

ROSETTACM: MULTI-TEMPLATE COMPARATIVE MODELING

Nica Marlow

brennica.marlow@vanderbilt.edu

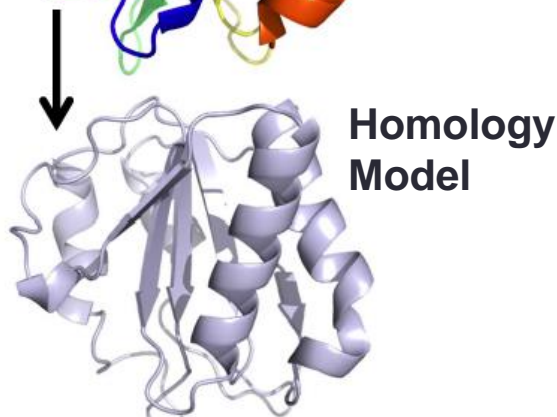
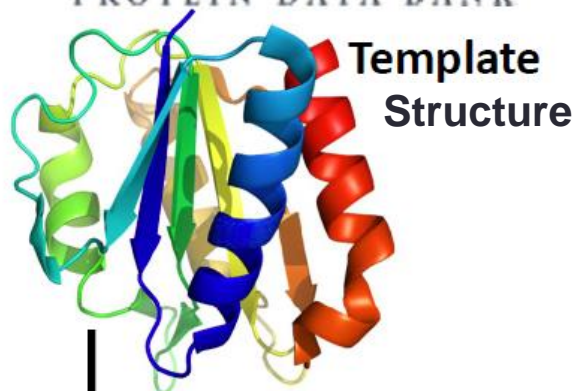
Georg Kuenze

georg.kuenze@vanderbilt.edu

Introduction: RosettaCM Homology Modeling

Target Sequence

```
MKIVYWSGTGNTERMAIAKGIIESGKDVNTI  
NVSDVNIDELLNEDILGCSAMGDEVLEESEF  
EPFIEEISTKISGKRALFGSYGWGDGKWMRDF  
EERMNGYGCVVETIVQNEPDEAEQDCIEFG  
KKIANI
```



• 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

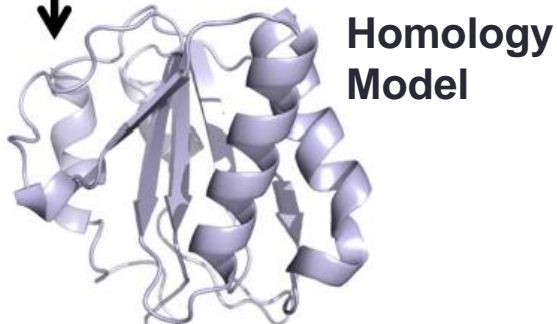
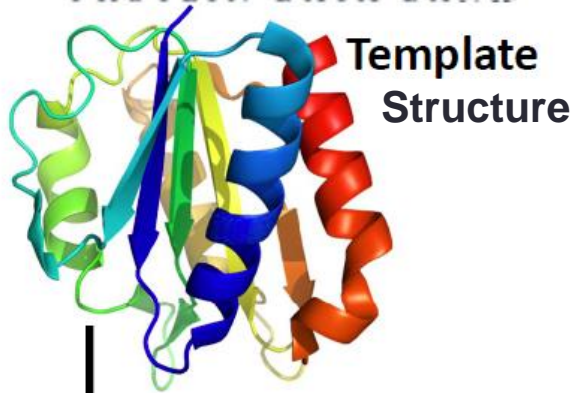
Identifying Template Structures

Target Sequence

```
MKIVYWSGTGNTERMAIAKGIIESGKDVNTI  
NVSDVNIDELLNEDIAGCSAMGDEVLEESEF  
EPFIEEISTKISGKIALFGSYGWGDGKWMRDF  
EERMNGYGCVVETIVQNEPDEAEQDCIEFG  
KKIANI
```



RCSB
PDB
PROTEIN DATA BANK



- **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--LAAYMFLLIIMLGFPINFLTLYVTVQHKKLRTPLNYYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIIFLTGIVGNGLVILVMGYQKKLRSMYDKYRLHLSVADLLF  
---DEVVVVGMGIVMS---LIVLAIIVFGNVLVITAIKFERLQTVTNYFITSLACADLVM  
-----IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVVTVNYFVVSLAAADIAV
```

Target Sequence

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

>d3d2_chimera

