Modeling of G Protein-Coupled Receptors with Rosetta

17th Annual Great Lakes GPCR Retreat

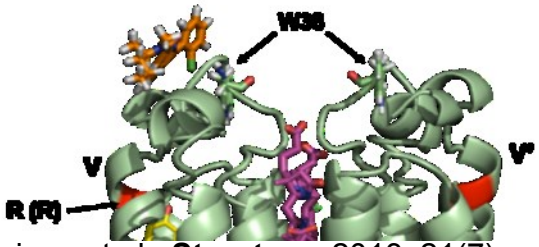
Rocco Moretti

Brian Bender

VANDERBILT UNIVERSITY

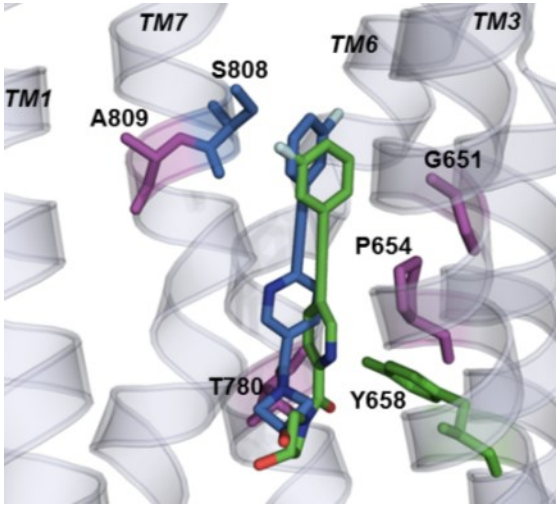
Center for Structural Biology and Institute of Chemical Biology
Departments of Chemistry, Pharmacology, and Biomedical Informatics

Computational Structural and Chemical Biology in the Meiler Lab

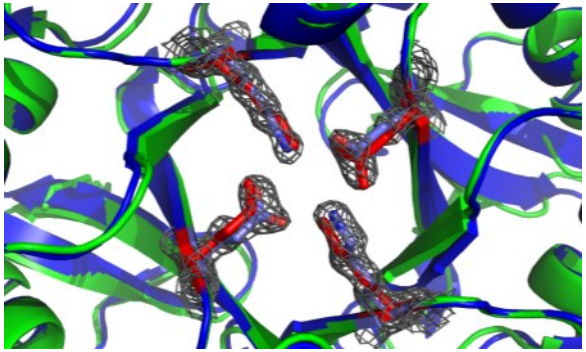


Weiner, et al., **Structure**, 2013, 21(7)

Protein structure prediction de novo and from limited experimental data

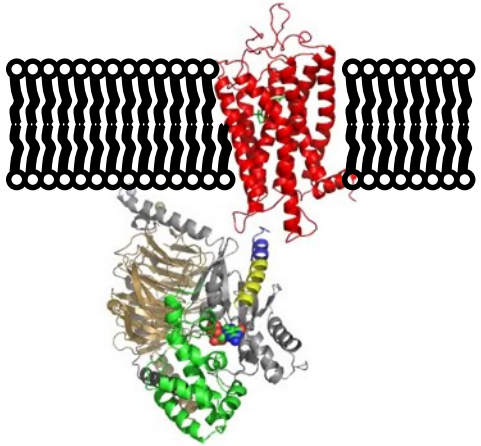


Dong, et al., **PLoS ONE**, 2013, 8(7)



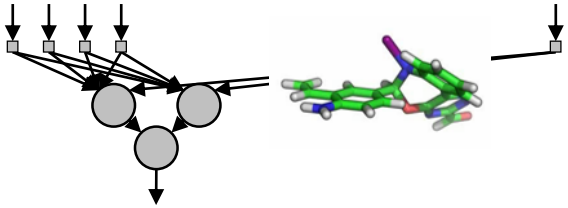
Fortenberry, et al., **JACS**, 2011, 133(45)

Design of large protein scaffolds, antibodies, and protein/ligand interfaces

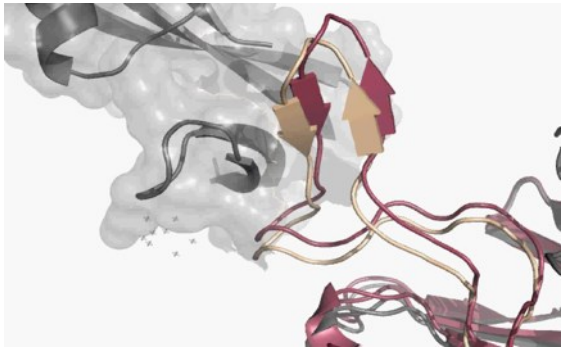


Alexander et al., **NSMB**, 2014, 21(1)

Merging ligand- and structure-based computer-aided drug discovery



Muller, et al., **ChemMedChem**, 2012, 7(3)



Willis, et al., **PLoS Comp. Biol.**, 2013, 9(4)

Acknowledgements – http://meilerlab.org

Alberto Cisneros
Alex Geanes
Alex Sevy
Alyssa Lokits
Amanda Duran
Amandeep Sangha
Axel Fischer
Ben Mueller
Bian Li
Brian Bender
Brittany Allison
Daniel Putnam
Carie Fortenberry
Darwin Fu
Gregory Sliwoski

Jessica Finn
John Lally
Karuna Gujar
Marcin Skwark
Marion Sauer
Michaela Fooksa
Sandeep Kothiwale
Soumya Ganguly
Sten Heinze
Steven Combs
Rocco Moretti
Heather Darling
Jeff Mendenhall
Yan Xia
Xuan Zhang

Selected Alumni: Brian Weiner, David Nannemann, Gordon Lemmon, Jordon Willis, Julia Koehler Leman, Liz Dong Nguyen, Mariusz Butkiewicz, Mert Karakas, Nathan Alexander, Nils Woetzel, Steffen Lindert, Will Lowe

VU Collaborators: Carlos Arteaga, Jim Crowe, Heidi Hamm, Randy Blakely, Jeff Conn, Christine Lovly, Craig Lindsley, Seva Gurevich, Hassane Mchaourab, Dave Weaver, Ambra Pozzi, Chuck Sanders, Roger Cone, Charles Manning, Steve Fesik

Active: NIH NIGMS R01GM080403, R01GM073151, NIH NIDDK R01DK097376, NSF CHE 1305874, NIH NIAID U19 AI117905, NIAID HHSN272201400024C, NIH NILBI R01HL122010, NIH NCI P50 CA098131

Completed: NSF MCB0742762 (CAREER), NIH NIGMS R01GM099842, NIH NIMH R21MH082254, R01MH090192, DTRA HDTRA1-10-0067

Live Demo Files (for later) @ http://meilerlab.org/index.php/rosetta-tutorials

- www.meilerlab.org
- Resources tab
  - Rosetta Tutorials

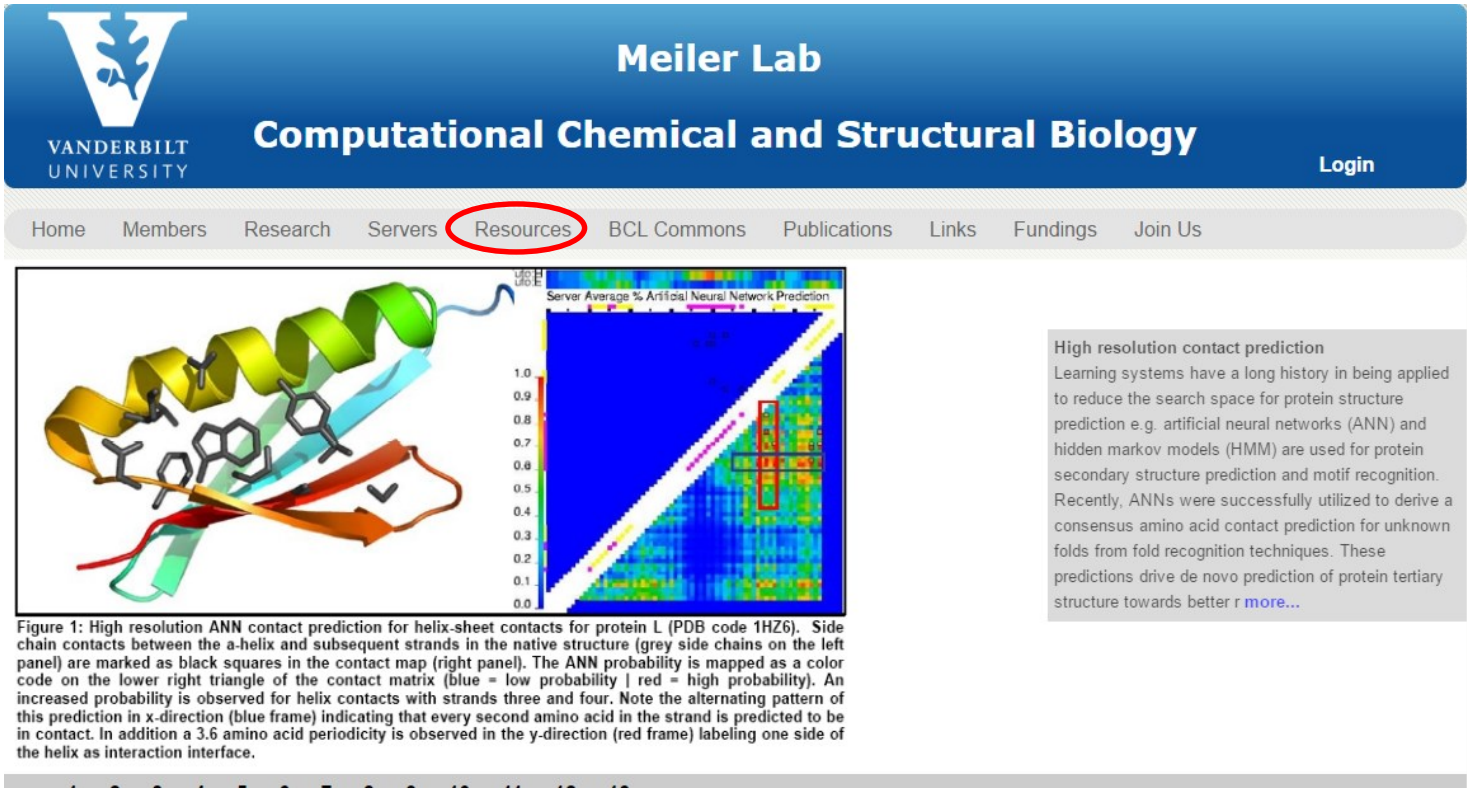
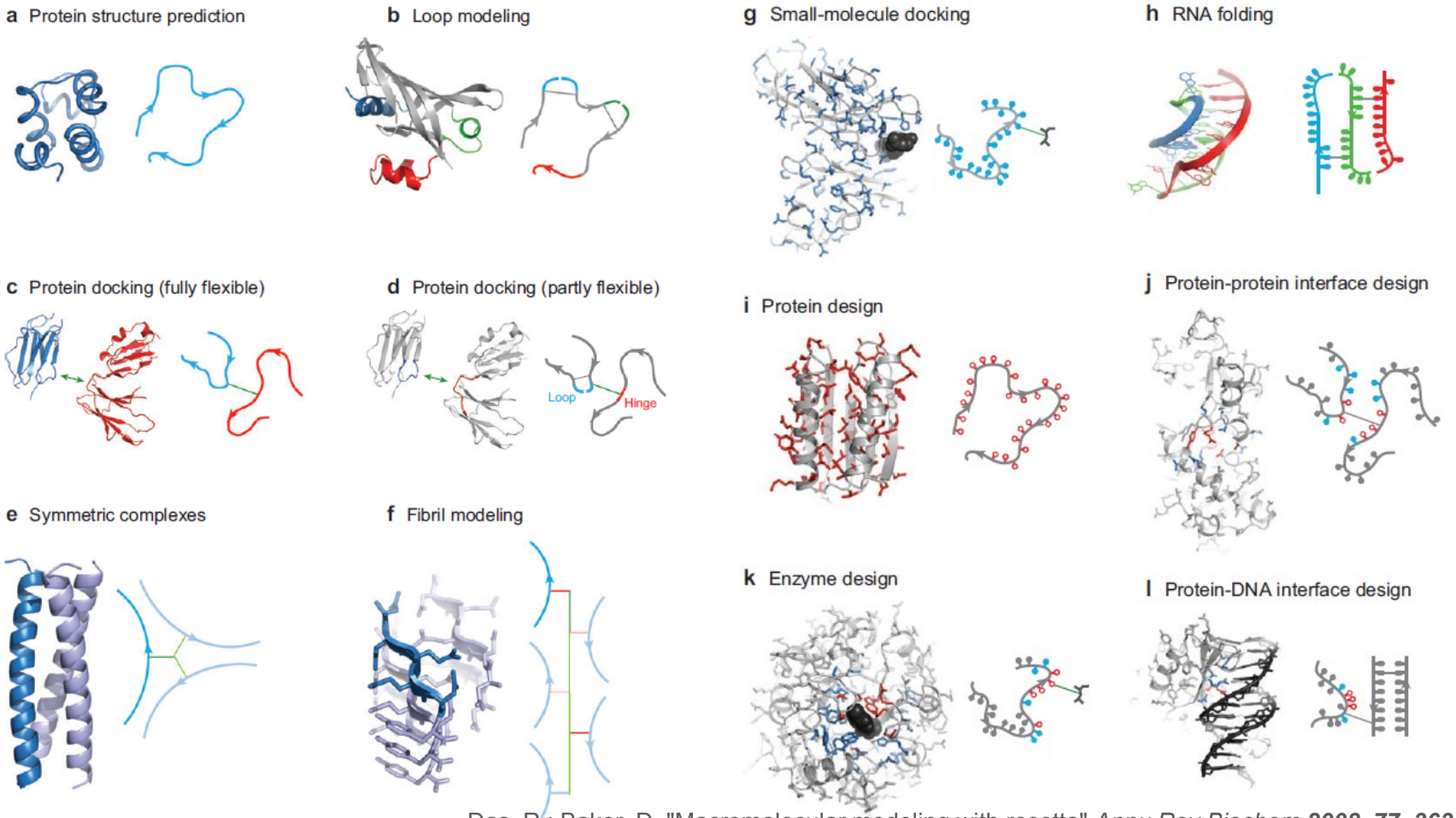


Figure 1: High resolution ANN contact prediction for helix-sheet contacts for protein L (PDB code 1HZ6). Side chain contacts between the a-helix and subsequent strands in the native structure (grey side chains on the left panel) are marked as black squares in the contact map (right panel). The ANN probability is mapped as a color code on the lower right triangle of the contact matrix (blue = low probability | red = high probability). An increased probability is observed for helix contacts with strands three and four. Note the alternating pattern of this prediction in x-direction (blue frame) indicating that every second amino acid in the strand is predicted to be in contact. In addition a 3.6 amino acid periodicity is observed in the y-direction (red frame) labeling one side of the helix as interaction interface.

