Modeling of G Protein-Coupled Receptors with Rosetta

17th Annual Great Lakes GPCR Retreat
Rocco Moretti
Brian Bender

VANDERBILT UNIVERSITY
Center for Structural Biology and Institute of Chemical Biology
Departments of Chemistry, Pharmacology, and Biomedical Informatics
Computational Structural and Chemical Biology in the Meiler Lab

Protein structure prediction de novo and from limited experimental data

Merging ligand- and structure-based computer-aided drug discovery

Design of large protein scaffolds, antibodies, and protein/ligand interfaces

Weiner, et al., *Structure*, 2013, 21(7)

Dong, et al., *PLoS ONE*, 2013, 8(7)

Alexander et al., *NSMB*, 2014, 21(1)

Muller, et al., *ChemMedChem*, 2012, 7(3)


Fortenberry, et al., *JACS*, 2011, 133(45)
Acknowledgements – [http://meilerlab.org](http://meilerlab.org)

VU Collaborators: Carlos Arteaga, Jim Crowe, Heidi Hamm, Randy Blakely, Jeff Conn, Christine Lovly, Craig Lindsley, Seva Gurevich, Hassane Mchaourab, Dave Weaver, Ambra Pozzi, Chuck Sanders, Roger Cone, Charles Manning, Steve Fesik

Active: NIH NIGMS R01GM080403, R01GM073151, NIH NIDDK R01DK097376, NSF CHE 1305874, NIH NIAID U19 AI117905, NIAID HHSN272201400024C, NIH NILBI R01HL122010, NIH NCI P50 CA098131

Completed: NSF MCB0742762 (CAREER), NIH NIGMS R01GM099842, NIH NIMH R21MH082254, R01MH090192, DTRA HDTRA1-10-0067


Jessica Finn
John Lally
Karuna Gujar
Marcin Skwark
Marion Sauer
Michaela Fooksa
Sandeep Kothiwale
Soumya Ganguly
Sten Heinze
Steven Combs
Rocco Moretti
Heather Darling
Jeff Mendenhall
Yan Xia
Xuan Zhang
Live Demo Files (for later) @ http://meilerlab.org/index.php/rosetta-tutorials

- www.meilerlab.org
- Resources tab
  - Rosetta Tutorials
Rosetta: A Unified Framework for Tackling Molecular Modeling

Rosetta consists of multiple modules: protein folding, comparative modeling, ligand docking, protein design, antibody/antigen interactions, etc.

- Rosetta is developed in a consortium of 23 laboratories by over 150 developers
- **PyRosetta** is a python interface that allows integration with Pymol
- **FoldIt** is the better video game for you and your kids
- **Rosetta@home** uses your computer for our research

**Rosetta is free for academic use; user guide/tutorials available at www.rosettacommmons.org**

Sampling and Scoring for Protein Folding Simulation

- **Local Sequence Bias**
  - Approximate local interactions using the distribution of conformations seen for similar sequences in known protein structures

- **Monte Carlo simulations**
  - Select broadest minima using cluster analysis

- **Energy evaluation of non-local interactions using knowledge-based energy function**
  - Steric overlap
  - Residue environment
  - Pair wise interactions
  - Strand pairing
  - Compactness
  - Secondary Structure Packing

Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

Lennard-Jones Potential

$$\sum_{i<j} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

van der Waals Energy

Ramachandran Plot

Statistical mining of Protein Databank (PDB)
Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

\[ Energy = w_1 \cdot \text{term}_1 + w_2 \cdot \text{term}_2 + w_3 \cdot \text{term}_3 + \ldots \]
Monte Carlo Selection allows rapid sampling

\[ E_{tot} = \sum_i w_i T_i \]
Criteria for accepting structures (Metropolis Criterion):
- If $E_{\text{new}} < E_{\text{old}}$ : Accept new structure
- If $E_{\text{new}} > E_{\text{old}}$
  - Pick a random number $p(0, 1)$
  - if $e^{\frac{-E_{\text{new}} - E_{\text{old}}}{kBT}} > p$, accept new structure

Score is Central to Monte Carlo Selection
While not every protein fold is present in the protein databank, all possible conformations of small peptides are!

