

Supplemental Figures:

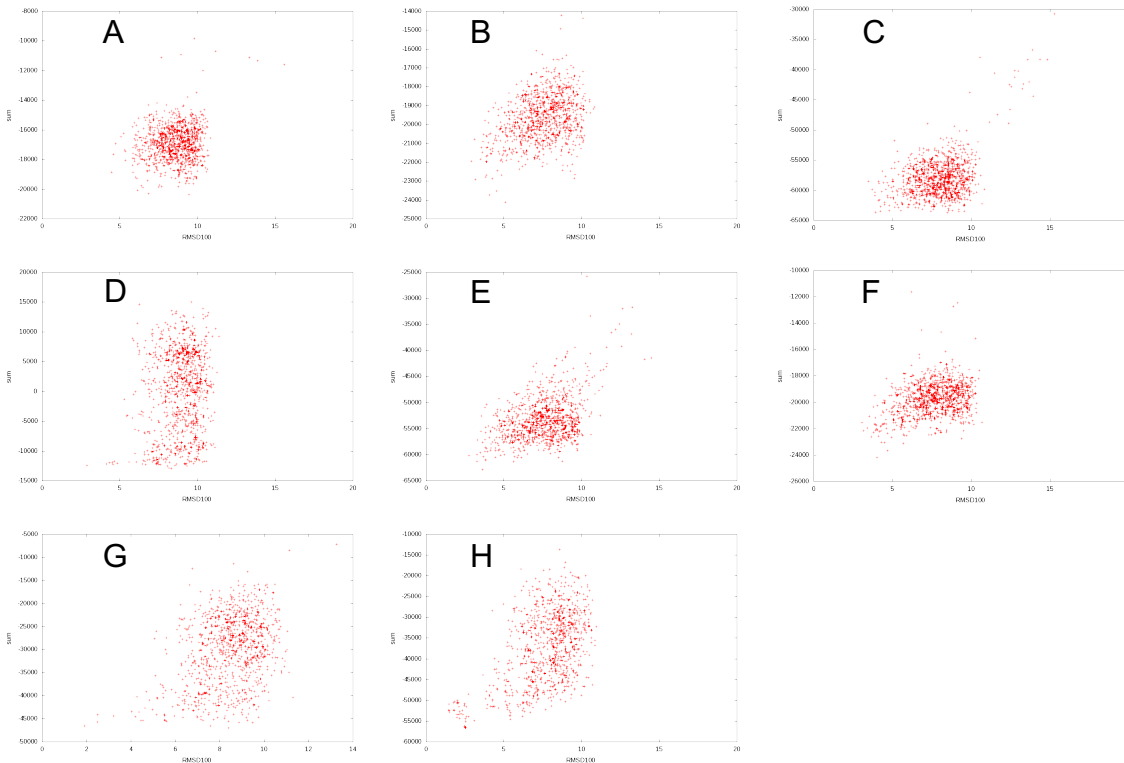


Figure S1. Related to Figure 1. Score versus RMSD100 plot for BCL::MP-Fold assembly of rhodopsin fold. (A) de novo; (B) NMR; (C) EPR; (D) EM; (E) EM_NMR; (F) NMR_EPR; (G) EM_EPR; (H) EM_NMR_EPR.