Rosetta: A Unified Framework for Tackling Molecular Modeling



Das, R.; Baker, D. "Macromolecular modeling with rosetta" *Annu Rev Biochem* ***2008, 77, 363-82.***

www.rosettacommons.org

- Rosetta consists of multiple modules: protein folding, comparative modeling, ligand docking, protein design, antibody/antigen interactions, etc.
- Rosetta is developed in a consortium of 23 laboratories by over 150 developers
- PyRosetta is a python interface that allows integration with Pymol
- FoldIt is the better video game for you and your kids
- Rosetta@home uses your computer for our research

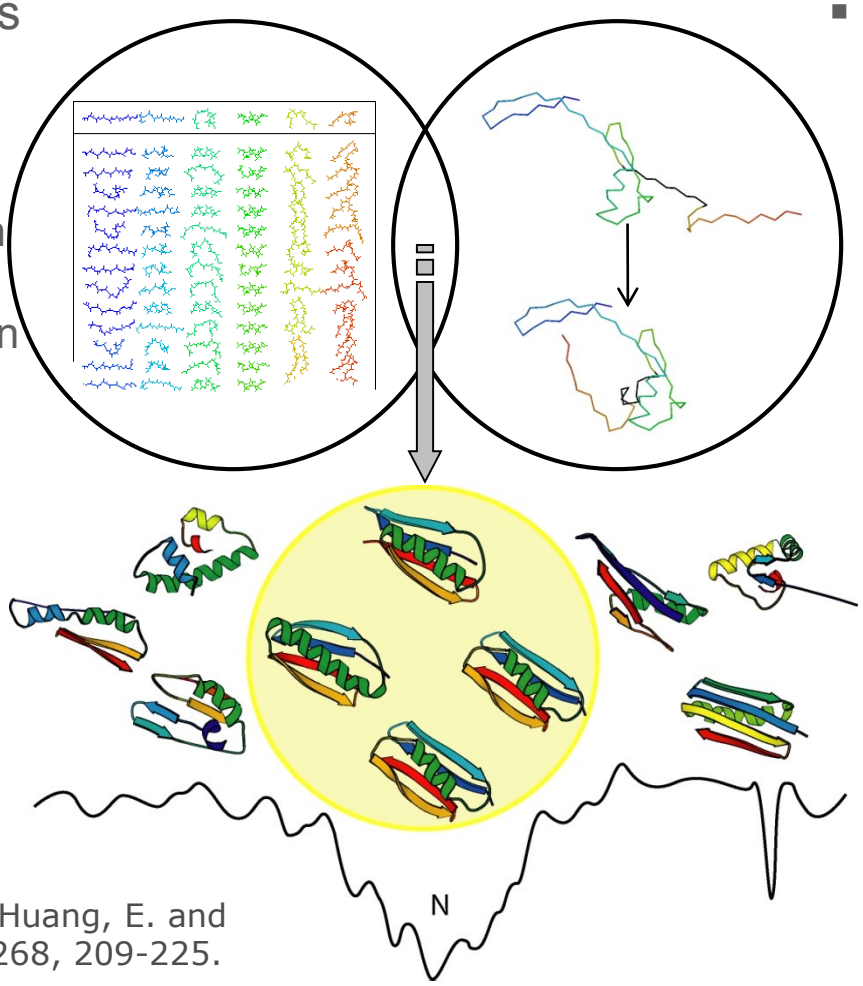
RosettaCon 2015, Leavenworth, WA, USA

Rosetta is free for academic use; user guide/tutorials available at www.rosettacommons.org

Bender, B.J.; et al. “Protocols for Molecular Modeling with Rosetta3 and RosettaScripts" *Biochemistry* ***2016.***

Sampling and Scoring for Protein Folding Simulation

- Local Sequence Bias
  - Approximate local interactions using the distribution of conformations seen for similar sequences in known protein structures
- Monte Carlo simulations
  - Select broadest minima using cluster analysis



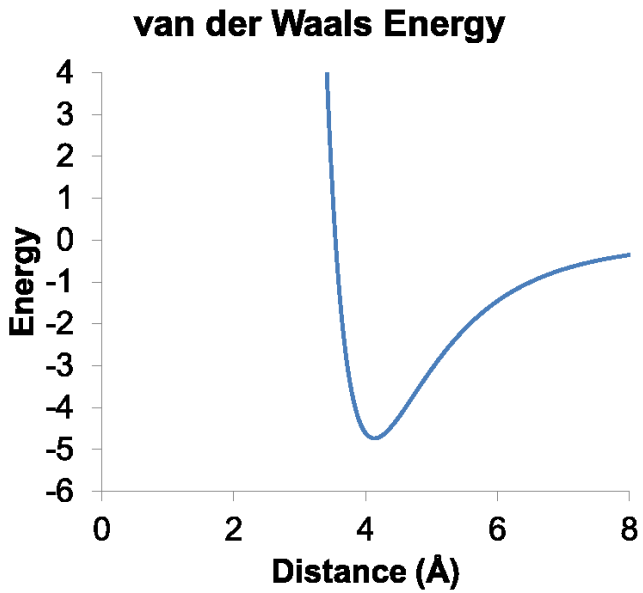
- Energy evaluation of non-local interactions using knowledge-based energy function
  - Steric overlap
  - Residue environment
  - Pair wise interactions
  - Strand pairing
  - Compactness
  - Secondary Structure Packing

Simons, K. T., Kooperberg, C., Huang, E. and Baker, D. (1997) *J. Mol. Biol.*, 268, 209-225.

Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

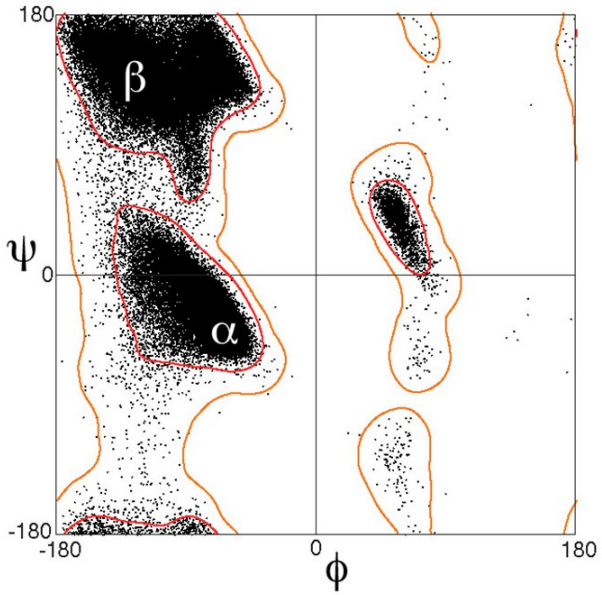
Lennard-Jones Potential

$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$



Ramachandran Plot

Statistical mining of Protein Databank (PDB)

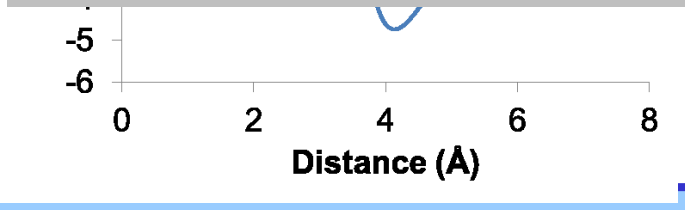


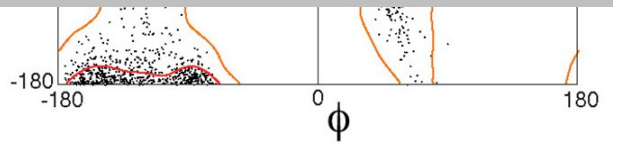
Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

Lennard-Jones

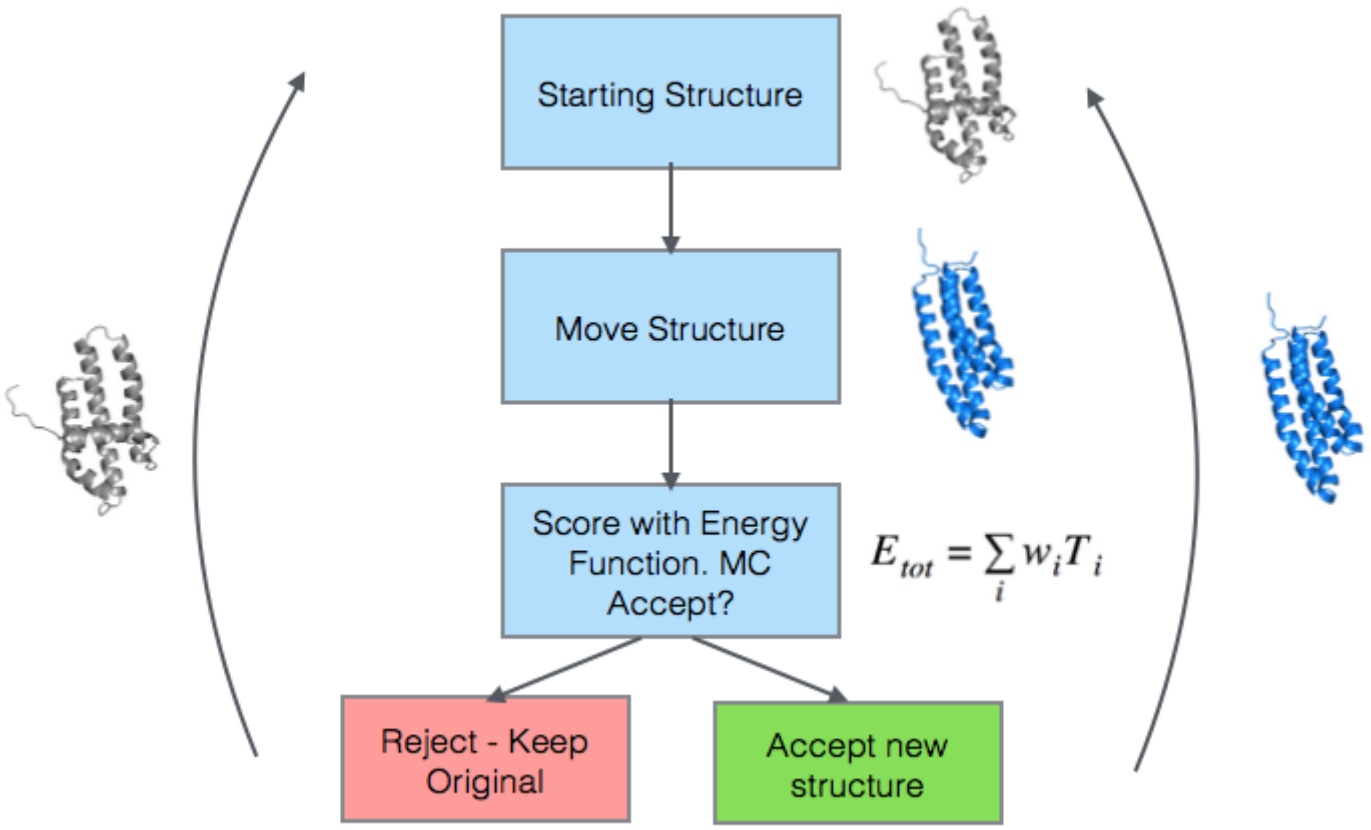
Ramachandran Plot

$$Energy = w_1 * term_1 + w_2 * term_2 + w_3 * term_3 + \ldots$$



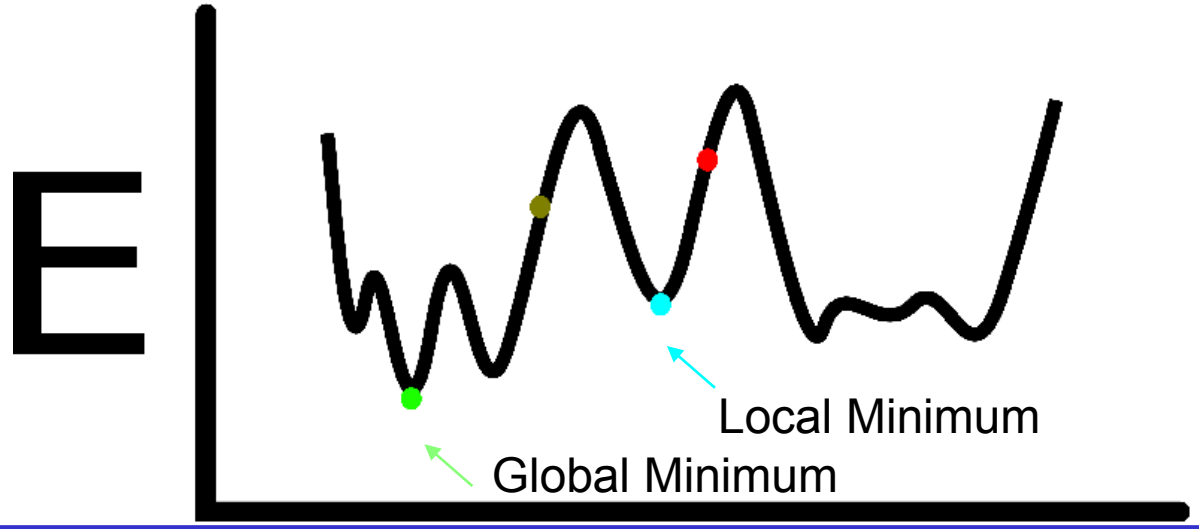


Monte Carlo Selection allows rapid sampling

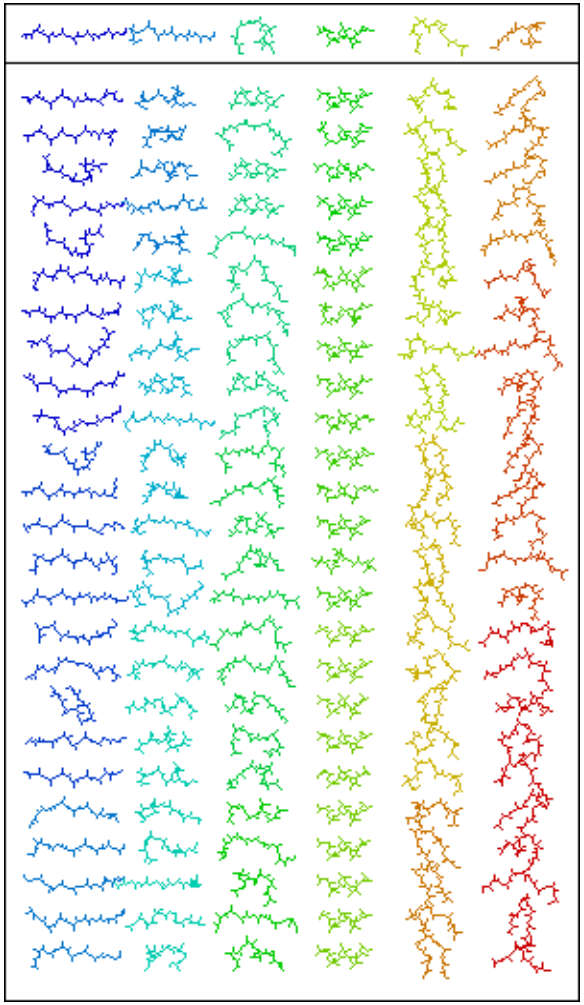


Score is Central to Monte Carlo Selection

- Criteria for accepting structures (Metropolis Criterion):
- If Enew < Eold : Accept new structure
- If Enew > Eold :
  - Pick a random number p(0, 1)
  - if $e^{[-(E_{new}-E_{old})/kBT]} > p$, accept new structure



Local Sequence Bias – Rapid Approximation of Local Interactions

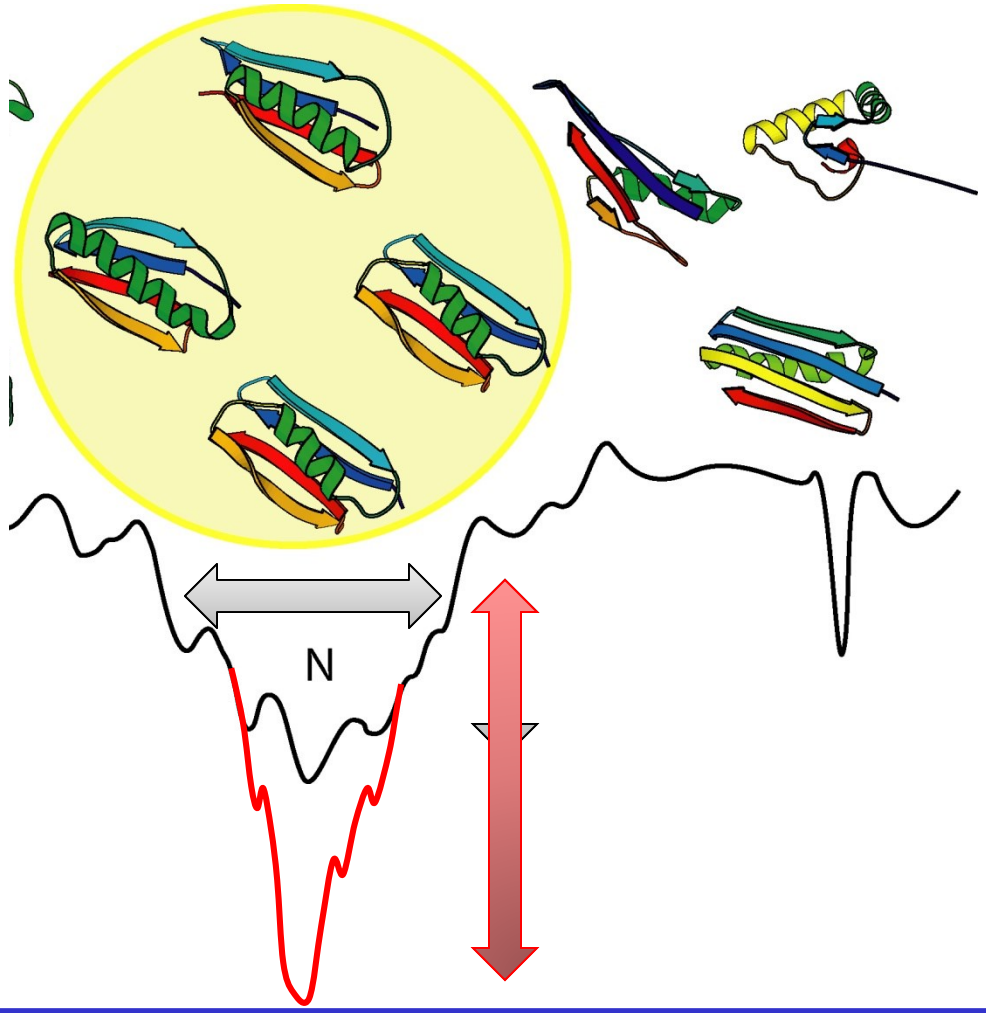


- **While not every protein fold is present in the protein databank, all possible conformations of small peptides are!**

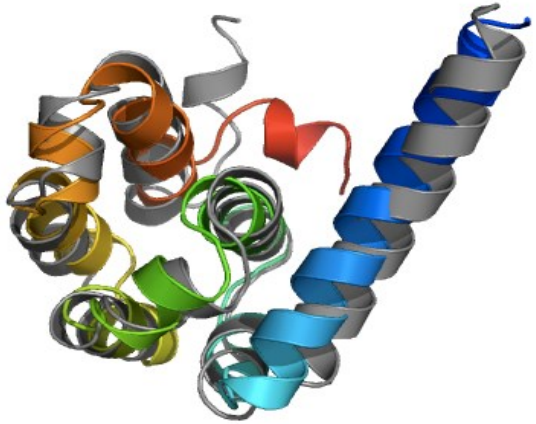
- Approximate local interactions using the distribution of conformations seen for similar sequences in known protein structures
- For each sequence window, select fragments that represent the conformations sampled during folding

Native-like Protein Models Form Large Clusters

- The free energy minimum corresponds (usually) to the native protein fold
- Its depth is obscured because of the simplified energy approximation
- However, the width of the funnel leading to the free energy minimum of the native protein fold is well preserved

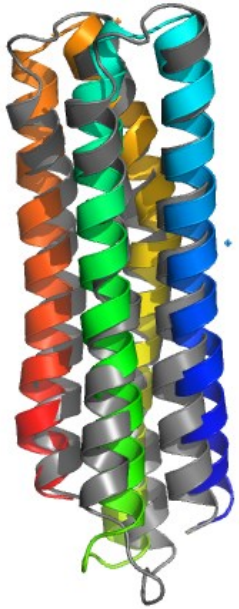


What can Rosetta actually fold?



