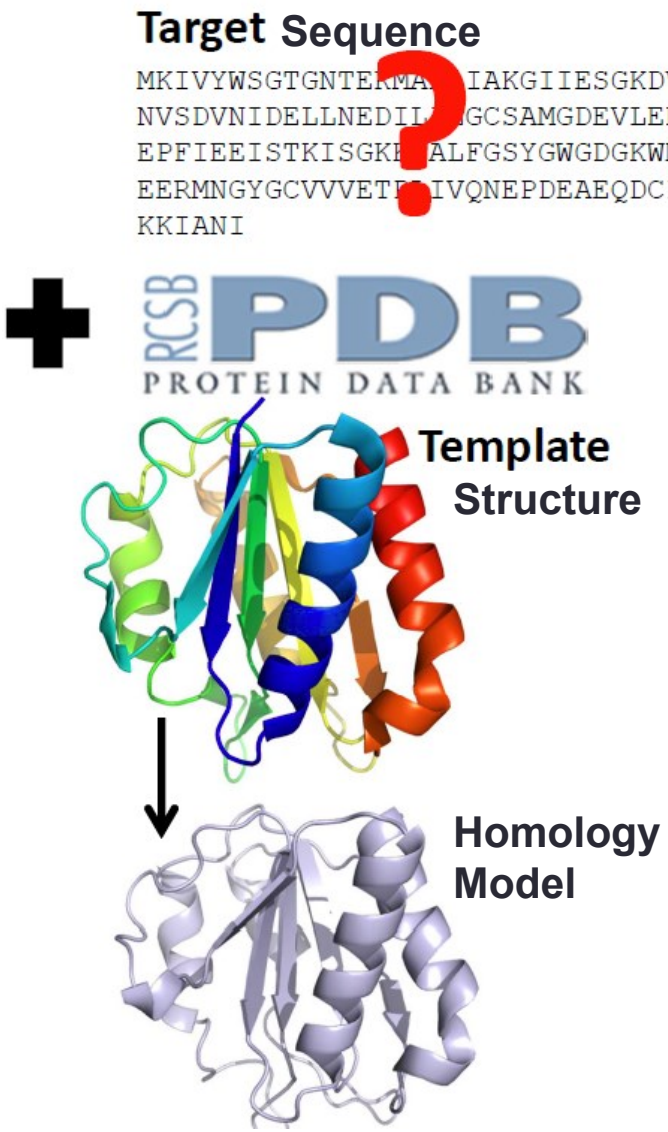
RosettaCM:
MULTI-TEMPLATE COMPARATIVE MODELING

Oanh Vu
oanh.t.vu.2@vanderbilt.edu
Georg Kuenze
georg.kuenze@vanderbilt.edu
Brian Bender
brian.j.bender@vanderbilt.edu

Introduction: RosettaCM Homology Modeling

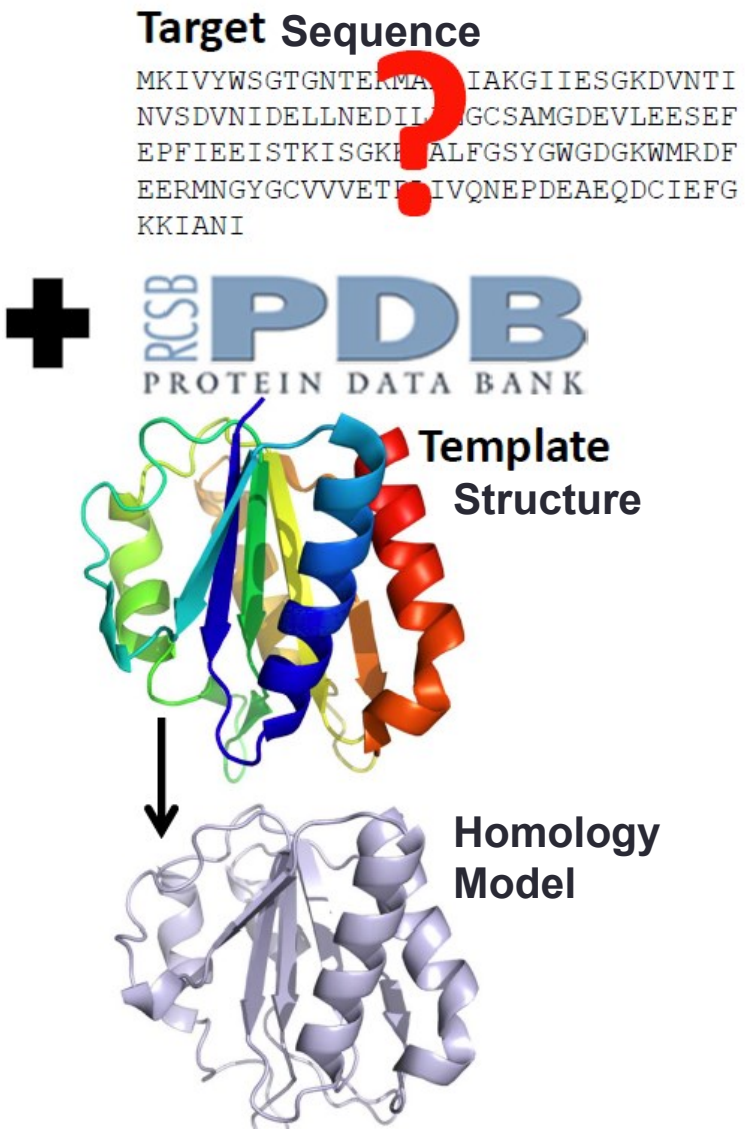


- **Single Template Modeling:**
  - Single template as input
  - Uses sequence and template derived fragments
  - Used when available templates have very high identity (>60%)
- **Multiple Template Modeling:**
  - Multiple templates as input
  - Combine sections of multiple threaded models and sequence derived fragments
  - Used when available templates have low identity (30-50%)

Nomenclature Note

- Comparative Modeling = Homology Modeling in the land of Rosetta

Identifying Template Structures



- **Similarity of Sequences :**
  - compare proteins based on amino acid sequences (BLASTP using PDB as search database)
  - 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: Dopamine D3 receptor

- PDB ID: 3pbl
- Class A G-protein coupled receptor (GPCR)
- No high identity templates

- 7 transmembrane helices
- 3 extracellular loops, 3 intracellular loops
- Highly conserved GPCR residues