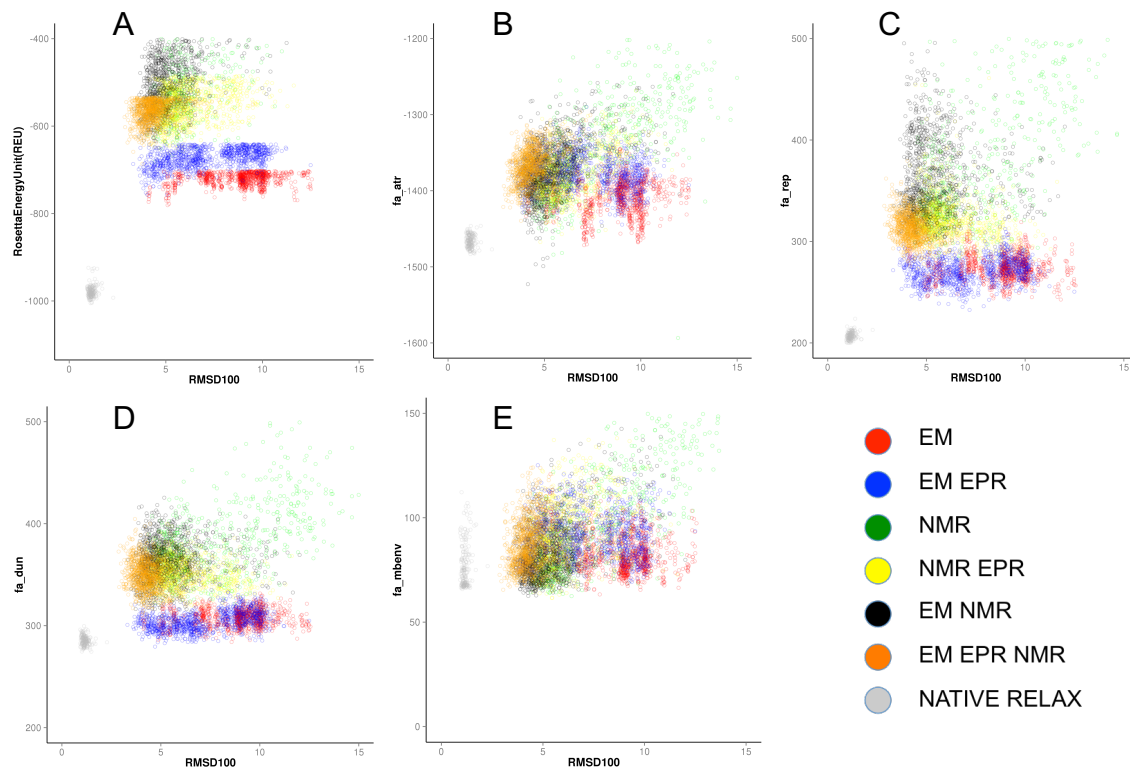


Figure S2. Related to Figure 4 and Discussion. Score versus RMSD100 in TMH for models rescored with original Rosetta membrane scores and individual score terms. fa_atr is the Lennard-Jones attractive energy between atoms in different residues; fa_rep is the Lennard-Jones repulsive energy between atoms in different residues; fa_dun is the internal energy of rotational tautomer of the sidechain, derived from Dunbrack's statistics of protein sidechains observed in pdb; fa_mbenv is the solvation energy for residues in membrane environment based on amino acids' hydrophobicity.