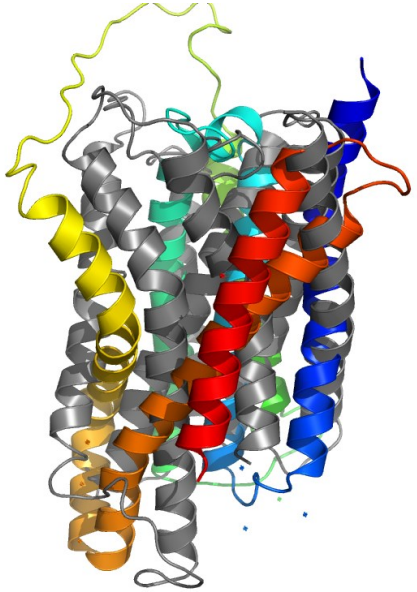
Small, globular,
soluble proteins

T4-lysozyme
C-terminal domain



Small, simple
membrane proteins

V-type Na^+ ATP synthase
subunit

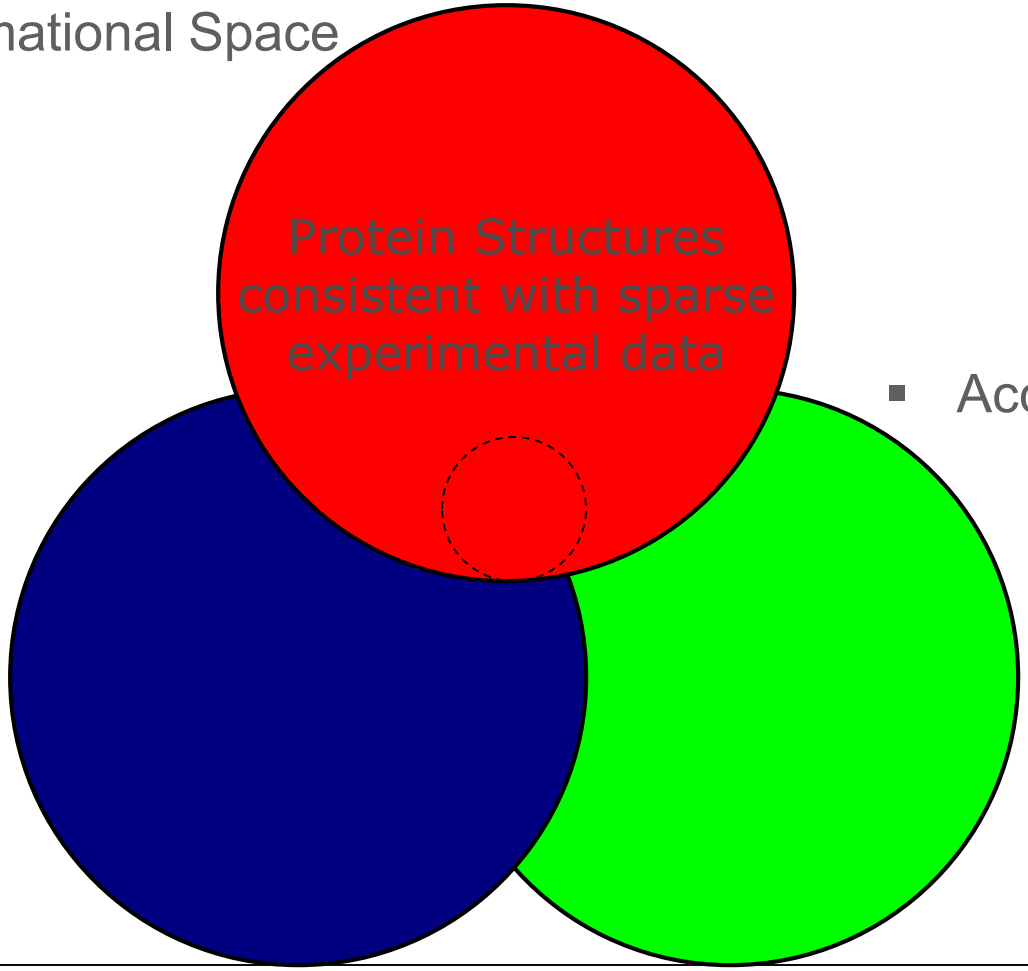


…but not large,
complex proteins

Rhodopsin

Combining Strengths: Building Accurate Models from Limited Data

- Complete Conformational Space
- Efficient Sampling Strategy
- Accurate Energy Function



Comparative modeling & docking of GPCRs

1) selection of template structure

↓

2) sequence alignment of the target to the template

↓

3) thread target sequence onto the template PDB file

↓

4) create fragments derived from PDB

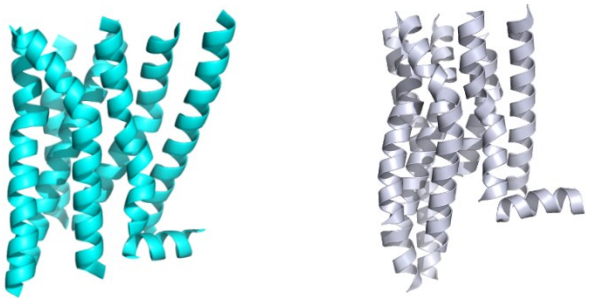
↓

5) comparative modeling: Hybridize: RosettaCM

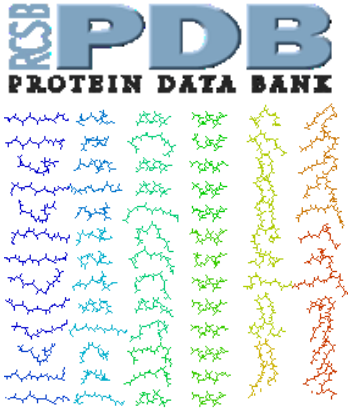
2

```
target   1 MCFSVSLSATVALGCMFVPKVY
template 1 KEVYILLNWIGYVNSGFNPLIT
```

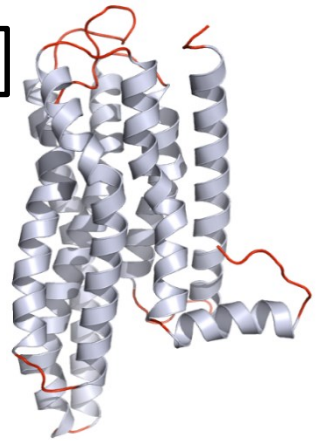
3



4



5



Combs, S.A., et al (2013) Nat Protoc, I(7), 1277-98.
Nguyen, E.D., et al (2013) PLOS ONE 8(7):e67302

Comparative modeling & docking of GPCRs

6) selection of comparative model for ligand docking

↓

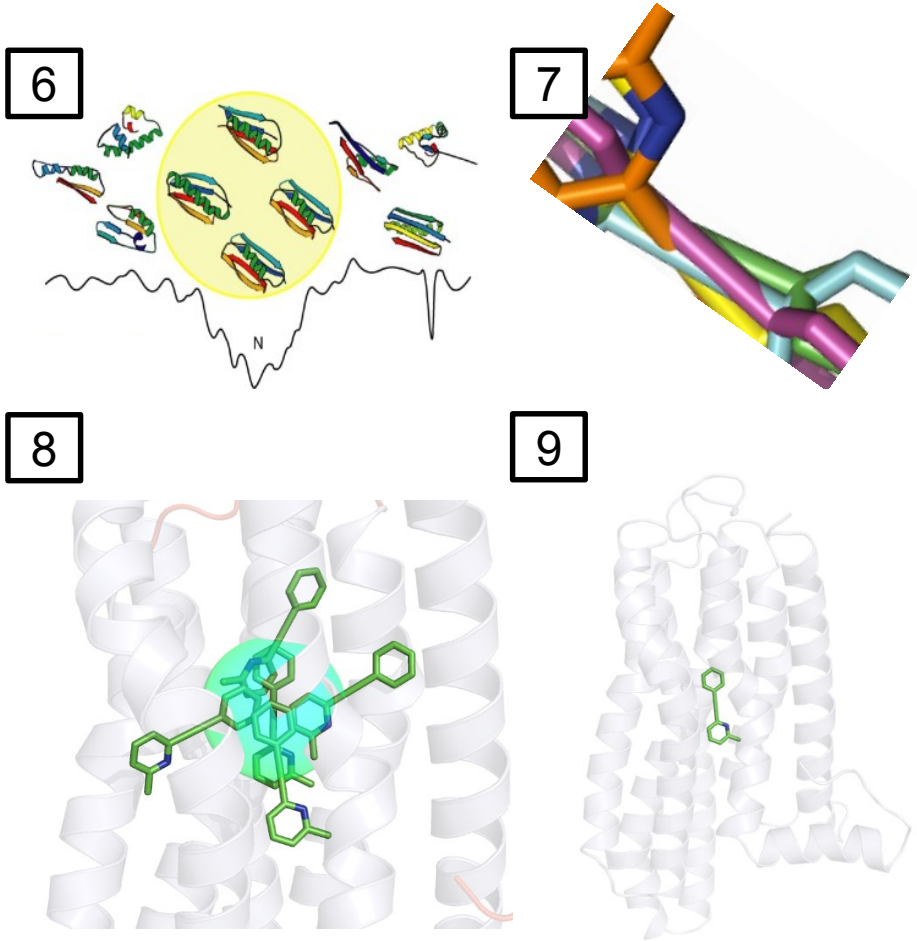
7) generate conformations of the ligand

↓

8) ligand docking via coarse/ fine sampling & all-atom refinement

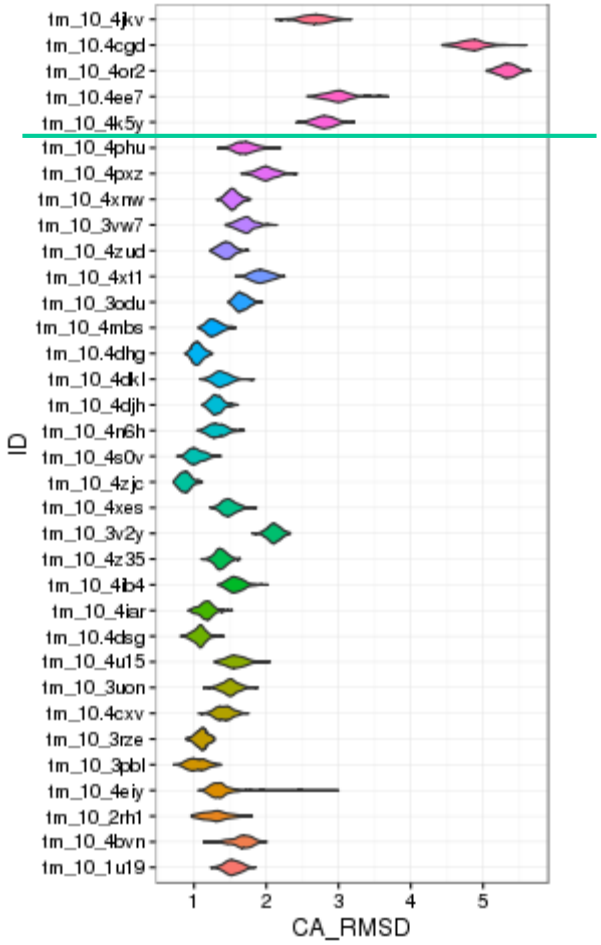
↓

9) data analysis and selection of final model



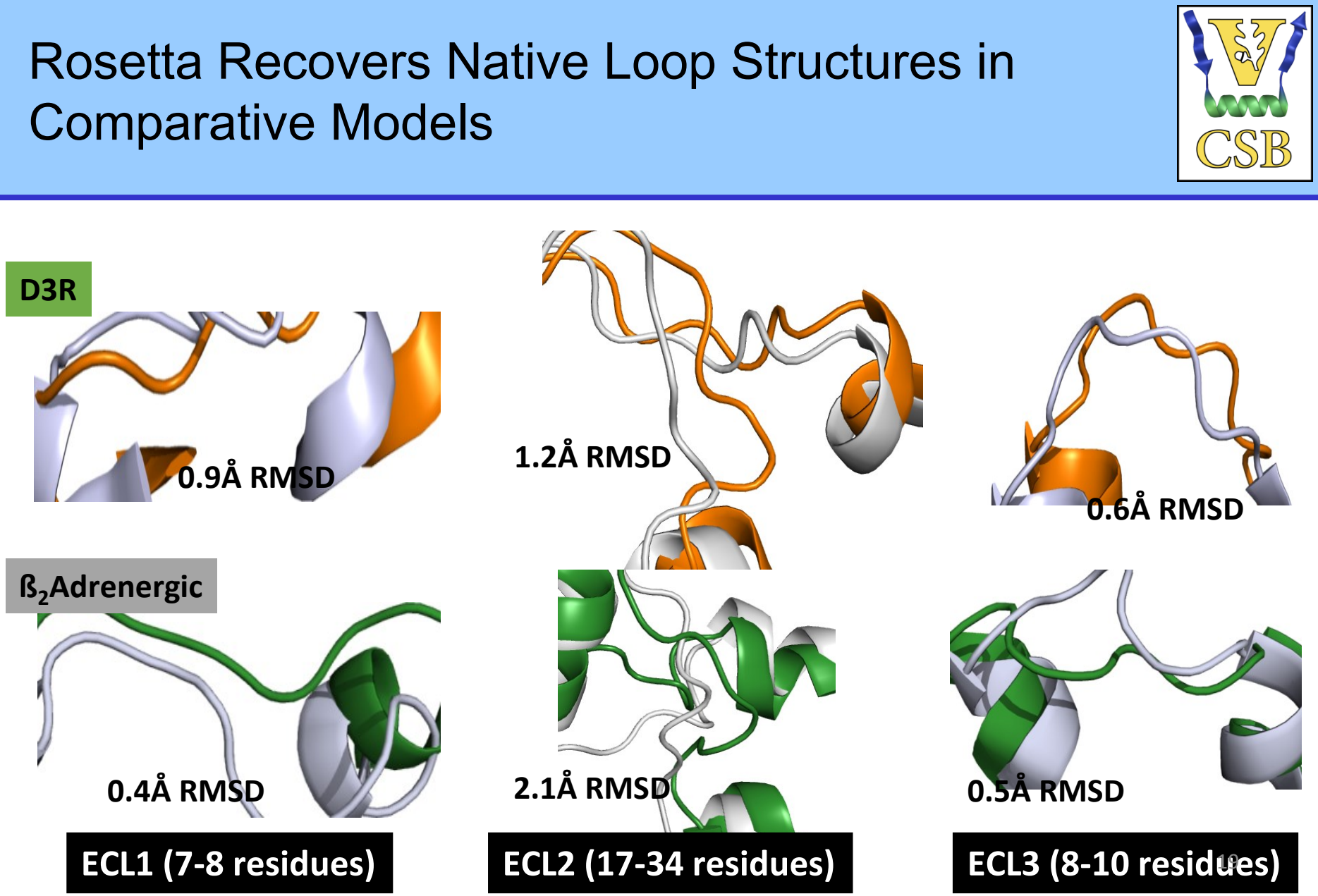
Combs, S.A., et al (2013) Nat Protoc, I(7), 1277-98.
Nguyen, E.D., et al (2013) PLOS ONE 8(7):e67302

