ROSETTACM:
MULTI-TEMPLATE COMPARATIVE MODELING

Harikrishna Sekar Jayanthan
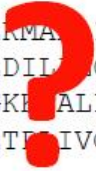
harikrishna.sekar@vanderbilt.edu

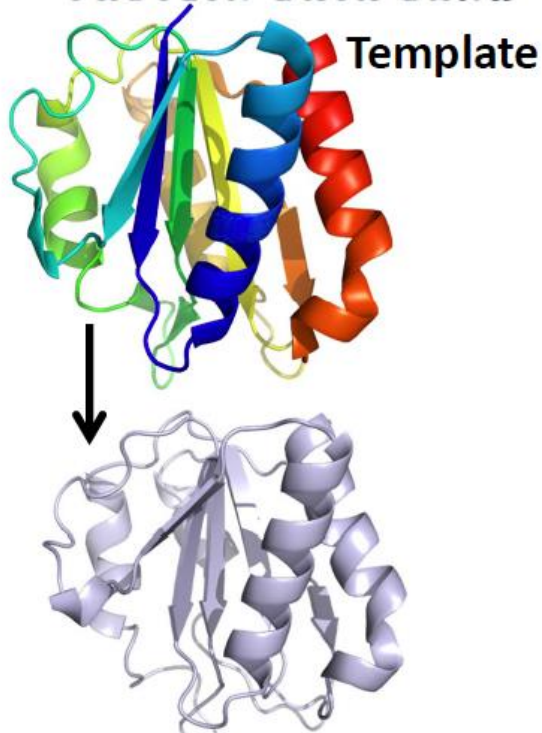Rosetta Workshop Nov 2018

Intro to Multi-template Comparative Modeling

Target

```
MKIVYWSGTGNTERMA   IAKGIIESGKDVNTI
NVSDVNIDELLNEDIL   GCSAMGDEVLEESEF
EPFIEEISTKISGKE    ALFGSYGWGDGKWMRDF
EERMNGYGCVVVETE    IVQNEPDEAEQDCIEFG
KKIANI
```

RCSB PDB
PROTEIN DATA BANK

Template

- **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
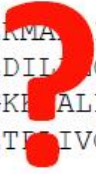 - Loop modeling protocol is internal

Nomenclature Note
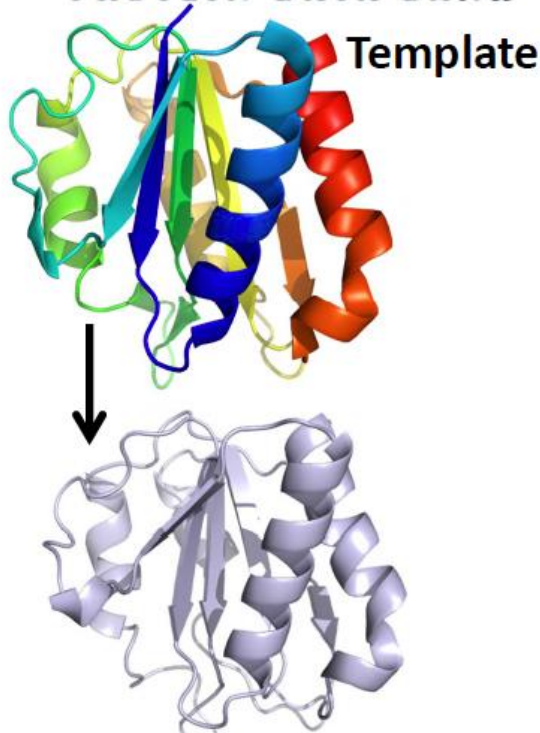- Comparative Modeling = Homology Modeling in the land of Rosetta

Identifying Template Structures

Target

MKIVYWSGTGNTEKMA   IAKGIIESGKDVNTI
NVSDVNIDELLNEDIL   GCSAMGDEVLEESEF
EPFIEEISTKISGKE ALFGSYGWGDGKWMRDF
EERMNGYGCVVVETI IVQNEPDEAEQDCIEFG
KKIANI

?

