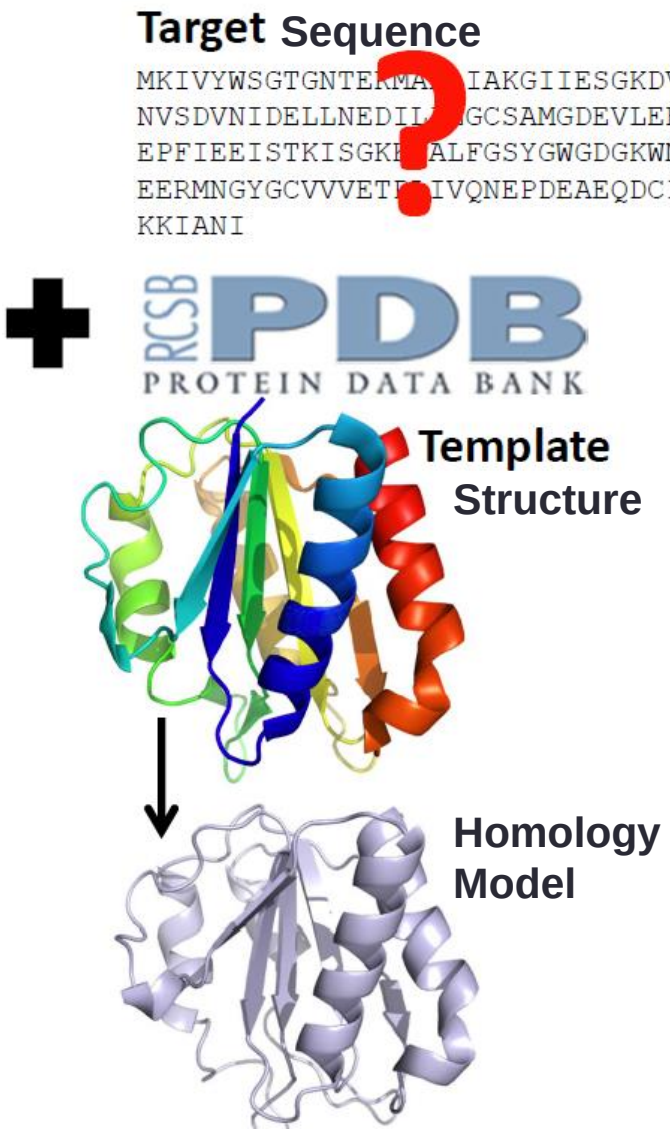
RosettaCM:
MULTI-TEMPLATE COMPARATIVE MODELING

Nica Marlow
brennica.marlow@vanderbilt.edu
Georg Kuenze
georg.kuenze@vanderbilt.edu

Introduction: RosettaCM Homology Modeling

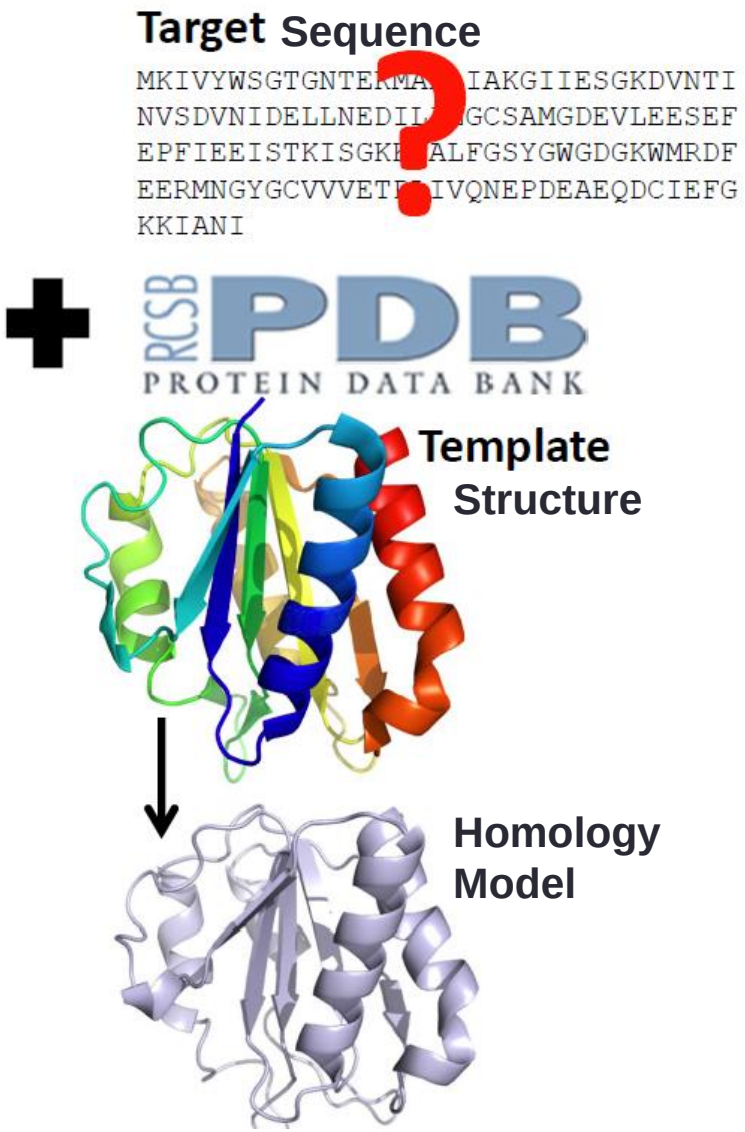


- **Single Template Modeling:**
  - Single template as input
  - Uses sequence and template derived fragments
  - Used when available templates have very high identity (>60%)
- **Multiple Template Modeling:**
  - Multiple templates as input
  - Combine sections of multiple threaded models and sequence 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, HHBlits, etc.)
  - 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/D2 chimera receptor