Low Identity Templates

- It is advisable to use multiple templates due to the low sequence identity in available templates

Template	PDB ID	% Seq id
β2-adrenoceptor	3SN6	36
5-HT1B receptor	4IAR	32
β2-adrenoceptor	3D4S	34
5-HT2B receptor	5TVN	32
M1 receptor	5CXV	32
H1 receptor	3RZE	31
M4 receptor	5DSG	29
A2A receptor	2YDO	28
A1 receptor	5N2S	27

Comparative Modeling Protocol

- **Step 1**: Align target sequence to template sequences
- **Step 2**: Partial-thread the target sequence onto template structures
- **Step 3**: Combine pieces from different templates using RosettaCM Hybridize

```
-----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM
ANFNKIFL------PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF
---DEVWVVGMGIVMS---LIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVM
-------IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
```

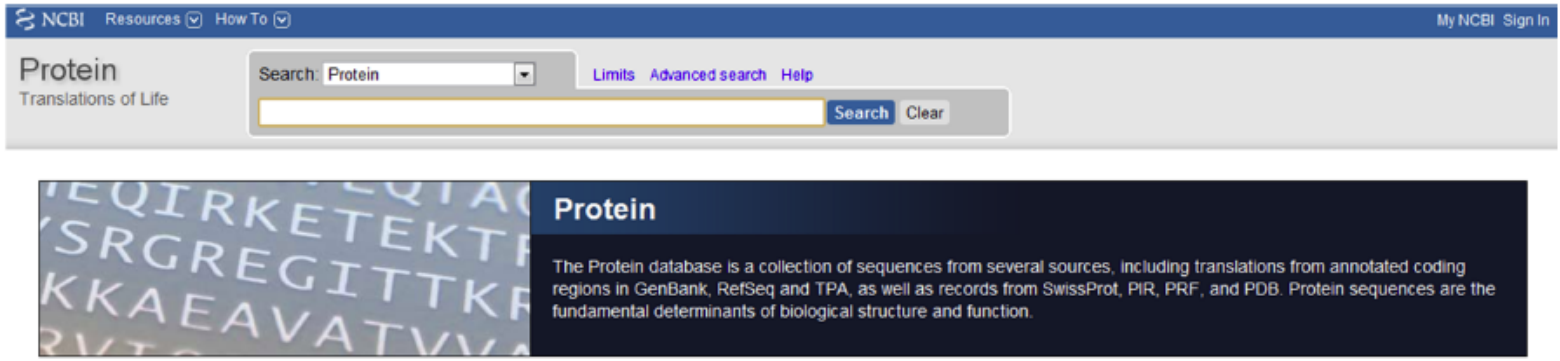
Target Sequence

(PDB: Dopamine D3 receptor 3pbl)

Find this file at */rosetta_cm/demo/input_files/3pbl.fasta*

>3pbl

YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATL
VMPWVVYLEVTGGVWNFSRICCDVFVTLDVMMCTASIWNLCAISIDR
YTAVVMPVHYQHGTGQSSCRRVALMITAVWVLAFAVSCPLLFGFNTT
GDPTVCSISNPDFVIYSSVVSFYLPFGVTVLVYARIYVVLKQRRRKAAA
AAAAAGVPLREKKATQMVAIVLGAFIVCWLPFFLTHVLNTHCQTCHVS
PELYSATTWLGYVNSALNPVIYTTFNIEFRKAFLKILSC



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

Template PDBs

Human 5HT-1B receptor (PDB: 4iar)

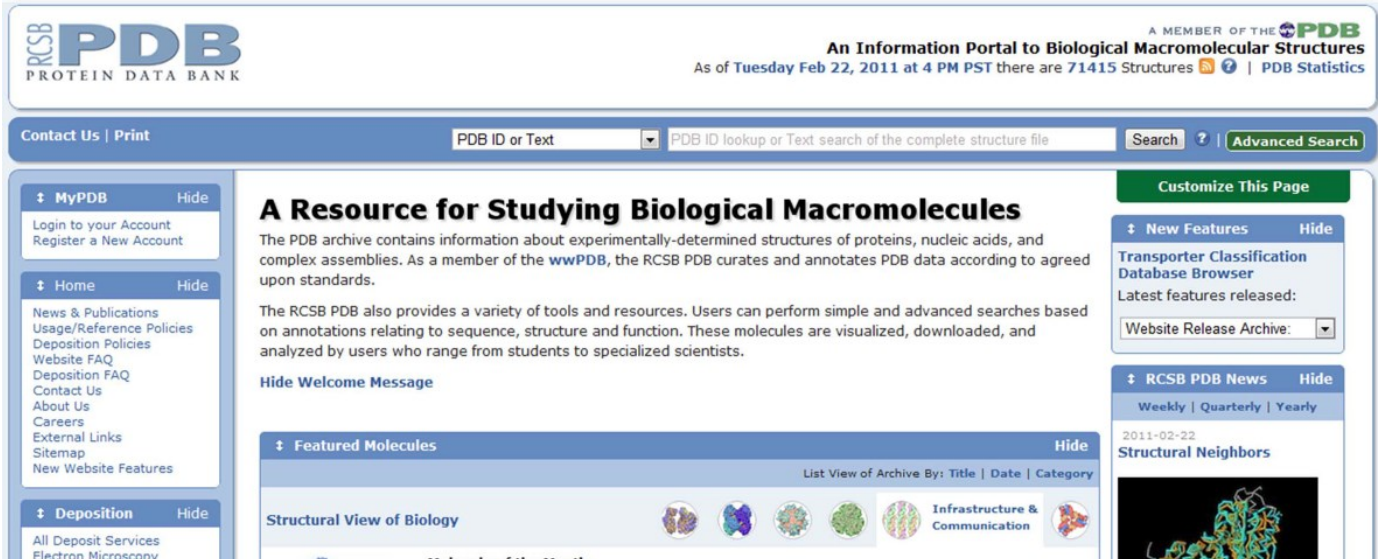
Human beta1-adrenoceptor (PDB: 4bvn)

Human B2-adrenergic receptor (PDB: 2rh1)

Human M4 muscarinic acetylcholine receptor (PDB: 5dsg)

Human M1 muscarinic acetylcholine receptor (PDB: 5cxv)

