

# ROSETTACM: MULTI-TEMPLATE COMPARATIVE MODELING

---

Harikrishna Sekar Jayanthan

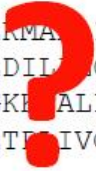
[harikrishna.sekar@vanderbilt.edu](mailto:harikrishna.sekar@vanderbilt.edu)

Rosetta Workshop Nov 2018

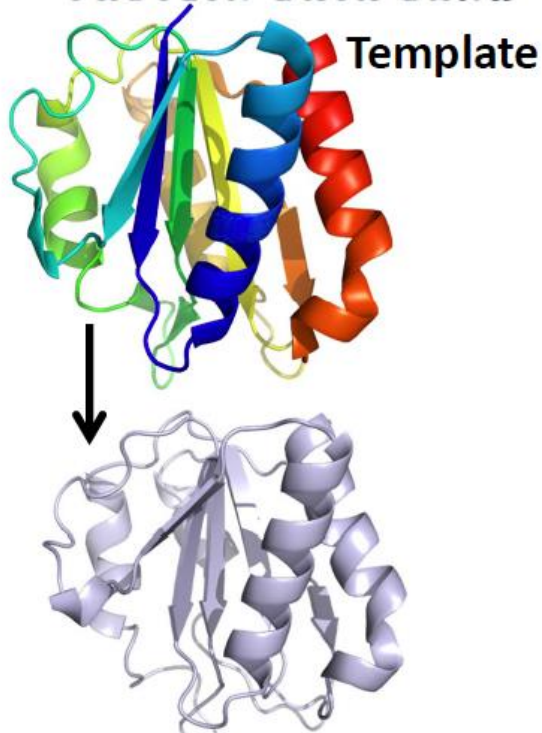
# Intro to Multi-template Comparative Modeling

## Target

MKIVYWSGTGNTERMAIAKGIIESGKDVNTI  
NVSDVNIDELLNEDIIIGCSAMGDEVLEESEF  
EPFIEEISTKISGKIALFGSYGWGDGKWMRDF  
EERMNGYGCVVETIVQNEPDEAEQDCIEFG  
KKIANI



RCSB **PDB**  
PROTEIN DATA BANK



## • 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

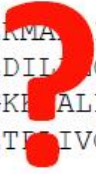
## \*Nomenclature Note\*

- Comparative Modeling = Homology Modeling in the land of Rosetta

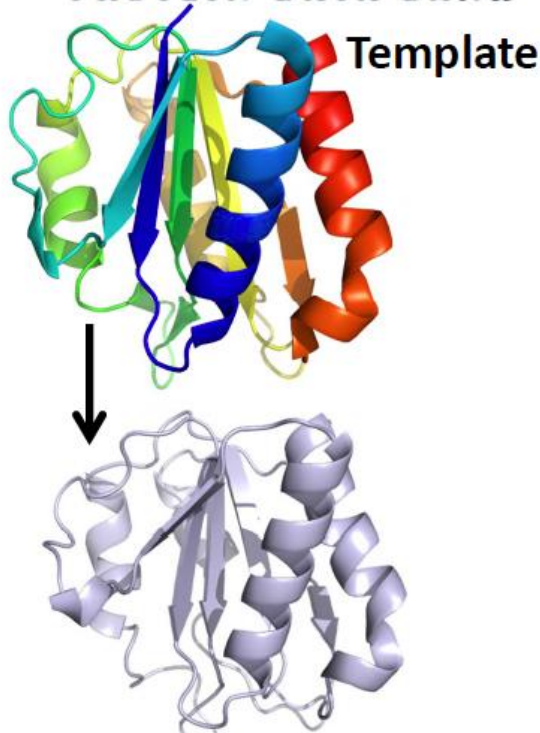
# Identifying Template Structures

## Target

MKIVYWSGTGNTERMAIAKGIIESGKDVNTI  
NVSDVNIDELLNEDIIIGCSAMGDEVLEESEF  
EPFIEEISTKISGKIALFGSYGWDGKWMRDF  
EERMNGYGCVVETIVQNEPDEAEQDCIEFG  
KKIANI