Approximate local interactions using the distribution of conformations seen for similar sequences in known protein structures

For each sequence window, select fragments that represent the conformations sampled during folding
Native-like Protein Models Form Large Clusters

- The free energy minimum corresponds (usually) to the native protein fold
- Its depth is obscured because of the simplified energy approximation
- However, the width of the funnel leading to the free energy minimum of the native protein fold is well preserved
What can Rosetta actually fold?

- Small, globular, soluble proteins
  - T4-lysozyme C-terminal domain
- Small, simple membrane proteins
  - V-type Na\(^+\) ATP synthase subunit
- ...but not large, complex proteins
  - Rhodopsin
Combining Strengths: Building Accurate Models from Limited Data

- Complete Conformational Space
- Efficient Sampling Strategy
- Accurate Energy Function

Protein Structures consistent with sparse experimental data
Comparative modeling & docking of GPCRs

1) selection of template structure

2) sequence alignment of the target to the template

3) thread target sequence onto the template PDB file

4) create fragments derived from PDB

5) comparative modeling: Hybridize: RosettaCM

Comparative modeling & docking of GPCRs

6) selection of comparative model for ligand docking

7) generate conformations of the ligand

8) ligand docking via coarse/fine sampling & all-atom refinement

9) data analysis and selection of final model

Rosetta Captures the TM Region with High Accuracy

- TM RMSD of most Class A receptors falls between 1-2 Å
- Classes B, C, and F suffer from lack of template structures
Rosetta Recovers Native Loop Structures in Comparative Models

D3R

ECL1 (7-8 residues) 0.9 Å RMSD

ECL2 (17-34 residues) 1.2 Å RMSD

ECL3 (8-10 residues) 0.6 Å RMSD

β₂Adrenergic

0.4 Å RMSD

2.1 Å RMSD

0.5 Å RMSD

© Jens Meiler
RosettaLigand Docking Recovers Native Ligand Binding Poses

- Retinal docked into rhodopsin (3.7Å RMSD)
- Carazolol docked into β₂AR (1.7Å RMSD)
- Cyanopindolol docked into β₁AR (1.9Å RMSD)
- 241385 docked into A2A denosine (2.5Å RMSD)
- IT1t docked into CXCR4 (3.7Å RMSD)
- Eticlopride docked dopamine D3 (2.6Å RMSD)
Comparative Models Aid in Drug Discovery - Docking mGlu5 PAMs
Rosetta Binding Energy Correlates with Binding Affinity

H. Wu, et al.; "Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator "; Science; 2013
Human Smoothened Receptor/LY-2940680

- <15% sequence identity with RMSD 3Å
- Long extracellular loops (up to 39 residues) with unique conformations
- Unique extracellular domain linker
- Ligand binding pocket has no resemblance to existing structures
Most Accurate Prediction of hSMO/ LY-2940680 Complex

- 88 submitted models
- Ligand RMSD 4.42Å
- Correct contacts 8.8%
- Correct prediction of the helical fold and disulfide bond in ECL3
Iterative feedback loop between experiment and model

- Build comparative model
- Dock ligand and propose residues important for interaction
- Experimental verification
Modeling of GPCR Dimers

H. Wu, et al.; "Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator"; Science; 2013
Xavier Pedragosa-Badia et al., “Pancreatic polypeptide is recognized by two domains of the human Y4 receptor”
G protein cycle: Formation of the G protein | R* complex

Overall structure of the b2AR|Gs complex omitting crystallization aids

Experimental Distance Measurements Show Domain Opening in Active State

Modeling of Inactive and Active G protein states using EPR Data as Conformational Ensembles
Single particle EM reconstructions of the β2AR|Gs complex

Modeling Reveals Energetics of R* | Gai interface in the R*-Gai complex
Energetic basis of signal transduction during Gai interaction with receptor R*
Basal and receptor-mediated nucleo-tide exchange rates match model.
Conclusions

- Rosetta is a suite of molecular modeling software that utilizes fragment-based assembly and knowledge-based scoring potentials to rapidly sample conformational space and derive low energy conformational states of proteins.

- GPCR comparative modeling can accurately model receptors in the absence of crystal structure.

- Ligand docking into comparative models is possible.

- All methods are increased in accuracy by the use of sparse experimental data.