+ RCSB **PDB**
PROTEIN DATA BANK

Template

- **<u>Sequence Similarity</u>:** compare proteins based on amino acid properties alone (BLAST, PSI-BLAST)

  - **Suitable Templates:** ideally have >30% sequence identity to target

- **<u>Fold Recognition</u>:** 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
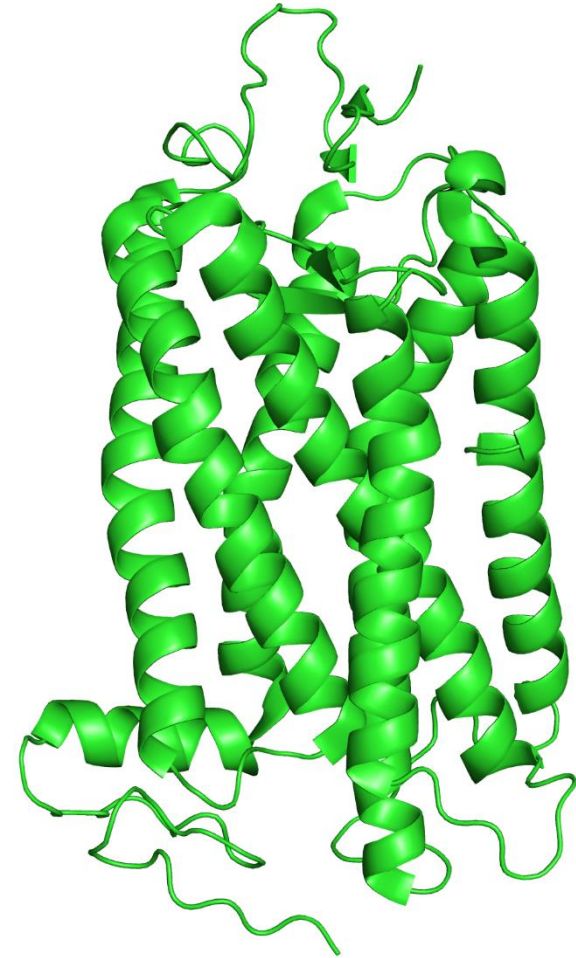
Low Identity Templates

template	PDB ID	Seq ID
β2 adrenergic receptor	2RH1	16
β1 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)

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--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM
ANFNKIFL------PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF
---DEVWVVGMGIVMS---LIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVM
-------IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
```
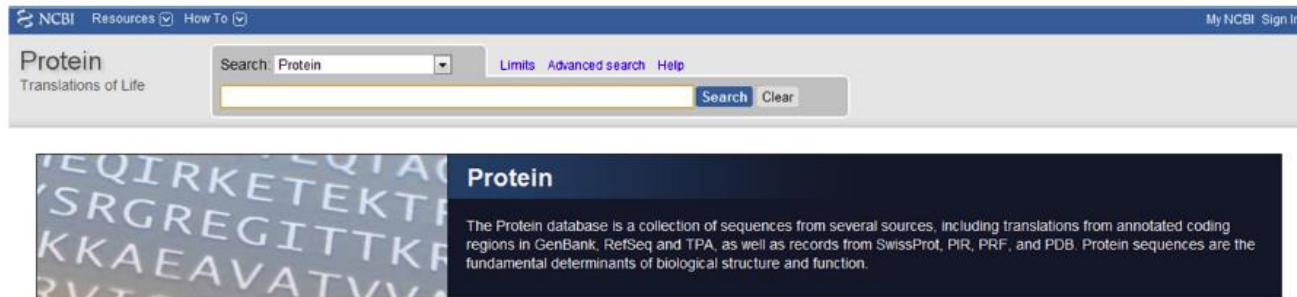
Target Sequence

Bovine rhodopsin (PDB: 1u19)

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

>1u19

PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVAD
LFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLGGEIALWSLVVL
AIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSR
YIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLV
FTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIF
THQGSDFGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCG

Template PDBs

Human β2-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/*

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---LIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVM
3EML_A        -------IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
                     :     :   : : .   * *.  .    ..*:.  :      *: **:  .
```


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:

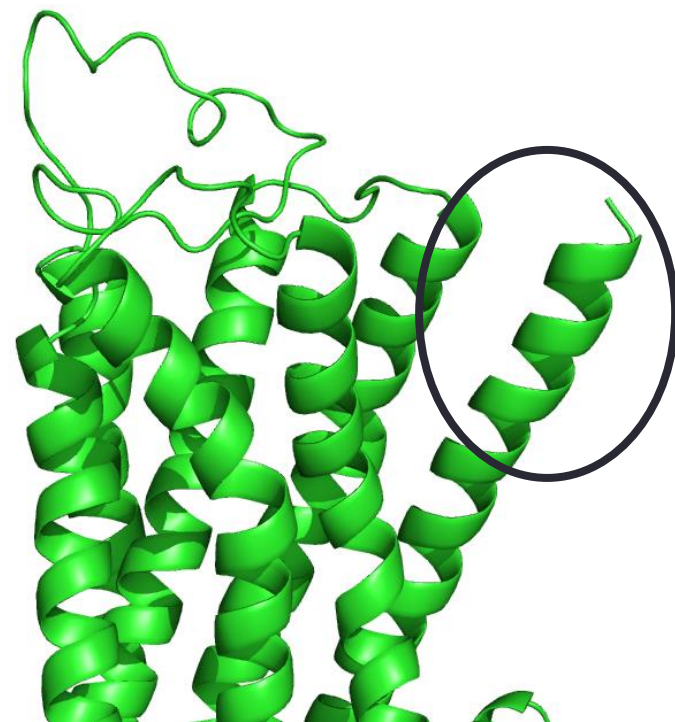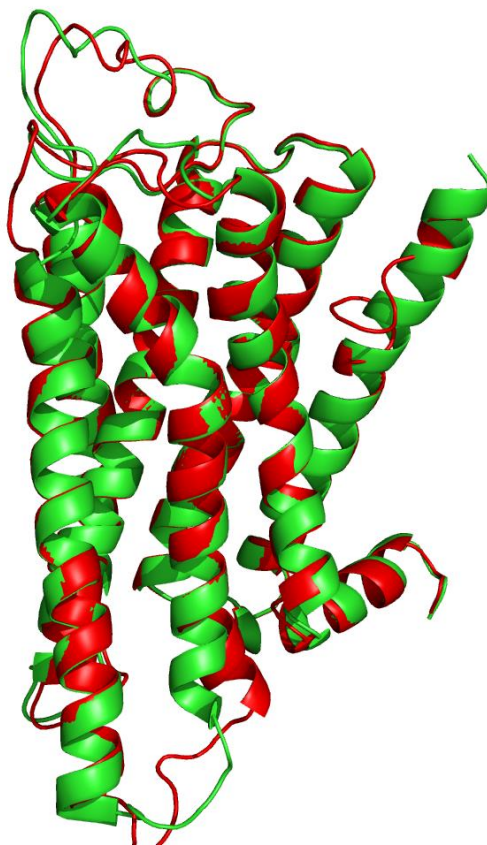|																																	
1u19	-	-	-	-	-	-	-	P	W	Q	F	S	M	L	A	A	Y	M	F	L	L	I	M	L	G	F	P	I	N	F	L	T	L	Y	V	T	V	Q	H
3ODU_A	A	N	F	-	-	-	-	-	N	K	I	F	L	P	T	I	Y	S	I	I	F	L	T	G	I	V	G	N	G	L	V	I	L	V	M	G	Y	Q
2RH1_A	-	-	-	D	-	-	-	E	V	W	V	V	G	M	G	I	V	M	S	L	I	V	L	A	I	V	F	G	N	V	L	V	I	T	A	I	A	K	F
3EML_A	-	-	-	-	-	-	-	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/*
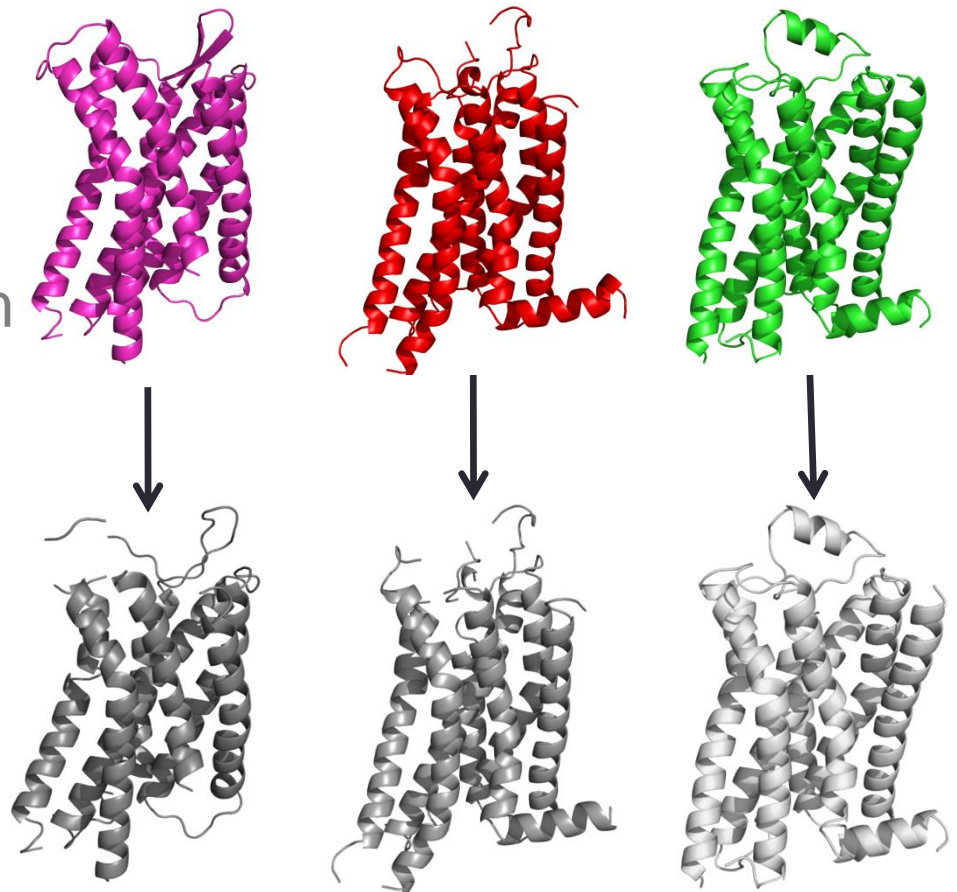
   (1u19_2rh1.grishin, 1u19_3eml.grishin, 1u19_3odu.grishin)