- PDB ID: 3pbl(D3), 6cm4(D2)
- Class A G-protein coupled receptor (GPCR)
- No high identity templates

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

High Identity Templates

- It is advisable to use multiple templates when there is low sequence identity in available templates

Template	PDB ID	% Seq id
D3 receptor	3PBL	
D2 receptor	6CM4	74
D4 receptor	5WIU	50

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

Find this file at */rosetta_cm/demo/input_files/d3d2_chimera.fasta*

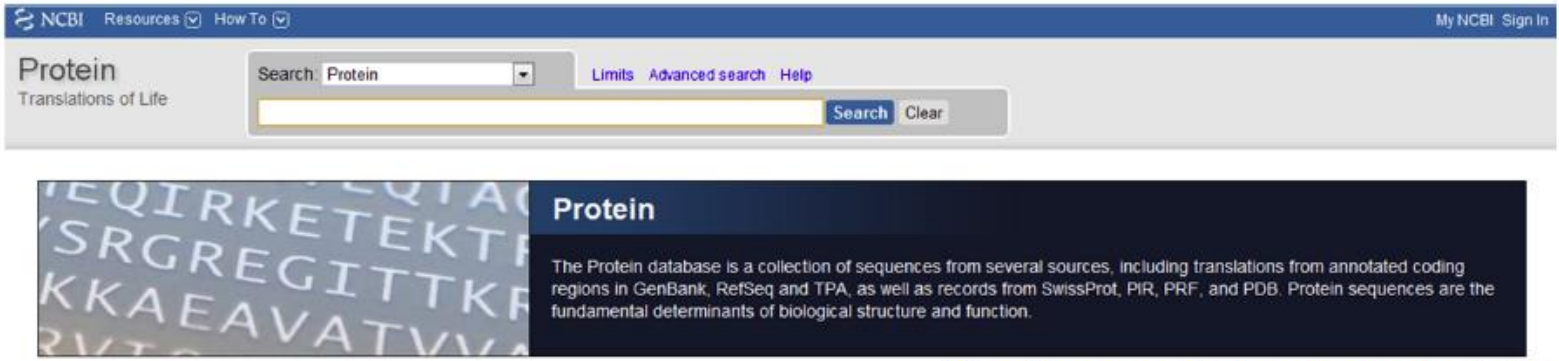
>d3d2_chimera

YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATL
VMPWVVYLEVTGGVWNFSRICCDVFVTLDVM

MCTASIWNLCAISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVL
AFAVSCPLLFGLNNADQNECIIANAPAFV

IYSSVVSFYLPFGVTVLVYARIYVVLKQRRRKAAAAAAAAGVPLREKK
ATQMVAIVLGAFIVCWLPFFLTHVLNTHC

QTCHVSPELYSATTWLGYVNSALNPVIYTTFNIEFRKAFLKILSC



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

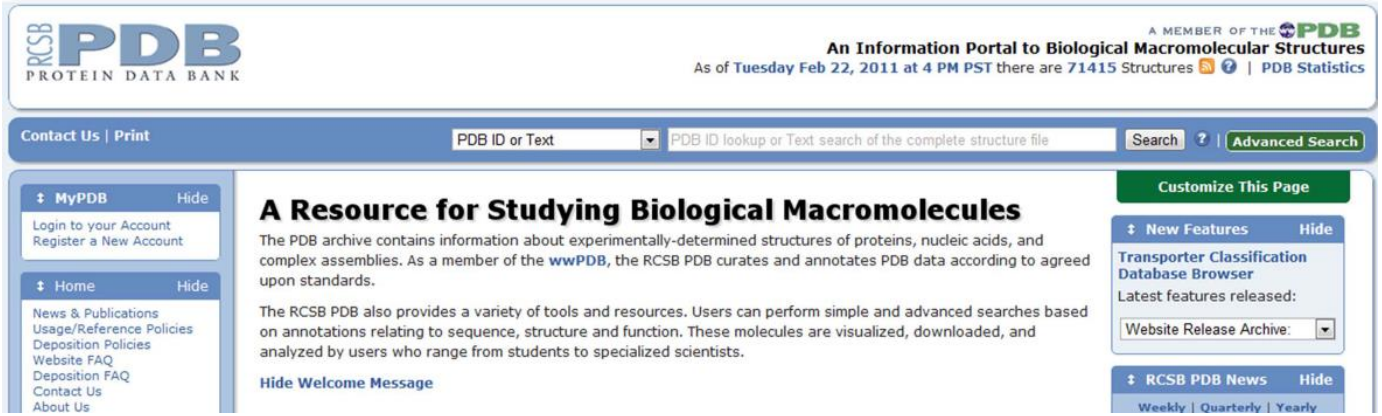
Template PDBs

Human D3 dopamine receptor (PDB: 3pbl)