```
YALSYCALILAIVFGNGLVCM AVLKERALQTTTNYLVVSLAVADLLVATL
VMPWVVYLEVTGGVWNFSRICCDVFVTL DVM
MCTASIWNLC AISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVL
AFAVSCPLL FGLNNADQNECIIANAPAFV
IYSSVVSFYLPFGVTVLVYARIYVVLKQRRRKAAAAAAAAAAGVPLREKK
ATQMVAIVLGAFIVCWLPFFLTHVLNTHC
QTCHVSP ELYSATTWLG YVNSALNPVIYTTFNIEFRKAFLKILSC
```



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 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/*

The screenshot shows the RCSB PDB website homepage. At the top left is the RCSB PDB logo. At the top right, it says "A MEMBER OF THE PDB" and "An Information Portal to Biological Macromolecular Structures". Below this, it states "As of Tuesday Feb 22, 2011 at 4 PM PST there are 71415 Structures" and provides links for "PDB Statistics". A search bar is located below the header, with a dropdown menu for "PDB ID or Text" and a search button. The main content area features a large heading "A Resource for Studying Biological Macromolecules" followed by a paragraph describing the PDB archive. To the left of the main content are navigation menus for "MyPDB" and "Home". To the right is a "Customize This Page" section with a "New Features" menu and a "Website Release Archive" dropdown.

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

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

Customize This Page

New Features Hide

Transporter Classification Database Browser

Latest features released:

Website Release Archive:

RCSB PDB News Hide

Weekly | Quarterly | Yearly

Multiple Sequence Alignment

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

CLUSTAL O(1.2.4) multiple sequence alignment

```
5wiu      GAAALVGGVLLIGAVLAGNSLVCVSVATERALQTP TNSFIVSLAAADLLLALLVLP LFVY
6cm4      -NYYATLLTLLIAVIVFGNVLVCM AVSREKALQTTTNYLIVSLAVADLLVATLVMPWVVY
d3d2_chimera ---YALSYCALILAIIVFGNGLVCM AVLKERALQTTTNYLVVSLAVADLLVATLVMPWVVY
3pb1      ---YALSYCALILAIIVFGNGLVCM AVLKERALQTTTNYLVVSLAVADLLVATLVMPWVVY
```

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

Adjusting multiple sequence alignment

Experimental expectations:

- Highly conserved residues
- Secondary structure elements

Raw ClustalO alignment:

```
d3d2_chimera  - - - - - - - - - - - - - - - - - - - Y A L S Y - - C A L I L A I V F G N G L V C M A V L K E
2rh1          - - - - - - - - - - - - - - - - - - - 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
4iar         - Y I Y Q D S I S L P W K V L L V M L L - - - - - A L I T L A T T L S N A F V I A T V Y R T
6cm4        - - - - - - - - - - - - - - - - - - - N Y Y A T L L - - - - - T L L I A V I V F G N V L V C M A V S R E
5wiu        - - - - - - - - - - - - - - - - - - - G A A A L V G G V L L I G A V L A G N S L V C V S V A T E
6bqg        - G G R F K F P D G V Q N W P A L S I - - - - - V I I I I M T I G G N I L V I M A V S M E
```

Adjusted alignment:

```
d3d2_chimera  - - - - - - - - - - - - - - - - - - - Y A L S Y C A L I L A I V F G N G L V C M A V L K E R A
2rh1          - - - - - - - - - - - - - - - - - - - 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
4iar         - Y I Y Q D S I S L P W K V L L V M L L A L I T L A T T L S N A F V I A T V Y R T R K
6cm4        - - - - - - - - - - - - - - - - - - - N Y Y A T L L T L L I A V I V F G N V L V C M A V S R E K A
5wiu        - - - - - - - - - - - - - - - - - - - G A A A L V G G V L L I G A V L A G N S L V C V S V A T E R A
6bqg        - - - - - - - - - - - - - - - - - - - G G R F K F P D G V Q N W P A L S I V I I I I M T I G G N I L V I M A V S M E K K
```

helix regions

highly conserved residues

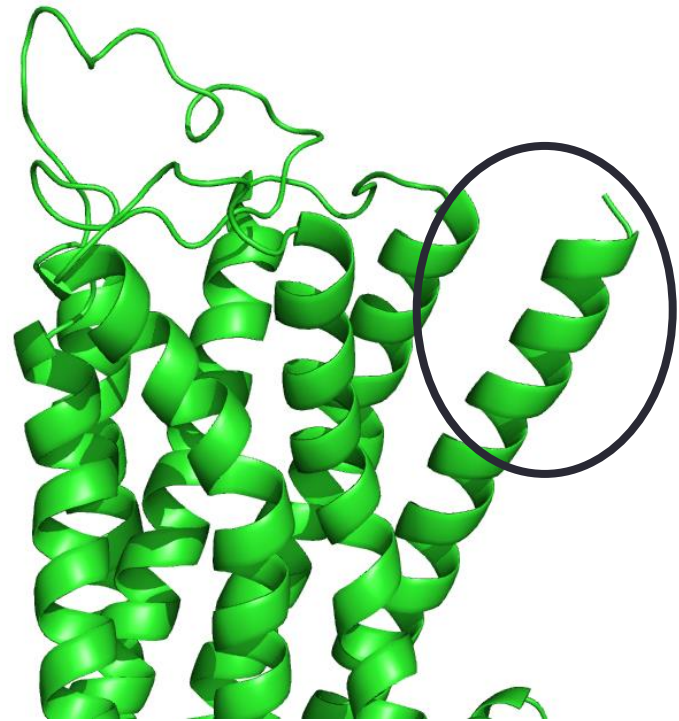
Alignment issues to be resolved

predicted membrane spanning region from OCTOPUS

Removing helix gaps



Example model using
raw alignment



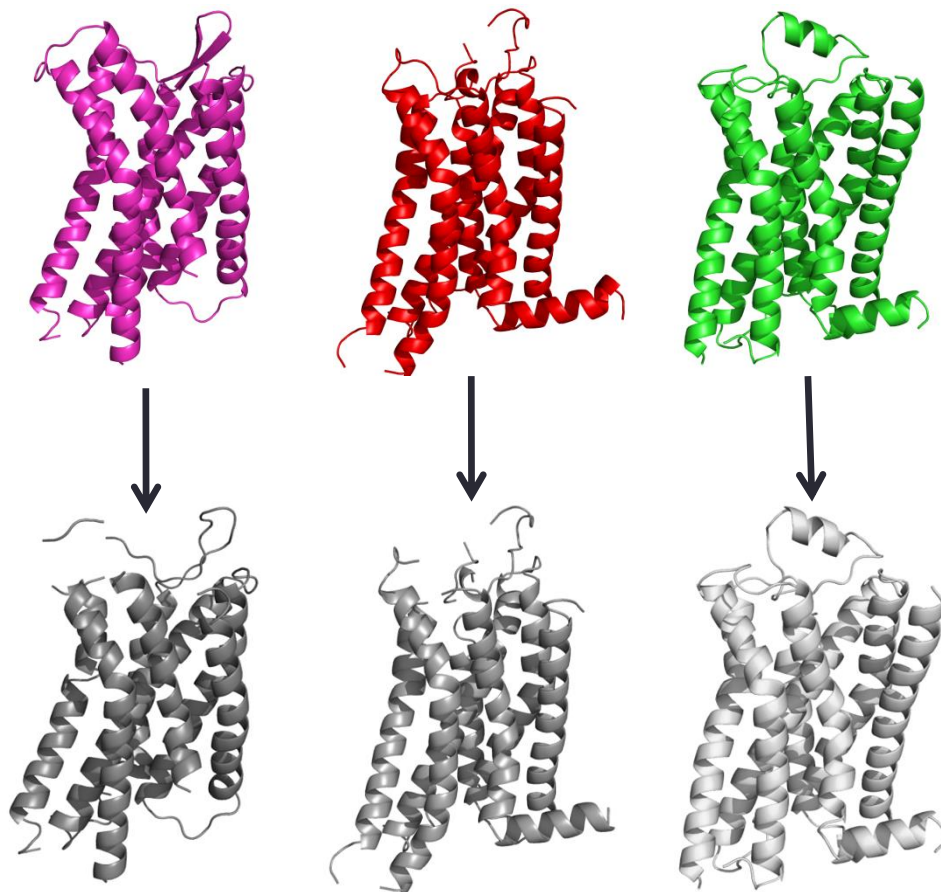
Example model using
adjusted alignment

Comparative Modeling Protocol

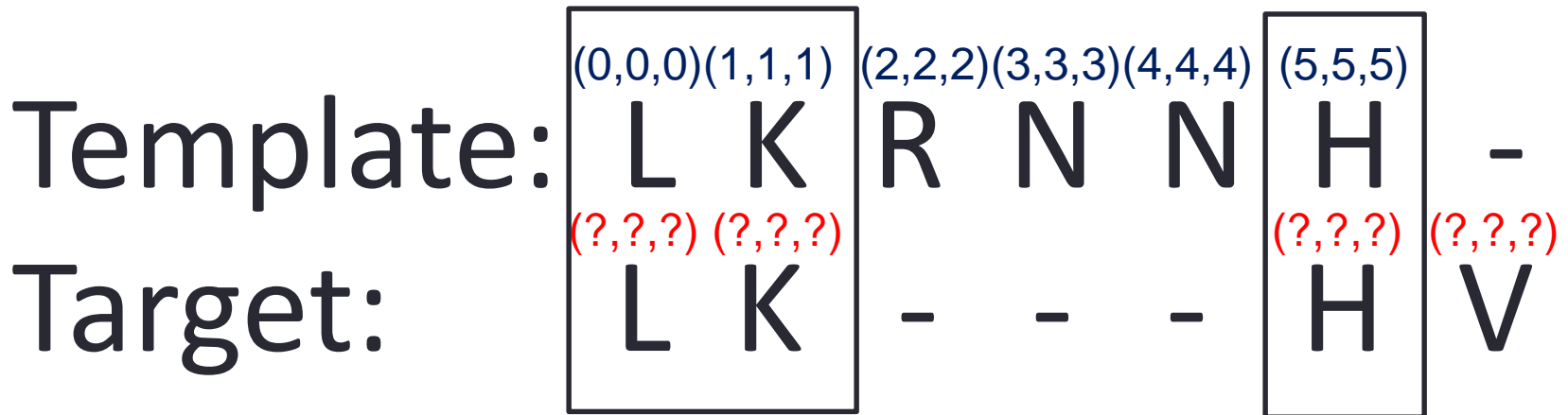
- **Step 1:** Align target sequence to template sequences
- **Step 2:** Partial-thread the target sequence onto template structures
- **Step 3:** Combine pieces from different templates using RosettaCM Hybridize