- With an increasing number of templates, computational methods become more accurate and will play an increasing role in therapeutic development.
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model
First description of multiple template homology modeling in Rosetta

High-Resolution Comparative Modeling with RosettaCM

Yifan Song,1,3 Frank DiMaio,1,3 Ray Yu-Ruei Wang,1 David Kim,1 Chris Miles,1 TJ Brunetta,1 James Thompson,1 and David Baker1,2,5
1Department of Biochemistry, University of Washington, Seattle, WA 98195, USA
2Howard Hughes Medical Institute, University of Washington, Box 357370, Seattle, WA 98195, USA
3These authors contributed equally to this work
*Correspondence: ddbaker@u.washington.edu
http://dx.doi.org/10.1016/j.str.2013.08.005

Application of RosettaCM to modeling GPCRs (basis of workshop)

Pancreatic Polypeptide Is Recognized by Two Hydrophobic Domains of the Human Y4 Receptor Binding Pocket*5

Xavier Pedragosa-Badia5, Gregory R. Sliwoski5, Elizabeth Dong Nguyen5, Diana Lindner5, Jan Stichel5, Kristian W. Kaufmann5, Jens Meiler5, and Annette G. Beck-Sickenger5

From the 1Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, Universität Leipzig, 04103 Leipzig, Germany and the 5Center for Structural Biology, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8725
Updates to Rosetta’s Comparative Modeling Method

- **Single Template Modeling:**
  - Single template
  - Thread single backbone as input
  - Use fragments
  - Extra step of Loop Modeling
    - Provide Loop file definitions

- **Multiple Template Modeling:**
  - Multiple templates
  - Thread multiple backbones as input
  - Uses sections of multiple threaded models + uses fragments
  - Loop modeling protocol is internal
  - Need final Dualspace Relax in Cartesian coordinate space
Multiple Templates Increase Model Accuracy Over Single Template

Large majority of models show an improvement when using multiple templates compared to single template.
Necessary Files for GPCR Comparative Modeling

User Input
- Target sequence
- Template PDBs
- Fragment files
- Membrane topology file
- Disulfide (other constraints) file

Organizational
- Options file
- RosettaScripts XML

Predefined
- Score function weights files
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1**: Align TARGET sequence with sequence of template structure
- **Step 2**: Thread the target sequence onto the backbone of the template structure
- **Step 3**: Hybridize template segments and fragments from PDB to generate full length models
- **Step 4**: Dock ligand into comparative model
• Need to obtain fasta for target sequence (NCBI Protein)
• Often our main interest is modeling of transmembrane region so we will choose not to model extra regions
  • Long unstructured termini
  • Class A GPCRs can have intracellular loops >100 residues
    • If deleting loops, need to replace with shorter poly-G or poly-A stretches to have continuous sequence

Target Sequence

> DRD3

**MASLSQLSSHLNTCGAENSTGASQARPHAY** YALSYCALILAI VFGNGLVCMAVLK ERALQTITNYLAVS HAVADLLL VAVL VMPWVV YLEV TGGVWNSHICCDVFVTLDVM MCTASILNLCAISIDRYTAVVMPVHQHTGQSSCR RVALMITAVWVL AFASCPL LF GFTTDPTVCSISNPDFVIYSSVV SYLPFGVT VLVYARIYEV LKQRR RK **RIL** TRQN SQCN SVRPG FPQQTLSPDPAHLEL KRYSSICQDTALGGPGFQER GGELKREE KTRN SLSTPIAPKL SLEVRKLSNRGLSTSLKLGGPQPR GVPLREKKATQMVAIVLGA IVCWLPFFLTTHVLNTHCQ TCHVSP ELYSAT TWLG YVNSAL NPVI TT FNIE FRK AFLKILSC

> DRD3 Modified

YALSYCALILAI VFGNGLVCMAVLK KERALQTITNYLAVS HAVADLLL VAVLVMPWVV YLEV TGGVWNSHICCDVFVTLD VMMCTASILNLCAISIDRYTAVVMPVHQHTG QSSCCR VALMITAVWVL AFASCPL LF GFTTDPTVCSISNPDFVIYSSVV SYLPFGVT VLVYARIYEV LKQRR RK GVPLREKKATQMVAIVLGA IVCWLPFFLTTHVLNTHCQ TCHVSP ELYSAT TWLG YVNSAL NPVI TT FNIE FRK AFLKILSC
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1**: Align target sequence with sequence of **TEMPLATE** structure
- **Step 2**: Thread the target sequence onto the backbone of the template structure
- **Step 3**: Hybridize template segments and fragments from PDB to generate full length models
- **Step 4**: Dock ligand into comparative model
**Identifying Template Structures**

