

ROSETTACM: MULTI-TEMPLATE COMPARATIVE MODELING

Georg Kuenze

georg.kuenze@vanderbilt.edu

Alyssa Lokits

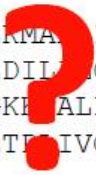
alyssa.d.lokits@vanderbilt.edu

Rosetta Workshop 2016

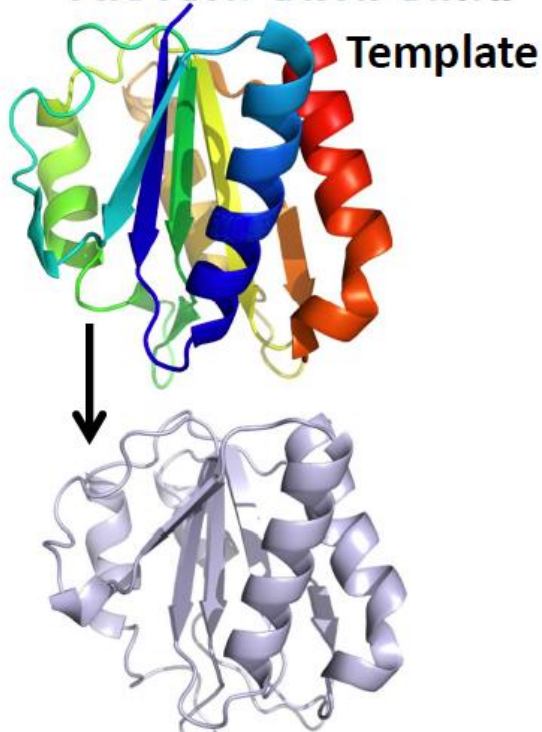
Introduction: RosettaCM multiple templates

Target

MKIVYWSGTGNTERMAIAKGIIESGKDVNTI
NVSDVNIDELLNEDIIGCSAMGDEVLEESEF
EPFIEEISTKISGKIALFGSYGWDGKWMRDF
EERMNGYGCVVETIVQNEPDEAEQDCIEFG
KKIANI



RCSB
PDB
PROTEIN DATA BANK



• Single Template Modeling:

- Single template
- Thread single backbone as input
- Uses sequence and template derived fragments

• Multiple Template Modeling:

- Multiple templates
- Thread multiple backbones as input
- Uses sections of multiple threaded models and sequence derived fragments
- Perform a final dualspace Relax

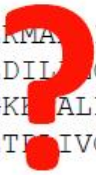
Nomenclature Note

- Comparative Modeling = Homology Modeling in the land of Rosetta

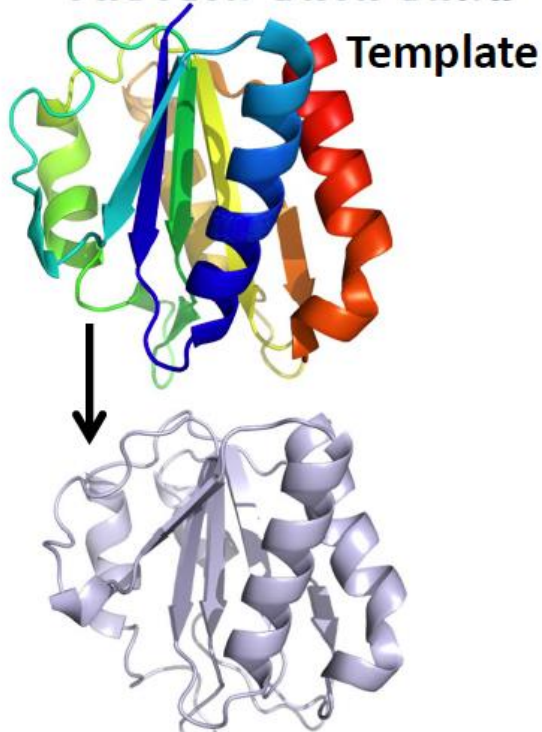
Identifying Template Structures

Target

MKIVYWSGTGNTERMAIAKGIIESGKDVNTI
NVSDVNIDELLNEDIIGCSAMGDEVLEESEF
EPFIEEISTKISGKIALFGSYGWDGKWMRDF
EERMNGYGCVVETIVQNEPDEAEQDCIEFG
KKIANI



RCSB **PDB**
PROTEIN DATA BANK



- **Sequence Similarity:**

- compare proteins based on amino acid properties alone (BLAST, PSI-BLAST)
- suitable templates have ideally >30% sequence identity to the target

- **Fold Recognition:**

- using predicted secondary structure information to detect proteins with similar 3D characteristics (**DALI**, **PHYRE**)

Practice Target: bovine Rhodopsin

- PDB ID: 1U19
- Class A G-protein coupled receptor (GPCR)
- No high identity or homologous templates
- 7 transmembrane helices
- 3 extracellular loops, 3 intracellular loops
- Highly conserved GPCR residues

Low Identity Templates

template	PDB ID	Seq ID
β 2 adrenergic receptor	2RH1	16
β 1 adrenergic receptor	2VT4	18
A2A adenosine receptor	3EML	20
CXCR4 chemokine receptor	3ODU	16
Dopamine D3 receptor	3PBL	26
Histamine H1 receptor	3RZE	18
M2 muscarinic acetylcholine receptor	3UON	21
Sphingosine 1-phosphate receptor 1	3V2W	19
M3 muscarinic acetylcholine receptor	4DAJ	22
κ -opioid receptor	4DJH	18
μ -opioid receptor	4DKL	21
N/OFQ opioid receptor	4EA3	21
δ -opioid receptor	4EJ4	19
5-HT1B receptor	4IAR	20
5-HT2B receptor	4IB4	20