Human D2 dopamine receptor (PDB: 6cm4)

Human D4 dopamine receptor (PDB: 5wiu)

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

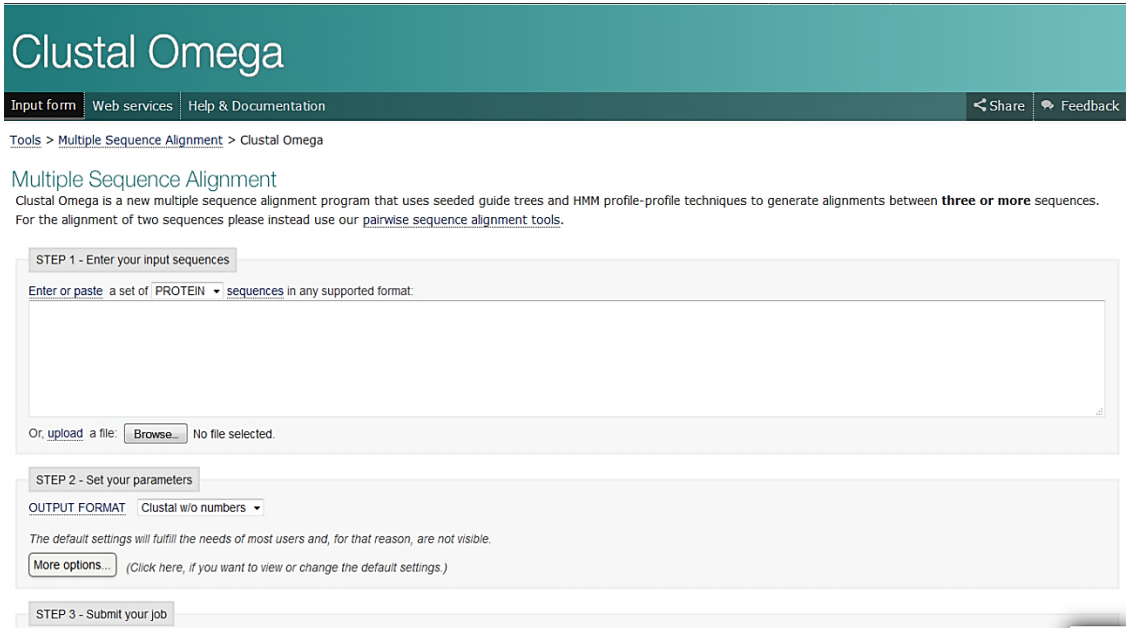


Multiple Sequence Alignment

Find this file at */demo/alignment_files/d3d2_chimera_alignments.txt*

```
CLUSTAL O(1.2.4) multiple sequence alignment


5wiu              GAAALVGGVLLIGAVLAGNSLVCVSVATERALQTPTNSFIVSLAAADLLLALLVLPLFVY
6cm4              -NYYATLLTLLIAVIVFGNVLVCMAVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVY
d3d2_chimera      ---YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATLVMPWVVY
3pbl              ---YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATLVMPWVVY
```



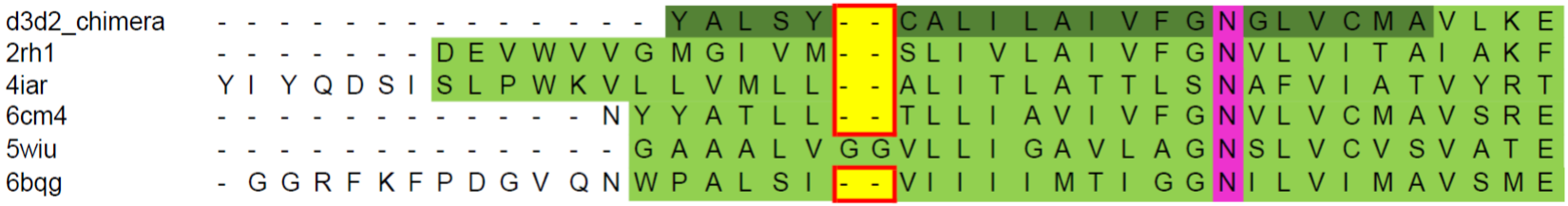
http://www.ebi.ac.uk/Tools/msa/clustalo/

Adjusting multiple sequence alignment

Experimental expectations:

- Highly conserved residues
- Secondary structure elements

Raw ClustalO alignment:



Adjusted alignment:

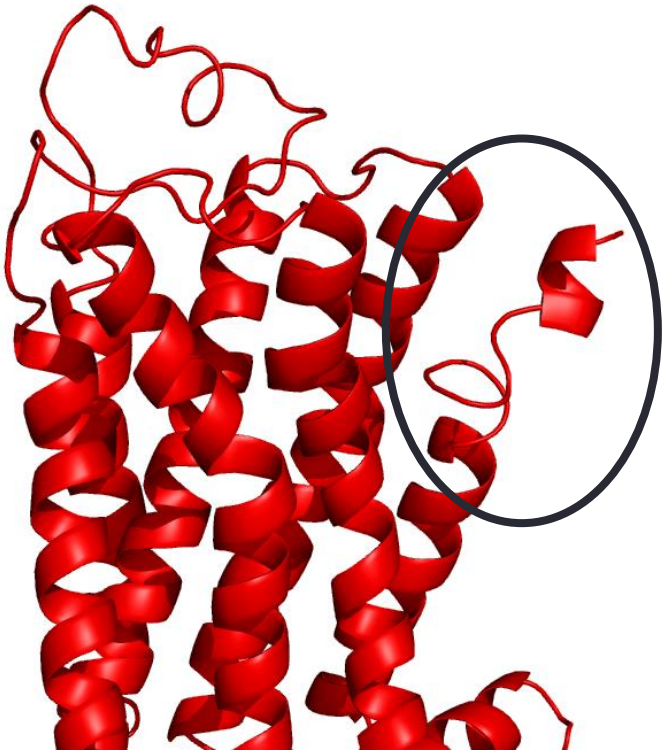
helix regions

highly conserved residues

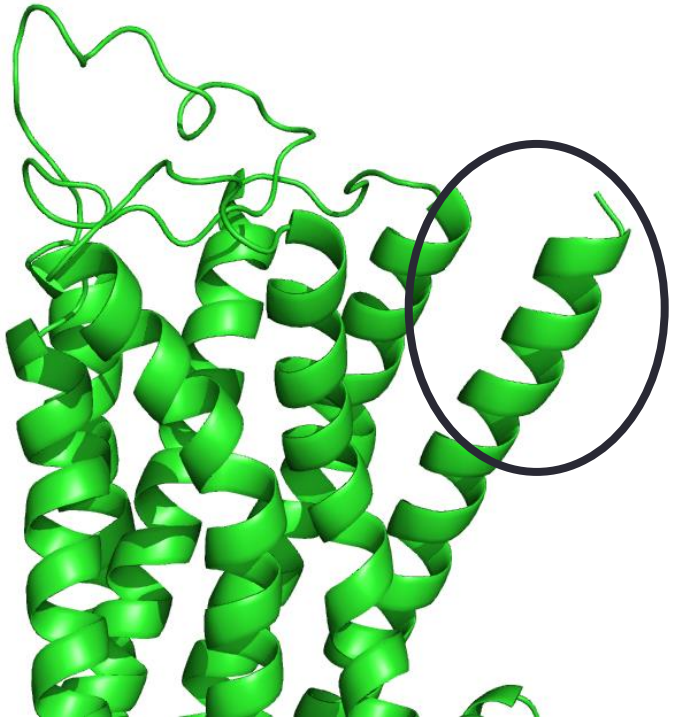
Alignment issues to be resolved

predicted membrane spanning region from OCTOPUS

Removing helix gaps



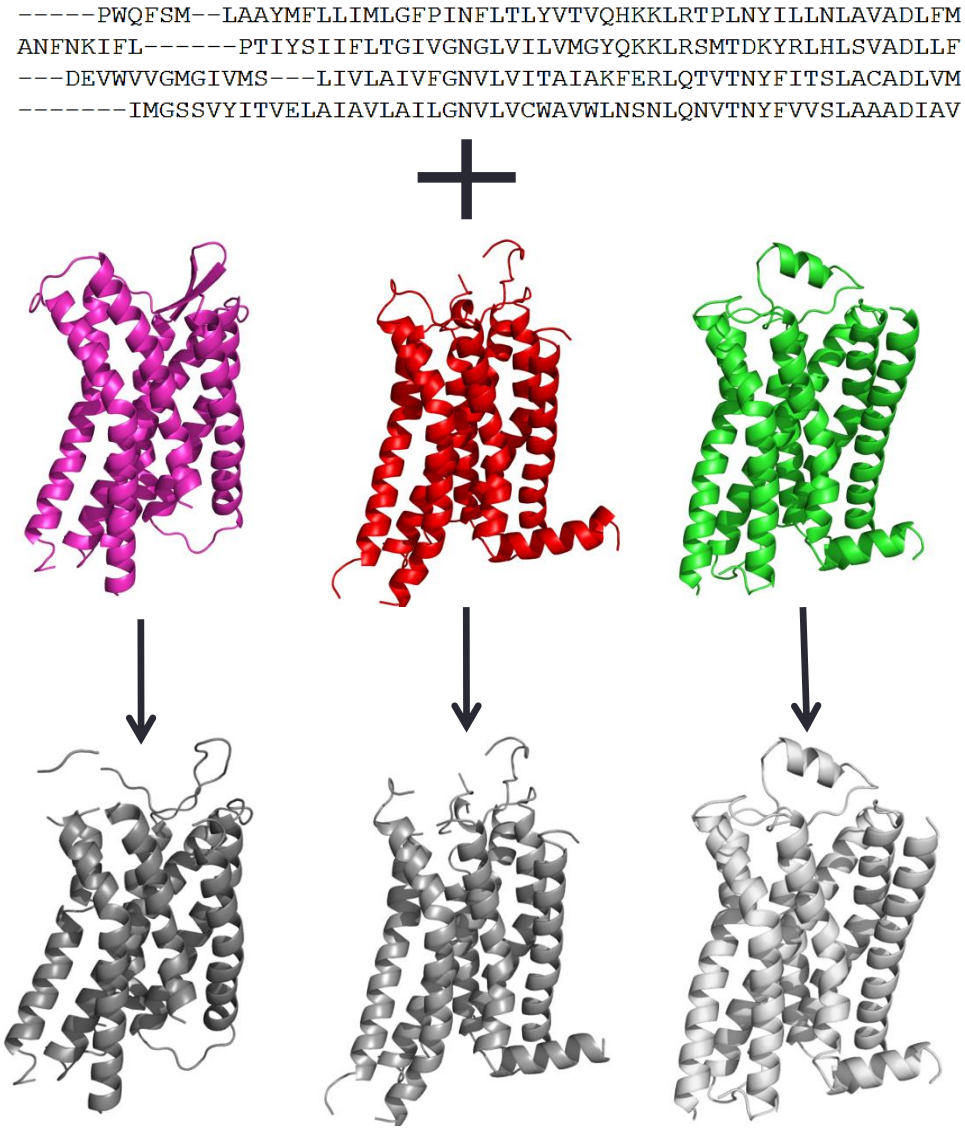
Example model using raw alignment



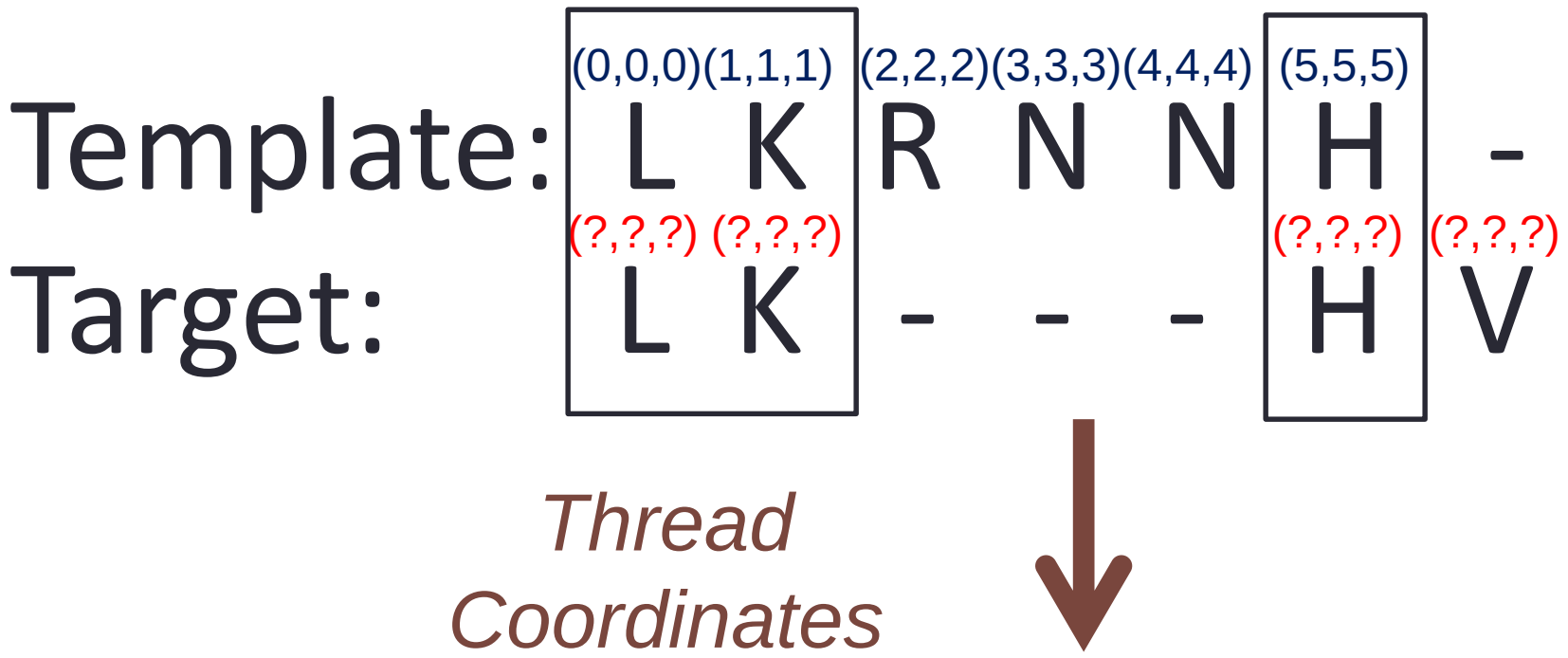
Example model using adjusted alignment

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

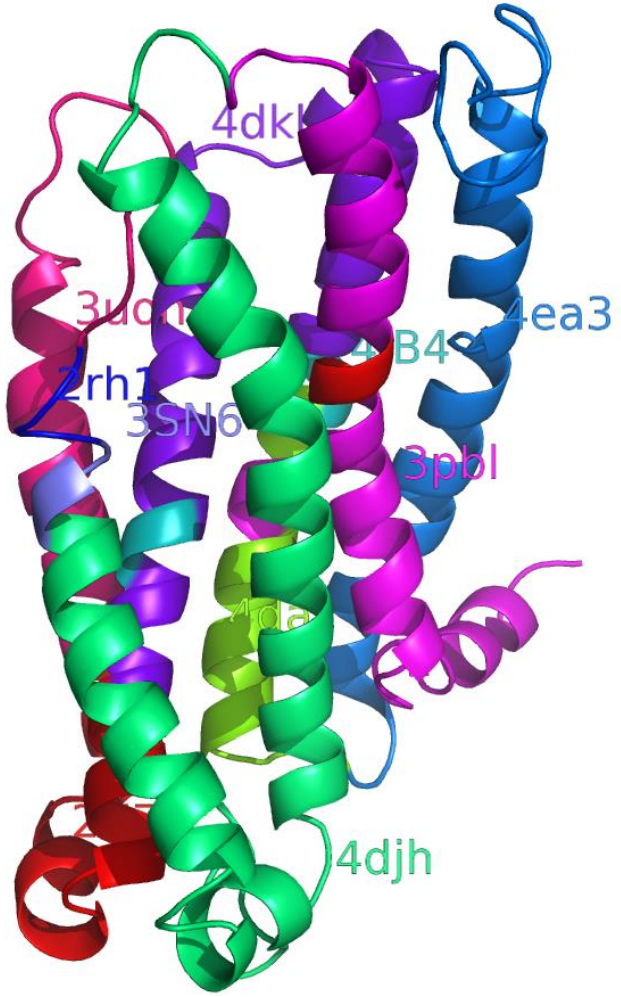
- ClustalO format:
  - All sequences in one file
  - Sequences broken up over several lines
- Grishin format:
  - One file per alignment pair
  - Sequences continuous over one line each
  - Contains header information

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

(3pbl.aln 6cm4.aln 5wiu.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

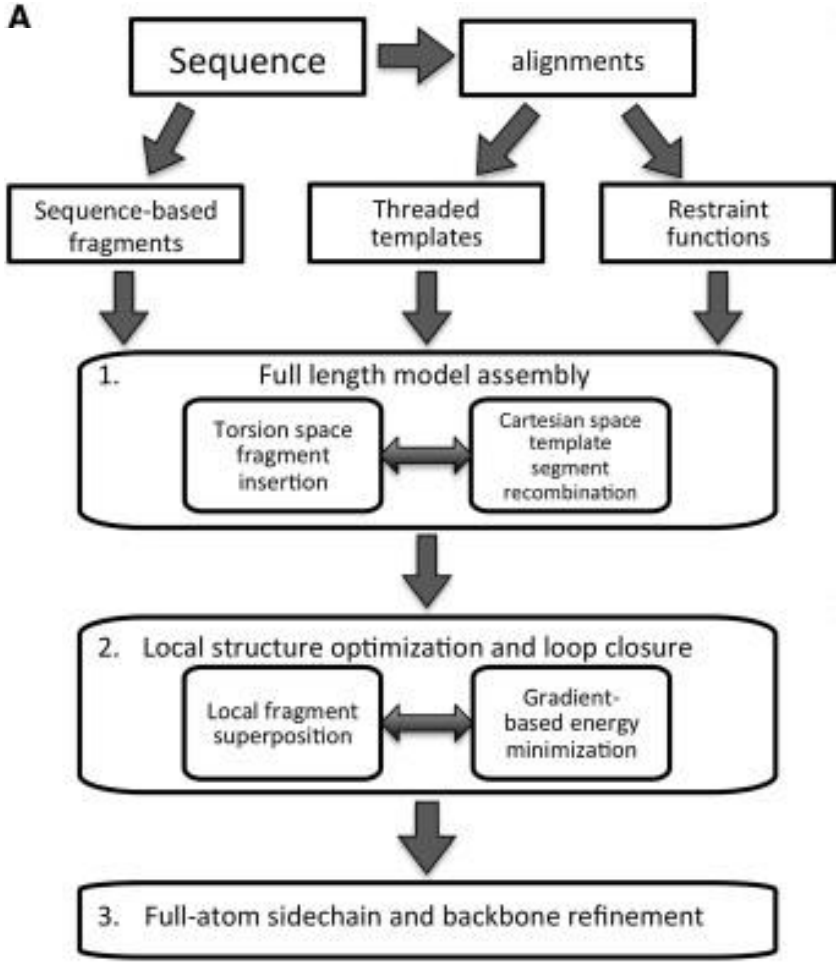


RosettaCM: Three Stages

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

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=                                   symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name="fullatom“ weights=
symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name=              weights=
symmetric="0">


        </ScoreFunction>
</SCOREFXNS>
```

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

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=
symmetric="0">