Find these files at */rosetta_cm/template_pdbs/original_files/*

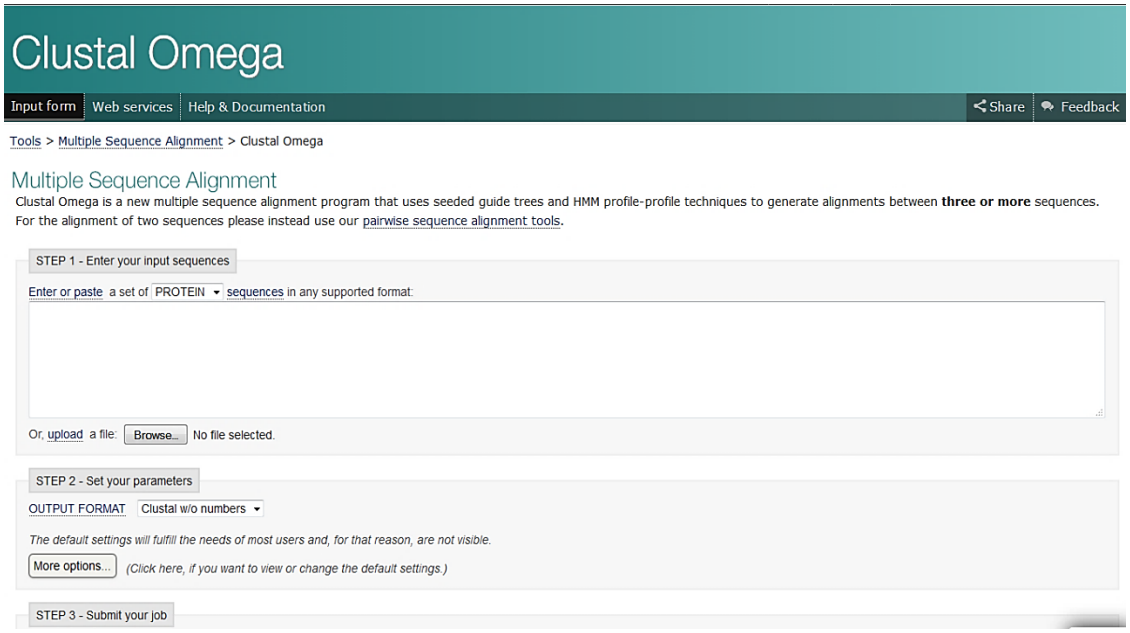


Multiple Sequence Alignment

Find this file at */demo/alignment_files/3pbl_alignments.txt*

```
CLUSTAL O(1.2.4) multiple sequence alignment


5cxv       -------KGPWQVAFIGITTGLLSLATVTGNLLVLISFKVNTELKTVNNYFLLSLACADL
5dsg       GPSSHNRYETVEMVFIATVTGSLSLVTVVGNILVMLSIKVNRQLQTVNNYFLFSLACADL
3pbl       ---------------YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADL
4iar       YIYQDSISLPWKV-LLVMLLALITLATTLSNAFVIATVYRTRKLHTPANYLIASLAVTDL
2rh1       -------DEVWVV-GMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADL
4bvn       ------LSQQWEA-GMSLLMALVVLLIVAGNVLVIAAIGSTQRLQTLTNLFITSLACADL
```



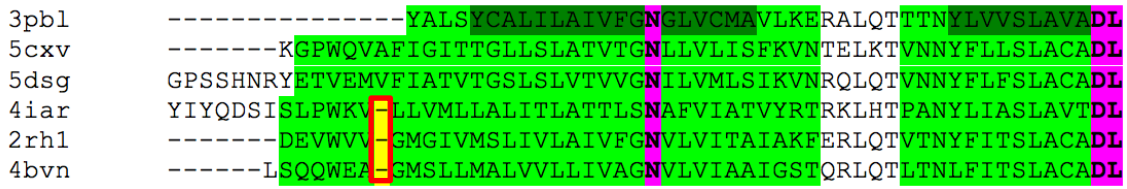
http://www.ebi.ac.uk/Tools/msa/clustalo/

Adjusting multiple sequence alignment

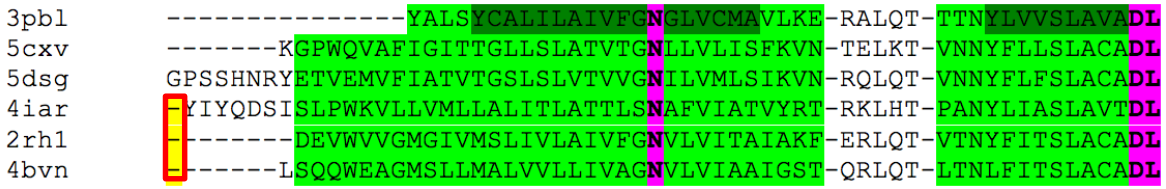
Experimental expectations:

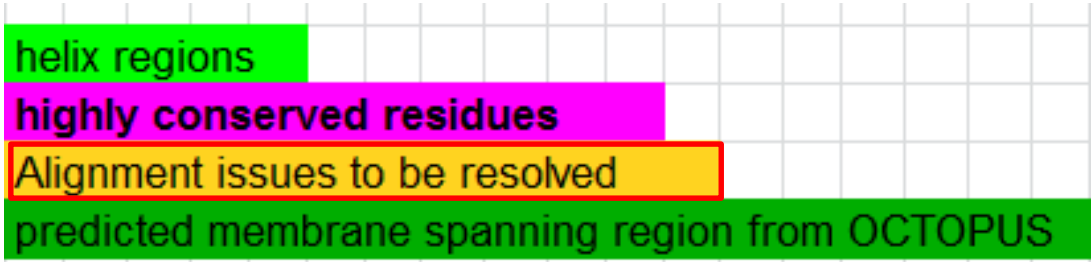
- Highly conserved residues
- Secondary structure elements

Raw ClustalO alignment:



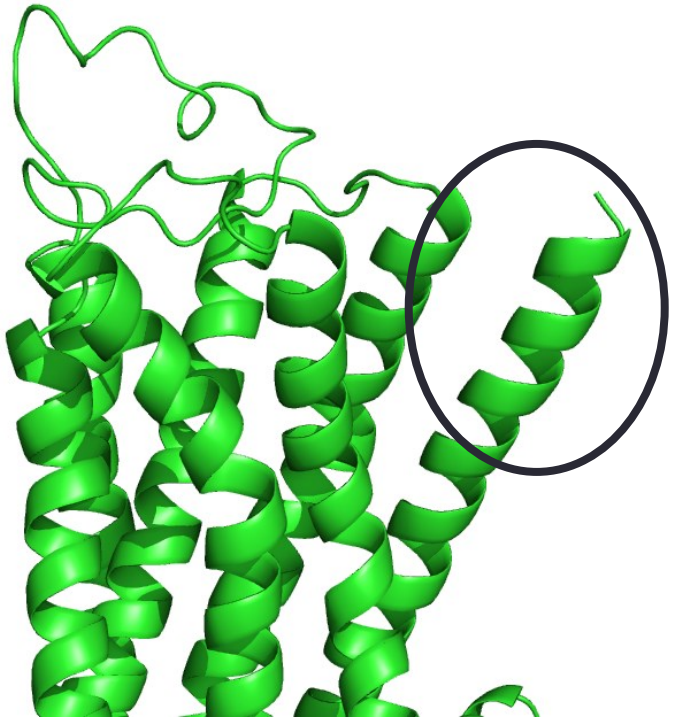
Adjusted alignment:





Removing helix gaps

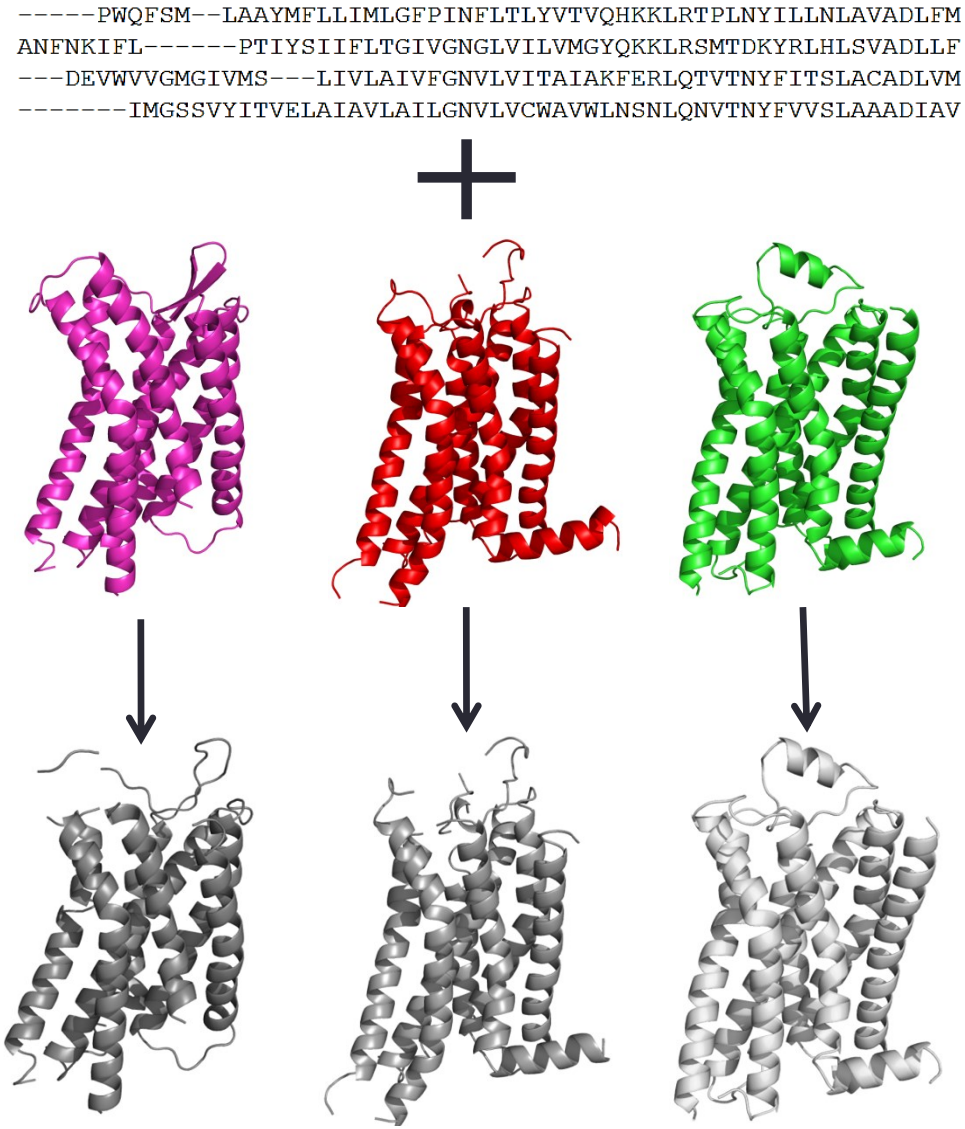
Example model using raw alignment



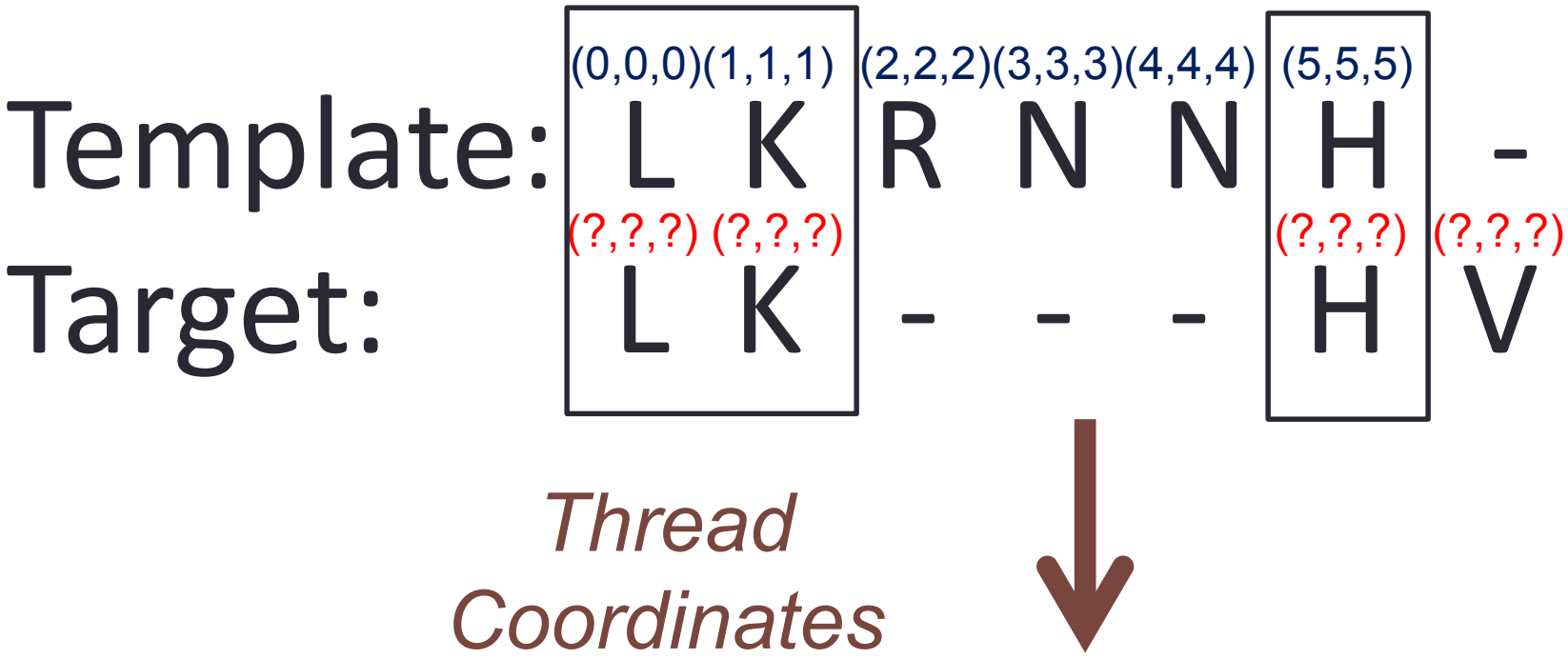
Example model using adjusted alignment

Comparative Modeling Protocol

- **Step 1**: Align target sequence to template sequences
- **Step 2**: Partial-thread the target sequence onto template structures
- **Step 3**: Combine pieces from different templates using RosettaCM Hybridize



Threading


(0,0,0)(1,1,1)
(2,2,2)(3,3,3)(4,4,4)
(5,5,5)
Template: L K R N N H -
(?,?,?) (?,?,?)
(?,?,?) (?,?,?)
Target: L K - - - H V
Thread Coordinates
(0,0,0) (1,1,1) (5,5,5)
Target: L K H V

Grishin Format Alignments Needed for Rosetta Threading

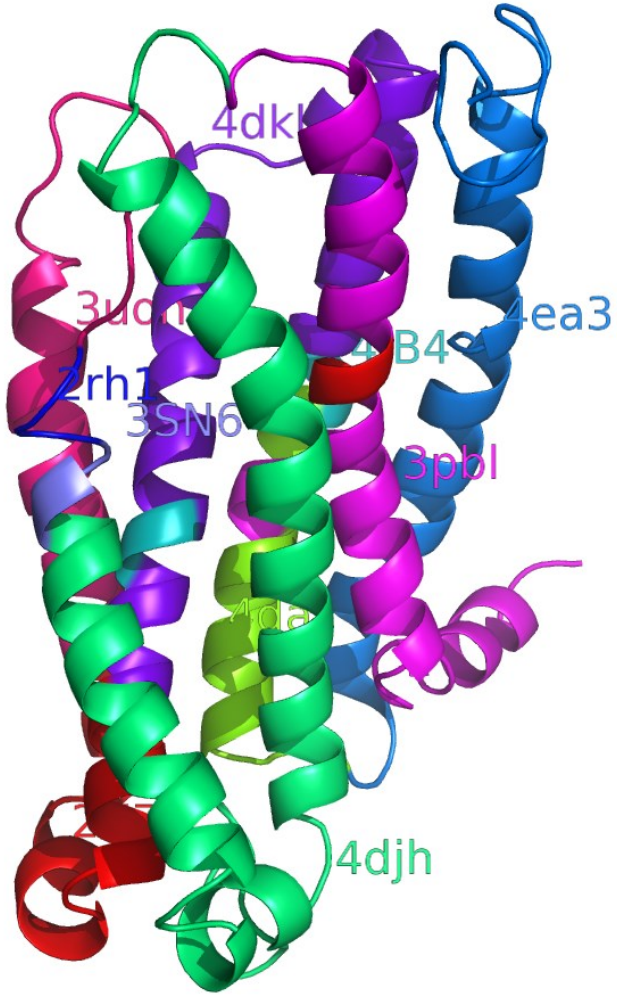
- ClustalO format:
  - All sequences in one file
  - Sequences broken up over several lines
- Grishin format:
  - One file per alignment pair
  - Sequences continuous over one line each
  - Contains header information
  - Due to complicated format, we have provided a script for conversion `make_alignment_files.sh` for your use back home

Find converted Grishin alignment files at */rosetta_cm/demo/alignment_files/*

(2rh1.aln 4bvn.aln 4iar.aln 5cxv.aln 5dsg.aln)

Comparative Modeling Protocol

- **Step 1**: Align target sequence to template sequences
- **Step 2**: Partial-thread the target sequence onto template structures
- **Step 3**: Combine pieces from different templates using RosettaCM Hybridize