- Class A GPCR's
- Highly conserved GPCR residues
- Similar fold profiles
- Low sequence identity (especially loops)

Comparative Modeling Protocol

- **Step 1:** Align target sequence with template sequences
- **Step 2:** Partial-thread the target sequence onto template structures
- **Step 3:** Combine pieces from different templates using RosettaCM Hybridize
- **Step 4:** Full-atom refinement and relax

```
-----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPPLNYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIFLTGIVGNGLVILVMGYQKKLRSM TDKYRLHLSVADLLF  
---DEVVVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSLACADLVM  
-----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNV TNYFVVSLAAADIAV
```

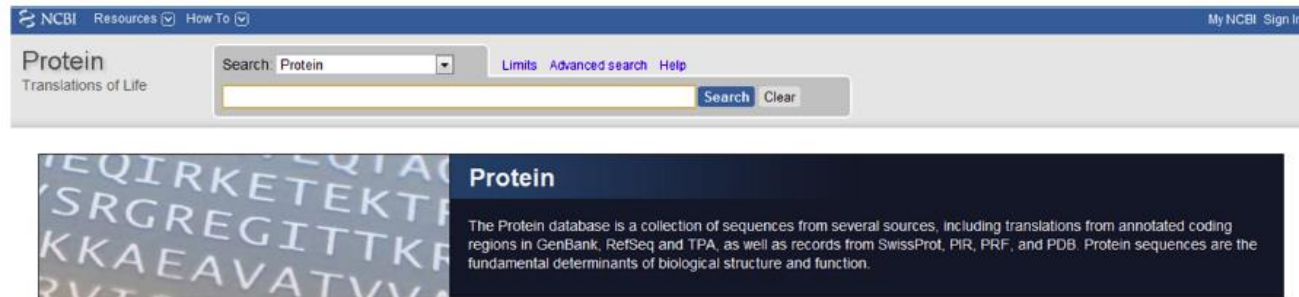
Target Sequence

Bovine rhodopsin (PDB: 1u19)

Find this file at */rosetta_cm/1_setup/1u19.fasta*

>1u19

```
PWQFSMLAAYMFLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVAD  
LFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVL  
AIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSR  
YIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIPLIVIFFCYGQLV  
FTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIF  
THQGSDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCG
```



NCBI Resources How To My NCBI Sign In

Protein
Translations of Life

Search: Protein Limits Advanced search Help

Search Clear

Protein

The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function.

<http://www.ncbi.nlm.nih.gov/protein>

Template PDBs

Human β 2-adrenergic receptor (PDB: 2rh1)

Human A2A adenosine receptor (PDB: 3eml)

Human CXC chemokine 4 receptor (PDB: 3odu)

Find these files at `/rosetta_cm/1_setup/`

RCSB PDB
PROTEIN DATA BANK

A MEMBER OF THE PDB
An Information Portal to Biological Macromolecular Structures
As of Tuesday Feb 22, 2011 at 4 PM PST there are 71415 Structures | PDB Statistics

Contact Us | Print

PDB ID or Text

PDB ID lookup or Text search of the complete structure file

Search

Advanced Search

MyPDB Hide

Login to your Account
Register a New Account

Home Hide

News & Publications
Usage/Reference Policies
Deposition FAQ
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

Search Hide

Advanced Search

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules Hide

List View of Archive By: Title | Date | Category

Structural View of Biology

Molecule of the Month:
Integrin

Our bodies are composed of approximately ten trillion cells, which poses challenging problems for structure and communication. All of these cells must be connected strongly together, to allow us to stand and walk. The infrastructure holding us together, however, must also be malleable enough to allow repairs, to allow us to heal from wounds. These many cells must also communicate with each other, ensuring that each plays its own proper part. Many different molecules in our bodies are involved in this complex

Infrastructure & Communication

Customize This Page

New Features Hide

Transporter Classification Database Browser

Latest features released:

Website Release Archive:

RCSB PDB News Hide

Weekly | Quarterly | Yearly

2011-02-22

Structural Neighbors

Explore structural neighbors lists to find connections between PDB

<http://www.rcsb.org>

Multiple Sequence Alignment

Find this file at `/rosetta_cm/2_threading/1u19_2rh1_3eml_3odu.aln`

CLUSTAL O(1.2.1) multiple sequence alignment

```
1u19      -----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM
3ODU_A    ANFNKIFL-----PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF
2RH1_A    ---DEVWVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSACADLVM
3EML_A    -----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVVTNYFVVSLAAADIAV

              :      :      :      .      *      *      .      .      .      *      :      :      *      :      **      :      .
```

Clustal Omega

[Input form](#) | [Web services](#) | [Help & Documentation](#)

[Share](#) [Feedback](#)

[Tools](#) > [Multiple Sequence Alignment](#) > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

STEP 1 - Enter your input sequences

Enter or paste a set of **PROTEIN** sequences in any supported format:

Or, upload a file: No file selected.

STEP 2 - Set your parameters

OUTPUT FORMAT **Clustal w/o numbers**

The default settings will fulfill the needs of most users and, for that reason, are not visible.

(Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

Adjusting multiple sequence alignment

Experimental expectations:

- Highly conserved residues
- Secondary structure elements

Raw ClustalO 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	A	N	F	N	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

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	
3ODU_A	A	N	F	-	-	-	-	-	-	-	N	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
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	
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	

helix regions

highly conserved residues

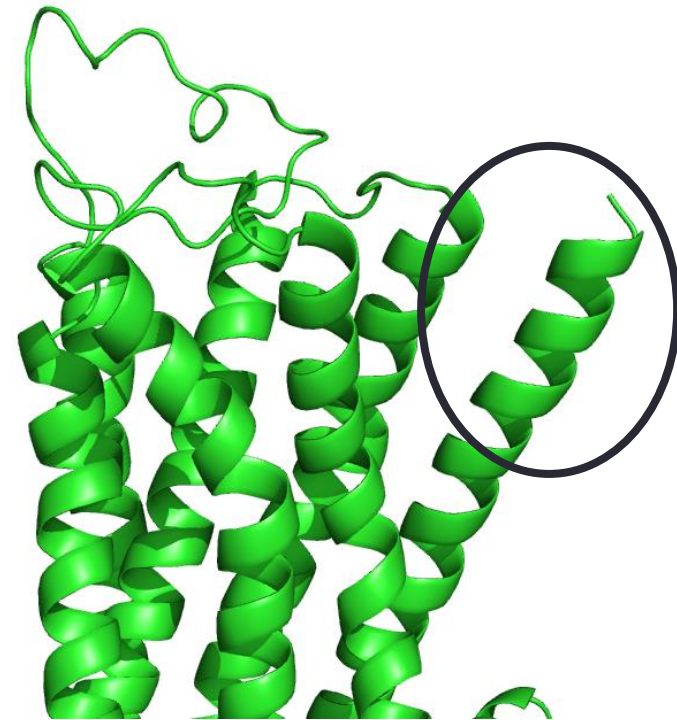
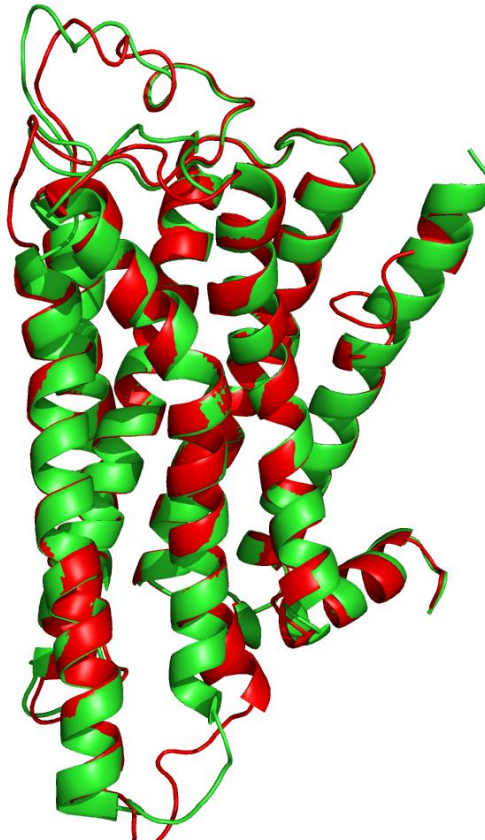
Alignment issues to be resolved

predicted membrane spanning region from OCTOPUS

Removing helix gaps



Raw alignment



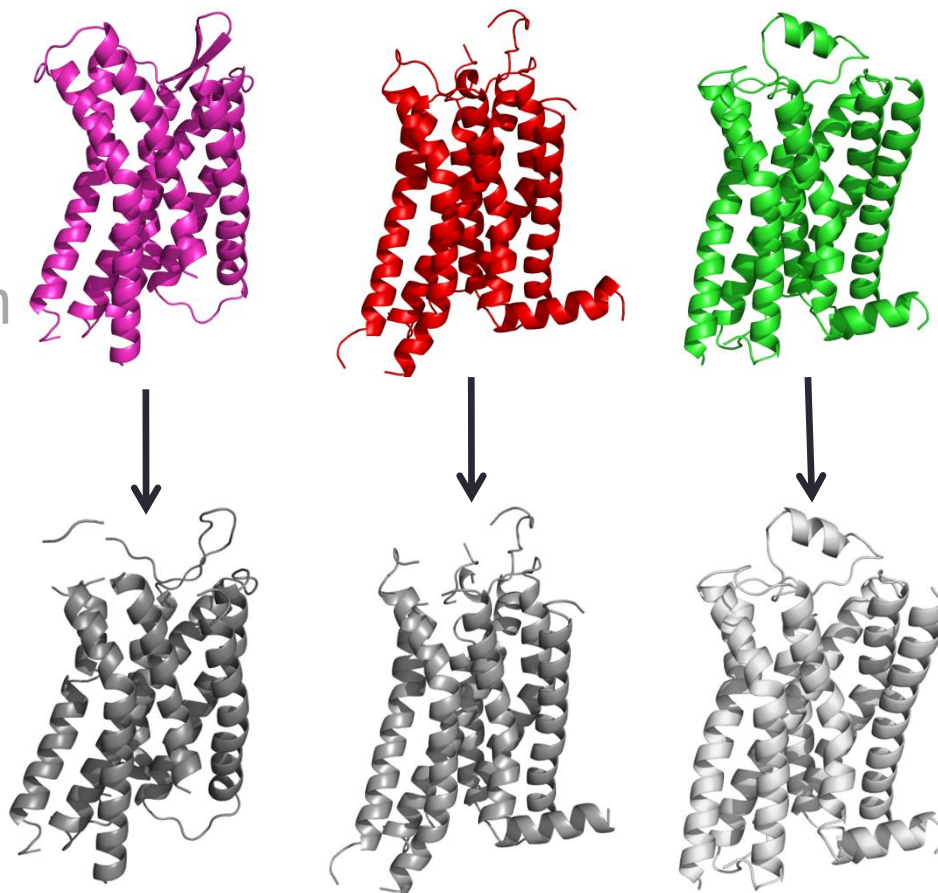
Adjusted alignment

Comparative Modeling Protocol

- **Step 1:** Align target sequence with template sequences
- **Step 2:** Partial-thread the target sequence onto template structures
- **Step 3:** Combine pieces from different templates using RosettaCM Hybridize
- **Step 4:** Full-atom refinement and relax

```
-----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF  
---DEVVVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSACADLVM  
-----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNTNYFVVSLAAADIAV
```