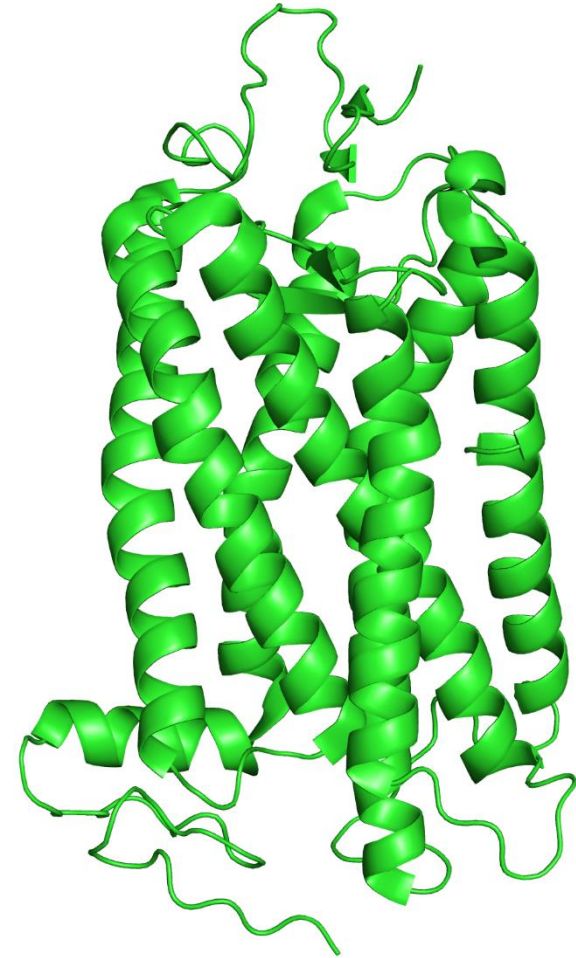
RCSB  
**PDB**  
PROTEIN DATA BANK



- **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**)

# Practice Target: bovine Rhodopsin

- PDB ID: 1U19
- Class A G-protein coupled receptor
- 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
$\beta$ 2 adrenergic receptor	2RH1	16
$\beta$ 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
$\kappa$ -opioid receptor	4DJH	18
$\mu$ -opioid receptor	4DKL	21
N/OFQ opioid receptor	4EA3	21
$\delta$ -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)

# Multi-Template 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--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPNLYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIFLTGIVGNGLVILVMGYQKKLRSM TDKYRLHLSVADLLF  
---DEVVVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSLACADLVM  
-----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNV TNYFVVSLAAADIAV
```

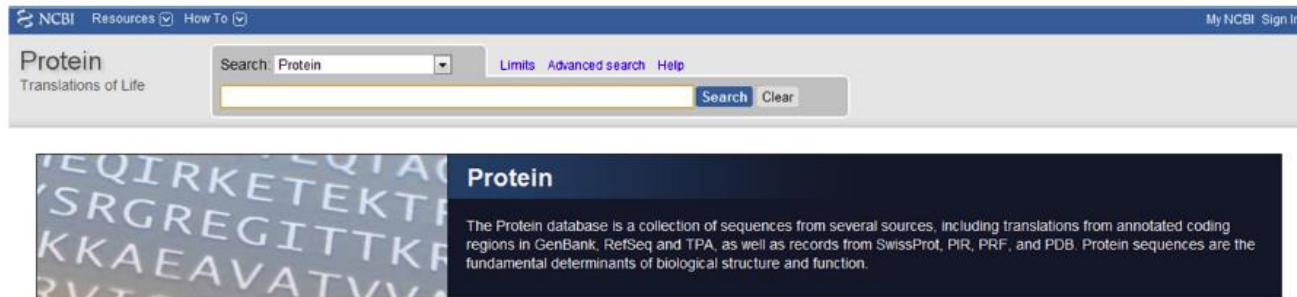
# Target Sequence

Bovine rhodopsin (PDB: 1u19)

Find this file at */rosetta\_cm/1\_setup/1u19.fasta*

>1u19

PWQFSMLAAYMFLLIMLGFPINFLTLYVTQVQHKKLRTPLNYILLNLAVAD  
LFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVL  
AIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSR  
YIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIPLIVIFFCYGQLV  
FTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIF  
THQGSDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLC



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  $\beta$ 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 Policies  
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    ---DEVVVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSACADLVM
3EML_A    -----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
           :      :  :  .  *  * .  .      ..*:.. :      *: ** :  .
```

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

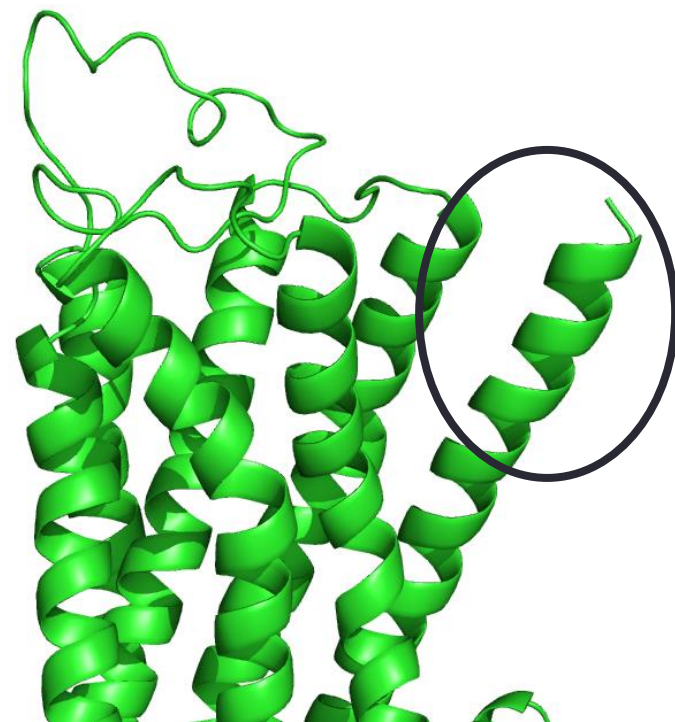
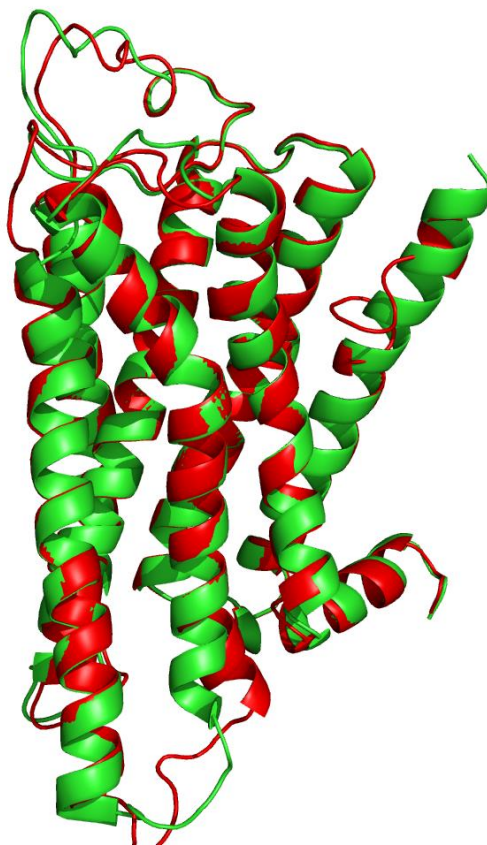
Alignment issues to be resolved