- **Sequence Similarity**: compare proteins based on amino acid properties alone (BLAST, PSI-BLAST)
  - Suitable Templates: ideally have >30% sequence identity to target

- **Fold Recognition**: using predicted secondary structure information to detect proteins with similar 3D characteristics (DALI, PHYRE)
GPCR Structure Determination Yields Many Structural Templates for Comparative Modeling
Current available crystal structures share identities often below 30%

Emphasizes need for multiple template modeling
### GPCR Template Selection

Within subfamilies identities increase.
## List of Templates Ranked by Sequence Identity

<table>
<thead>
<tr>
<th>Receptor</th>
<th>PDB ID</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HT-1B</td>
<td>4iar</td>
<td>39</td>
</tr>
<tr>
<td>B1AR</td>
<td>4bvn</td>
<td>38</td>
</tr>
<tr>
<td>B2AR</td>
<td>2rh1</td>
<td>36</td>
</tr>
<tr>
<td>M4R</td>
<td>5dsg</td>
<td>35</td>
</tr>
<tr>
<td>M1R</td>
<td>5cvv</td>
<td>34</td>
</tr>
<tr>
<td>5HT-2B</td>
<td>4lb4</td>
<td>34</td>
</tr>
<tr>
<td>H1R</td>
<td>3rze</td>
<td>34</td>
</tr>
<tr>
<td>M2R</td>
<td>3uon</td>
<td>33</td>
</tr>
<tr>
<td>M3R</td>
<td>4u15</td>
<td>31</td>
</tr>
<tr>
<td>A2AAR</td>
<td>4eiy</td>
<td>30</td>
</tr>
<tr>
<td>OX2R</td>
<td>4s0v</td>
<td>28</td>
</tr>
<tr>
<td>LPA1</td>
<td>4z35</td>
<td>28</td>
</tr>
<tr>
<td>MOR</td>
<td>4dkl</td>
<td>27</td>
</tr>
<tr>
<td>NOP</td>
<td>5dhg</td>
<td>26</td>
</tr>
<tr>
<td>S1P1R</td>
<td>3v2y</td>
<td>26</td>
</tr>
<tr>
<td>DOR</td>
<td>4n6h</td>
<td>26</td>
</tr>
<tr>
<td>OX1R</td>
<td>4jzc</td>
<td>26</td>
</tr>
<tr>
<td>NTSR1</td>
<td>4xes</td>
<td>25</td>
</tr>
<tr>
<td>Rho</td>
<td>1u19</td>
<td>25</td>
</tr>
<tr>
<td>CXCR4</td>
<td>3odu</td>
<td>25</td>
</tr>
<tr>
<td>KOR</td>
<td>4djh</td>
<td>25</td>
</tr>
<tr>
<td>AT1R</td>
<td>4zud</td>
<td>23</td>
</tr>
<tr>
<td>US28</td>
<td>4xt1</td>
<td>22</td>
</tr>
<tr>
<td>P2Y1</td>
<td>4xnw</td>
<td>22</td>
</tr>
<tr>
<td>CCR5</td>
<td>4mbx</td>
<td>22</td>
</tr>
<tr>
<td>P2Y12</td>
<td>4pxz</td>
<td>22</td>
</tr>
<tr>
<td>PAR1</td>
<td>3vw7</td>
<td>21</td>
</tr>
<tr>
<td>FFAR1</td>
<td>4phu</td>
<td>18</td>
</tr>
<tr>
<td>GCGR</td>
<td>5ee7</td>
<td>14</td>
</tr>
<tr>
<td>SMO</td>
<td>4jkv</td>
<td>13</td>
</tr>
<tr>
<td>GRM1</td>
<td>4or2</td>
<td>12</td>
</tr>
<tr>
<td>CRF1R</td>
<td>4k5y</td>
<td>12</td>
</tr>
<tr>
<td>GRM5</td>
<td>5cgd</td>
<td>11</td>
</tr>
</tbody>
</table>

- Search query in blastp was D3R sequence
- D3R is a Class A Subclass α Receptor
- Top hits include other A α receptors
  - Adrenergic
  - Serotonin
  - Muscarinic
- Top hits are >30% identical
- Template PDBs are easily obtained from RSCB
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** ALIGN target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model