Restraint type	Atom 1	Residue 1	Atom 2	Residue 2	Evaluation Function	lb	ub	sd	Tag /rswitch
AtomPair	H	75	1HG1	130	BOUNDED	0	5.94786	1	NOE
AtomPair	3HD1	59	3HD2	77	BOUNDED	0	5.33137	1	NOE
AtomPair	2HD1	131	1HG1	254	BOUNDED	0	4.62996	1	NOE
AtomPair	3HD1	76	1HD1	131	BOUNDED	0	4.93543	1	NOE
AtomPair	3HG2	63	3HD2	77	BOUNDED	0	5.92617	1	NOE
AtomPair	2HD1	133	1HG1	218	BOUNDED	0	4.05534	1	NOE
AtomPair	1HD2	125	3HD2	262	BOUNDED	0	5.8083	1	NOE
AtomPair	3HG1	139	3HG2	230	BOUNDED	0	4.66986	1	NOE
AtomPair	2HG2	129	H	219	BOUNDED	0	3.95145	1	NOE
AtomPair	1HG1	130	H	156	BOUNDED	0	5.00824	1	NOE
AtomPair	3HG1	209	3HD1	214	BOUNDED	0	5.63787	1	NOE
AtomPair	H	260	1HD1	305	BOUNDED	0	4.45914	1	NOE
AtomPair	1HG2	129	H	219	BOUNDED	0	5.48106	1	NOE
AtomPair	1HD1	128	3HD1	219	BOUNDED	0	4.94971	1	NOE
AtomPair	1HD2	72	3HG2	250	BOUNDED	0	5.42105	1	NOE
AtomPair	2HD2	131	3HG2	254	BOUNDED	0	4.32513	1	NOE
AtomPair	3HD1	131	H	254	BOUNDED	0	5.35469	1	NOE
AtomPair	2HG1	139	H	230	BOUNDED	0	4.91498	1	NOE
AtomPair	2HD1	75	2HG2	130	BOUNDED	0	5.69671	1	NOE
AtomPair	1HD2	72	1HG1	250	BOUNDED	0	4.5216	1	NOE
AtomPair	H	51	1HG2	87	BOUNDED	0	4.4161	1	NOE
AtomPair	2HD2	50	1HG2	304	BOUNDED	0	4.84477	1	NOE
AtomPair	H	47	1HG2	300	BOUNDED	0	5.80309	1	NOE
AtomPair	3HD1	219	2HG2	258	BOUNDED	0	3.0854	1	NOE

Table S1. Sample NMR restraints used in Rhodopsin fold determination. Related to STAR Methods section, “Integral membrane protein structure refinement using combined experimental restraints and Rosetta”. NMR restraints were specified in a line-based file. AtomPair specifies the type of restraint to be used, followed by atom name and residue number of the first and second atom. BOUNDED term is used to set the restraint evaluation for calculating a bounded penalty. The lb and ub define the lower and upper bound of the NOE distances, while sd defines the standard deviation of distances. The tag is optional input to specify a rswitch term, when tag is not numeric, the rswitch is set to default of 0.5.