predicted membrane spanning region from OCTOPUS

# Removing helix gaps



Raw alignment



Adjusted alignment

# 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/*

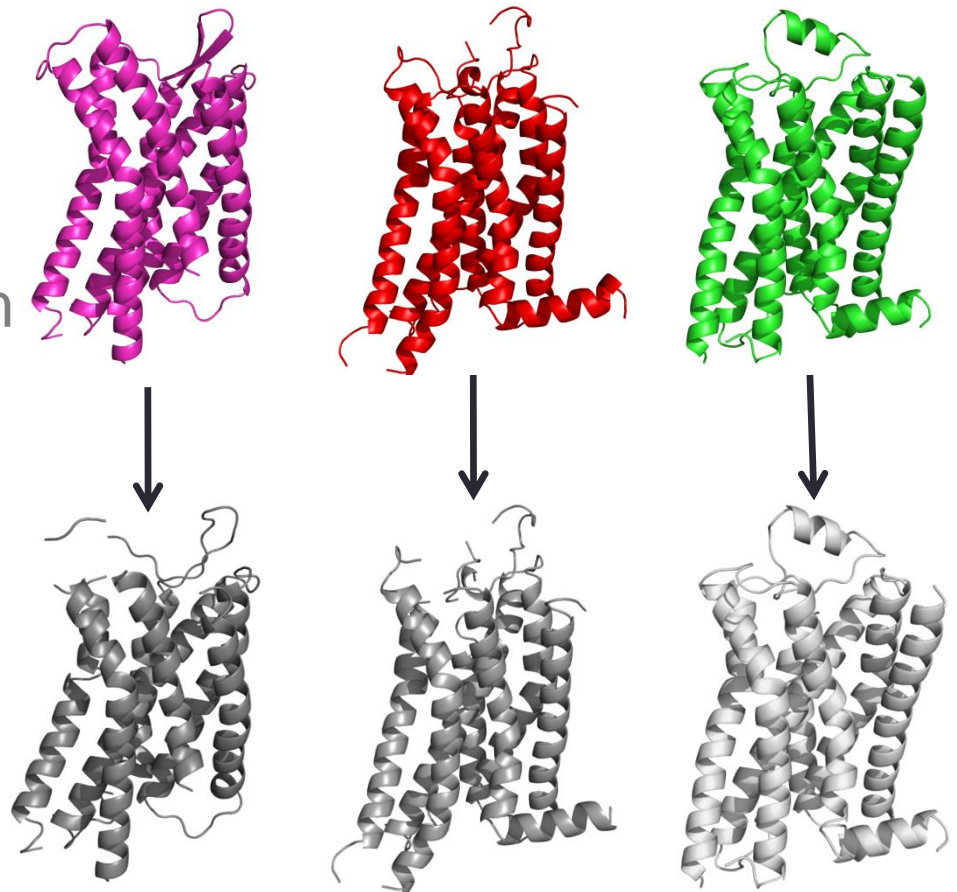
(1u19\_2rh1.grishin, 1u19\_3eml.grishin, 1u19\_3odu.grishin)

# 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

```
-----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPINYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF  
---DEVWVVGMGIVMS---LIVLAIVFGNVLVITAIKFERLQTVTNYFITSLACADLVM  
-----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
```

+



# 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	-
(?,?,?)	(?,?,?)				(?,?,?)	(?,?,?)

Target: 