+



Grishin Format

- ClustalO:
 - all sequences in one file
 - Sequences broken up over several lines
- Grishin:
 - one file per alignment pair
 - sequences continuous over one line each
 - Contains header information

Find converted alignment files at */rosetta_cm/2_threading/*

(1u19_2rh1.grishin, 1u19_3eml.grishin, 1u19_3odu.grishin)

Threading

Template:

(0,0,0)	(1,1,1)	(2,2,2)	(3,3,3)	(4,4,4)	(5,5,5)	
L	K	R	N	N	H	-
(?,?,?)	(?,?,?)				(?,?,?)	(?,?,?)
L	K	-	-	-	H	V

*Thread
Coordinates*

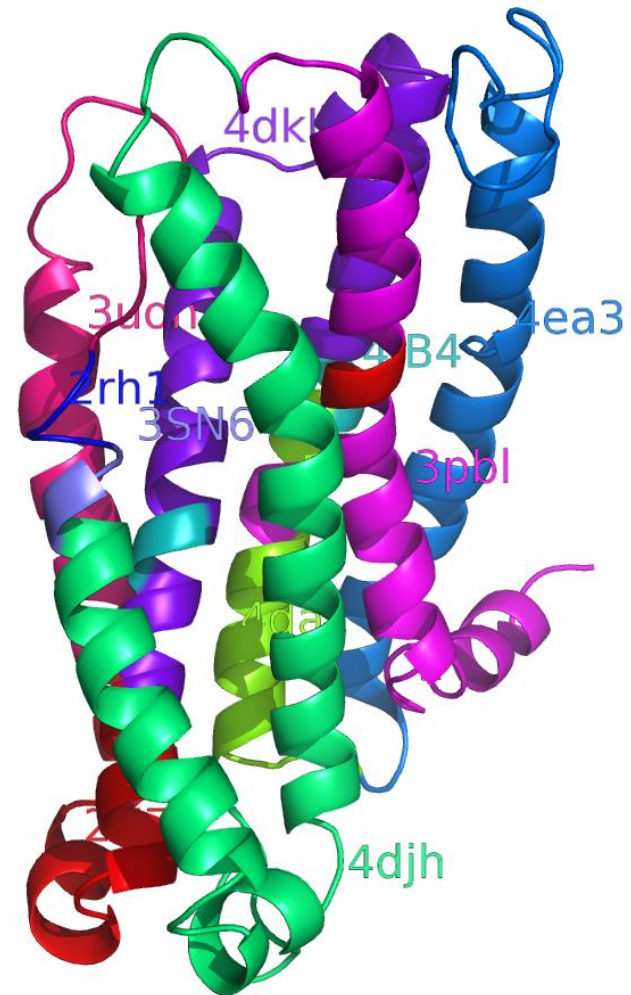


Target:

(0,0,0)	(1,1,1)	(5,5,5)	
L	K	H	V

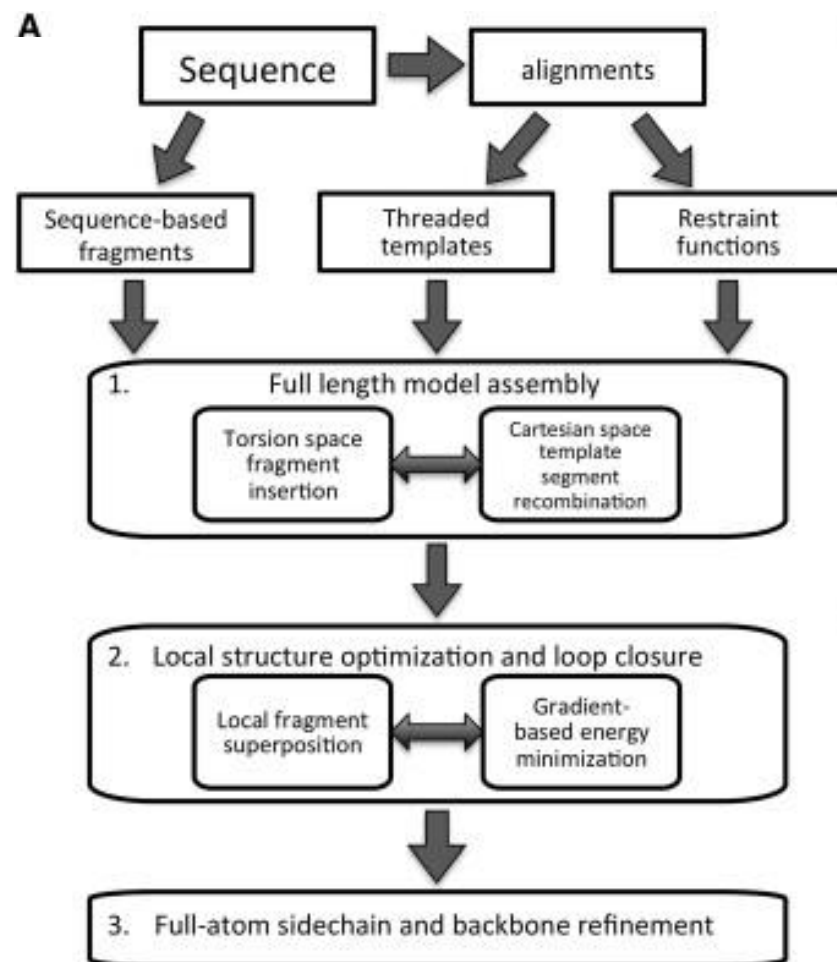
Comparative Modeling Protocol

- **Step 1:** Align target sequence with template sequences
- **Step 2:** Partial-thread the target sequence onto template structures
- **Step 3:** Combine pieces from different templates using RosettaCM Hybridize
- **Step 4:** Full-atom refinement and relax