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

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
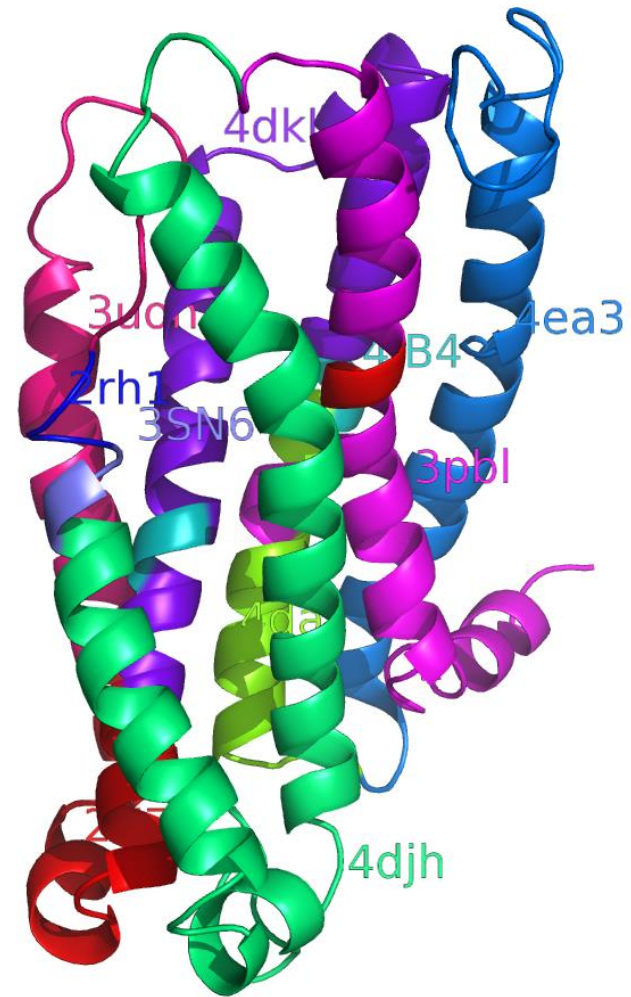
Thread Coordinates

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
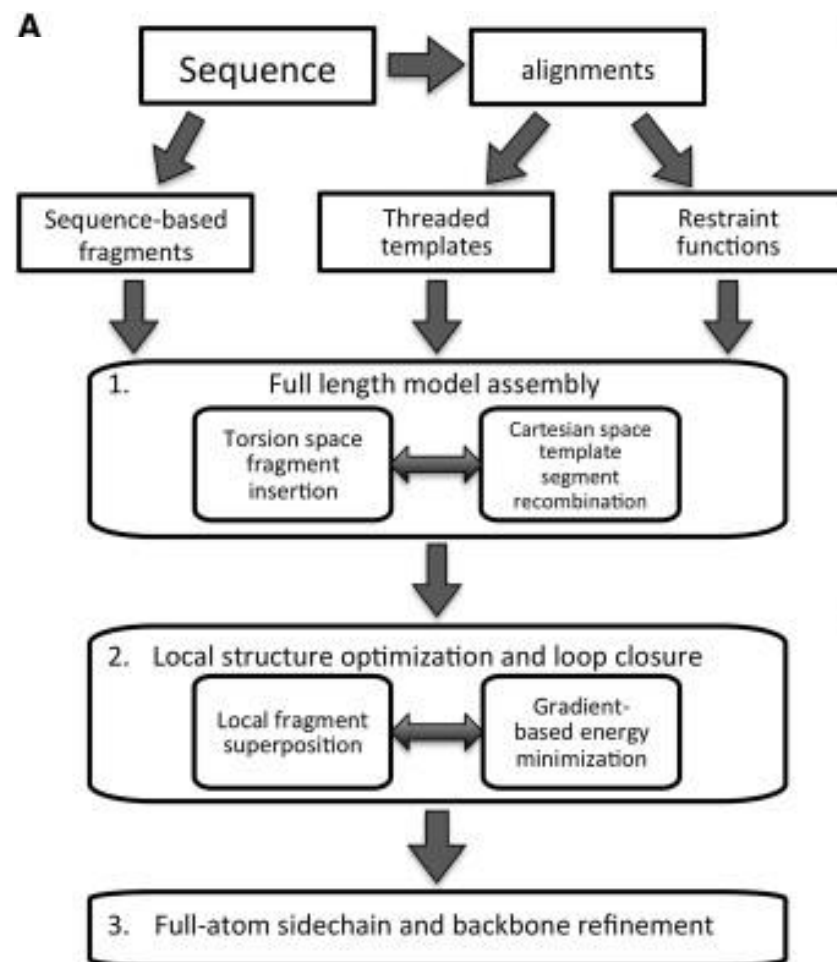
- **Step 4**: Full-atom refinement