Rosetta Captures the TM Region with High Accuracy



- TM RMSD of most Class A receptors falls between 1-2 Å
- Classes B, C, and F suffer from lack of template structures

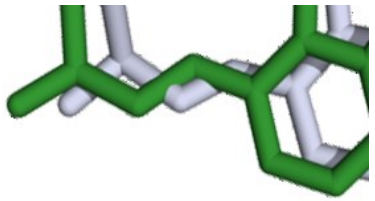
Rosetta Recovers Native Loop Structures in Comparative Models



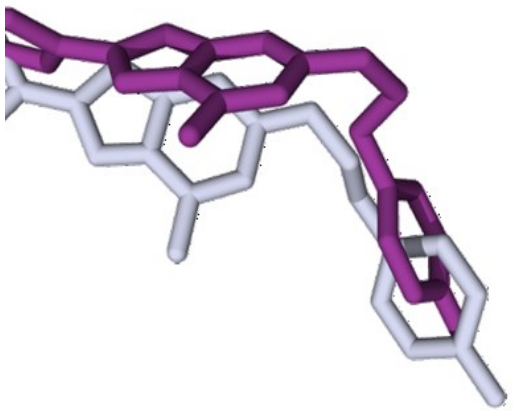
RosettaLigand Docking Recovers Native Ligand Binding Poses

Retinal docked into
hodopsin (3.7Å RMSD)

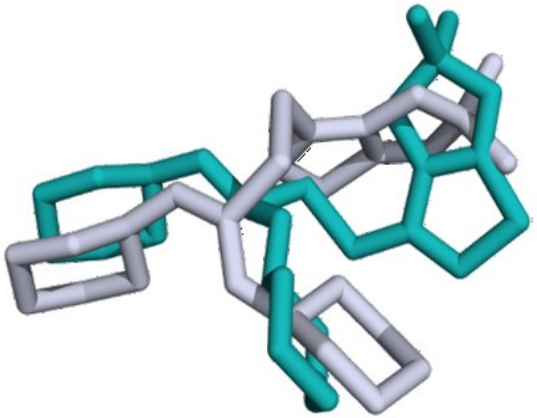
Carazolol docked into $ß_2$AR
(1.7Å RMSD)



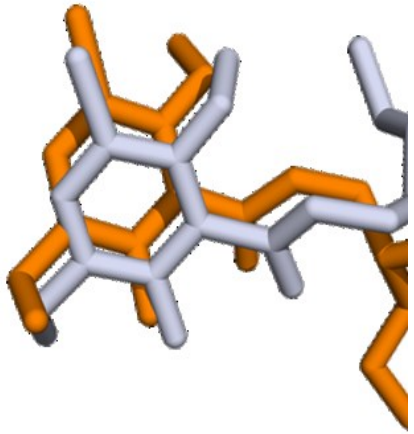
Cyanopindolol doc
into $ß_1$AR (1.9Å RM



241385 docked into A2A
denosine (2.5Å RMSD)

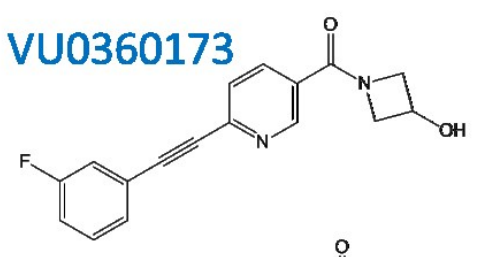


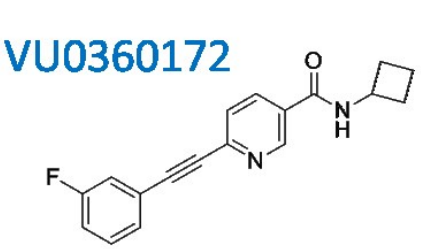
IT1t docked into CXCR4
(3.7Å RMSD)

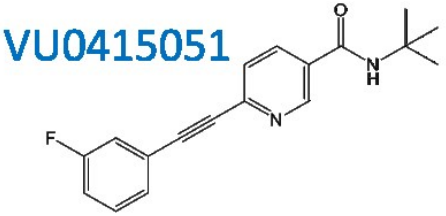


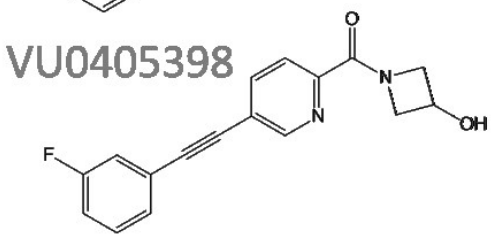
Eticlopride docked
dopamine D3 (2.6Å

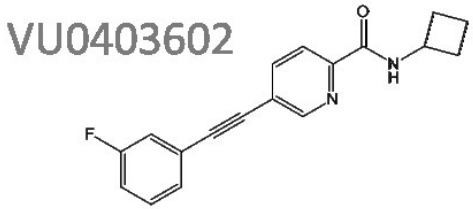
Comparative Models Aid in Drug Discovery - Docking mGlu5 PAMs

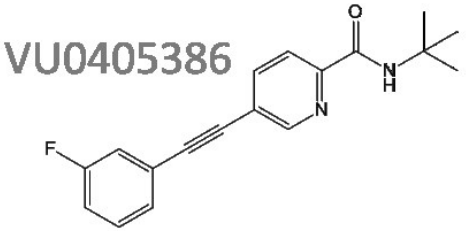


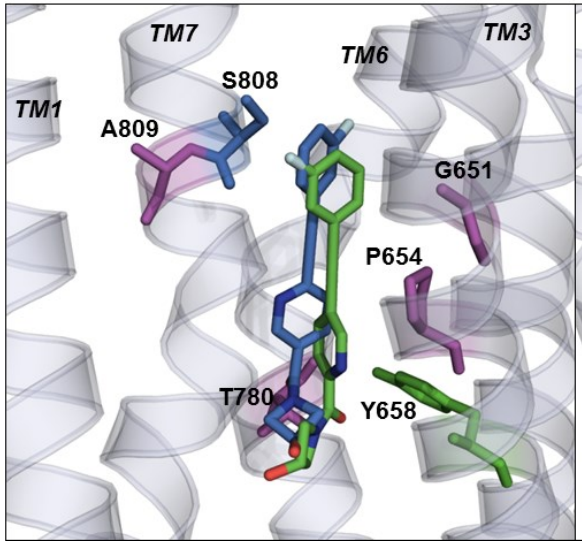


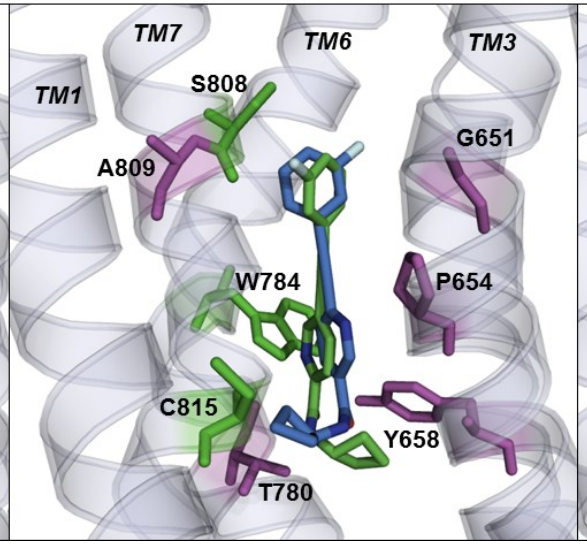


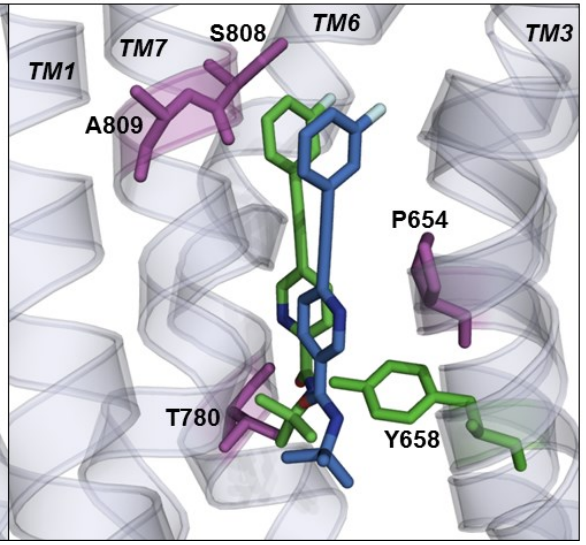




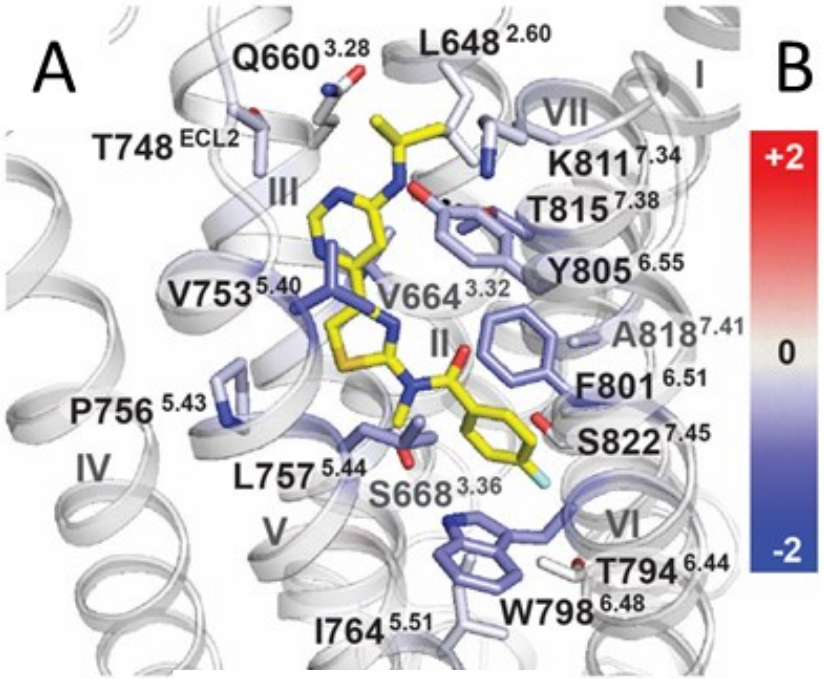


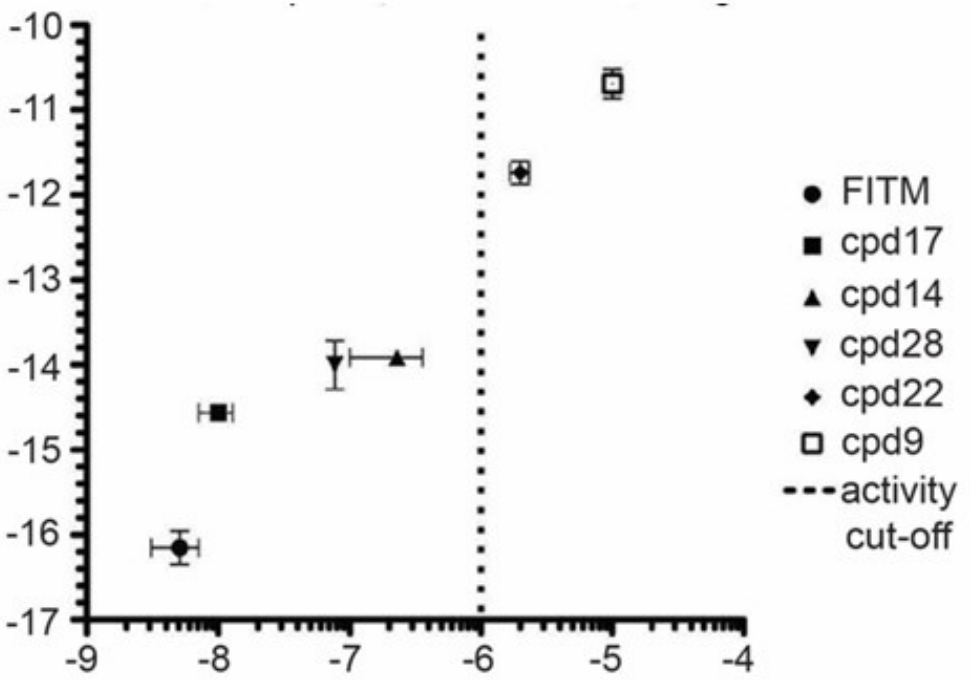


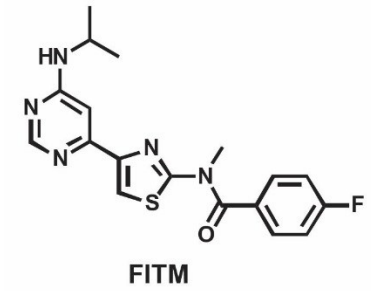


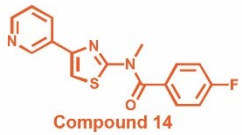


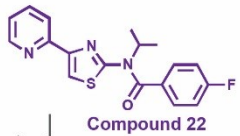
Rosetta Binding Energy Correlates with Binding Affinity

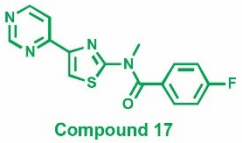


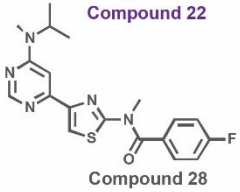










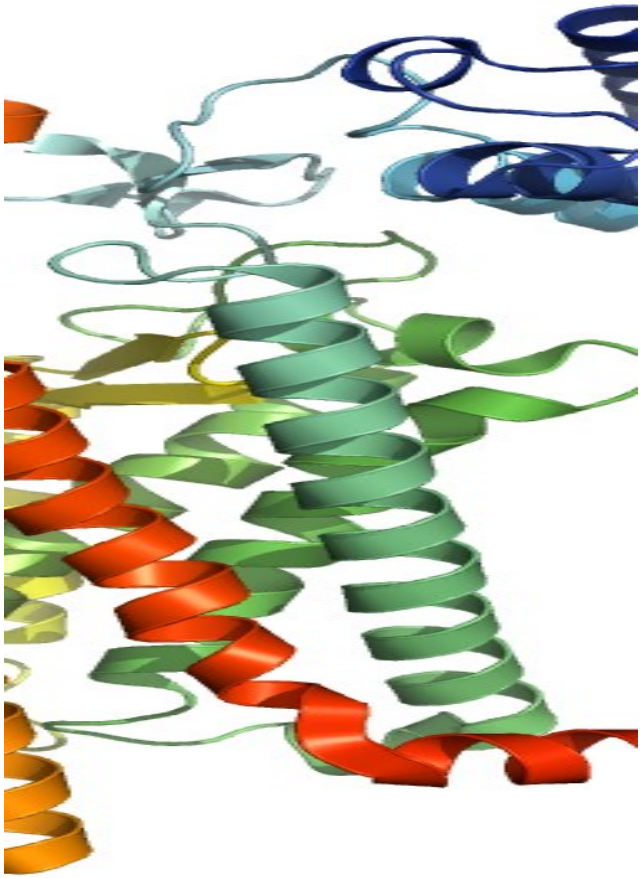


H. Wu, et al.; "Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator "; *Science*; **2013**

GPCR Dock 2013: Blind Assessment of Modeling Accuracy

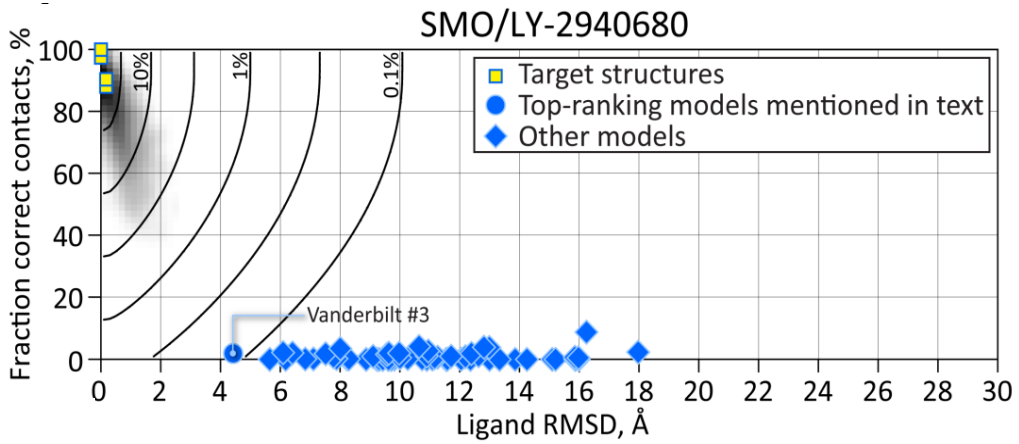
Human Smoothened Receptor/LY-2940680

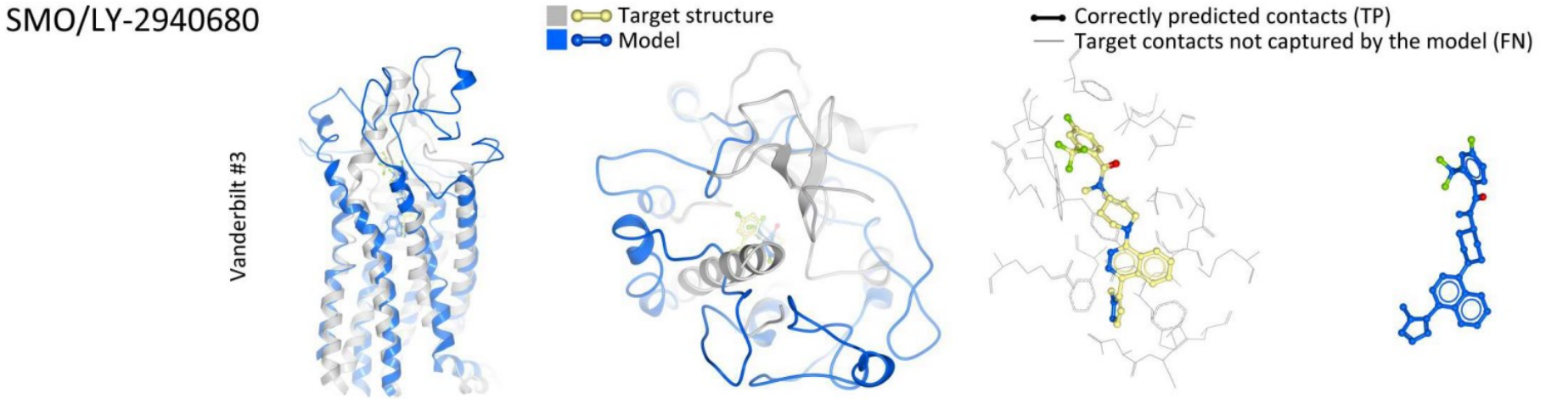
- <15% sequence identity with RMSD 3Å
- Long extracellular loops (up to 39 residues) with unique conformations
- Unique extracellular domain linker
- Ligand binding pocket has no resemblance to existing structures