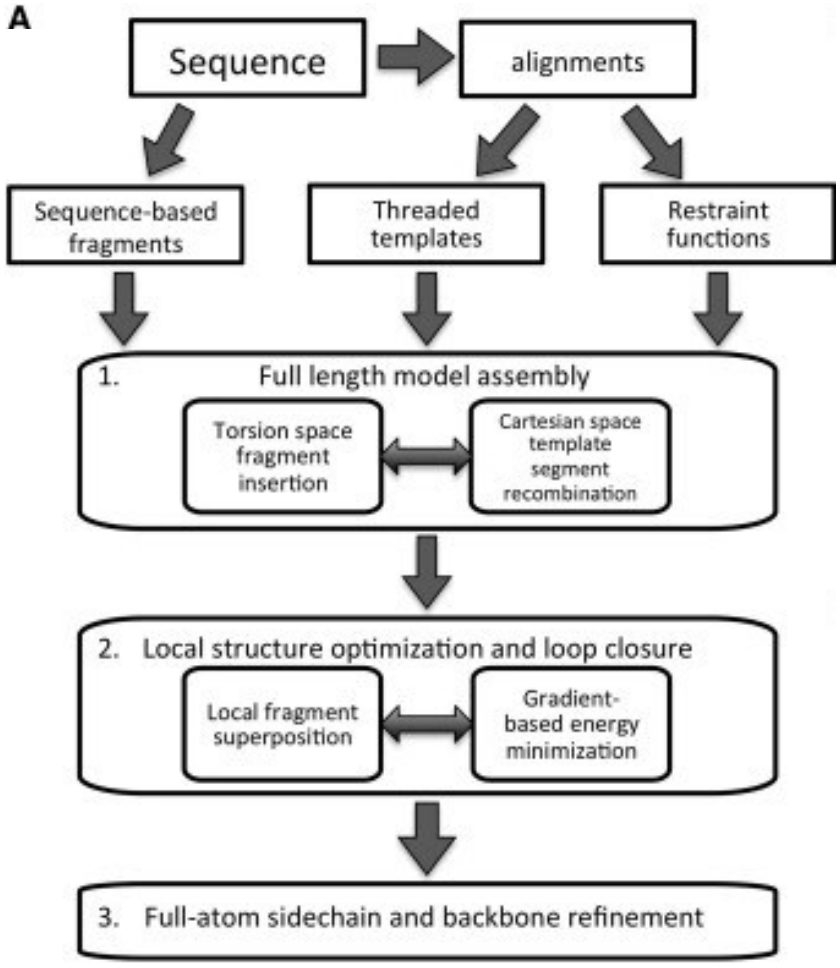


RosettaCM: Three Stages

1. Generate initial models from template alignments

2. **Explore deviations from templates and close loops in 2 steps :**
   - 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

Bare minimum:

- Partial-threaded structures
- Mover definition and options

Specific to membrane proteins (not needed if modeling soluble proteins):

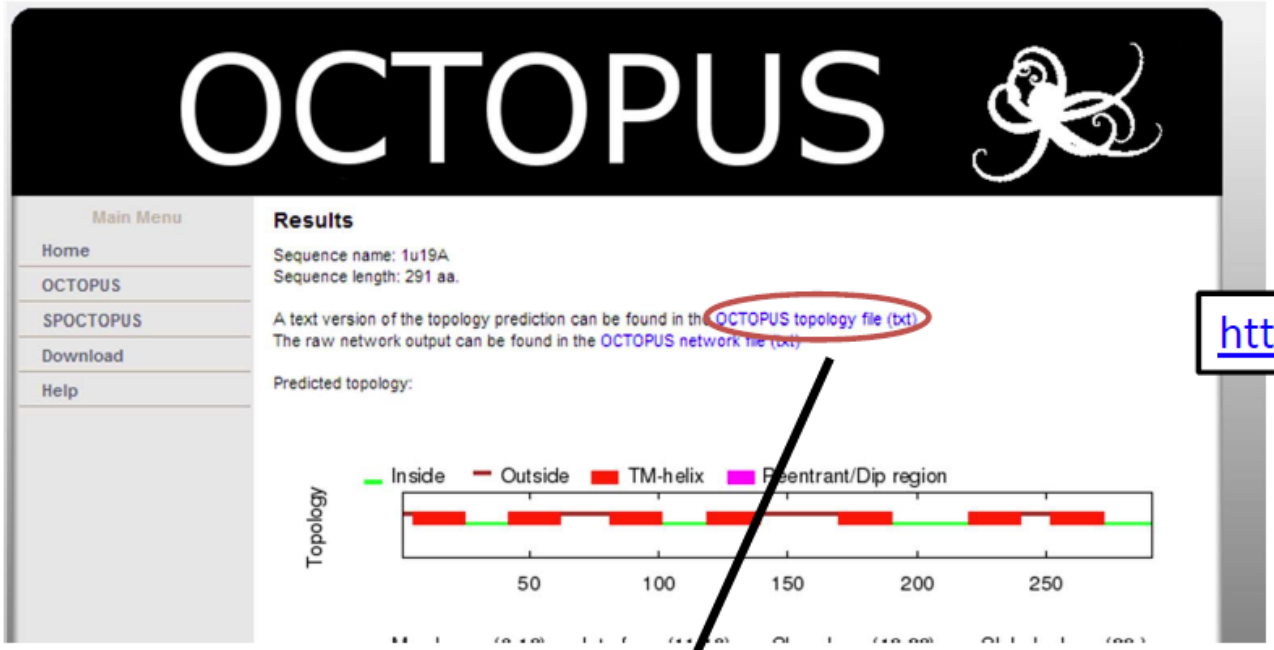
- Membrane spanning regions (span file)
- Membrane weight patches

Optional files based on available information:

- Constraint information (eg. atom pair connectivity)
- Disulfide Connectivity

Membrane spanning regions

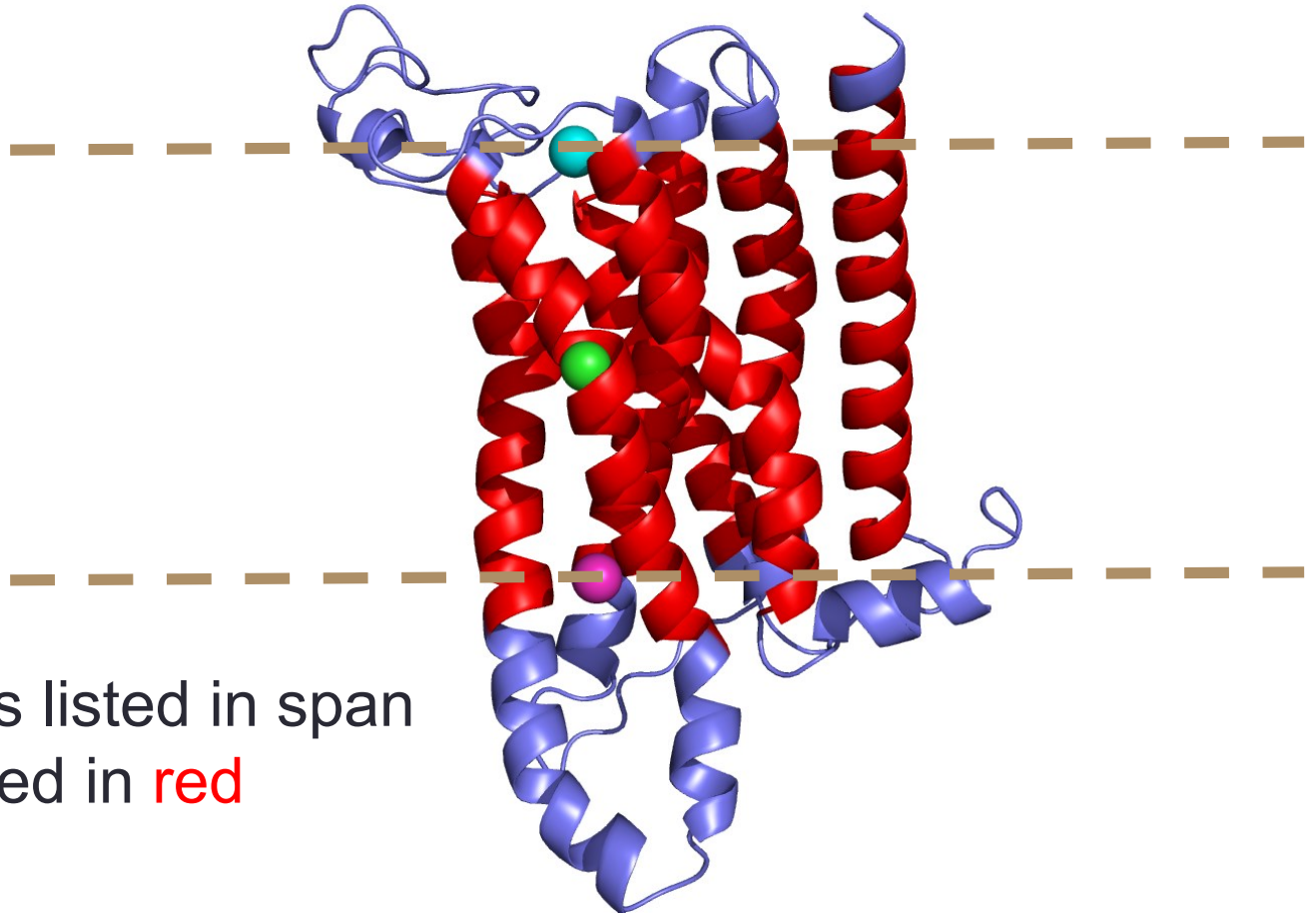
Find this file at */rosetta_cm/demo/input_files/3pbl.span*



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

octopus2span.pl 3pbl.octopus

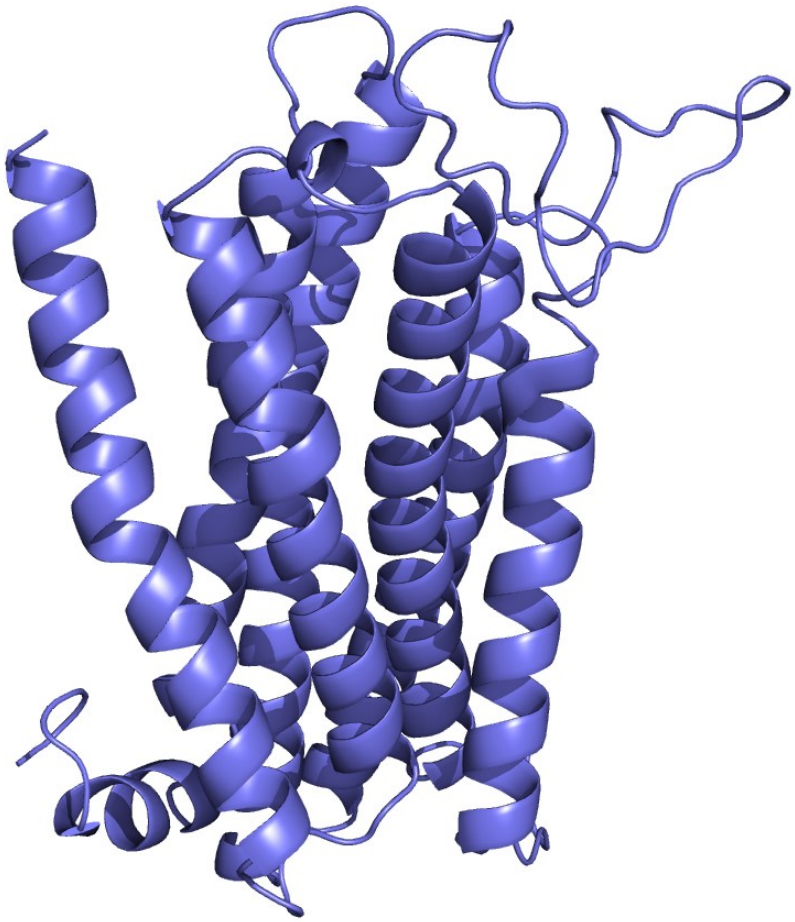
Rosetta Membrane



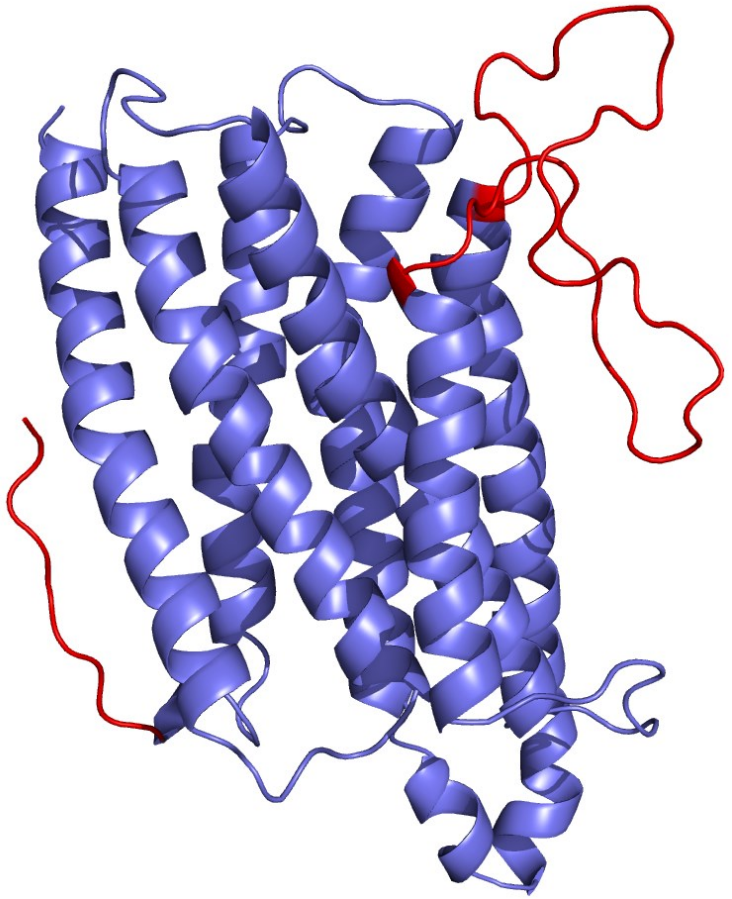
Residues listed in span file colored in red

Why use membrane scoring terms?