Restraint type	Atom 1	Residue 1	Atom 2	Residue 2	Evaluation Function	EPR descript	Distance value	Weight	Bin size
AtomPair	CB	5	CB	215	SPLINE	EPR_DISTANCE	31.6286	1	0.5
AtomPair	CB	13	CB	210	SPLINE	EPR_DISTANCE	32.5877	1	0.5
AtomPair	CB	35	CB	241	SPLINE	EPR_DISTANCE	56.7064	1	0.5
AtomPair	CB	39	CB	293	SPLINE	EPR_DISTANCE	9.27139	1	0.5
AtomPair	CB	41	CB	159	SPLINE	EPR_DISTANCE	35.8508	1	0.5
AtomPair	CB	42	CB	136	SPLINE	EPR_DISTANCE	41.8808	1	0.5
AtomPair	CB	57	CB	295	SPLINE	EPR_DISTANCE	22.5476	1	0.5
AtomPair	CB	58	CB	223	SPLINE	EPR_DISTANCE	25.2647	1	0.5
AtomPair	CB	70	CB	221	SPLINE	EPR_DISTANCE	22.281	1	0.5
AtomPair	CB	72	CB	128	SPLINE	EPR_DISTANCE	11.9059	1	0.5
AtomPair	CB	78	CB	299	SPLINE	EPR_DISTANCE	13.5557	1	0.5
AtomPair	CB	81	CB	263	SPLINE	EPR_DISTANCE	22.6959	1	0.5
AtomPair	CB	91	CB	165	SPLINE	EPR_DISTANCE	19.359	1	0.5
AtomPair	CB	97	CB	285	SPLINE	EPR_DISTANCE	13.7844	1	0.5
AtomPair	CB	115	CB	291	SPLINE	EPR_DISTANCE	18.7978	1	0.5
AtomPair	CB	116	CB	205	SPLINE	EPR_DISTANCE	23.0765	1	0.5
AtomPair	CB	131	CB	261	SPLINE	EPR_DISTANCE	11.0256	1	0.5
AtomPair	CB	131	CB	291	SPLINE	EPR_DISTANCE	25.1033	1	0.5
AtomPair	CB	135	CB	167	SPLINE	EPR_DISTANCE	23.5482	1	0.5
AtomPair	CB	140	CB	226	SPLINE	EPR_DISTANCE	7.02843	1	0.5
AtomPair	CB	158	CB	212	SPLINE	EPR_DISTANCE	21.0541	1	0.5
AtomPair	CB	162	CB	249	SPLINE	EPR_DISTANCE	29.4457	1	0.5
AtomPair	CB	200	CB	301	SPLINE	EPR_DISTANCE	30.2289	1	0.5
AtomPair	CB	213	CB	267	SPLINE	EPR_DISTANCE	16.8667	1	0.5
AtomPair	CB	230	CB	290	SPLINE	EPR_DISTANCE	37.2075	1	0.5
AtomPair	CB	247	CB	306	SPLINE	EPR_DISTANCE	16.5986	1	0.5
AtomPair	CB	274	CB	300	SPLINE	EPR_DISTANCE	26.0892	1	0.5