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
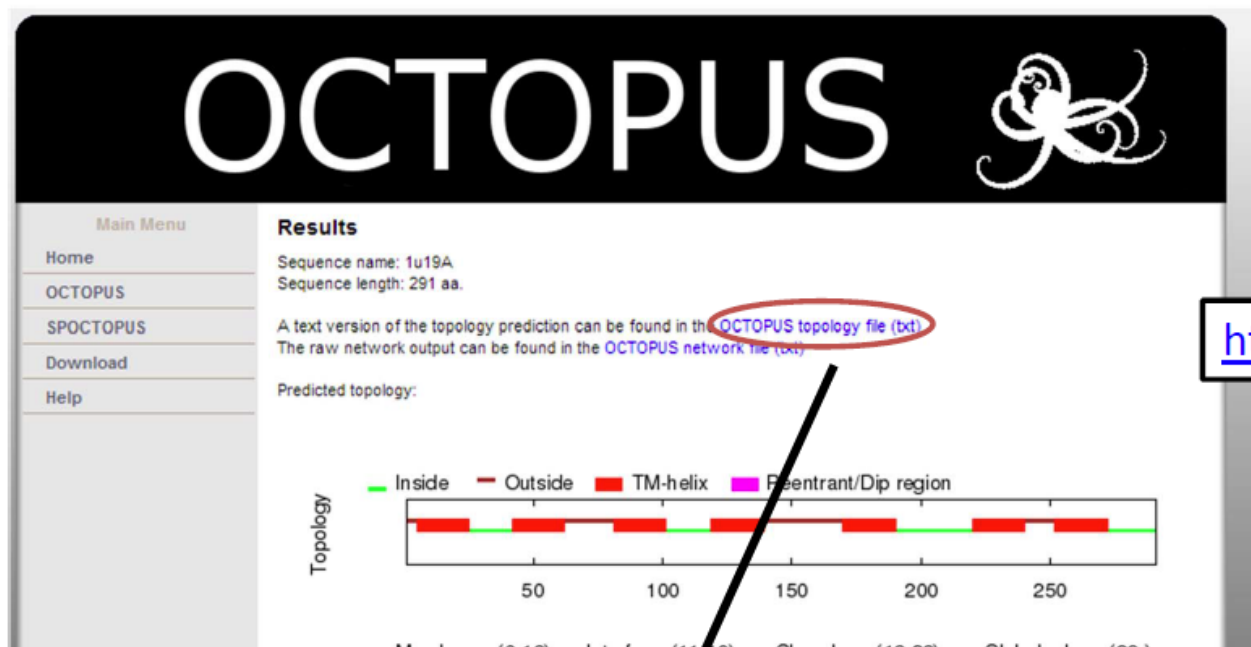