D3R 1 MCFSVSLATVALGCMFVPKVVII
B2AR 1 KEVYILLNWIGYVNSGFNPLITCR
Multiple Sequence Alignment of Target with Template Sequences

- Obtain template FASTAs from RSCB or pull from PDB file
- Remove fusion proteins
- Generate multiple sequence alignment (Clustal, MUSCLE) including target and templates

```
1u19        ----PWQFSMLAAYMFLLIMLGFPINFTLYVTQHKKLRTPLNYILLNLAVADLFM
3ODU_A      ANFNKIFL-------PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF
2RH1_A      ----DEVVVVGMGIVMSS---LIVLAIVFCNVLVITAIKEFLQT vendingTVNYFITSLACADLVM
3EML_A      ----------IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTYFVVSSLADIAV
```

```
```
Experimental expectations:

- Highly conserved residues
- Secondary structure elements

Raw ClustalO alignment:

Adjusted alignment:

1u19  - - - - P W Q F S M - - L A A Y M F L L I M L G F P I N F L T L Y V T V Q H K K 3ODU_A N F K I F L - - - - - - - - - - P T I Y S I I F L T G I V G N G L V I L V M G Y Q K K 2RH1_A - - D E V W V V G M G I V M S - - L I V L A I V F G N V L V I T A I A K F E R 3EML_A - - - - - - - - - - I M G S S V Y I T V E L A I A V L A I L G N V L V C W A V W L N S N

helix regions
highly conserved residues
Alignment issues to be resolved
predicted membrane spanning region from OCTOPUS
Optimizing the Alignment is Important for Structural Integrity
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** THREAD the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model
Using the sequence alignment between target and template, thread the aligned target onto the 3D coordinates of the target structures.
Threading

Template:

```
L K R N N H -
LGK - - HHV
```

Target:

```
L K H V
```

Thread Coordinates:

```
(0,0,0) (1,1,1) (5,5,5)
```
Rosetta Requires Grishin Format for Threading

ClustalO:
• All sequences in one file
• Sequences broken up over several lines

FASTA:
• All sequences in one file
• Each sequence on individual line

Grishin:
• One file per alignment pair
• Each sequence on individual line
• Specific header information

```#
# 3pbl 2rh1_out
#
```
```scores_from_program: 0
0 ------YALSYCALILAIVFNGGLVCAMVLKE-RALQT-TTYLYVSLAVADLLVATLVMPPWVVYLEV-TGGVWNFS-
0 DEVVVVMGIVMMLVLAIVFNGNVLVITAIAKF-ERLQT-VTYFITSLACADLVMGLAVVPGAAHIL-MK-MWTFG-
```
Examining results after threading can give clues about validity of alignment.

Sequence Alignment is first step of the “Structural Alignment”
- No major breaks in Secondary Structure Elements
- Proper amino acids are on the right “face” of a helix
- Make sure amino acid placement along the template backbone “makes sense”
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** HYBRIDIZE template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model

D3R 1 MCFSVSLSATVALGCMFVPKVYII
B2AR 1 KEVYILLNWIGYVNSGFNPLITCR
Hybridization

- Combination of template fragments and PDB-derived fragments to generate a full length model
- Broken into a series of smaller steps from coarse sampling to fine sampling to prevent chain breaks
- Resulting model has various combinations of starting template segments
Hybridize: Full Model Assembly

1. Full length model assembly
   - Torsion space fragment insertion
   - Cartesian space template segment recombination

2. Local structure optimization and loop closure
   - Local fragment superposition
   - Gradient-based energy minimization

3. Full-atom sidechain and backbone refinement
Hybridize: Loop Closure

1. Full length model assembly
   - Torsion space fragment insertion
   - Cartesian space template segment recombination

2. Local structure optimization and loop closure
   - Local fragment superposition
   - Gradient-based energy minimization

3. Full-atom sidechain and backbone refinement

D. Local structure optimization and loop closure

Local fragment superposition
Hybridize: Energy Minimization

1. Full length model assembly
   - Torsion space fragment insertion
   - Cartesian space template segment recombination

2. Local structure optimization and loop closure
   - Local fragment superposition
   - Gradient-based energy minimization

3. Full-atom sidechain and backbone refinement

Pre-Relax

Post-Relax
Sequence, alignment, and threading account for initial manual stages of RosettaCM.