Table S2. Sample EPR restraints used in Rhodopsin fold determination. Related to STAR Methods section, “Integral membrane protein structure refinement using combined experimental restraints and Rosetta”. EPR DEER distance restraints were specified in a line-based file, AtomPair specifies the type of restraint to be used, followed by atom name and residue number of the first and second atom. SPLINE and EPR_DISTANCE would specify the program read in a histogram file and create cubic spline for the RosettaEPR knowledge-based potential. Experimental distances are set in the next column. Weight sets the numerical multiplier for the score term when linearly adding it to the total energy evaluation. Bin size set the histogram bins, in this case, distances are evaluated by a 0.5Å bin.

Method S1, related to STAR Methods

We documented the protocol for structure determination using hybrid experimental restraints using BCL::MP-Fold and Rosetta framework. command lines used for generating the test restraint sets for structure determination and model production using the two stage BCL::MP-Fold and Rosetta structure prediction suite. The following steps were taken to prepare the simulated restraints files used in each stage of the pipeline and running structural prediction with experimental data in BCL::MP-Fold and Rosetta

S	Text	Commands	Comments
1. Simulate EPR DEER distance restraints	EPR distance restraints were simulated using BCL by first predicting the optimum a.a. pairs, then simulating spin label distances with uncertainty added.	<p>1.1 bcl.exe restraint:OptimizeDataSetPairwise -fasta 1GZM.fasta -pool secondary_structure.pool -exclude_residue_types GLYCINE -restraint_distance_structures native.ls -read_mutates_start mutate_start.table -read_mutates_optimization mutate_opt.table -read_scores_optimization score_opt_bipolar.table -read_mutates_end mutate_end.table -nmodels 100 -mc_number_iterations 10000</p> <p>1.2 bcl.exe restraint:SimulateDistances -pdb 1GZM.pdb -skip_undefined_aas -simulate_distance_restraints -add_distance_uncertainty sl-cb_distances.histograms -output_file 1GZM.epr_cst_bcl -restraint_list restraint.ls</p>	<p>1.1 The OptimizeDataSetPairwise outputs the set of amino acid pairing in the protein sequence and a score for the set. Input: 1GZM.fasta #fasta file secondary_structure.pool #secondary structure pool native.ls #directory to native pdb mutate_start.table #specify number of restraint set Output: bcl.data #table of optimized set of EPR spin labeling a.a. pairs</p> <p>1.2 SimulateDistances outputs the EPR distance restraints of the given sets of a.a. pairs. Input: 1GZM.pdb #native pdb sl-cb_distances.histograms #spin label to C_{beta} distance histogram used to simulate uncertainty restraint.ls #table of a.a. pairs from step 1.1. Output: 1GZM.epr_cst_bcl #EPR distance file</p>
2. Simulate NMR restraints	Spars e NMR restraints were simulated as NOEs as 1 restraint per a.a. residue using BCL. When NMR restraints were used in folding, secondary structure information predicted from backbone chemical shift simulated by TALOS+ was also incorporated.	<p>2.1 bcl.exe restraint:SimulateDistances -pdb 1GZM.pdb -simulate_nmr_distance_restraints -num_restraint_fraction 1 -output_file 1GZM.noe_star -aaclass AAComplete -add_distance_uncertainty noe_knowledge_based.histogram</p> <p>2.2.1 chemical shift simulation were done using SPARTA+ server</p> <p>2.2.2 bcl.exe protein:CreatSSEPool -prefix 1GZM -pool_min_sse_length 9 5 999 -ssmethod TALOS -factory SSPredThreshold -join_separate</p>	<p>2.1 SimulateDistances outputs the NOE distance restraints by randomly picking 1 restraint per residue. Input: 1GZM.pdb #native pdb (protonated) noe_knowledge_based.histogram # histogram used to add NOE distance uncertainty. Output: 1GZM.noe_star #restraint in NMR-STAR 3.1 file format 2.2.1 Input: 1GZM.pdb 1GZM.fasta Output: 1GZMSS.tab 2.2.2 Input: 1GZM.fasta 1GZMSS.tab from step 2.2.1 Output: 1GZM.TALOS.pool #SSE pool for BCL folding</p>
3. Prepare input files for BCL fold	Other input files such as TM and secondary structure prediction pool were prepared using octopus and JUFO9D. Stage file were used to guide prediction with different experimental data in each BCL phases of assembly and refinement. Score weights files were adjusted to	<p>3.1.1 Perform SSE prediction using Octopus and JUFO9D</p> <p>3.1.2 bcl.exe protein:CreatSSEPool -prefix 1GZM -pool_min_sse_length 9 5 999 -ssmethod SSE_prediction_method -factory SSPredThreshold</p> <p>3.2 Generate Stage file and corresponding scoring function weight set file for each stage specified by stage file</p>	<p>3.1.1 Input: 1GZM.fasta Output: SSE prediction file such as *.jufo9d And *.octo_topo 2.1.2. Input: SSE prediction files with the same prefix '1GZM' Output: *.pool #SSE pool 3.2 See sample file format below.</p>