L	K	-	-	-	H	V
---	---	---	---	---	---	---

*Thread  
Coordinates*



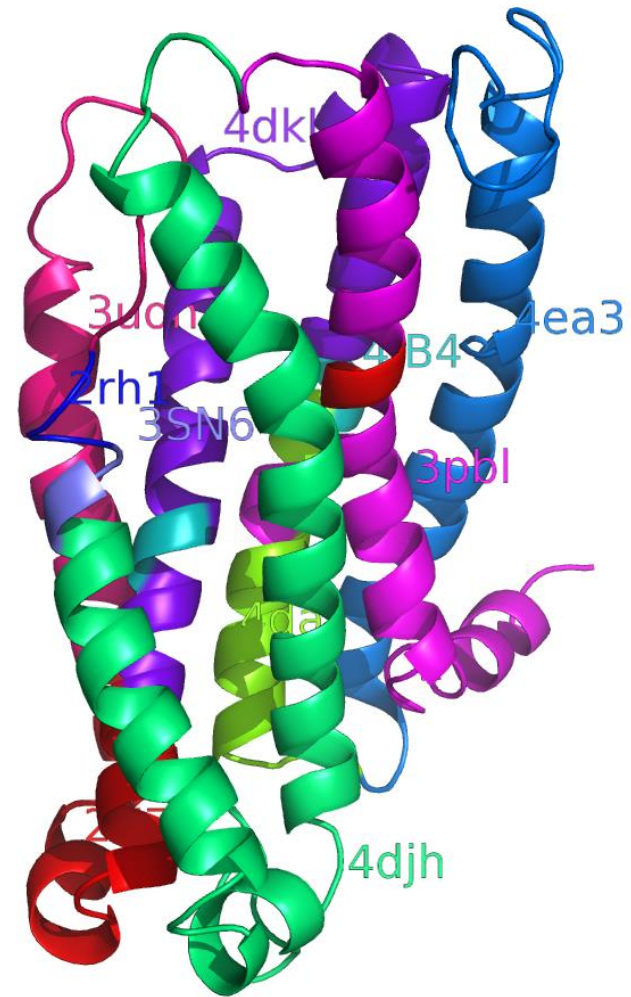
Target: 