Song, et al. 2013

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*

http://octopus.cbr.su.se/

octopus2span.pl
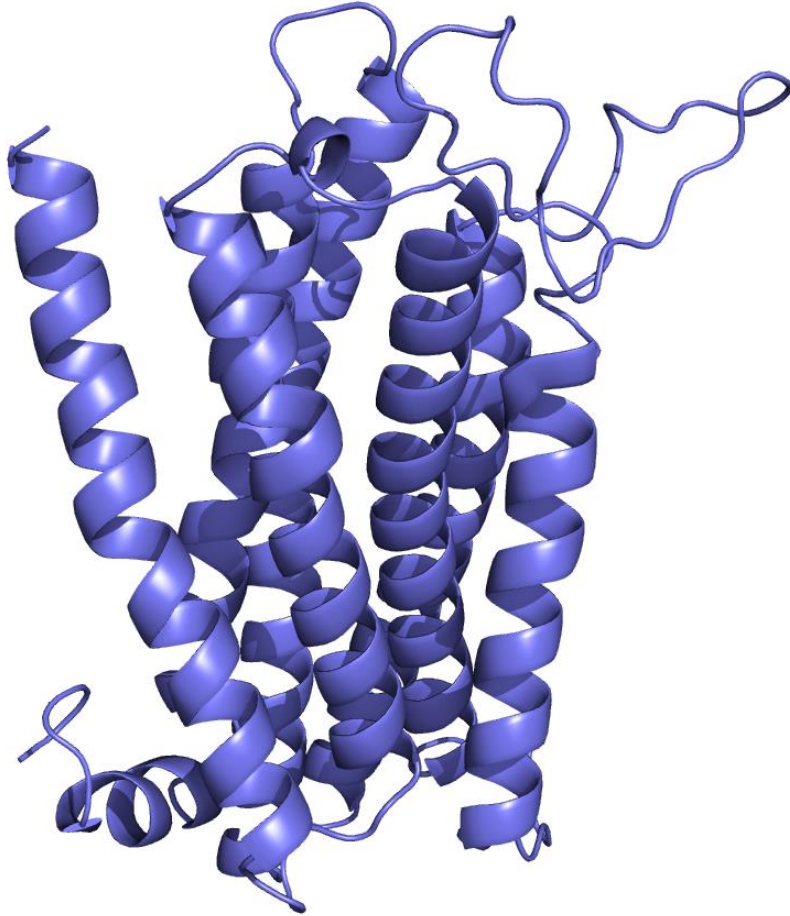1u19A.octopus
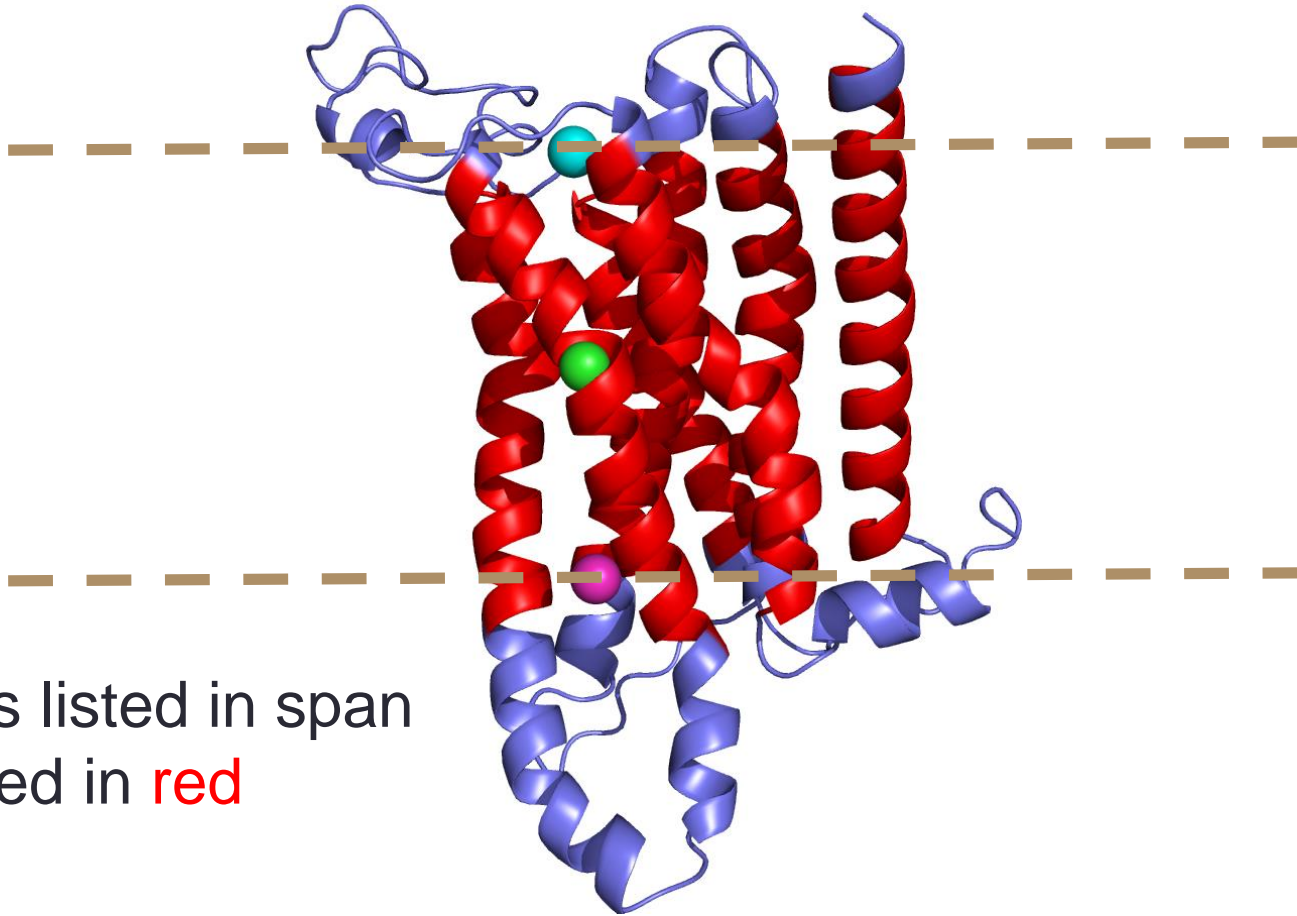
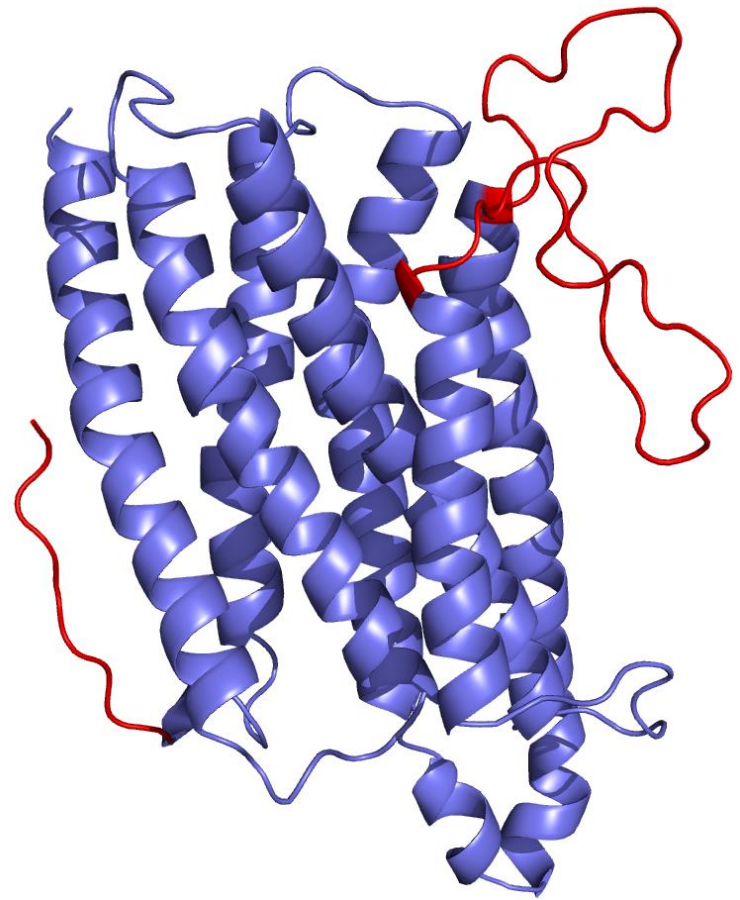# Rosetta Membrane