	account for different restraint's presence.		
4. Running BCL::MP-fold with hybrid experimental restraints	Running BCL::MP-Fold using the membrane environment and experimental restraints. The command line utilizes the files prepared in previous steps. Sections in '-restraint_types' and '-body_restraint' points to the BCL modules that could be mixed and matched to use different types of restraints.	<pre> bcl.exe protein:Fold -stages_read stages.txt -mc_temperature_fraction 0.25 0.05 -native 1GZM.pdb -quality RMSD GDT_TS -superimpose RMSD -pool_separate 1 -pool_min_sse_lengths 5 999 -sspred JUFO9D OCTOPUS TALOS -sequence_data /directory_to_SSE_prediction_data/ 1GZM -pool_prefix 1GZM -pool 1GZM.SSPredHighest_TALOS.pool -membrane -tm_helices 1GZM.SSPredHighest_OCTOPUS.pool -restraint_types NOE DistanceEPR -restraint_prefix 1GZM -body_restraint 1GZM_body.pdb 2.5 2.5 5.0 5.0 -1.0 -print_body_assignment -score_density_connectivity 1GZM.mrc -nmodels * -prefix prefix_for_output_pdbs -protein_storage /directory_to_save_models/ -random_seed </pre>	<pre> Input: stage.txt #Stage file 1GZM.pdb #native pdb 1GZM.SSPredHighest_OCTOPUS .pool #TM helices to guide membrane positioning 1GZM_body.pdb #body restraint pdb to guide EM fold movers 1GZM.mrc #EM density file Output: prefix_for_output_pdbs.pdb </pre>
5. Running Rosetta loop modeling with hybrid experimental restraints	<p>Rosetta takes the coarse-grained models produced in previous BCL stage and models their loops and atomic details.</p> <p>Several preparations steps needs to be taken to generate fragment files and restraints</p>	<pre> 5.1.1 fragment_picker.default.linuxgccrelease -database /rosetta/main/database/ -in::file::vall /rosetta/tools/fragment_tools/vall.jul19.2011.gz -frags::n_frag 200 -frags::frag_sizes 3 9 -frags::sigmoid_cs_A 2 -frags::sigmoid_cs_B 4 -out::file::frag_prefix.score -frags::describe_fragments 1GZM.fsc.score -frags::scoring::config scores.score.cfg -in::file::fasta 1GZM.fasta -in::file::talos_cs 1GZM.talos -frags::ss_pred predSS.tab talos -in::file::talos_phi_psi pred.tab 5.1.2 Prepare loop definition files 5.1.3 Prepare TM definition files /rosetta/tools/membrane_tools/octopus2span.pl 1GZM.octo_topo 5.1.4 Prepare Rosetta Format restraint files bcl.exe restraint:NmrFileConvert -pdb_file 1GZM.pdb -input_file 1GZM.noe_star star noe -output_file 1GZM_nmr.cst ROSETTA 5.2 loopmodel.linuxgccrelease -database /rosetta/main/database/ -in::file::s 1GZM_bcl.pdb -loops:loop_file 1GZM.loops -in::file::native 1GZM.pdb -evaluation:rmsd NATIVE_FULL FULL -evaluation:gdtmm -loops::frag_sizes 9 3 1 </pre>	<pre> 5.1.1 Input: vall database installed in rosetta 1GZM.fasta 1GZM.talos #chemical shift predicted from Sparta+ Output: aa1GZM09_05.200_v1_3 #3mer fragments aa1GZM09_05.200_v1_9 #9mer fragments 5.1.2 example loop file format: LOOP 1 22 0 0.0 0 #1st and 2nd column indicates start and end of loop residues to be modeled 5.1.3 Input: 1GZM.octo_topo #from step 3.1.1 Output: 1GZM.span 5.1.4 Input: 1GZM.noe_star #step 2.1 Output: 1GZM_nmr.cst # restraints in Rosetta format 5.2 Input: 1GZM_bcl.pdb #bcl models that will be further refined by rosetta 1GZM.loops #step 5.1.2 1GZM.pdb #native pdb aa1GZM09_05.200_v1_3 #step 5.1.1 aa1GZM09_05.200_v1_9 #step 5.1.1 1GZM_centroid_restraint.cst #step 5.1.4 1GZM_full_atom.cst #step 5.1.4 1GZM.mrc #EM density file </pre>