(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



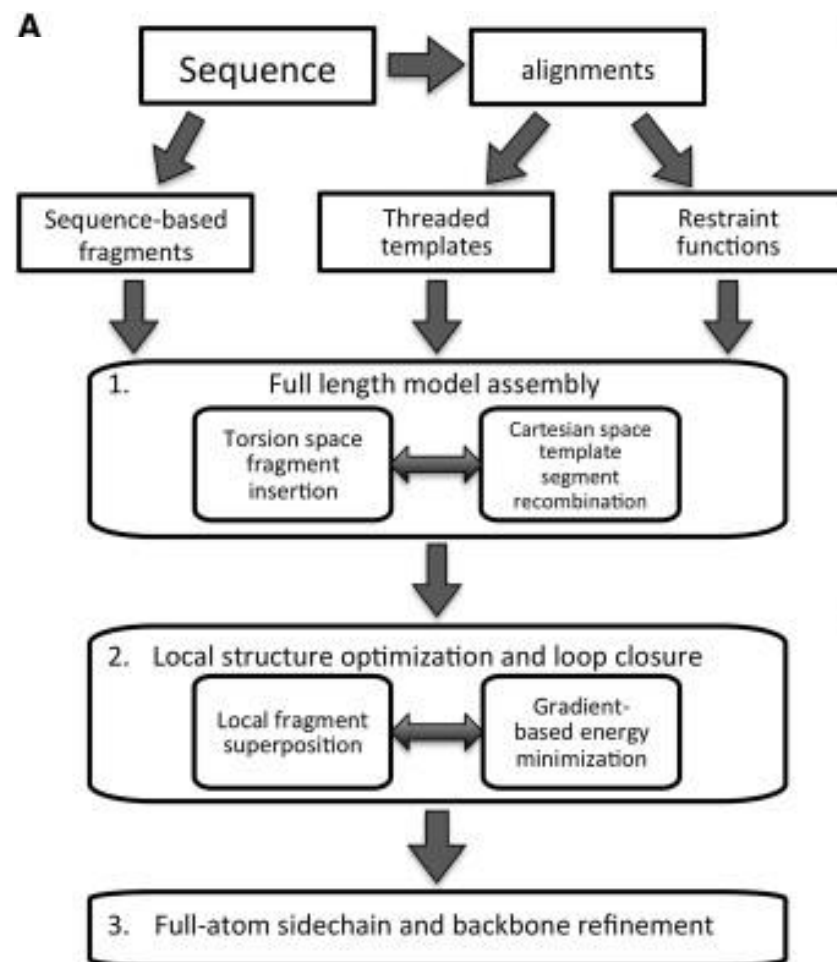
# RosettaCM: Three Stages

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 refinement

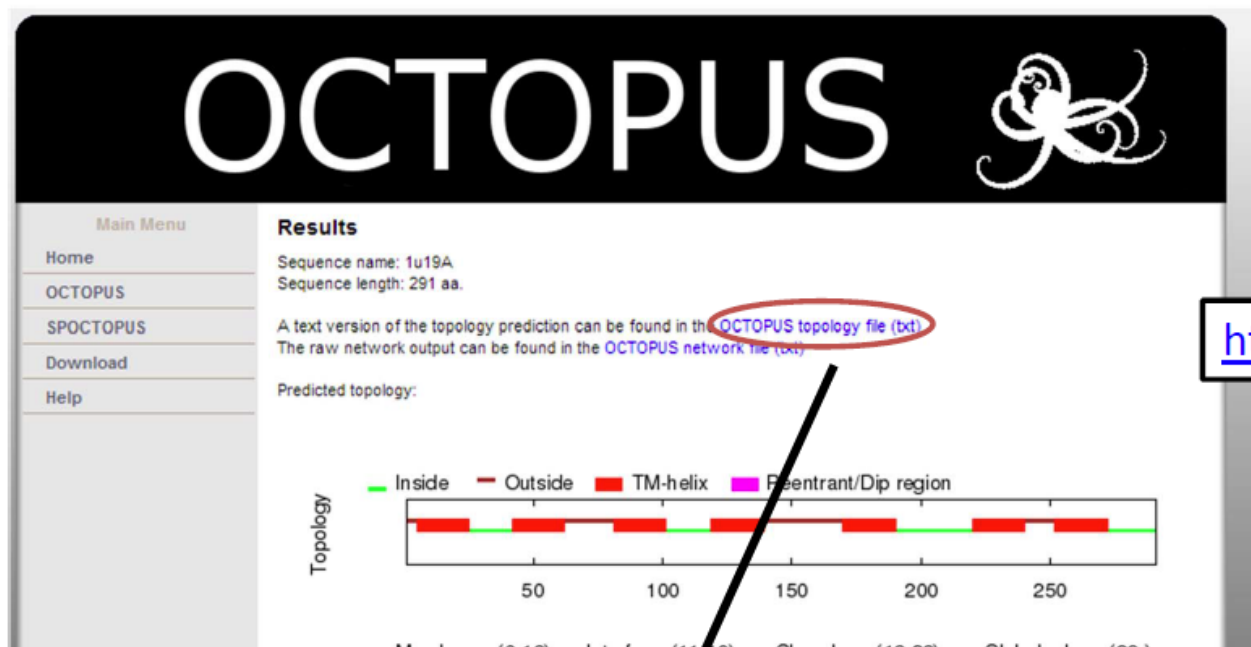


# Input Files for RosettaCM Hybridize

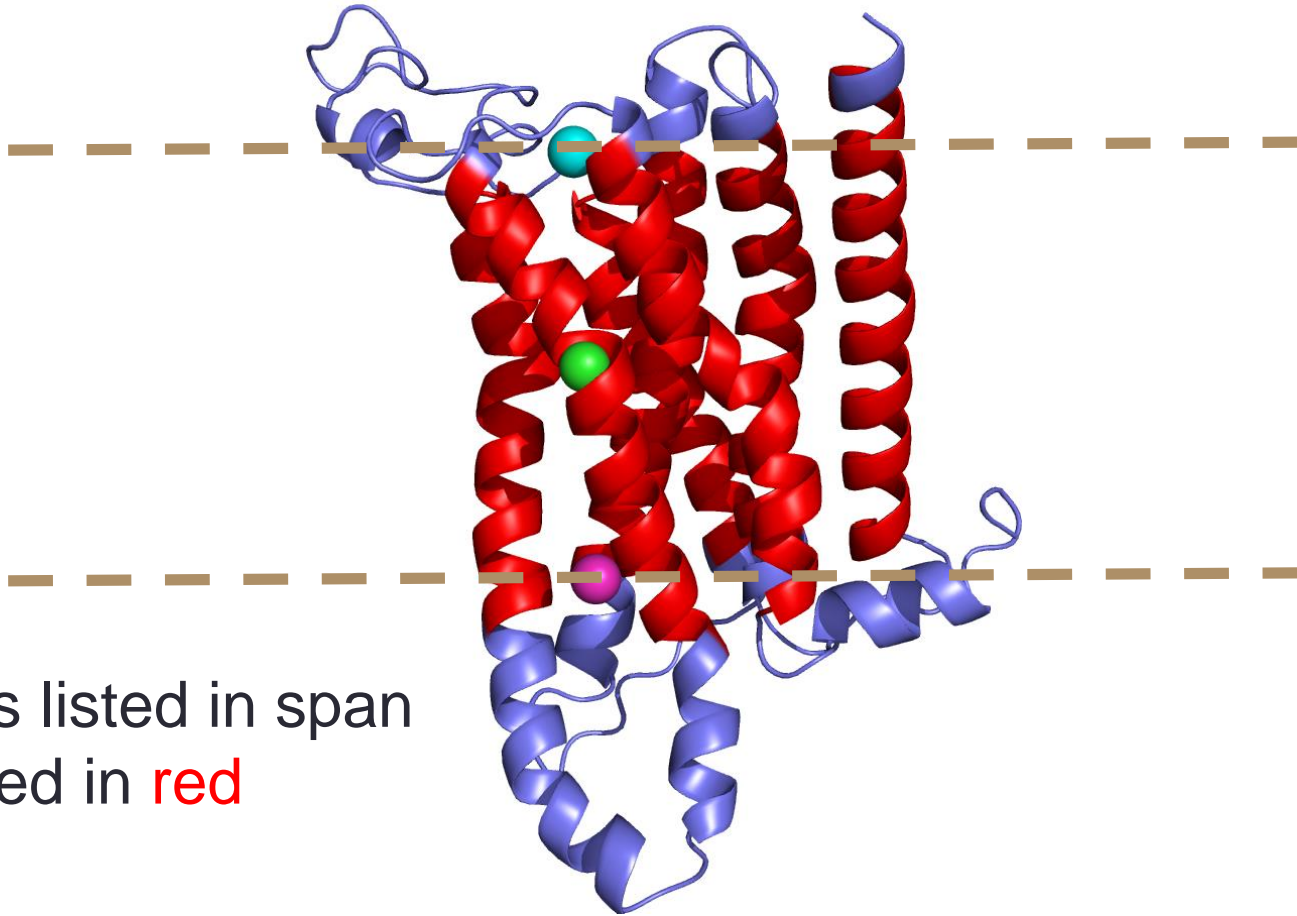
- Partial-threaded structures
- Membrane spanning regions (span file)
- Mover definition and options
- Score Weights