Fragment files are required.

Restraint functions are helpful but not necessary:
- Can assist in narrowing the conformational search space
- I.e. disulfide files, experimentally-derived constraints
Fragment Files

- 3- and 9-mer fragments derived from the PDB

- The target sequence is broken into these fragments and a fragment library is generated based on these segments
Making fragments with Robetta

http://robetta.bakerlab.org/
An important restraint file to use in modeling of GPCRs is a disulfide file

- Promotes the accurate formation of the conserved disulfide between TM3 and ECL2

3pbl.disulfide

72 158
Restraint Files: Membrane Span Files

- Predict membrane spanning regions using Octopus server
  - http://octopus.cbr.su.se/
- Octopus predictions are converted to a format compatible with Rosetta with the octopus2span.pl script
- Span file list start and stop residues of each TM region
Restraint Files: Membrane Span Files

With membrane penalties/weights

Without membrane penalties/weights
Additional Restraint Files

- Other data that can be helpful in accurately modeling your receptor include:
  - NMR
  - EPR
  - EM
  - Mutational analysis
  - H/D Exchange
  - Fluorescence
  - Mass Spec

- All can be incorporated in Rosetta but require specific handling of the constraint type.

Example restraint file:

```
AtomPair N 256 CA 76 SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AmbiguousConstraint
AtomPair CA 258 CA 207 SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AtomPair CA 258 CA 211 SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AtomPair CA 258 CA 227 FLAT_HARMONIC 0 2 5
AtomPair CA 258 CA 230 FLAT_HARMONIC 0 2 5
END
```
Hybridize: Full Model Assembly

Files needed for RosettaCM
- Partial-threaded structures
- Fragment files (3mer and 9mer lengths)
- Membrane spanning regions (span file)
- Mover definition and options
- Weight patches (pre-generated)
Submitting All Files to RosettaCM Hybridization

- RosettaCM is accessed through the RosettaScripts protocol
  - Relies on input being in XML format
- Additionally controls can be passed in through the command line
  - In general, it is preferred to move all command line options into an option file
RosettaScripts XML

Specifies:

- Score functions to be used

- Methods

- Order of execution
Score Function

- Score functions are set
- The first three weights are files that need to be copied into a directory for running
- The last score function “membrane_highres_Menv_smooth” is a default score function in Rosetta and does not to be moved
Hybridize mover specifies which score function to use at different stages of modeling.

- The disulfide file is pointed to in the main mover.
- Within the mover we point to the fragment files and threaded template files.
- The ClearConstraintsMover and FastRelax movers are additional steps to assist in energy minimization.
Options File Directs Rosetta to XML and Additional Inputs

```plaintext
-database /home/benderb/Rosetta/main/database/

# i/o
-in:file:fasta input_files/3pbl.fasta
-parser:protocol input_files/rosetta_cm.xml
-out:path:all output_files/

# output styles
-out:pdb
-out:file:scorefile 3pbl_scores.out
-nstruct 1

# membrane options
-in:file:spanfile input_files/3pbl.span
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-rg_reweight .1

# relax options
-relax:minimize_bond_angles
-relax:minimize_bond_lengths
-relax:jump_move true
-default max_cycles 200
-relax:min_type lbfgs_armijo_nonmonotone
-score:weights input_files/stage3_rlx_membrane.wts
-use_bicubic_interpolation
-hybridize:stagel_probability 1.0
-sog_upper_bound 15

# reduce memory footprint
-chemical:exclude_patches LowerDNA  UpperDNA CTerm amidation SpecialRotamer VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated
-thr phosphorlated tyr phosphorlated tyr sulfated lys dimethyalted lys monomethylated lys trimethylated
-lys acetylated glu carboxylated cys acetylated tyr diiodinated N acetylated C methylamidated MethylatedProteinCterm

#### path to Rosetta database

#### fasta of final sequence to be modeled

#### path to XML script

#### designates where to put pdb/silent files/scorefiles/etc

#### specifies output format as pdb

#### gives specific name for scorefile (default is scores.sc)

#### specifies number of models to be created

#### lists transmembrane spanning regions for membrane scoring

#### path to membrane weights file
```
What goes in, What comes out

