinitio application.

1. Make sure all the filenames and paths in the options file are correct!

2. Go to the `~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/my_files` directory
3. Execute the following command line.

```
~/rosetta_workshop/Rosetta/main/source/bin/membrane_abinitio2.default.linuxgccrelease\
-database ~/rosetta_workshop/Rosetta/main/database @BRD4_mem_abrlx.options
```

- NOTE: This will take 3-5 minutes per structure

3. Analyze your data.

Example data is provided for you in the

`~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/example_data/`
directory.

1. For practice, we will be analyzing data that has already been generated.

1. Create a directory for analysis of your data and switch into that directory.

```
mkdir data_analysis
cd data_analysis
```

2. Copy the files from the example_data directory.

```
cp ~/rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/example_data/* .
```

2. Combining and extracting silent files and extracting PDBs of membrane proteins is slightly different than for soluble proteins, but is a very similar process as that described in the de novo folding tutorial. To combine silent files:

```
~/rosetta_workshop/Rosetta/main/source/bin/combine_silent.default.linuxgccrelease\
-database ~/rosetta_workshop/Rosetta/main/database -in:file:silent BRD4A*.out\
-in:file:silent_struct_type binary -in:file:residue_type_set centroid\
-in:file:spanfile BRD4A.span\
-score:weights ~/rosetta_workshop/Rosetta/main/database/scoring/weights/score_membrane.wts\
-out:file:silent BRD4A_mem_abrlx_all.out -out:file:silent_struct_type binary\
-out:file:residue_type_set centroid
```

1. Find the lowest-scoring models.

```
python ~/rosetta_workshop/Rosetta/tools/protein_tools/scripts/score_scatter_plot.py\
--x_axis rms_TM --y_axis score --silent BRD4A_mem_abrlx_all.out\
BRD4A_mem_abrlx_all_score_vs_rmsd.table
```

```
sort -nk3 BRD4A_mem_abrlx_all_score_vs_rmsd.table | head -n10
```

2. To extract PDBs:

```
~/rosetta_workshop/Rosetta/main/source/bin/score_jd2.default.linuxgccrelease\
-database ~/rosetta_workshop/Rosetta/main/database\
-in:file:silent BRD4A_mem_abrlx_all.out -in:file:tags S_00000058_1\
-in:file:silent_struct_type binary -in:file:residue_type_set centroid\
-in:file:spanfile BRD4A.span\
-score:weights ~/rosetta_workshop/Rosetta/main/database/scoring/weights/score_membrane.wts\
-out:output -out:pdb -out:file:residue_type_set centroid\
```

3. Generate score vs. RMSD plots etc. The scripts to make an XY-scatter plot for membrane proteins are called `score_vs_rmsTM.R` and `score_vs_rmsTM_low.R`. These scripts are found in the `~/rosetta_workshop/tutorials/protein_folding/scripts` directory. If you want to rescore your models and compute the RMSD against the lowest-scoring model, repeat the previous step with the following changes:

```
~/rosetta_workshop/Rosetta/main/source/bin/score_jd2.default.linuxgccrelease\
-database ~/rosetta_workshop/Rosetta/main/database -in:file:native S_00000044_1_0001_0001.pdb\
-in:file:silent BRD4A_mem_abrlx_all.out -in:file:silent_struct_type binary\
-in:file:residue_type_set centroid -in:file:spanfile BRD4A.span\
-score:weights ~/rosetta_workshop/Rosetta/main/database/scoring/weights/score_membrane.wts\
-out:file:residue_type_set centroid -out:file:silent BRD4A_mem_abrlx_all_rescore.out\
-out:file:silent_struct_type binary -evaluation:rmsd NATIVE TM_low BRD4A_TM_rms.txt
```

4. *OPTIONAL: Relaxing Membrane Proteins in Rosetta (will not be covered in tutorial)*

1. *The options and command line for performing full atom refinement of a membrane protein in ROSETTA is provided in the `~rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/input_files/mem_rlx` directory. The example output can be found in `~rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/analyze_membrane_relax`.*
2. *To obtain atomic information of membrane proteins after folding them de novo, we often choose certain low-energy centroid models to refine using the relax protocol. For each input model, we usually generate at least 10, but sometimes up to 1,000 relaxed models.*
3. *Analysis of relaxed models is the same as discussed above. The when passing a scoring weights file, use `membrane_highres_Menv_smooth.wts` instead of `score_membrane.wts`.*

Tutorial 2: Membrane Protein Folding with Restraints

1. Folding membrane proteins with the topology broker with restraints

The example input and output files, as well as the command line and analysis files, can be found in `~rosetta_workshop/tutorials/protein_folding/MembraneAbinitio/membrane_broker_with_constraints`.

1. This process is very similar to folding soluble proteins with restraints except that the ROSETTAMEMBRANE-specific options, scoring weights, and spanfile are needed.
2. Based on what you've learned from previous steps, try to set up, run, and analyze the data from folding bacteriorhodopsin with simulated EPR distance restraints.