# Membrane spanning regions

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



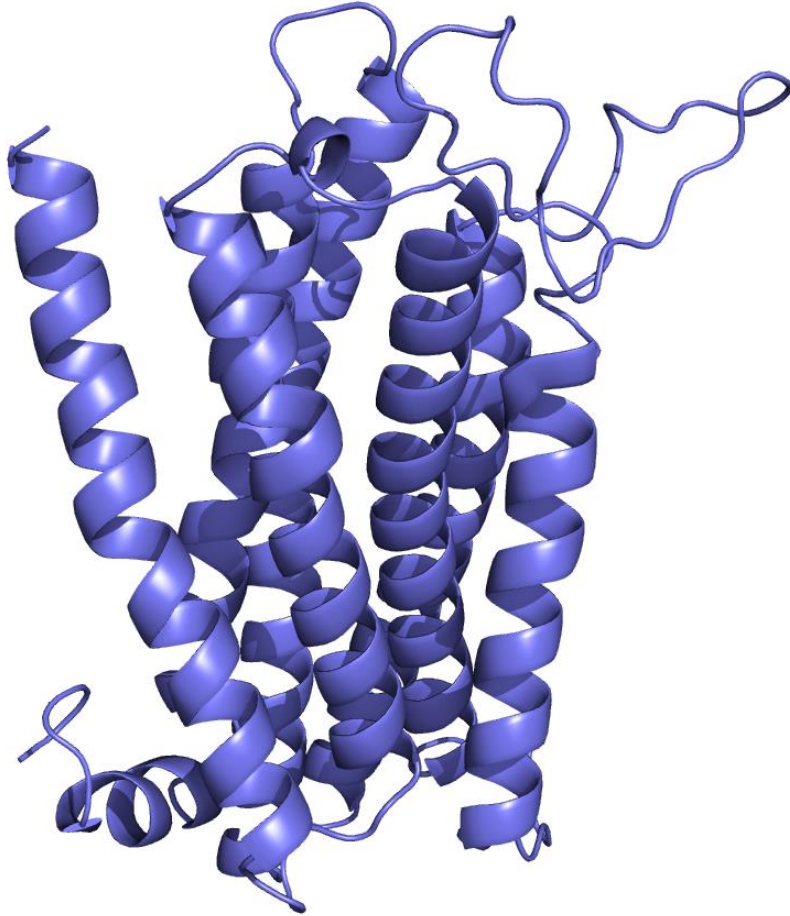
# Rosetta Membrane



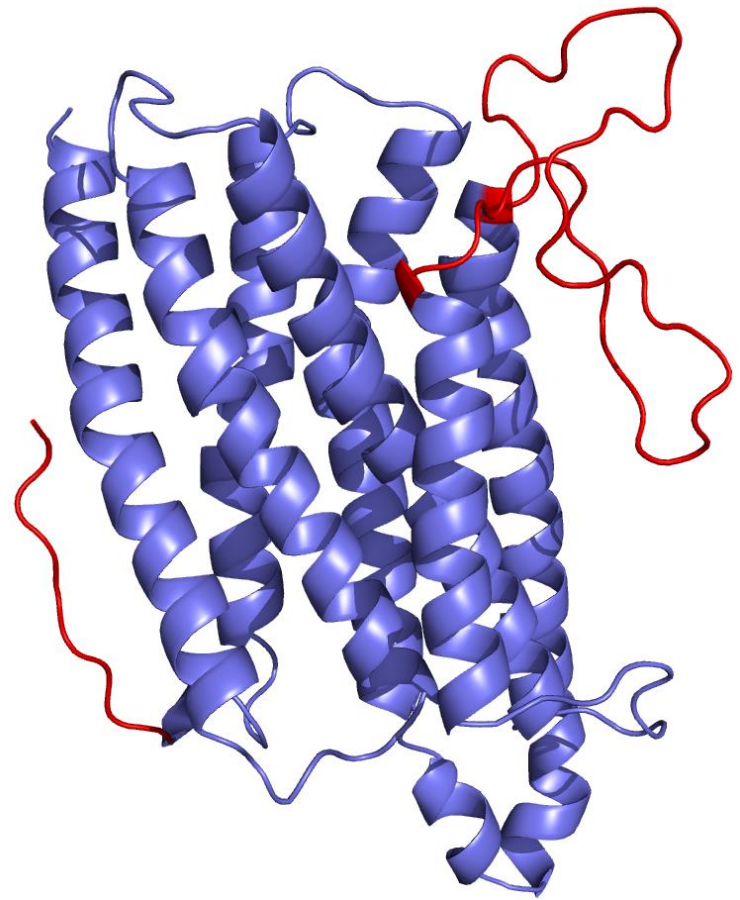
Residues listed in span  
file colored in red