```
-----PWQFSM--LAAYMFLIMLGFPINFLTYVTVQHKKLRTPLNYYILLNLAVADLFM  
ANFNKIFL-----PTIYSIIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF  
---DEVVVVGMGIVMS---LIVLAIIVFGNVLVITAIKFERLQTVTNYFITSLACADLVM  
-----IMGSSVYITVELAIAVLAAILGNVLVCWAVWLNSNLQVNTNYFVVSLAAADIAV
```

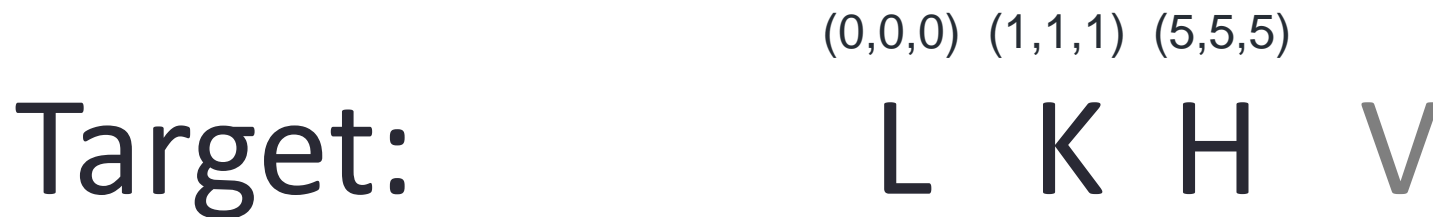
+



Threading



*Thread
Coordinates*



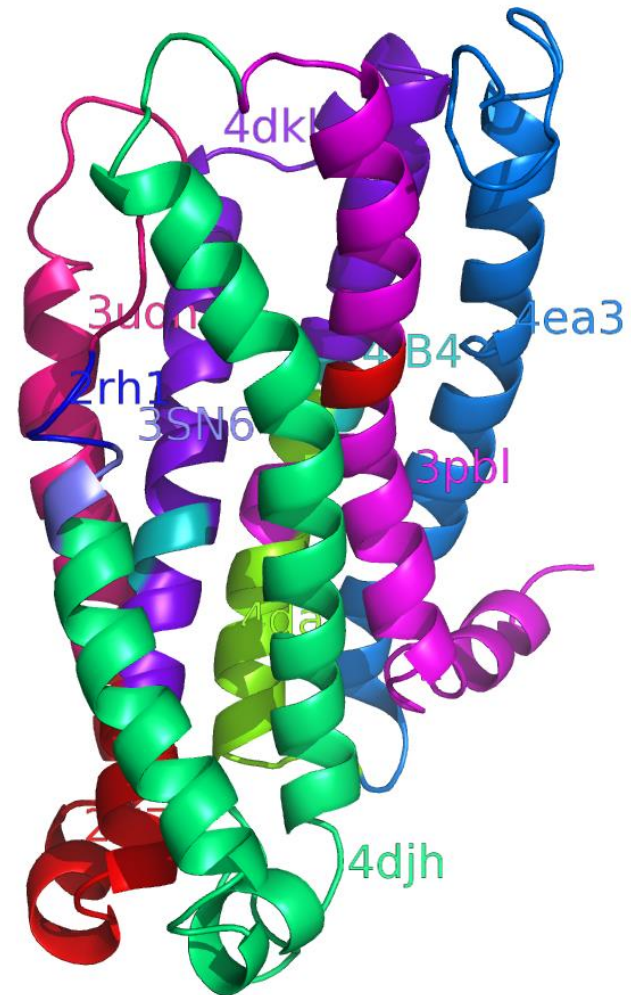
Grishin Format Alignments Needed for Rosetta Threading

- 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)

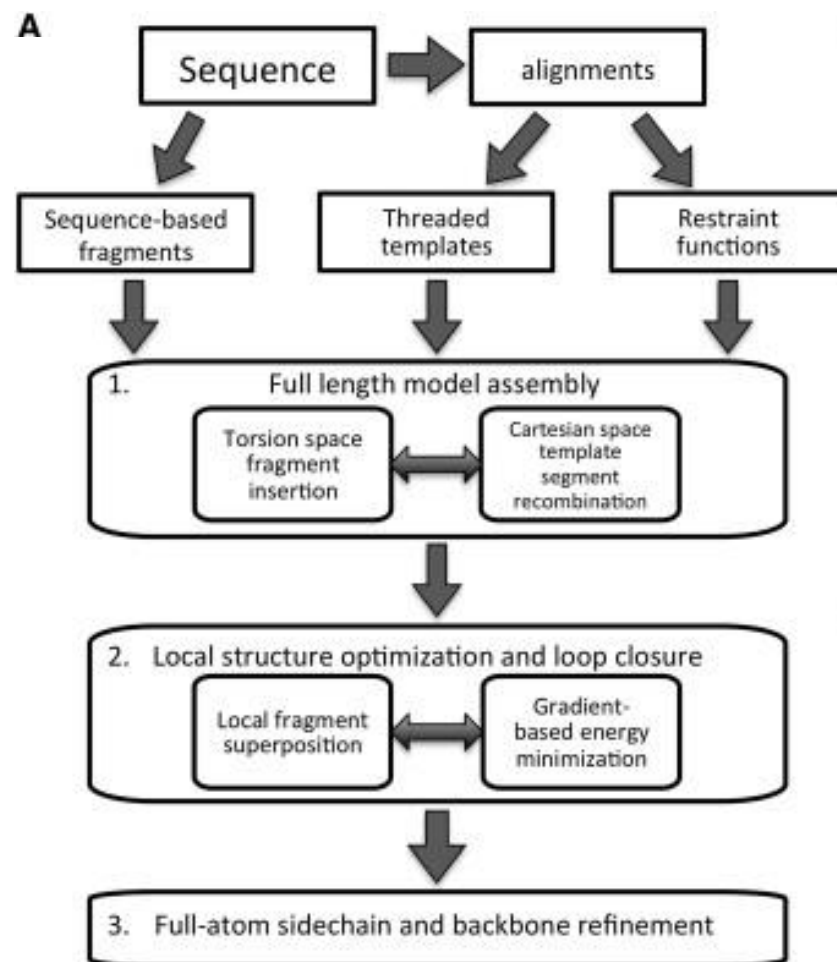
Comparative Modeling Protocol

- **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



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_pdb/3pbl_out.pdb" cst_file="AUTO" weight="1.000" />
    <Template pdb="threaded_pdb/6cm4_out.pdb" cst_file="AUTO" weight="1.000" />
    <Template pdb="threaded_pdb/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**