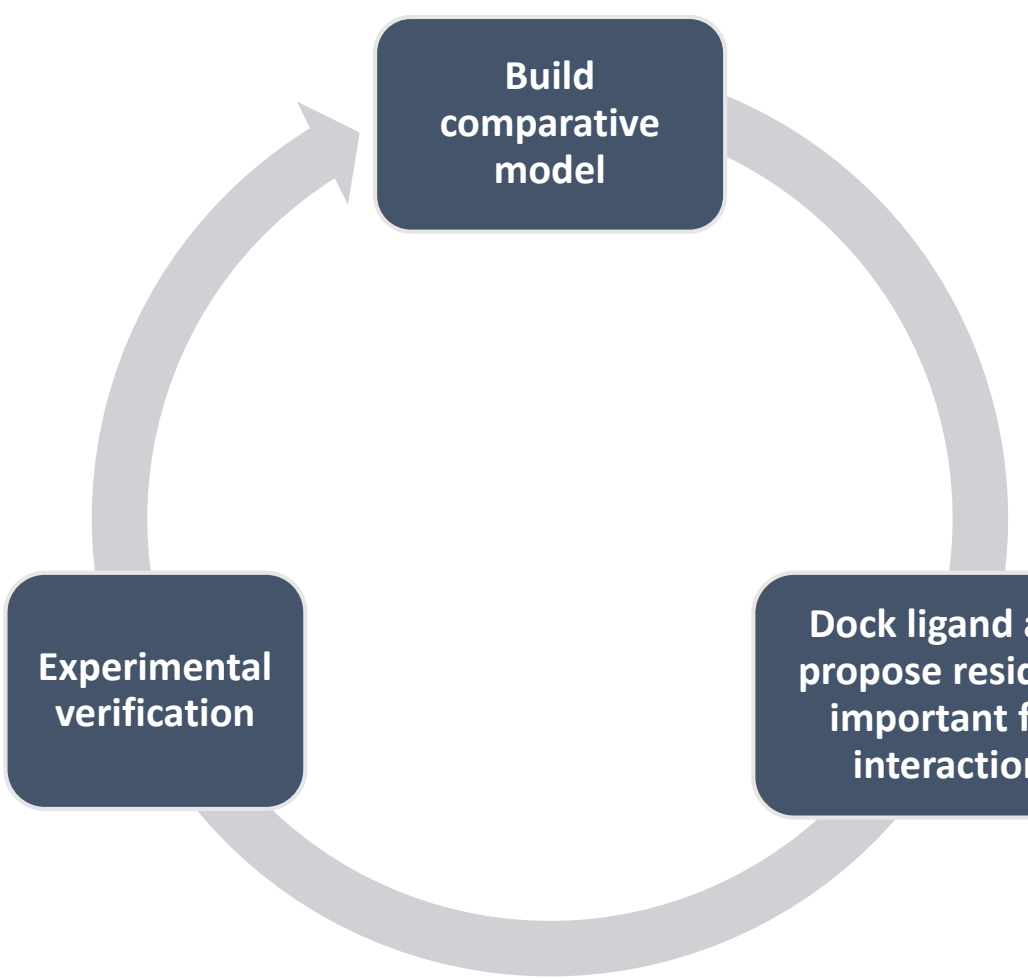
Most Accurate Prediction of hSMO/ LY-2940680 Complex

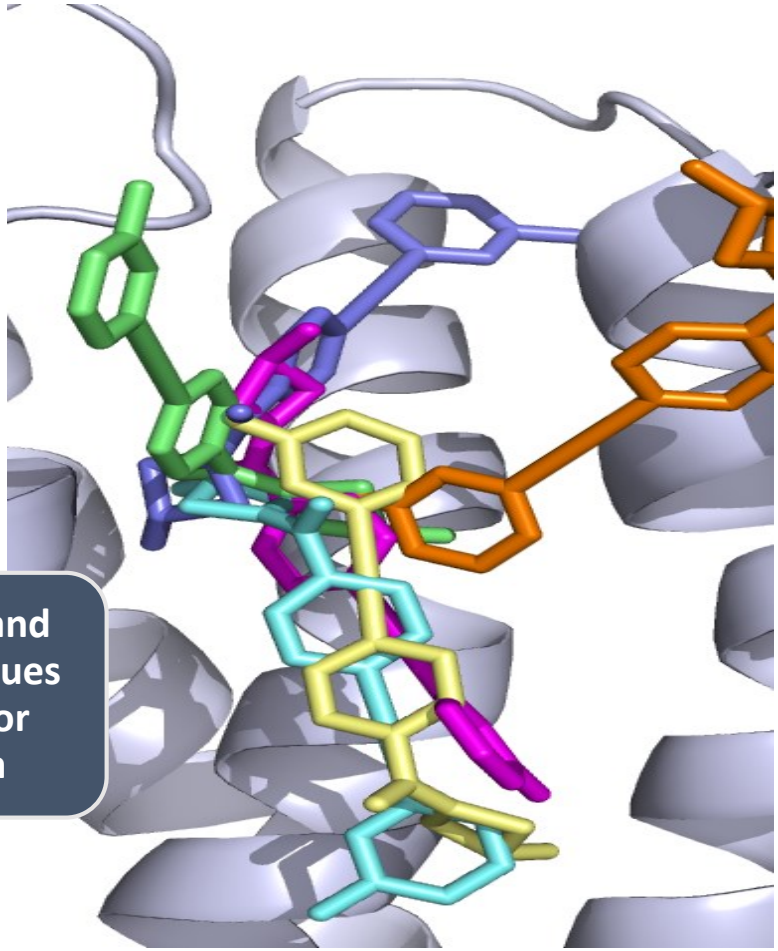
- 88 submitted models
- Ligand RMSD 4.42Å
- Correct contacts 8.8%
- Correct prediction of the helical fold and disulfide bond in ECL3



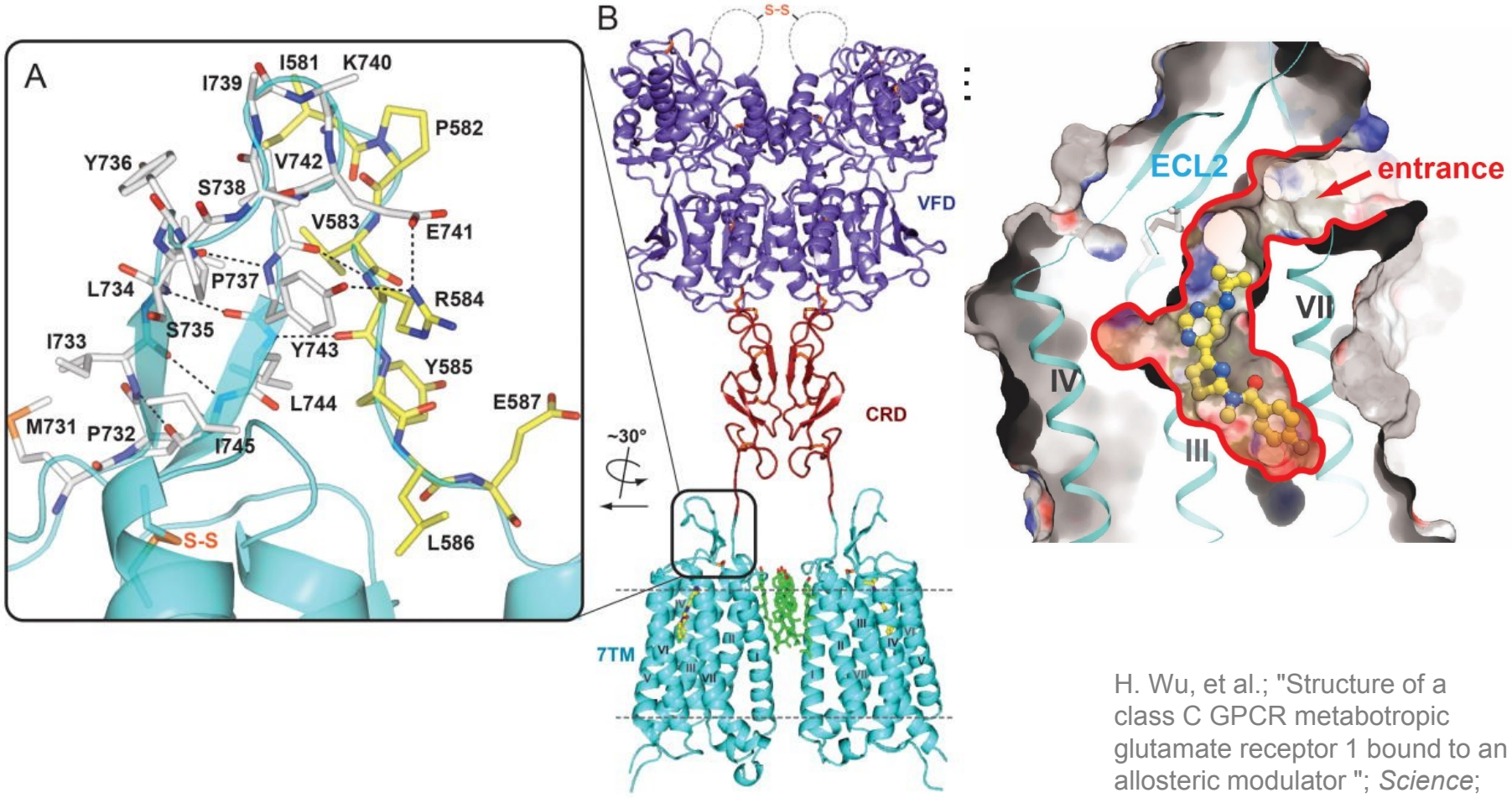


Iterative feedback loop between experiment and model



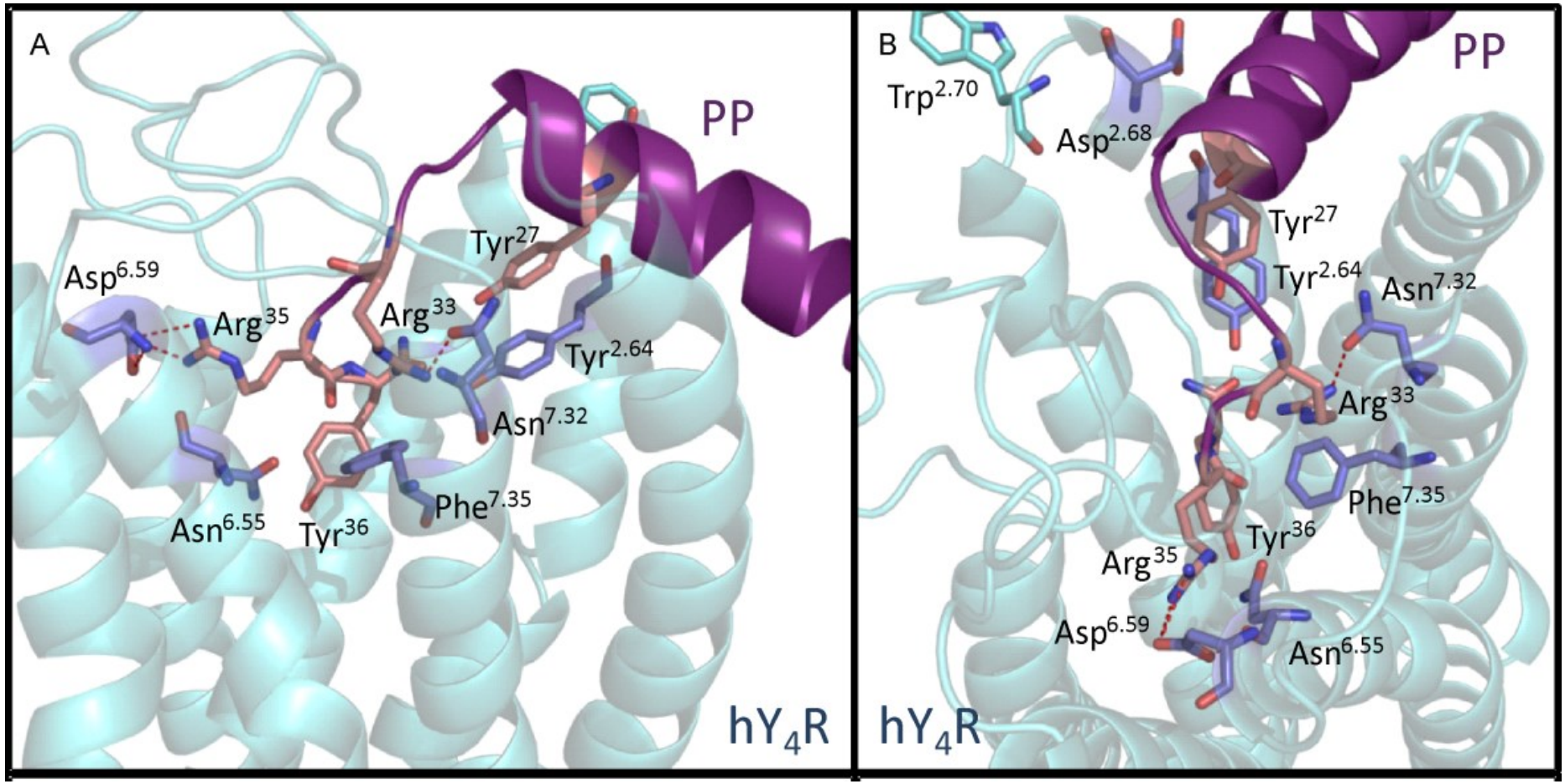


Modeling of GPCR Dimers



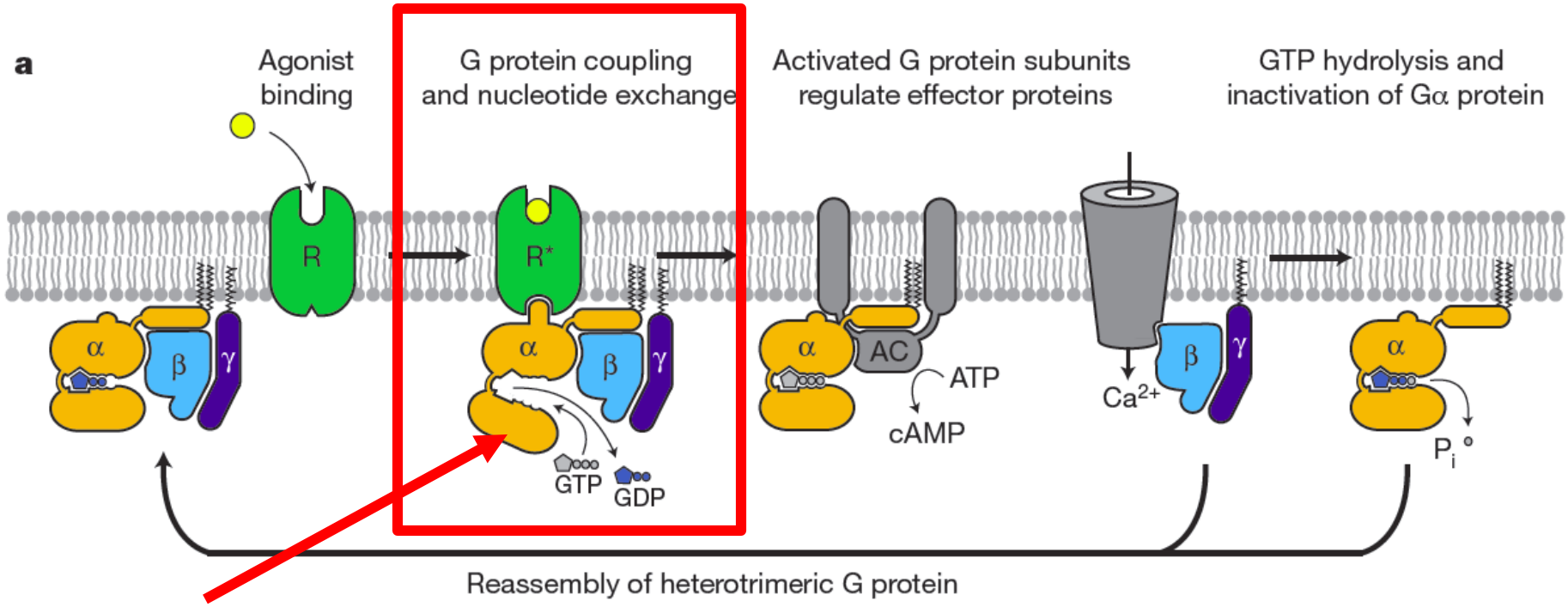
H. Wu, et al.; "Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator "; *Science*; **2013**