# Why use membrane scoring terms?

**With** membrane penalties/weights



**Without** membrane penalties/weights





# Hybridize Mover

*/rosetta\_cm/3\_hybridize/rosetta\_cm.xml*

```
<SCOREFXNS>
  <ScoreFunction name="stage1" weights="stage1_membrane.wts" symmetric="0">
    <Reweight scoretype="atom_pair_constraint" weight="1"/>
  </ScoreFunction>
  <ScoreFunction name="stage2" weights="stage2_membrane.wts" symmetric="0">
    <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
  </ScoreFunction>
  <ScoreFunction name="fullatom" weights="stage3_rlx_membrane.wts" symmetric="0">
    <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
  </ScoreFunction>
  <ScoreFunction name="membrane" weights="membrane_highres_Menv_smooth" symmetric="0">
  </ScoreFunction>
</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" disulf_file="1u19.disulfide">
```

```
    <Template pdb="1u19_on_2rh1.pdb" cst_file="AUTO" weight="1.0" />
```

```
    <Template pdb="1u19_on_3eml.pdb" cst_file="AUTO" weight="1.0" />
```

```
    <Template pdb="1u19_on_3odu.pdb" cst_file="AUTO" weight="1.0" />
```

```
  </Hybridize>
```

```
    <ClearConstraintsMover name="clearconstraints"/>
```

```
</MOVERS>
```

# RosettaCM Options

*/rosetta\_cm/3\_hybridize/rosetta\_cm.options*

## # i/o

-in:file:fasta **1u19.fasta**  
-parser:protocol input\_files/rosetta\_cm.xml  
-out:pdb

#### your target sequence

## #Initialize membrane

-in:file:spanfile **1u19.span**  
-membrane:no\_interpolate\_Mpair  
-membrane:Menv\_penalties  
-rg\_reweight .1  
-restore\_talaris\_behavior

#### only if modeling a membrane protein

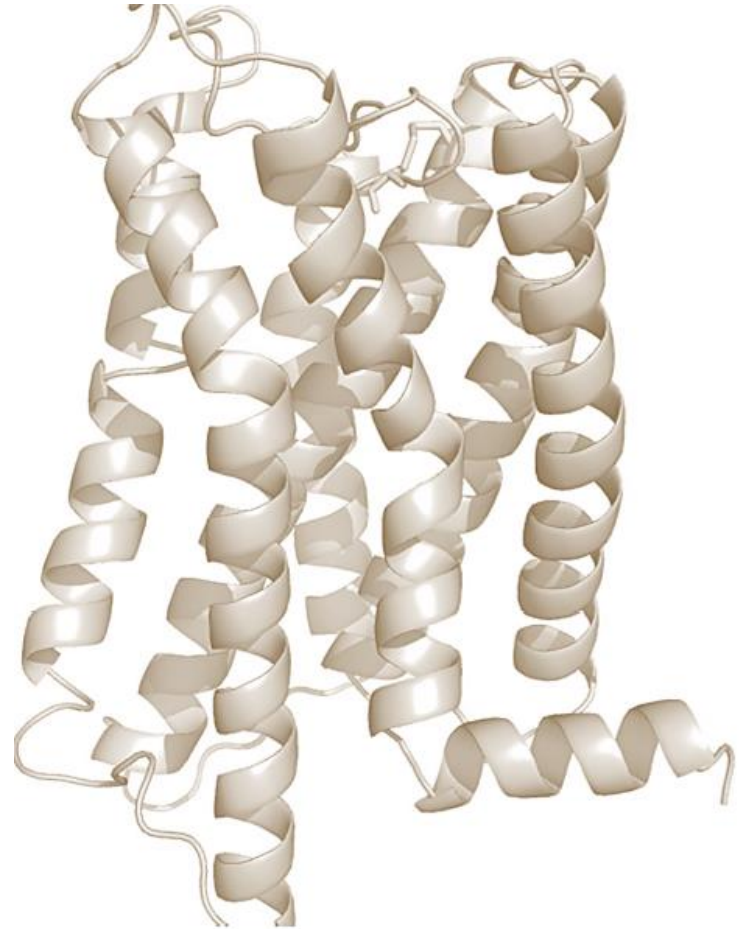
## # relax options

-relax:minimize\_bond\_angles  
-relax:minimize\_bond\_lengths  
-relax:jump\_move true  
-default\_max\_cycles 200  
-relax:min\_type lbfgs\_armijo\_nonmonotone  
-score:weights **membrane\_highres\_Menv\_smooth**  
-use\_bicubic\_interpolation  
-hybridize:stage1\_probability 1.0  
-sog\_upper\_bound 15