		<pre> -loops::frag_files aa1GZM09_05.200_v1_3 aa1GZM03_05.200_v1_3 none -loops::remodel quick_ccd -loops::intermedrelax no -loops::refine refine_ccd -loops::relax fastrelax -ex1 -ex2 -constraints:cst_file 1GZM_centroid_restraint.cst -constraints:cst_weight 5.0 -constraints:cst_fa_file 1GZM_full_atom.cst -constraints:cst_fa_weight 5 -constraints:epr_distance -constraints:viol -constraints:viol_level 101 -edensity::mapfile 1GZM.mrc -edensity::sliding_window 9 -edensity::mapreso 5.5 -edensity::grid_spacing 4.0 -whole_structure_allatom_wt 0.1 -score:weights membrane_highres_Menv_smooth.wts -membrane:no_interpolate_Mpair -membrane:Menv_penalties -in.file:spanfile 1GZM.span -out:pdb -out:output -out:file:scorefile score_file.sc -out:nstruct * -out:prefix /directory_to_save_models/prefix </pre>	<pre> 1GZM.span #step 5.1.3 Output: score_file.sc #score file /directory_to_save_models/prefix.p db #models produced by loopmodeling </pre>
<p>6. Running Rosetta refinement with hybrid experimental restraints</p>	<p>A final refinement step was used to relax the predicted model to their energy minimum using the Rosetta relax application. Top scoring models from loop modeling were taken</p>	<pre> 6 relax.linuxgccrelease -database /rosetta/main/database/ -in.file:s 1GZM_loopmodeled.pdb -in.file:fullatom -in.file:native 1GZM.pdb -evaluation:rmsd NATIVE_FULL FULL -evaluation:gdtmm -relax:fast -relax:membrane -ex1 -ex2 -constraints:cst_fa_file 1GZM_full_atom.cst -constraints:cst_fa_weight 5 -constraints:epr_distance -constraints:viol -constraints:viol_level 101 -edensity::mapfile 1GZM.mrc -edensity::sliding_window 9 -edensity::mapreso 5.5 -edensity::grid_spacing 4.0 -whole_structure_allatom_wt 0.1 -score:weights membrane_highres_Menv_smooth.wts -membrane:no_interpolate_Mpair -membrane:Menv_penalties -in.file:spanfile 1GZM.span -out:pdb -out:output -out:file:scorefile score_file.sc -out:nstruct * -out:prefix /directory_to_save_models/prefix </pre>	<pre> 6 Input: 1GZM_loopmodeled.pdb #best scoring models after loop modeling 1GZM.pdb #native pdb aa1GZM09_05.200_v1_3 #step 5.1.1 aa1GZM09_05.200_v1_9 #step 5.1.1 1GZM_full_atom.cst #step 5.1.4 1GZM.mrc #EM density file 1GZM.span #step 5.1.3 Output: score_file.sc #score file /directory_to_save_models/prefix.p db #models produced by relax </pre>

Example BCL Stage file format:

The BCL stage files sets the parameters for BCL::MP-Fold for the specific restraints used in structure prediction. In the example file, SCORE_PROTOCOLS specifies the membrane environment and the used of distance restraint and EM density; MUTATE_PROTOCOLS specifies the particular mutate moves in the program that are tailored to efficiently sample protein conformations for the given environment.

```
NUMBER_CYCLES 1
STAGE Stage_assembly_1
    SCORE_PROTOCOLS Default Membrane Restraint EM
    SCORE_WEIGHTSET_FILE assembly_1.scoreweights
    MUTATE_PROTOCOLS Default Assembly Membrane Restraint EM
    NUMBER_ITERATIONS 2000 400
STAGE_END
STAGE Stage_assembly_2
    SCORE_PROTOCOLS Default Membrane Restraint EM
    SCORE_WEIGHTSET_FILE assembly_2.scoreweights
    MUTATE_PROTOCOLS Default Assembly Membrane Restraint EM
    NUMBER_ITERATIONS 2000 400
STAGE_END
STAGE Stage_assembly_3
    SCORE_PROTOCOLS Default Membrane Restraint EM
    SCORE_WEIGHTSET_FILE assembly_3.scoreweights
    MUTATE_PROTOCOLS Default Assembly Membrane Restraint EM
    NUMBER_ITERATIONS 2000 400
STAGE_END
STAGE Stage_assembly_4
    SCORE_PROTOCOLS Default Membrane Restraint EM
    SCORE_WEIGHTSET_FILE assembly_4.scoreweights
    MUTATE_PROTOCOLS Default Assembly Membrane Restraint EM
    NUMBER_ITERATIONS 2000 400
STAGE_END
STAGE Stage_refinement_1
    SCORE_PROTOCOLS Default Membrane Restraint EM
    SCORE_WEIGHTSET_FILE refinement_1.scoreweights
    MUTATE_PROTOCOLS Default Refinement Membrane Restraint EM
    NUMBER_ITERATIONS 2000 400
    PRINT_END_MODEL true
STAGE_END
```