RosettaCM: Three Stages

1. Generate initial models from template alignments
- 2. Explore deviations from templates and close loops with 2-step MC:**
 - Randomly select de novo or template-based fragment and substitute into current conformation
 - Cartesian space full-backbone minimization
3. Full atom backbone and side chain refinement and final relax



Song, et al. 2013

Input Files for RosettaCM Hybridize

- Partial-threaded structures
- Fragment files (3mer and 9mer lengths)
- Membrane spanning regions (span file)
- Mover definition and options
- Weight patches

Fragment Files

Find these files at
*/rosetta_cm/3_hybridize/1u19_3.frag*s &
*/rosetta_cm/3_hybridize/1u19_9.frag*s



www.bakerlab.org

Structure Prediction Fragment Libraries Alanine Scanning DNA Interface Scan
[Queue] [Submit] [Queue] [Submit] [Queue] [Submit] [Queue] [Submit]
[Register / Update] [Docs / FAQs] [Login]

Submit a job to the Fragment Server

*Please submit one job at a time

Required

[Registered Username:](#) or [Registered Email Address:](#)

Target Name:

Paste [Fasta](#)

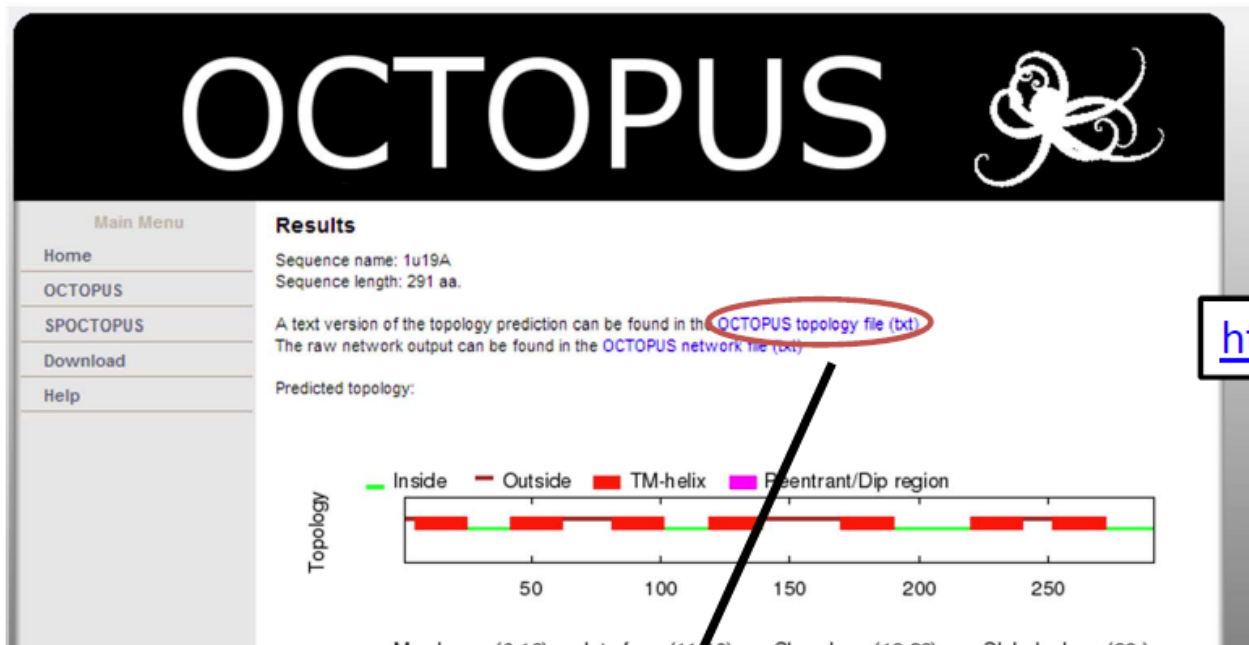
or Upload [Fasta](#): No file chosen

<http://rosetta.bakerlab.org/fragmentsubmit.jsp>

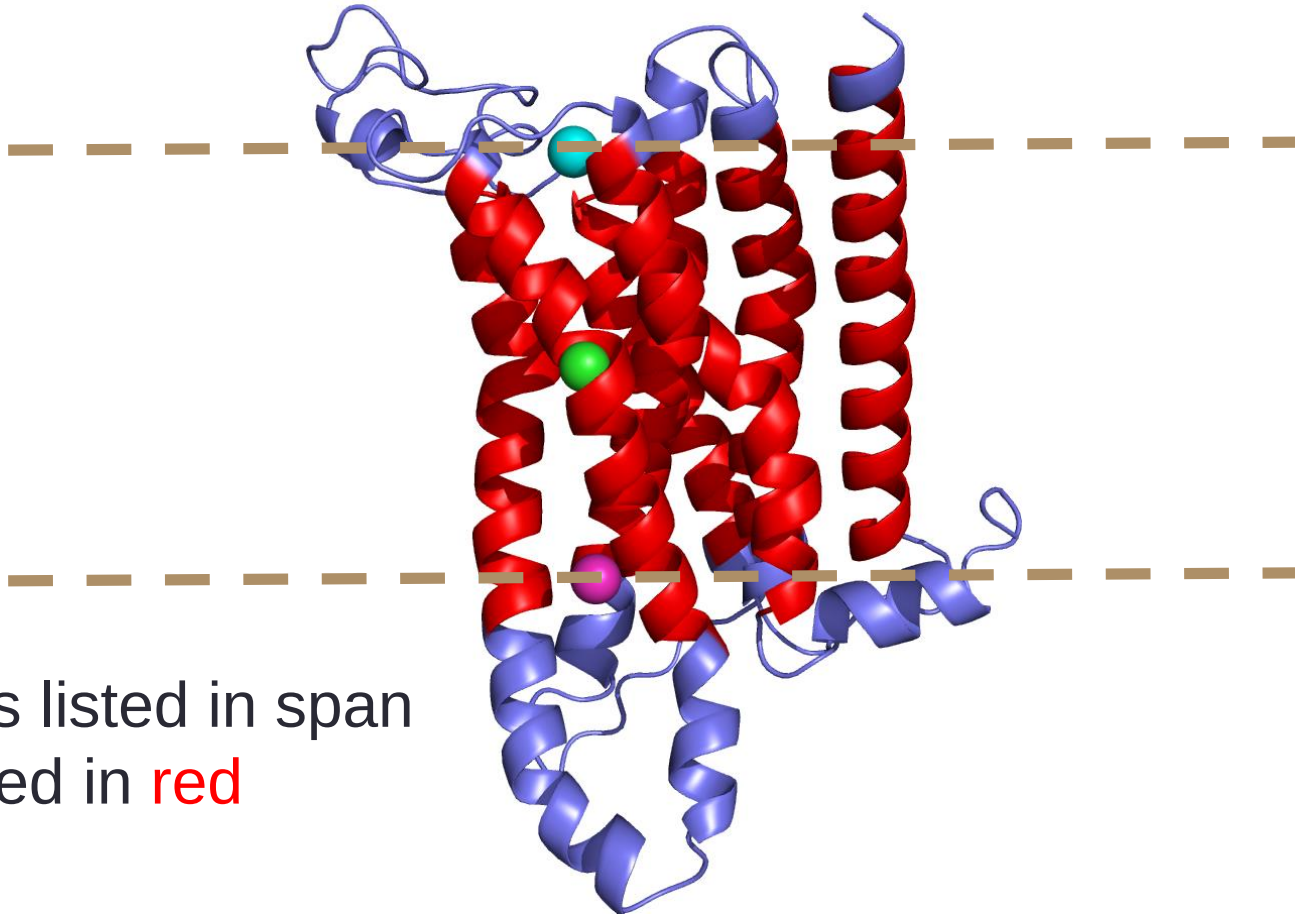


Membrane spanning regions

Find this file at `/rosetta_cm/3_hybridize/1u19.span`



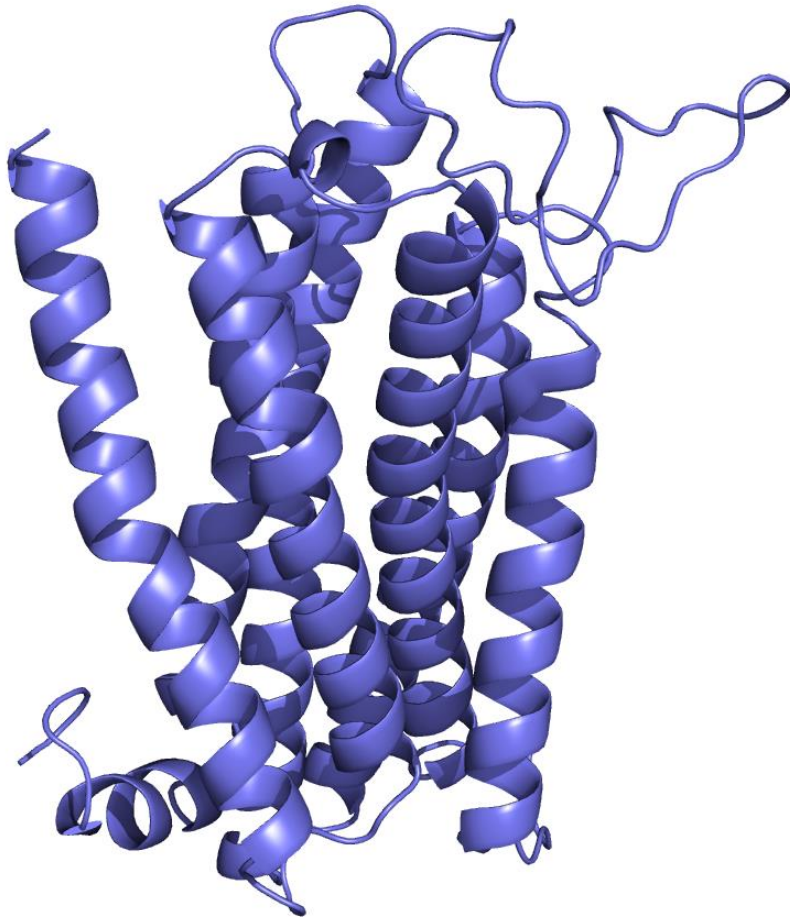
Rosetta Membrane



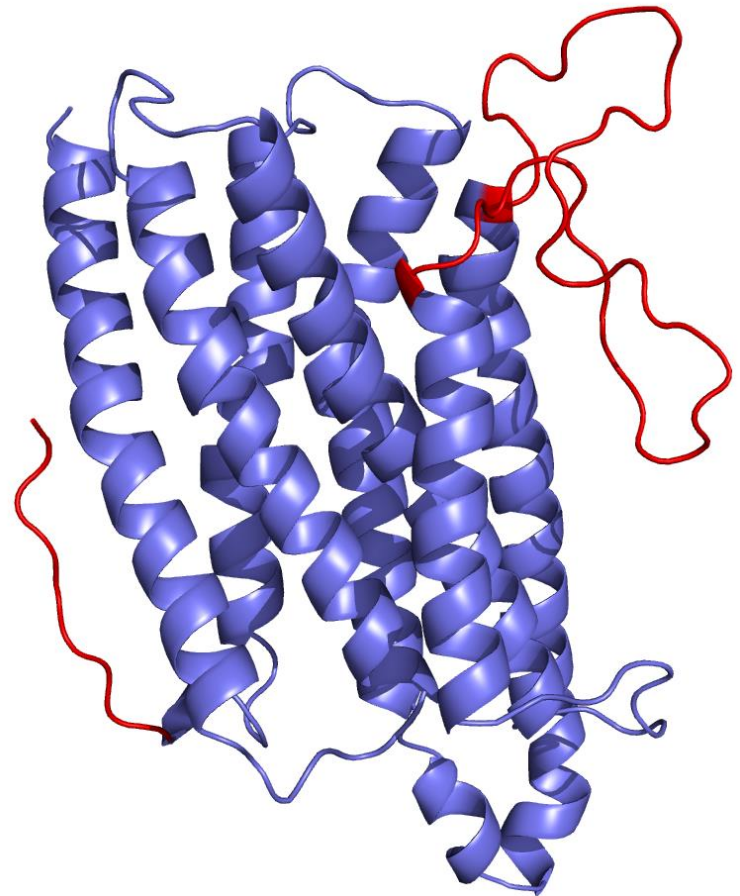
Residues listed in span
file colored in red

Why use membrane scoring terms?

With membrane penalties/weights



Without membrane penalties/weights



Hybridize Mover

/rosetta_cm/3_hybridize/rosetta_cm.xml

```
<SCOREFXNS>
  <stage1 weights="stage1_membrane.wts" symmetric=0>
    <Reweight scoretype=atom_pair_constraint weight=1/>
  </stage1>
  <stage2 weights="stage2_membrane.wts" symmetric=0>
    <Reweight scoretype=atom_pair_constraint weight=0.5/>
  </stage2>
  <fullatom weights="stage3_rlx_membrane.wts" symmetric=0>
    <Reweight scoretype=atom_pair_constraint weight=0.5/>
  </fullatom>
</SCOREFXNS>