#### use ref2015\_cart if soluble protein

# 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.** Model selection

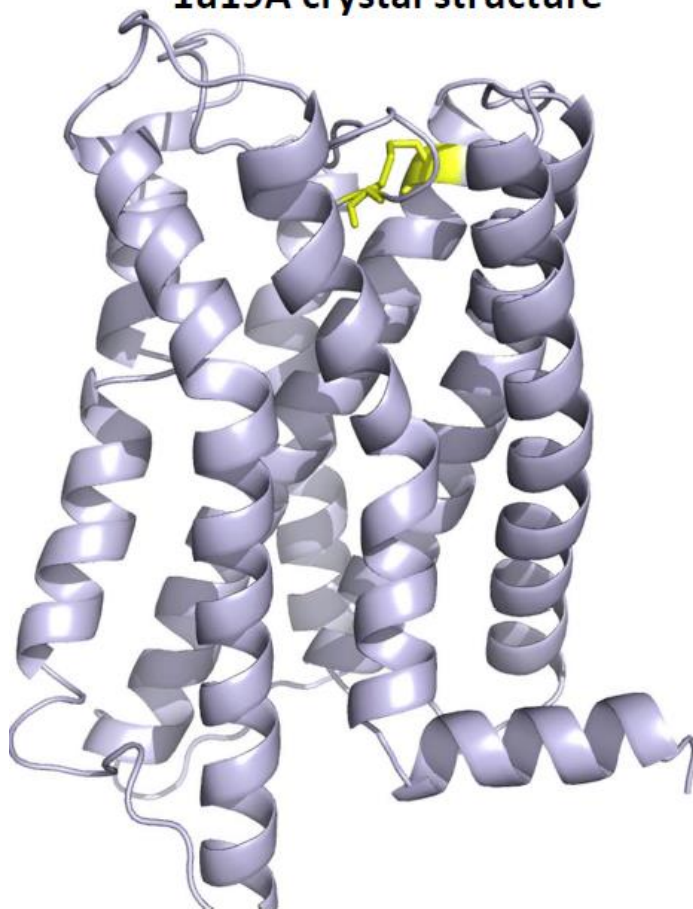


# Disulfide constraints

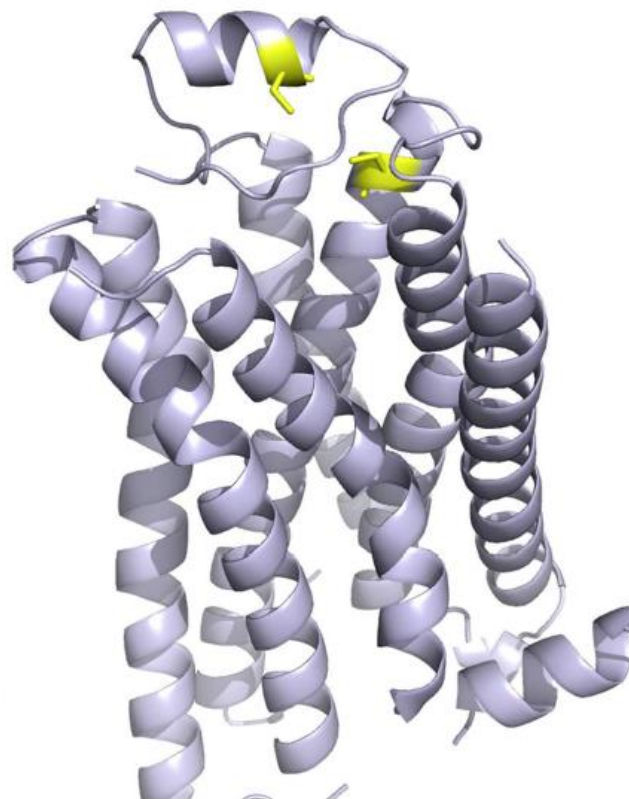
Find this file at `/rosetta_cm/4_relax/1u19.disulfide`

77 154

**1u19A crystal structure**



**1u19A threaded onto 2rh1A**



# Tutorial

Comparative modeling of Rhodopsin with three class A GPCR's

Four stages:

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



# 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](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.

[https://www.rosettacommons.org/demos/latest/tutorials/rosetta\\_cm/rosetta\\_cm\\_tutorial](https://www.rosettacommons.org/demos/latest/tutorials/rosetta_cm/rosetta_cm_tutorial)