            <Reweight scoretype="atom_pair_constraint" weight="0.5"/>
        </ScoreFunction>
        <ScoreFunction name=                weights=
symmetric="0">


        </ScoreFunction>
</SCOREFXNS>
```

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

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=                weights=
symmetric="0">


        </ScoreFunction>
</SCOREFXNS>
```

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

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">


        </ScoreFunction>
</SCOREFXNS>
```

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

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/d3d2_chimera.disulfide“
fa_cst_file=“fullatom.cst”>

            <Template pdb=                                  cst_file="AUTO" weight="1.000" />
            <Template pdb=                                   cst_file="AUTO" weight="1.000" />
            <Template pdb=                                  cst_file="AUTO" weight="1.000" />

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

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/d3d2_chimera.disulfide“
fa_cst_file=“fullatom.cst”>

            <Template pdb="threaded_pdbs/3pbl_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/6cm4_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/5wiu_out.pdb" cst_file="AUTO" weight="1.000" />

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

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/d3d2_chimera.disulfide“
fa_cst_file=“fullatom.cst”>

            <Template pdb="threaded_pdbs/3pbl_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/6cm4_out.pdb" cst_file="AUTO" weight="1.000" />
            <Template pdb="threaded_pdbs/5wiu_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/demo/input_files/rosetta_cm.options

```
# i/o
-in:file:fasta input_files/d3d2_chimera.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
-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                                        #### use ref2015_cart if soluble protein
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15
```

RosettaCM Options

/rosetta_cm/demo/input_files/rosetta_cm.options

```
# i/o
-in:file:fasta input_files/d3d2_chimera.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/d3d2_chimera.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                                        #### use ref2015_cart if soluble protein
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15
```

RosettaCM Options

/rosetta_cm/demo/input_files/rosetta_cm.options

```
# i/o
-in:file:fasta input_files/d3d2_chimera.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/d3d2_chimera.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
```

Tutorial

Comparative modeling of D3/D2 chimera 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.