With membrane penalties/weights



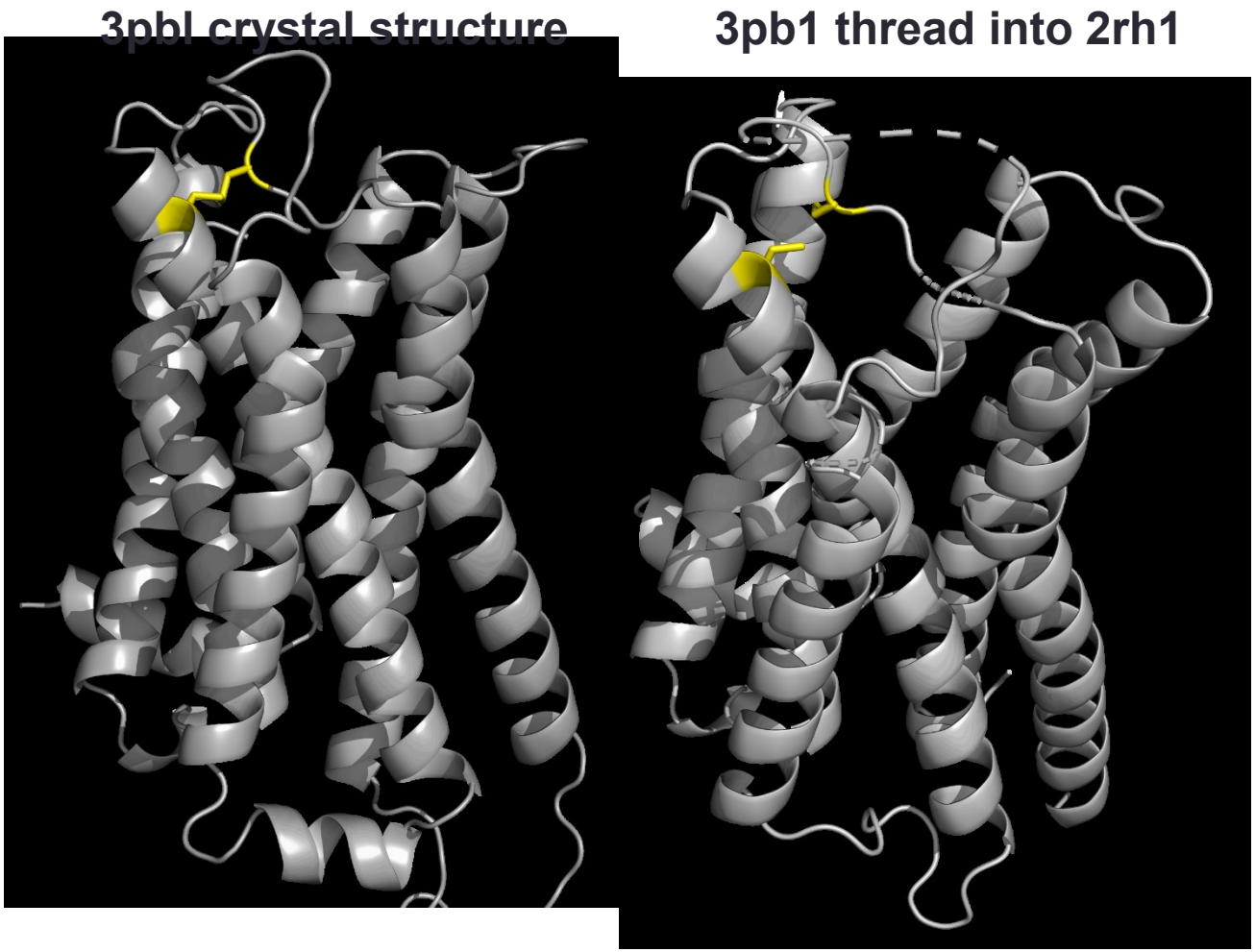
Without membrane penalties/weights



Disulfide constraints

Find this file at */rosetta_cm/demo/input_files/3pbl.disulfide*

72 150



RosettaCM XML

/rosetta_cm/demo/input_files/rosetta_cm.xml

```
<SCOREFXNS>
        <ScoreFunction name="stage1" weights="input_files/stage1_membrane.wts" symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="1"/>
        </ScoreFunction>
        <ScoreFunction name="stage2" weights="input_files/stage2_membrane.wts" symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name="fullatom" weights="input_files/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">
            <Reweight scoretype="cart_bonded" weight="0.5"/>
            <Reweight scoretype="pro_close" weight="0"/>
        </ScoreFunction>
</SCOREFXNS>
```

*Find all **.wts** files in */rosetta_cm/ demo/input_files*

RosettaCM XML

/rosetta_cm/demo/input_files/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" realign_domains="0"  disulf_file="input_files/3pbl.disulfide“
fa_cst_file=“fullatom.cst”>

            <Template pdb="threaded_pdbs/4iar_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/4bvn_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/2rh1_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/5dsg_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/5cxv_out.pdb" cst_file="AUTO" weight="1.000" />

        </Hybridize>
        <ClearConstraintsMover name="clearconstraints"/>
        <FastRelax name="relax" scorefxn="membrane" repeats=”1" dualspace="1"
bondangle="1"/>
</MOVERS>
<OUTPUT scorefxn=“membrane”/>
```

RosettaCM Options

/rosetta_cm/3_hybridize/rosetta_cm.options

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

```
#Initialize membrane                                     #### only if modeling a membrane protein
-in:file:spanfile input_files/3pbl.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 input_files/stage3_rlx_membrane.wts       #### use ref2015_cart if soluble protein
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15
```

Tutorial

Comparative modeling of D3 receptor with five class A GPCR templates

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.