-parser:protocol input_files/rosetta_cm.xml
-out:path:all output_files/

your target sequence

#Initialize membrane

-in:file:spanfile
-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
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15

use ref2015_cart if soluble protein

RosettaCM Options

/rosetta_cm/demo/input_files/rosetta_cm.options

i/o

-in:file:fasta **input_files/d3d2_chimera.fasta**
-parser:protocol input_files/rosetta_cm.xml
-out:path:all output_files/

your target sequence

#Initialize membrane

-in:file:spanfile **input_files/d3d2_chimera.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
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15

use ref2015_cart if soluble protein

RosettaCM Options

/rosetta_cm/demo/input_files/rosetta_cm.options

i/o

-in:file:fasta **input_files/d3d2_chimera.fasta**
-parser:protocol input_files/rosetta_cm.xml
-out:path:all output_files/

your target sequence

#Initialize membrane

-in:file:spanfile **input_files/d3d2_chimera.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

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`

OCTOPUS

Main Menu

- Home
- OCTOPUS
- SPOCTOPUS
- Download
- Help

Results

Sequence name: 1u19A
Sequence length: 291 aa.

A text version of the topology prediction can be found in the [OCTOPUS topology file \(txt\)](#).
The raw network output can be found in the [OCTOPUS network file \(txt\)](#).

Predicted topology:

— Inside — Outside — TM-helix — Reentrant/Dip region

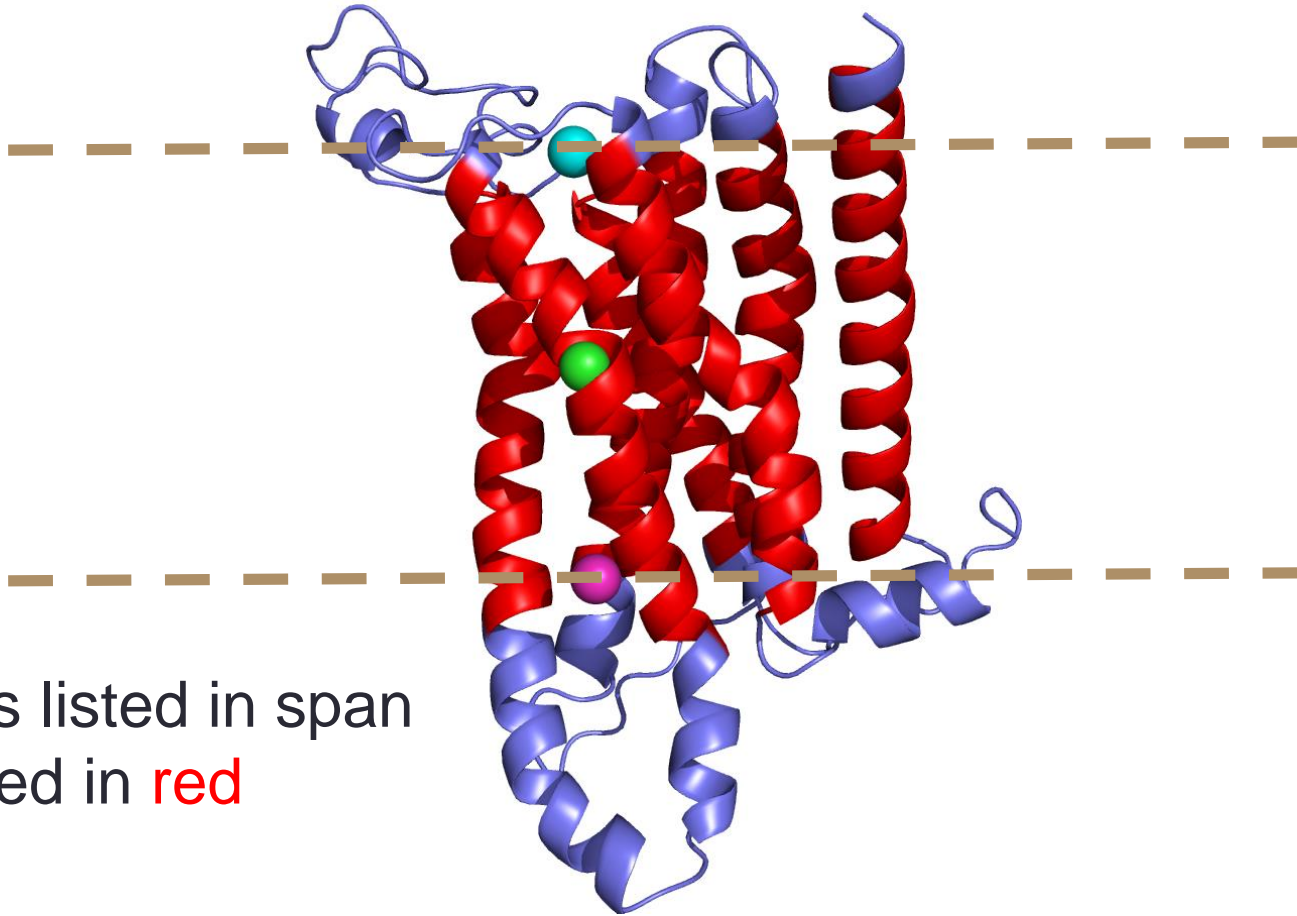
Topology

50 100 150 200 250

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

octopus2span.pl
d3d2_chimera.octopus

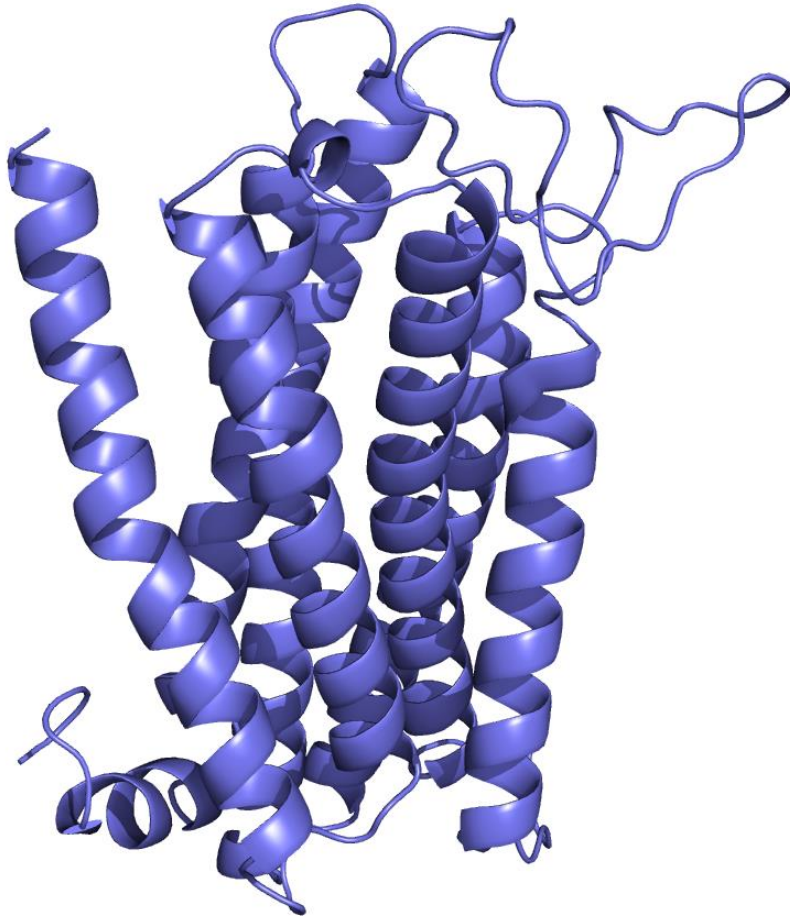
Rosetta Membrane



Residues listed in span
file colored in **red**

Why use membrane scoring terms?

With membrane penalties/weights



Without membrane penalties/weights