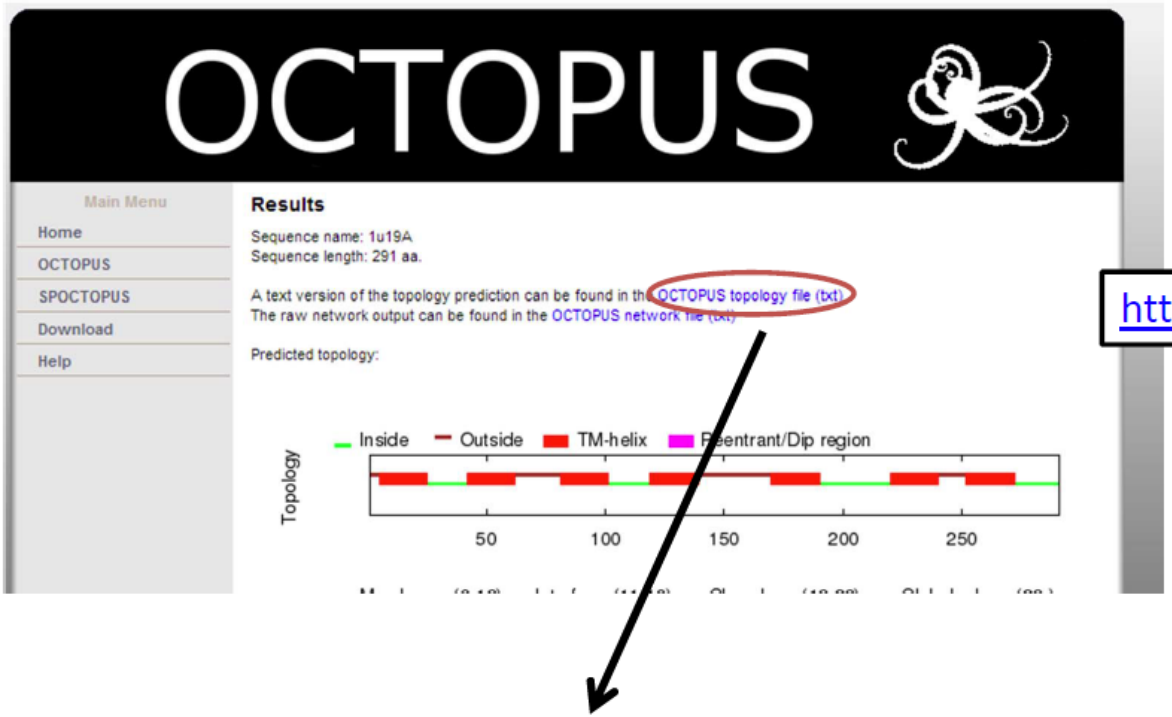
High Identity Templates

- It is advisable to use multiple templates when there is low sequence identity in available templates

Template	PDB ID	% Seq id
D2 receptor	6CM4	74
D4 receptor	5WIU	50
5-HT1B receptor	4IAR	38
5-HT2C receptor	6BQG	38
M1 receptor	5CXV	38
A1 receptor	5UEN	37
B2-adrenoreceptor	2RH1	34
H1 receptor	3RZE	31
M3 receptor	4U14	30