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                           #### your target sequence
-parser:protocol input_files/rosetta_cm.xml
-out:pdb
```

```
#Initialize membrane                                #### only if modeling a membrane protein
-in:file:spanfile 1u19.span
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-rg_reweight .1
-restore_talaris_behavior
```

```
# 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 ref2015_cart if soluble protein
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15
```
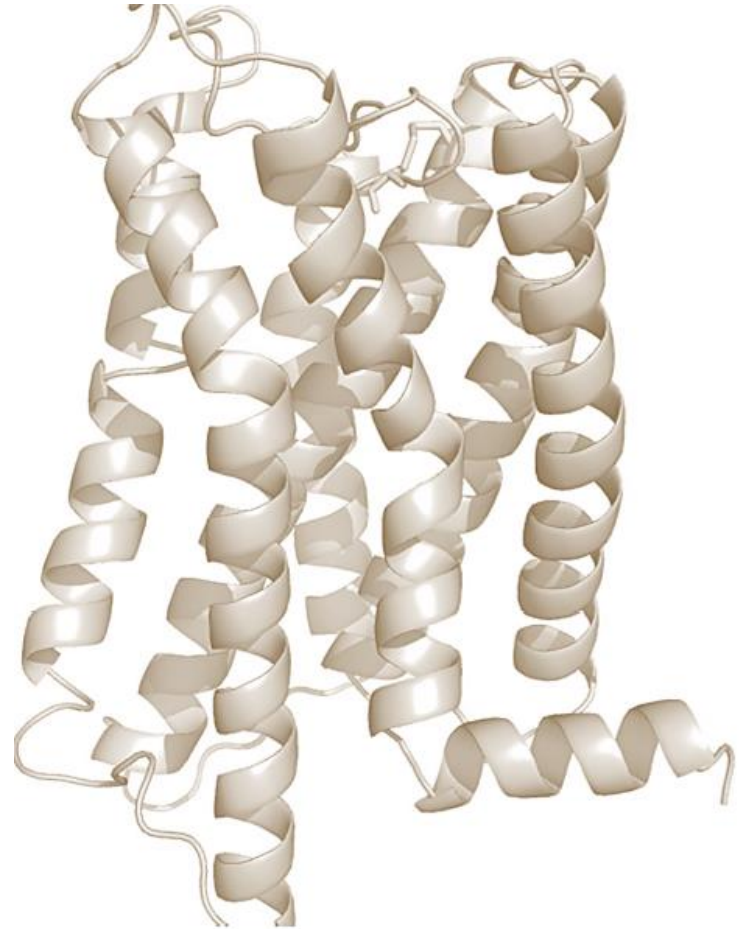
Comparative Modeling Protocol

- **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
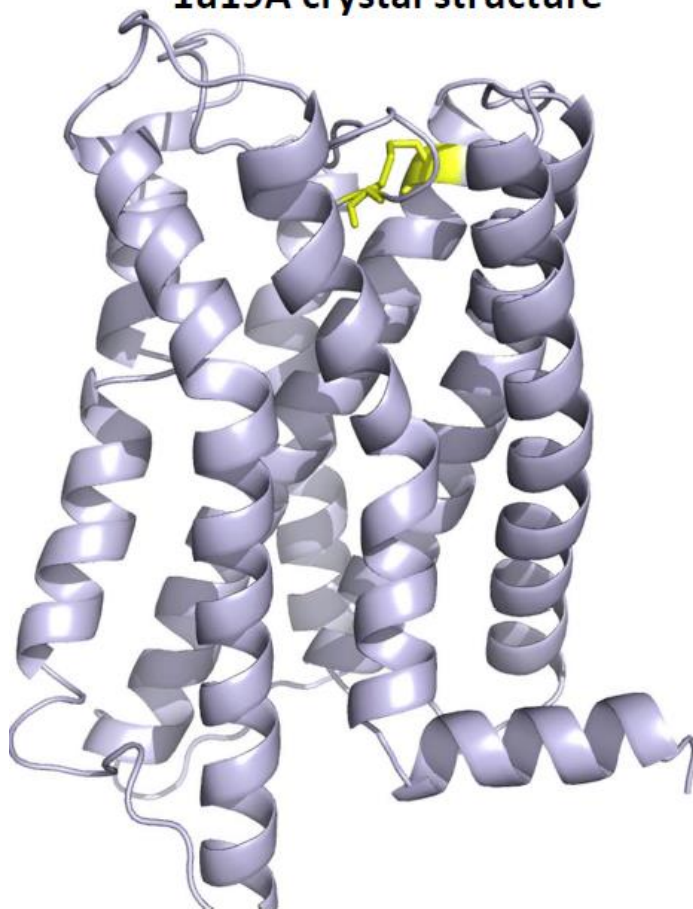- **Step 4**. Model selection
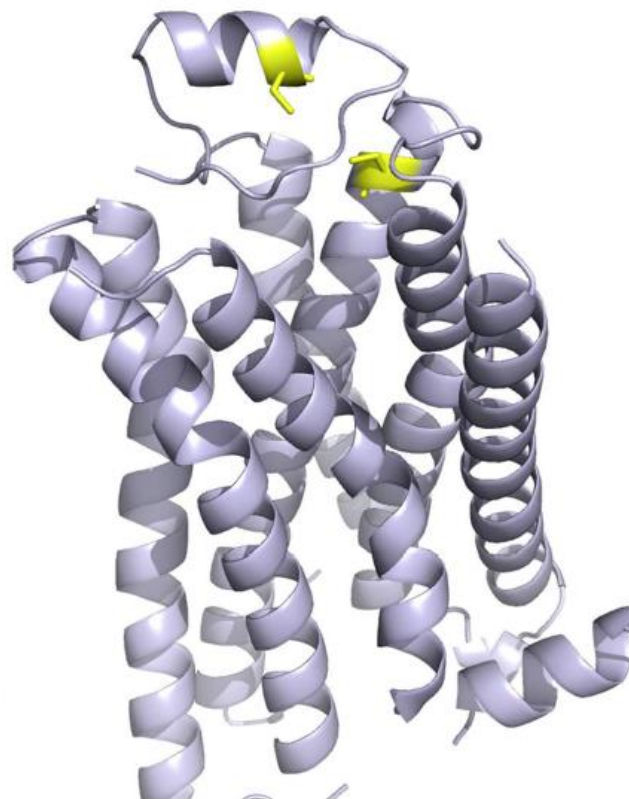
Disulfide constraints

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

- **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