Modeling of Peptide Ligands at GPCRs



Xavier Pedragosa-Badia et al., “Pancreatic polypeptide is recognized by two domains of the human Y_4 receptor”

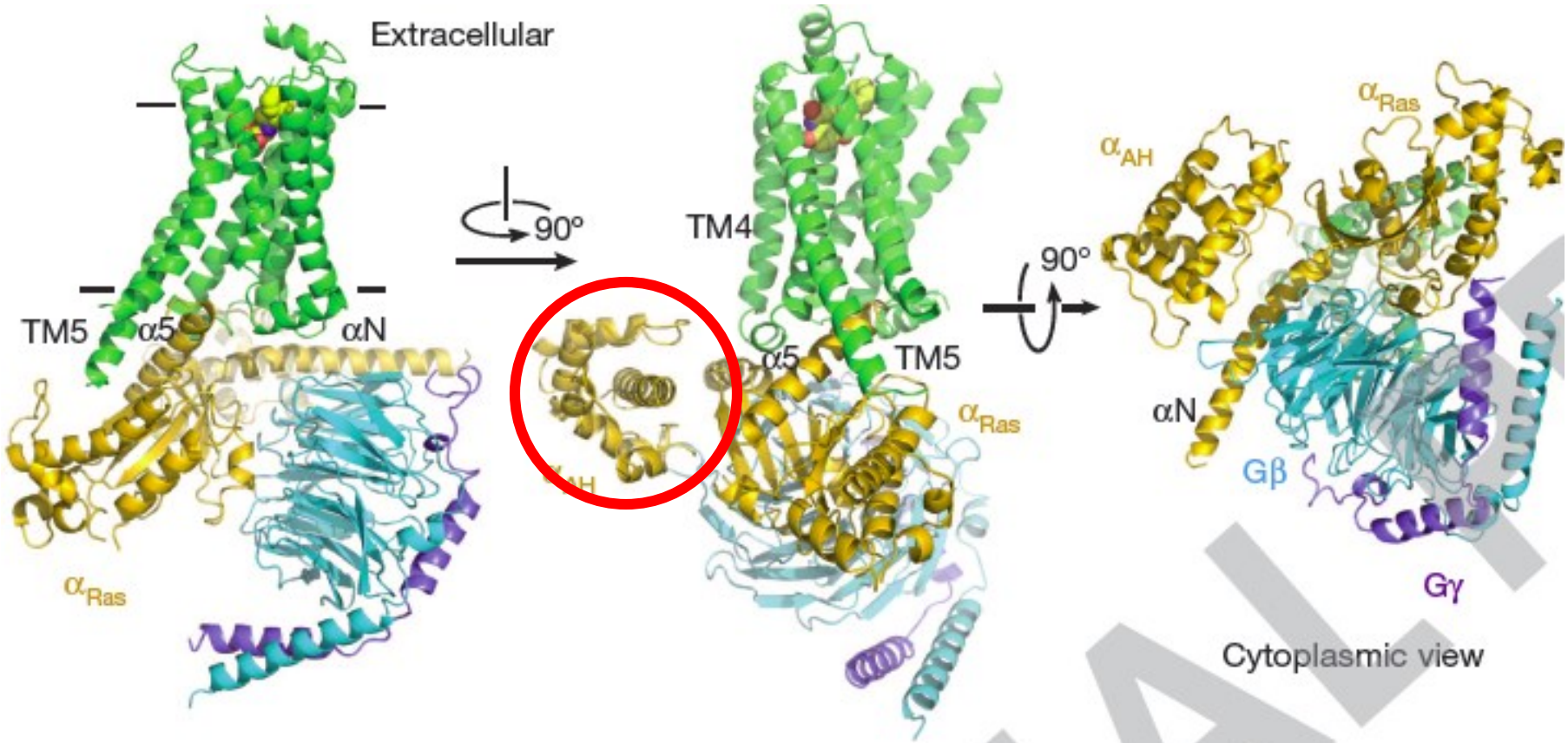
G protein cycle: Formation of the G protein | R* complex



Helical domain

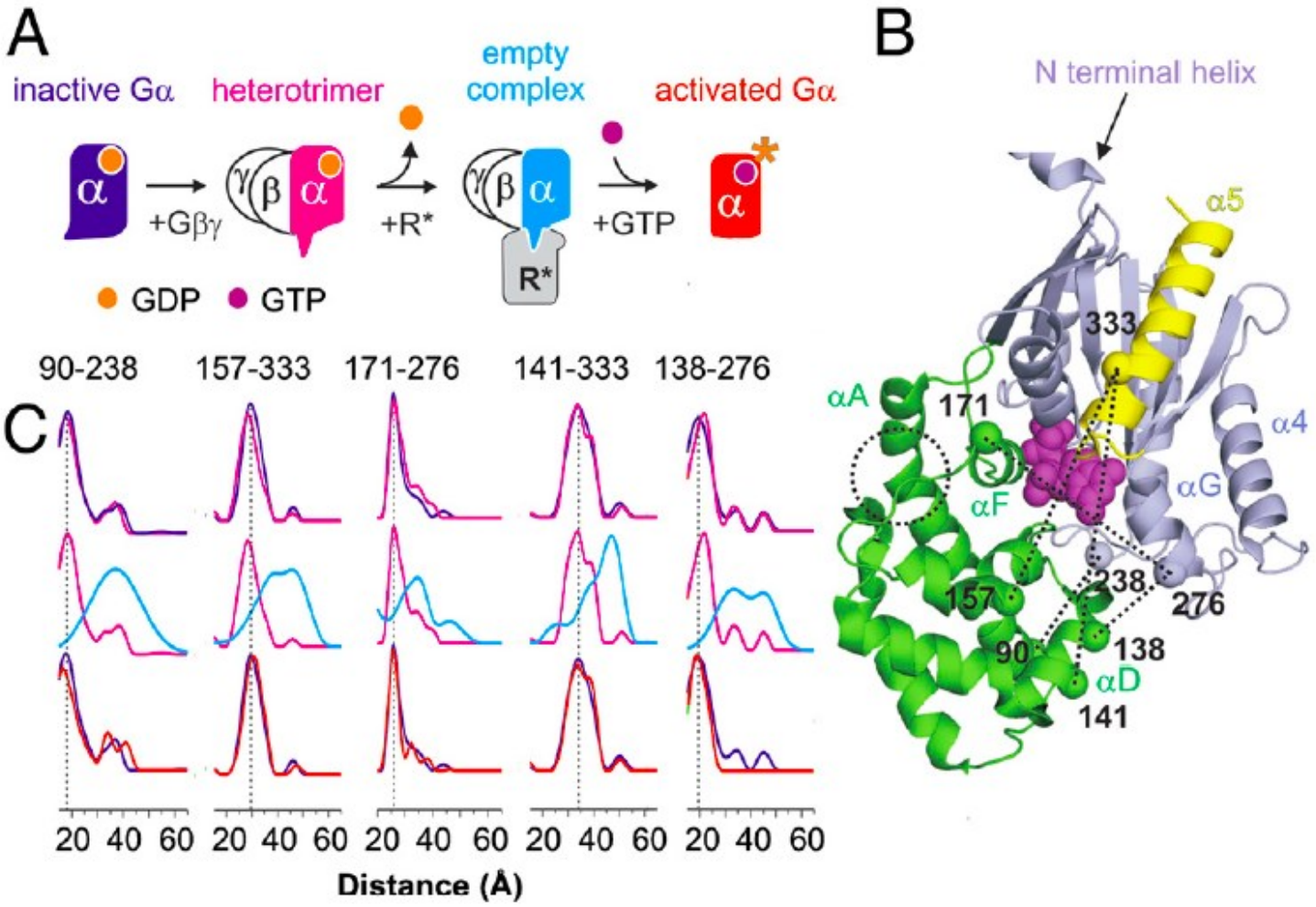
Rasmussen, S. G., et. al.; "Crystal structure of the beta(2) adrenergic receptor-Gs protein complex"; *Nature*; **2011**; Vol. p.

Overall structure of the b2AR|Gs complex omitting crystallization aids



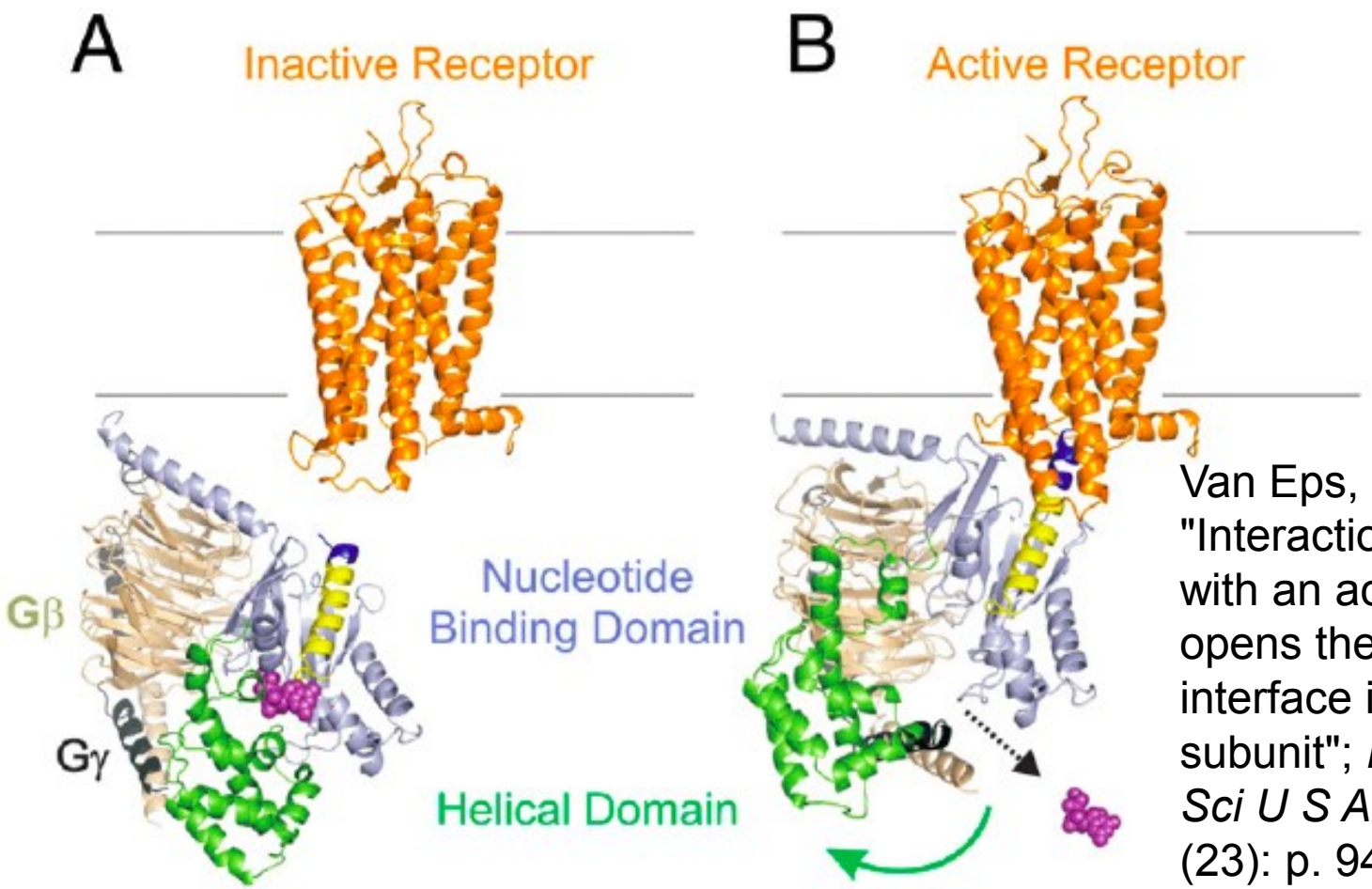
Rasmussen, S. G., et. al.; "Crystal structure of the beta(2) adrenergic receptor-Gs protein complex"; *Nature*; **2011**; Vol. p.

Experimental Distance Measurements Show Domain Opening in Active State



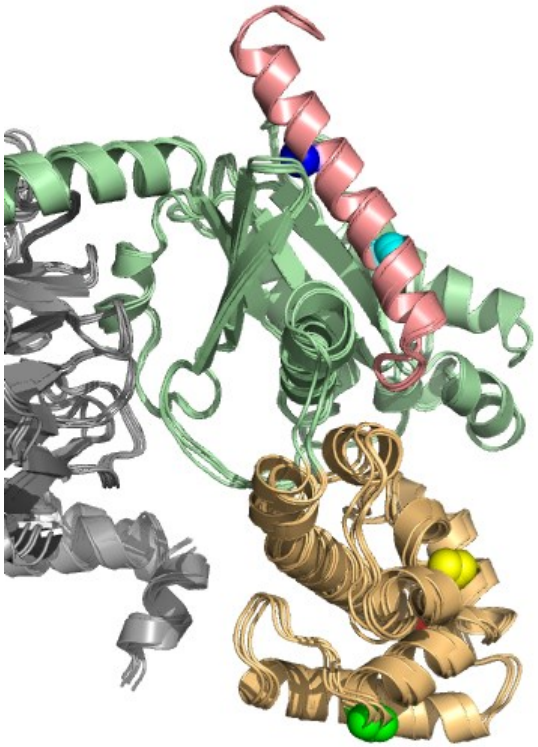
Van Eps, N., et. al.; "Interaction of a G protein with an activated receptor opens the interdomain interface in the alpha subunit"; *Proc Natl Acad Sci U S A*; **2011**; Vol. 108 (23): p. 9420-4.