```

*Find all **.wts** files in */rosetta_cm/3_hybridize/*

Hybridize Mover

/rosetta_cm/3_hybridize/rosetta_cm.xml

```
<MOVERS>
  <Hybridize name=hybridize stage1_scorefxn=stage1 stage2_scorefxn=stage2
    fa_scorefxn=fullatom batch=1 stage1_increase_cycles=1.0
    stage2_increase_cycles=1.0 linmin_only=1>
    <Fragments 3mers="1u19_3.frag" 9mers="1u19_9.frag"/>
    <Template pdb="1u19_on_2rh1.pdb" cst_file="AUTO" weight= 1.000 />
    <Template pdb="1u19_on_3eml.pdb" cst_file="AUTO" weight= 1.000 />
    <Template pdb="1u19_on_3odu.pdb" cst_file="AUTO" weight= 1.000 />
  </Hybridize>
</MOVERS>
```

RosettaCM Options

/rosetta_cm/3_hybridize/rosetta_cm.options

i/o

-in:file:fasta 1u19.fasta
-parser:protocol rosetta_cm.xml

relax options

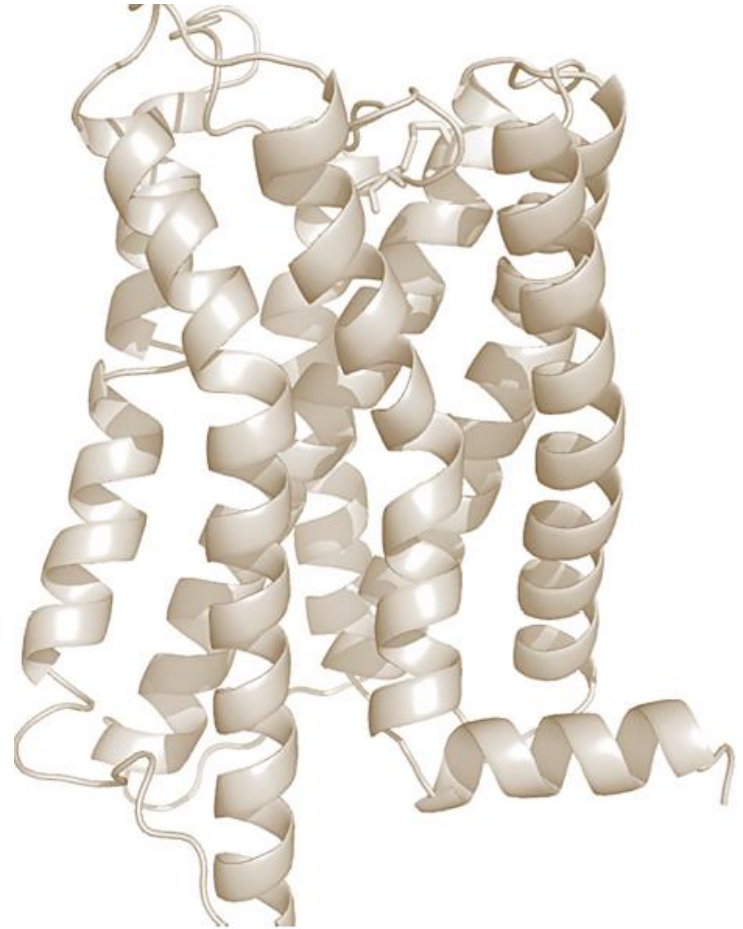
-relax:minimize_bond_angles
-relax:minimize_bond_lengths
-relax:jump_move true
-default_max_cycles 200
-relax:min_type lbfgs_armijo_nonmonotone
-score:weights stage3.wts
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0