Membrane spanning regions

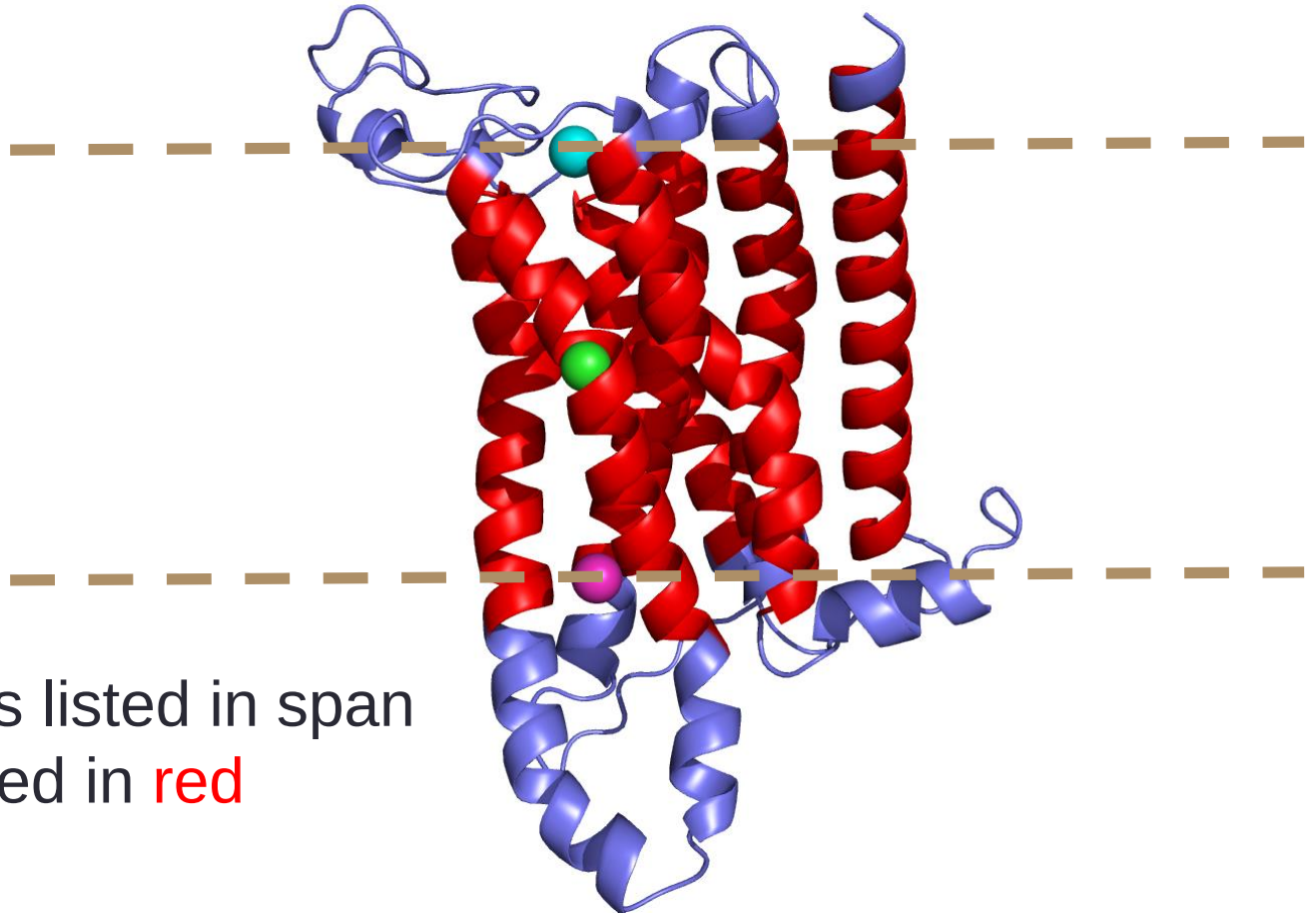
Find this file at */rosetta_cm/demo/input_files/d3d2_chimera.span*



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

octopus2span.pl
d3d2_chimera.octopus

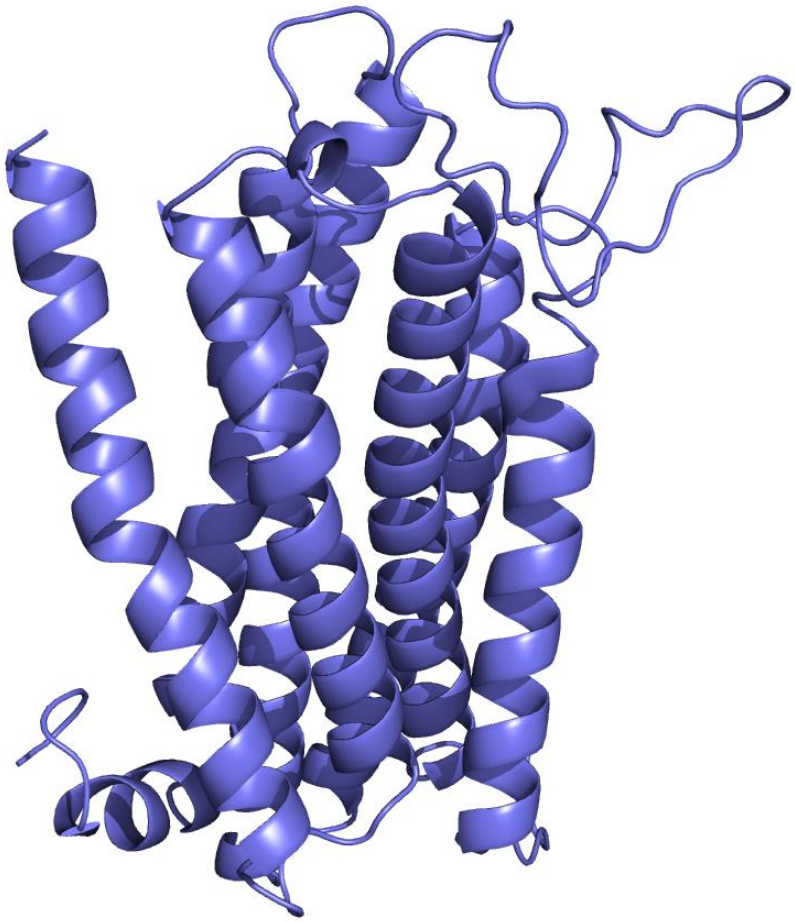
Rosetta Membrane



Residues listed in span file colored in red

Why use membrane scoring terms?

With membrane penalties/weights



Without membrane penalties/weights