Options

Rosetta Applications

RosettaScripts

XML

Sequence file
Span File
Path to XML

Fragment file
Disulfide file
Threaded templates
Weight files

Structure

Score File
Running RosettaCM

- If all files are set correctly the command line for running RosettaCM is:

  ```
  /path/to/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease @rosetta_cm.options
  ```

- By using the options file all command line options are contained in one, manageable file
- The options file points to the XML file with directs RosettaCM
- This protocol directly outputs pdb files which are ready for analysis/docking studies

- A single GPCR model typically takes 20-40 minutes depending on length
- It is advised to run 1000-5000 models
- Model selection is specific to goals of modeling but is often based on energy and structure-based clustering
Modeling Results of D3R

- The sampling of RosettaCM places important regions within 1-2 Å RMSD
- The overall RMSD including all loop regions averages just over 2 Å
- This level of accuracy is critical for having reliable models for ligand docking
Using RosettaCM online

- ROBETTA web-based application uses RosettaCM to generate 3D models from sequence.
- Completely automated.
- Generates alignments with Hhsearch, SPARKS-X, and RaptorX
- Does not include final loop building and membrane relax

http://robetta.bakerlab.org/submit.jsp
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

• **Step 1:** Align target sequence with sequence of template structure
• **Step 2:** Thread the target sequence onto the backbone of the template structure
• **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
• **Step 4:** DOCK ligand into comparative model
Ligand Docking
Ligands = Small molecules

Ligand Docking (RosettaLigand)

Peptide Docking (FlexPepDock)

Protein-Protein Docking (RosettaDock)
RosettaLigand Past

- Meiler and Baker 2006
  - Protein and ligand ensembles
  - Side chain flexibility
- Kaufmann et. al. 2008
  - Rotatable bond angle sampling
- Davis and Baker 2009
  - Backbone flexibility near binding site
- Lemmon and Meiler 2012
  - XML Scripts

HIV-1 PR homodimer (green/wheat) with acetylpepstatin (yellow) in binding site (red)

RosettaLigand Now

- Combs et. al. 2013
  - Docking into comparative models
- Allison et. al. 2013
  - Dock/design of interface
- Lemmon and Meiler 2013
  - Docking with interface waters
- Deluca and Meiler 2015
  - High throughput screening and improved low resolution sampling

ROSIE Ligand Docking Server: rosie.rosettacommons.org
Rosetta Protein – Ligand – Docking

- Protocol

- Ensemble of Ligand Conformations
- Ensemble of Protein Backbones

Pick Random Ligand and place in random Orientation and Position

Random Perturbation of Ligand Position and Orientation

Repack Side Chains (Rotamer Trial or Complete Combinatorial Optimization)

Gradient Based Rigid Body Minimization

Monte Carlo Accept?

50th cycles

Model of Protein-Ligand Complex
Taxol-Tubulin Complex

"Cross-Docking" Experiment

“Cross-Docking” Benchmark of 20 Protein-Ligand Complexes

- 10 proteins in complex with 2 ligands each
- 70 (80)% best RMSD below 2.0Å
- 75 (90)% below 2.0Å in first percentile
RosettaLigand Algorithm

1. Ensemble selection
2. Initial placement
3. Transform
4. Docking Cycles
5. High res refinement
RosettaLigand Algorithm

- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement
Ensemble selection → Initial placement → Transform → Docking Cycles → High res refinement

RosettaLigand Algorithm
RosettaLigand Algorithm

1. Ensemble selection
2. Initial placement
3. Transform
4. Docking Cycles
5. High res refinement

Grid based Monte-Carlo Translation, Rotation, and Conformation Sampling
Transform Algorithm

Start

Random Conformer Change

Flip a Coin

Did the Grid Score Improve?

Yes, Keep Move

No, Monte Carlo

Random Translation + Rotation
Simple Shape Complementarity

Red = Repulsion  Teal = Attraction
RosettaLigand Algorithm

- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement
- Move Ligand
- Rotamer Trials
- Full Repack
- Gradient-based Minimization (side chain, ligand)
- Accept/Reject
**RosettaLigand Algorithm**

- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement

**Gradient Based Minimization of side-chain and backbone torsion angles**
Methodology


Applications

- K. J. Gregory et al., Probing the metabotropic glutamate receptor 5 (mGlu5) positive allosteric modulator (PAM) binding pocket: discovery of point mutations that engender a “molecular switch” in PAM pharmacology., Molecular pharmacology 83, 991–1006 (2013).