#Initialize membrane

-in:file:spanfile 1u19.span
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-rg_reweight .1

Comparative Modeling Protocol

- **Step 1:** Align target sequence with template sequences
- **Step 2:** Partial-thread the target sequence template structures
- **Step 3:** Combine pieces from different templates using RosettaCM Hybridize
- **Step 4:** Full-atom refinement and relax



Final Relax Input Files

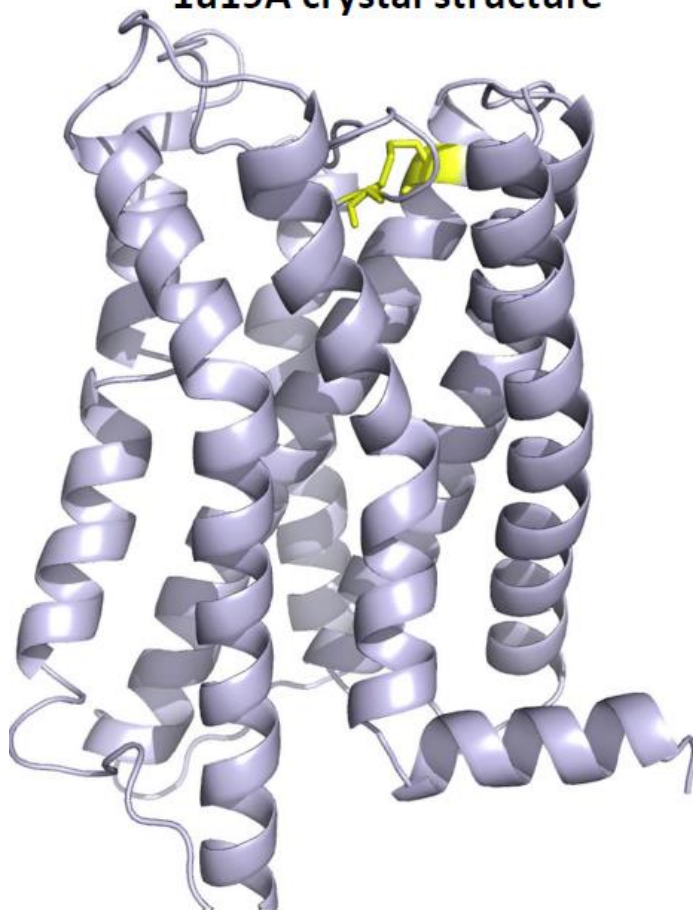
- Span file
- Relax options
- Optional constraints (atom pair distances, disulfide bonds, etc.)

Disulfide constraints

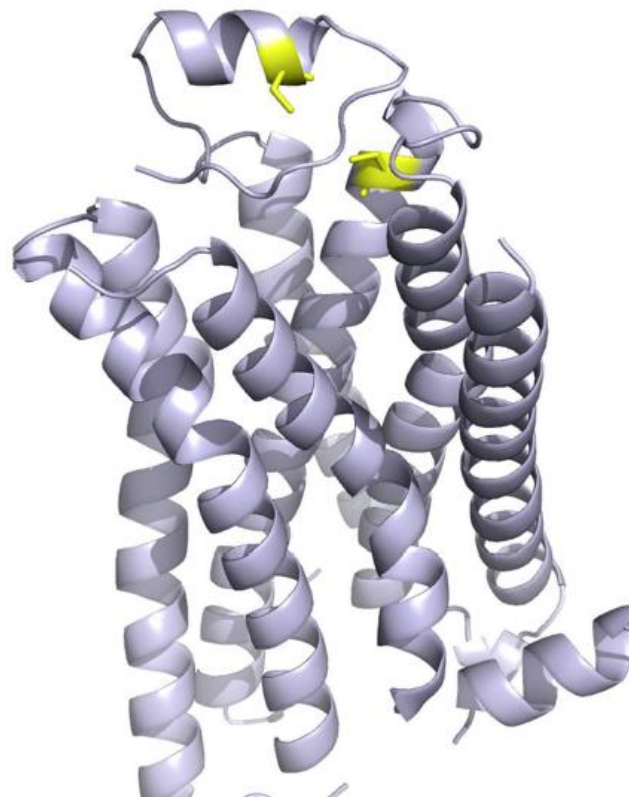
Find this file at `/rosetta_cm/4_relax/1u19.disulfide`

77 154

1u19A crystal structure



1u19A threaded onto 2rh1A



Relax options

/rosetta_cm/4_relax/relax.options

i/o

-in:fix_disulf 1u19.disulfide
-out:file:fullatom
-out:pdb

Dualspace settings

-relax:dualspace
-relax:minimize_bond_angles
-set_weights cart_bonded .5 pro_close 0
-default_max_cycles 200

#Initialize membrane

-in:file:spanfile 1u19.span
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-score:weights membrane_highres_Menv_smooth.wts

Tutorial

Comparative modeling of Rhodopsin with three class A GPCR's

Five stages:

- I. Setup
- II. Threading
- III. RosettaCM hybridize
- IV. Dualspace relax
- V. Final model selection

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 use Rosetta Membrane and final Relax.

<http://robetta.bakerlab.org/submit.jsp>

References

- **Rosetta User Guide & Documentation**

<https://www.rosettacommons.org/docs/latest/Home>

- **Membrane Proteins Documentation**

https://www.rosettacommons.org/docs/latest/application_documentation/Application%20Documentation#Membrane-Proteins

- **RosettaCM: Multi-template**

Yifan Song, et al. (2013). High-Resolution Comparative Modeling with RosettaCM. *Structure*, 21(10), 1735-1742.