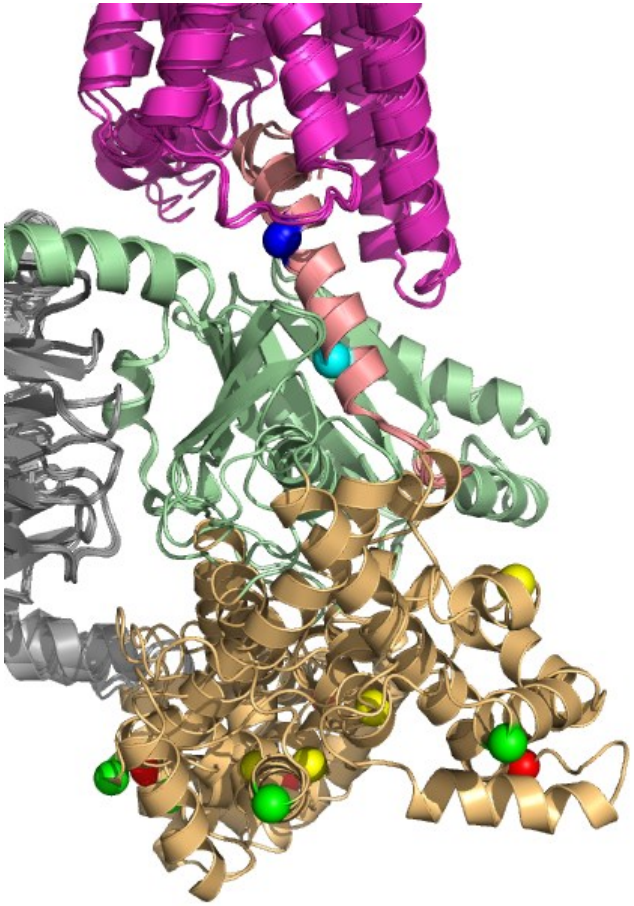
A model showing the opening of the inter-domain



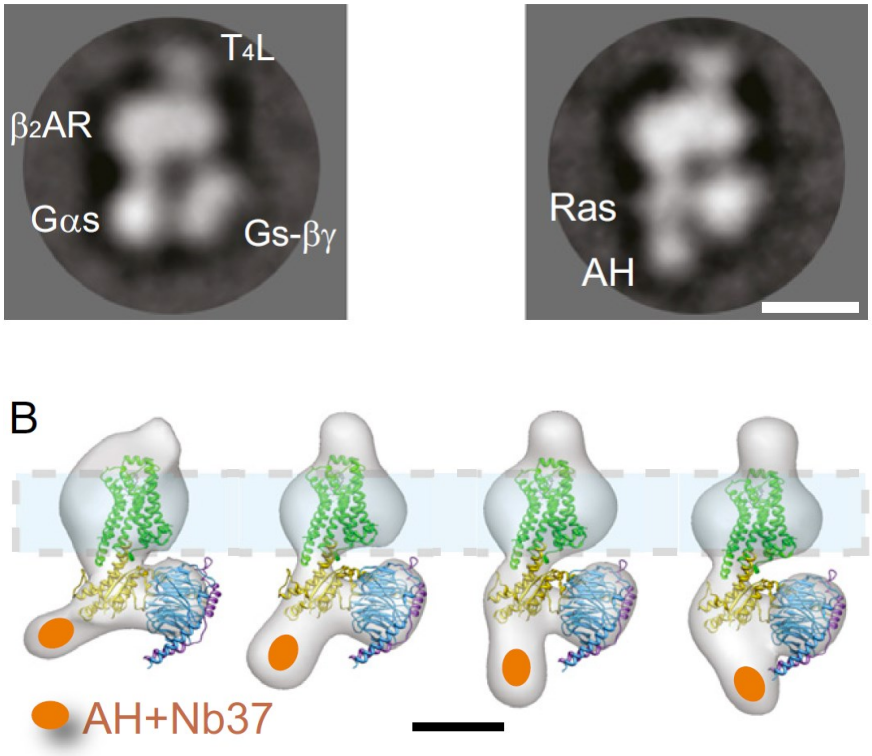
Van Eps, N., et. al.; "Interaction of a G protein with an activated receptor opens the interdomain interface in the alpha subunit"; *Proc Natl Acad Sci U S A*; **2011**; Vol. 108 (23): p. 9420-4.

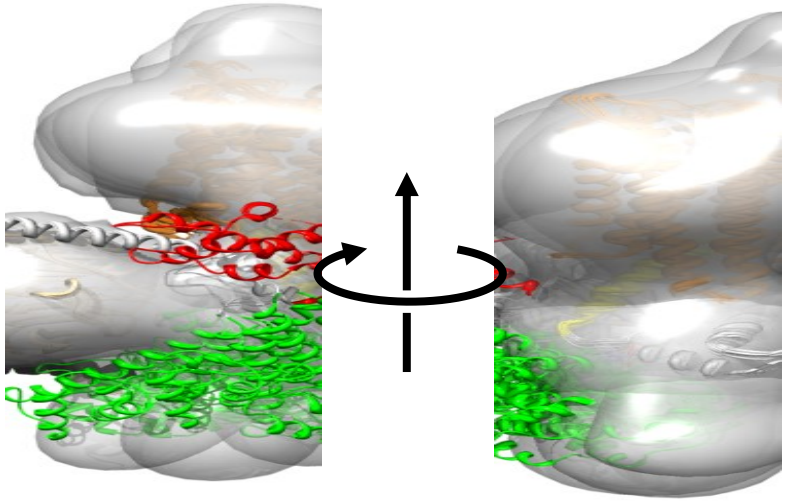
Modeling of Inactive and Active G protein states using EPR Data as Conformational Ensembles





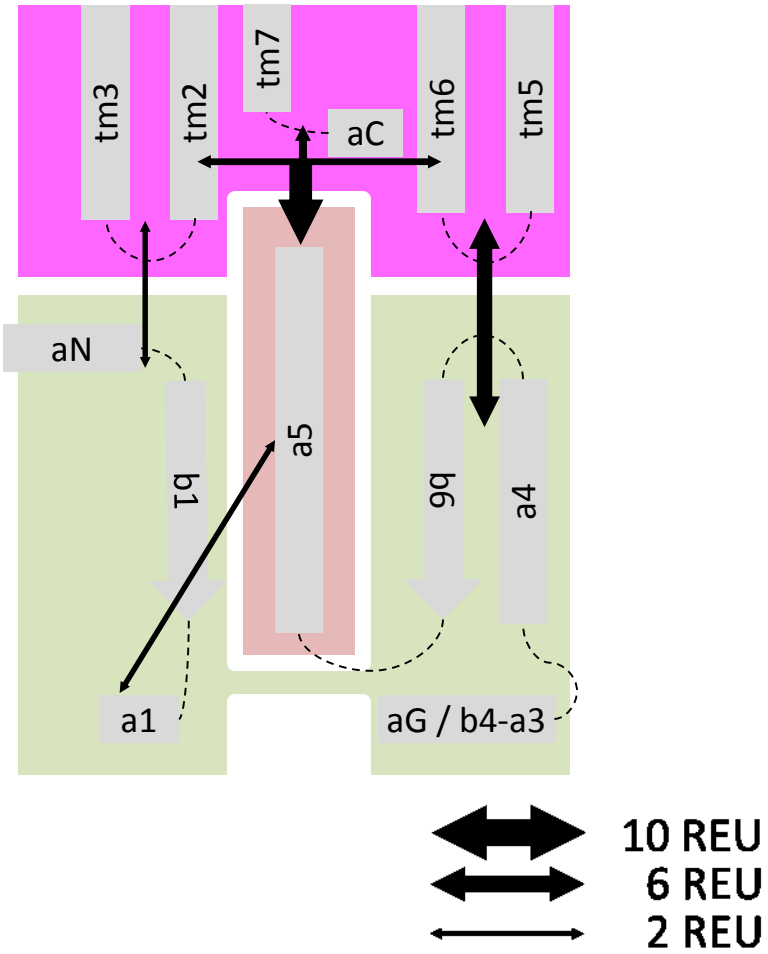
Single particle EM reconstructions of the β2AR|Gs complex

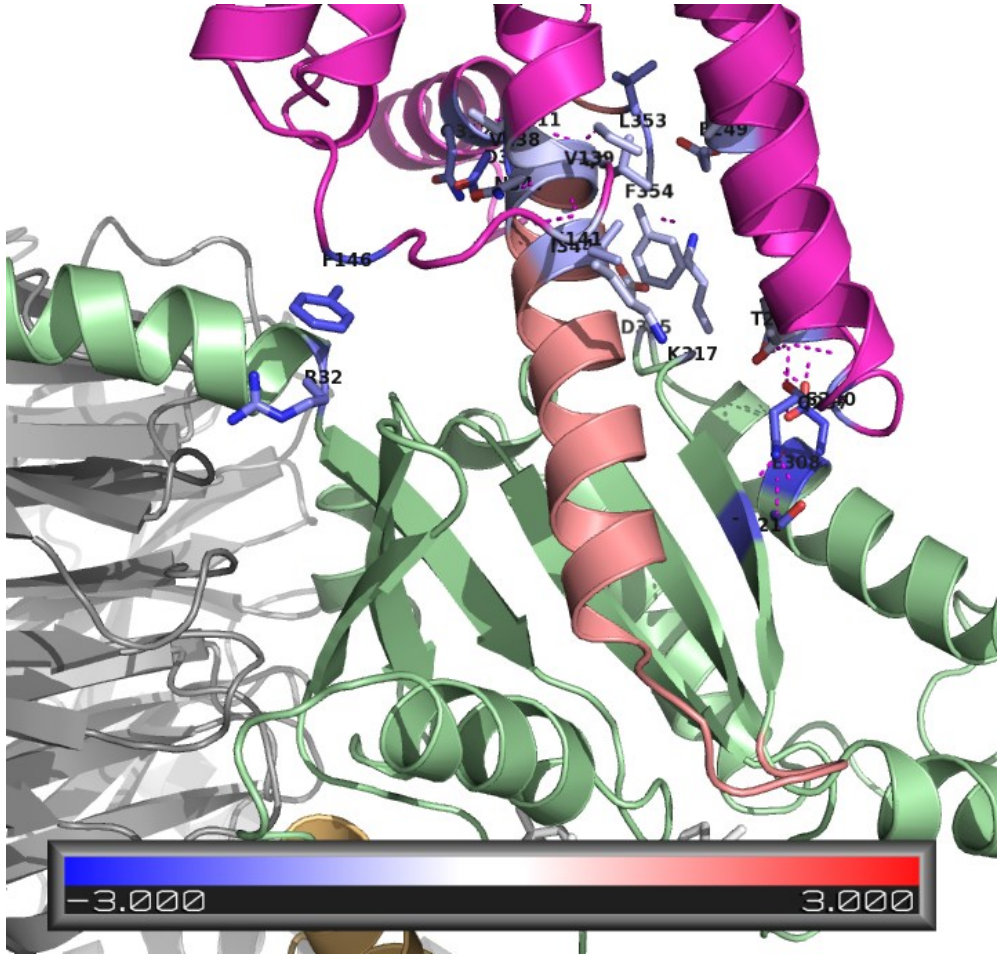




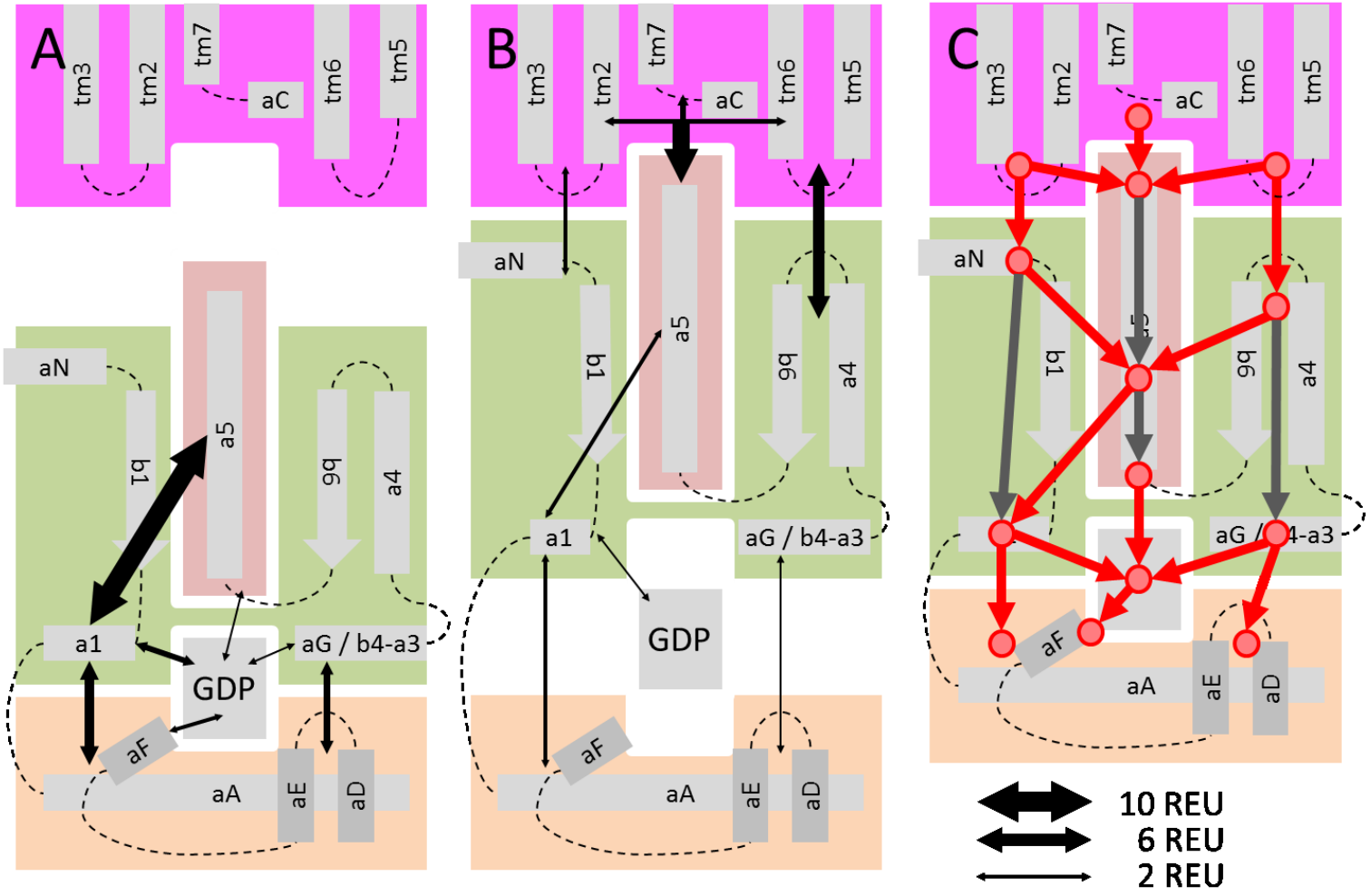
Westfield, G. H.; "Structural flexibility of the G alpha s alpha-helical domain in the beta2-adrenoceptor Gs complex"; *Proc Natl Acad Sci U S A*; **2011**; Vol. 108 (38): p. 16086-91.

Modeling Reveals Energetics of R* | Gai interface in the R*-Gai complex

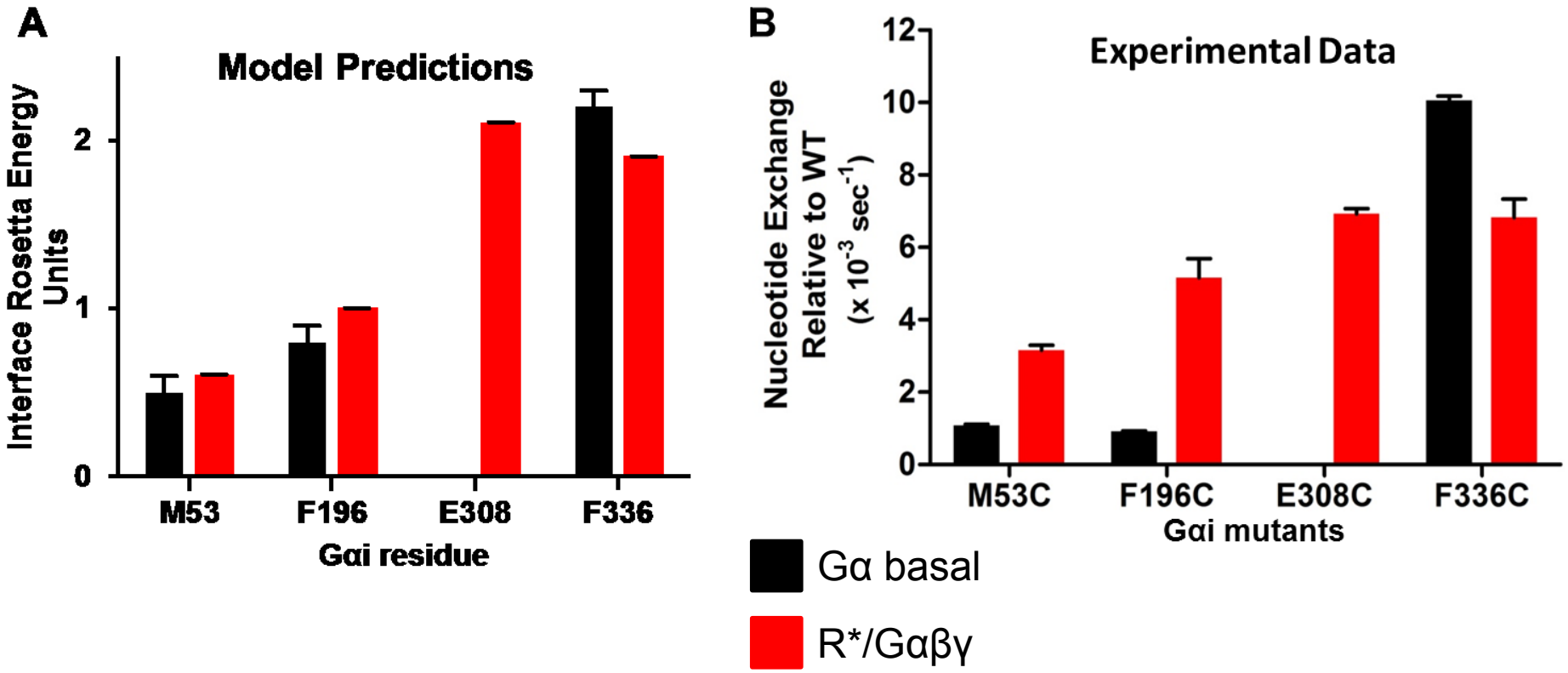




Energetic basis of signal transduction during Gai interaction with receptor R*



Basal and receptor-mediated nucleo-tide exchange rates match model



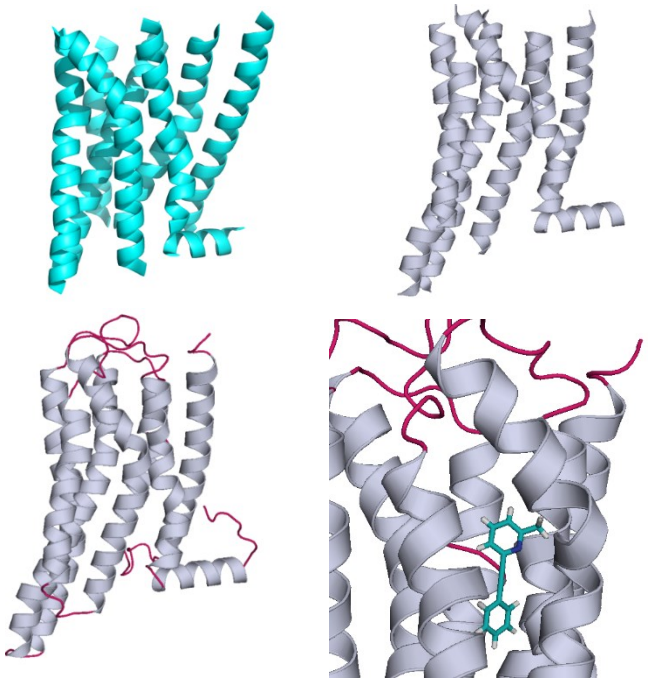
Conclusions

- Rosetta is a suite of molecular modeling software that utilizes fragment-based assembly and knowledge-based scoring potentials to rapidly sample conformational space and derive low energy conformational states of proteins.
- GPCR comparative modeling can accurately model receptors in the absence of crystal structure.
- Ligand docking into comparative models is possible.
- All methods are increased in accuracy by the use of sparse experimental data.
- With an increasing number of templates, computational methods become more accurate and will play an increasing role in therapeutic development.

Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model

```
D3R   1 MCFSVSLSATVALGCMFVPKVYII
B2AR  1 KEVYILLNWIGYVNSGFNPLITCR
```



Comparative Modeling of a GPCR and Ligand Docking using Rosetta

First description of multiple template homology modeling in Rosetta

Structure
Technical Advance

High-Resolution Comparative Modeling with RosettaCM

Yifan Song,[1,3] Frank DiMaio,[1,3] Ray Yu-Ruei Wang,[1] David Kim,[1] Chris Miles,[1] TJ Brunette,[1] James Thompson,[1] and David Baker[1,2,*]
[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA
[2]Howard Hughes Medical Institute, University of Washington, Box 357370, Seattle, WA 98195, USA
[3]These authors contributed equally to this work
*Correspondence: dabaker@u.washington.edu
http://dx.doi.org/10.1016/j.str.2013.08.005

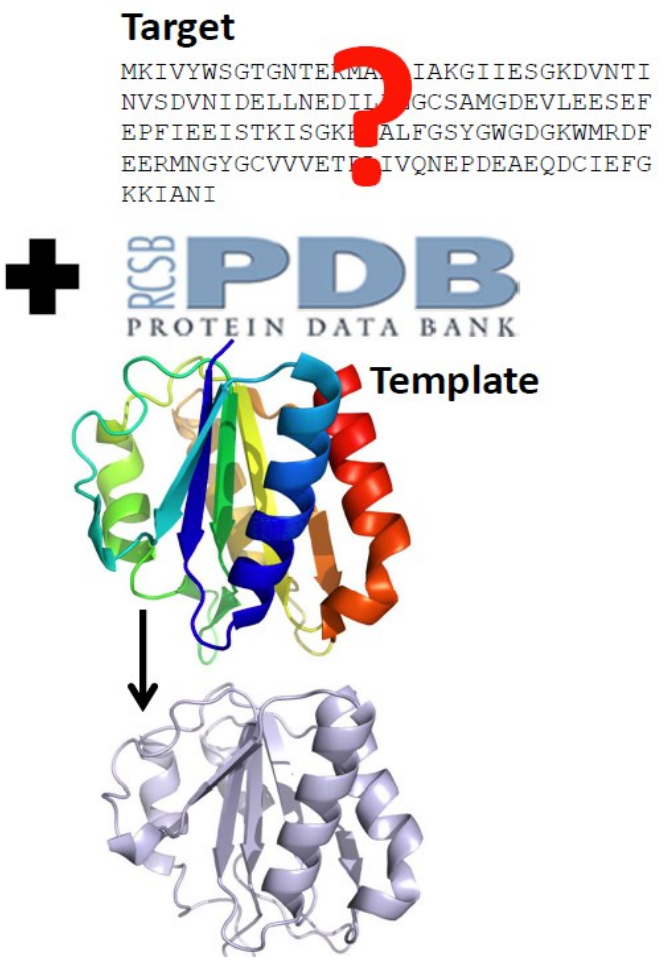
Application of RosettaCM to modeling GPCRs (basis of workshop)

Pancreatic Polypeptide Is Recognized by Two Hydrophobic Domains of the Human Y_4 Receptor Binding Pocket*[S]

Received for publication, July 28, 2013, and in revised form, December 21, 2013 Published, JBC Papers in Press, December 27, 2013, DOI 10.1074/jbc.M113.502021

Xavier Pedragosa-Badia[‡], Gregory R. Sliwoski[§], Elizabeth Dong Nguyen[§], Diana Lindner[‡1], Jan Stichel[‡], Kristian W. Kaufmann[§2], Jens Meiler[§3], and Annette G. Beck-Sickinger[‡4]
From the [‡]*Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, Universität Leipzig, 04103 Leipzig, Germany and the* [§]Center for Structural Biology, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8725

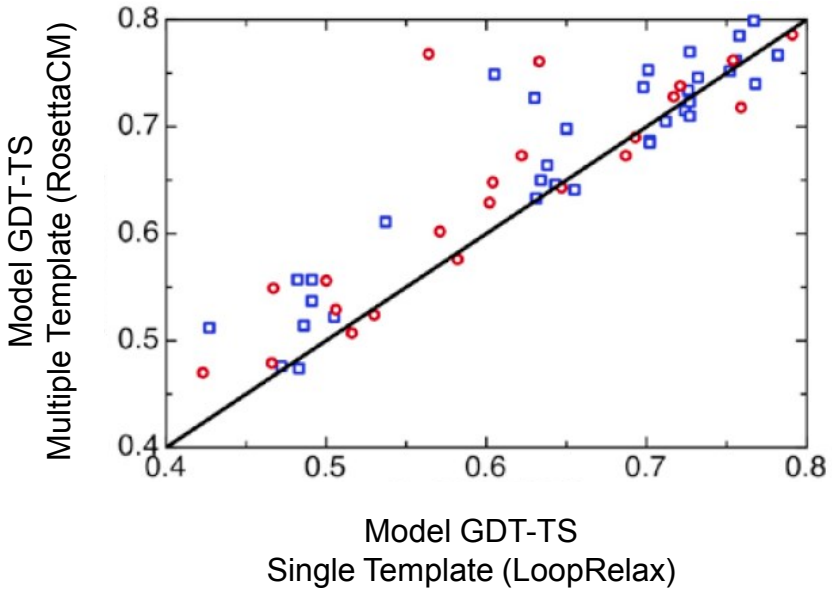
Updates to Rosetta's Comparative Modeling Method



- **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
  - Need final Dualspace Relax in Cartesian coordinate space

Multiple Templates Increase Model Accuracy Over Single Template



GDT-TS: Global Distance Test-Total Score
Higher value = better model

Large majority of models show an improvement when using multiple templates compared to single template

Necessary Files for GPCR Comparative Modeling

User Input

- Target sequence
- Template PDBs
- Fragment files
- Membrane topology file
- Disulfide (other constraints) file

Organizational

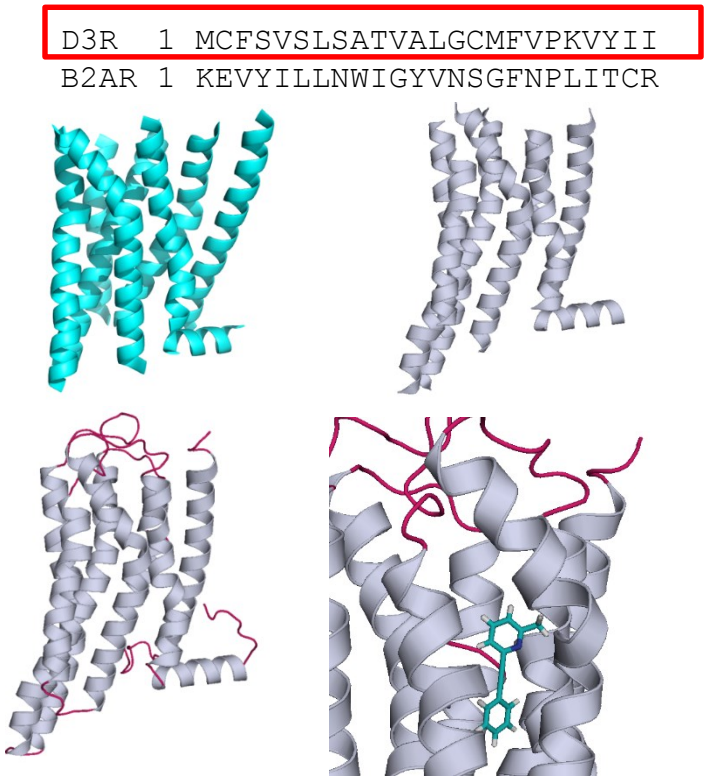
- Options file
- RosettaScripts XML

Predefined

- Score function weights files

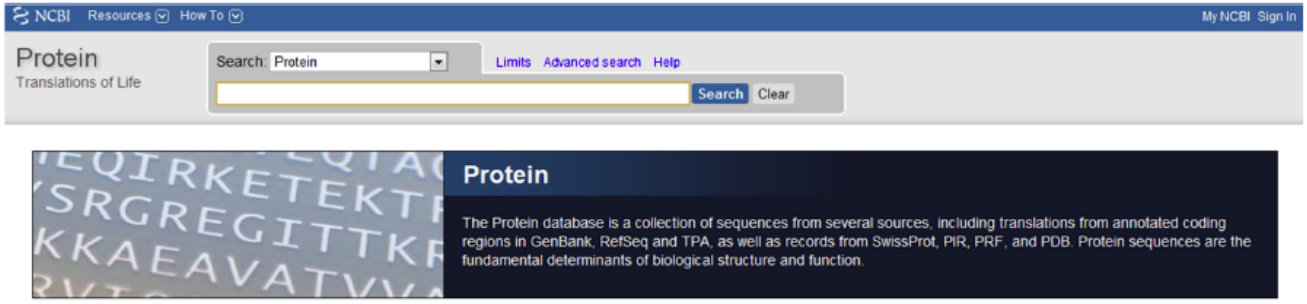
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align **TARGET** sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model



Target Sequence

- Need to obtain fasta for target sequence (NCBI Protein)
- Often our main interest is modeling of transmembrane region so we will choose not to model extra regions
  - Long unstructured termini
  - Class A GPCRs can have intracellular loops >100 residues
    - If deleting loops, need to replace with shorter poly-G or poly-A stretches to have continuous sequence



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

Target Sequence

> DRD3

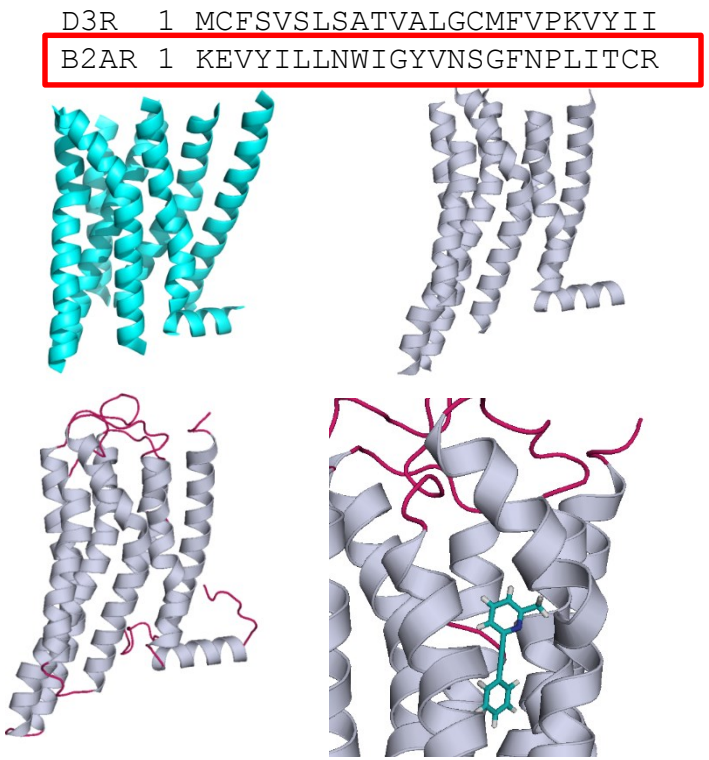
MASLSQLSSHLNYTCGAENSTGASQARPHAYYALSYCALILAIVFGNGLVCMAVLK
ERALQTTTNYLVVSLAVADLLVATLVMPWVVYLEVTGGVWNFSHICCDVFVTLDVM
MCTASILNLCAISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVLAFAVSCPL
LFGFNTTGDPTVCSISNPDFVIYSSVVSFYLPFGVTVLVYARIYEVLKQRRRK**RIL**
TRQNSQCNSVRPGFPQQTLSPDPAHLELKRYYSICQDTALGGPGFQERGGELKREE
KTRNSLSPTIAPKLSLEVRKLSNGRLSTSLKLGPPQPRGVPLREKKATQMVAIVLG
AFIVCWLPFFLTHVLNTHCQTCHVSPELYSATTWLGYVNSALNPVIYTTFNIEFRK
AFLKILSC

> DRD3 Modified

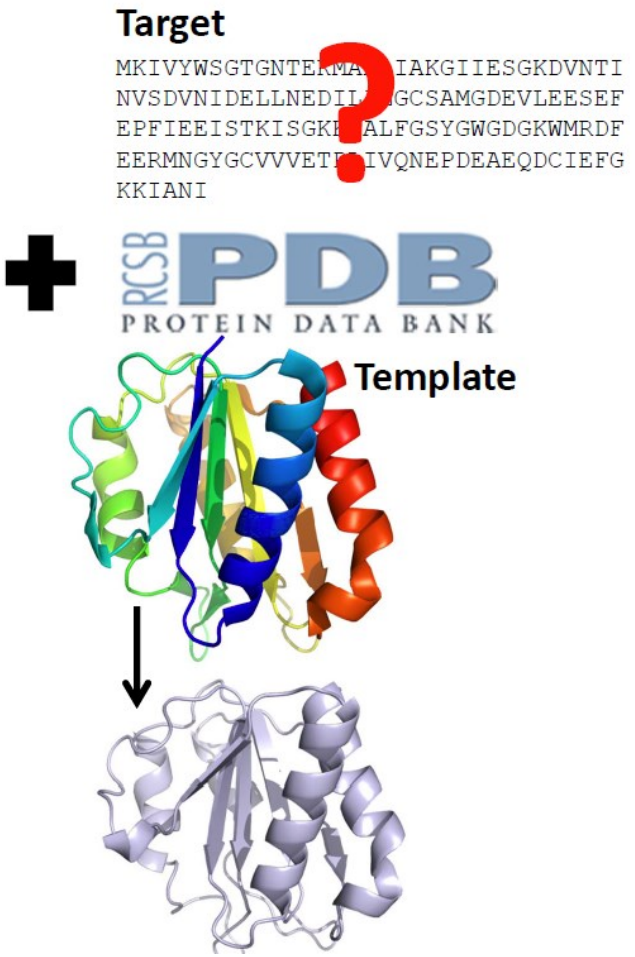
YALSYCALILAIVFGNGLVCMAVLKERALQTTTNYLVVSLAVADLLVATLVMPWVV
YLEVTGGVWNFSHICCDVFVTLDVMMCTASILNLCAISIDRYTAVVMPVHYQHGTG
QSSCRRVALMITAVWVLAFAVSCPLLFGFNTTGDPTVCSISNPDFVIYSSVVSFYL
PFGVTVLVYARIYEVLKQRRRKGVPLREKKATQMVAIVLGAFIVCWLPFFLTHVLN
THCQTCHVSPELYSATTWLGYVNSALNPVIYTTFNIEFRKAFLKILSC

Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of **TEMPLATE** structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model

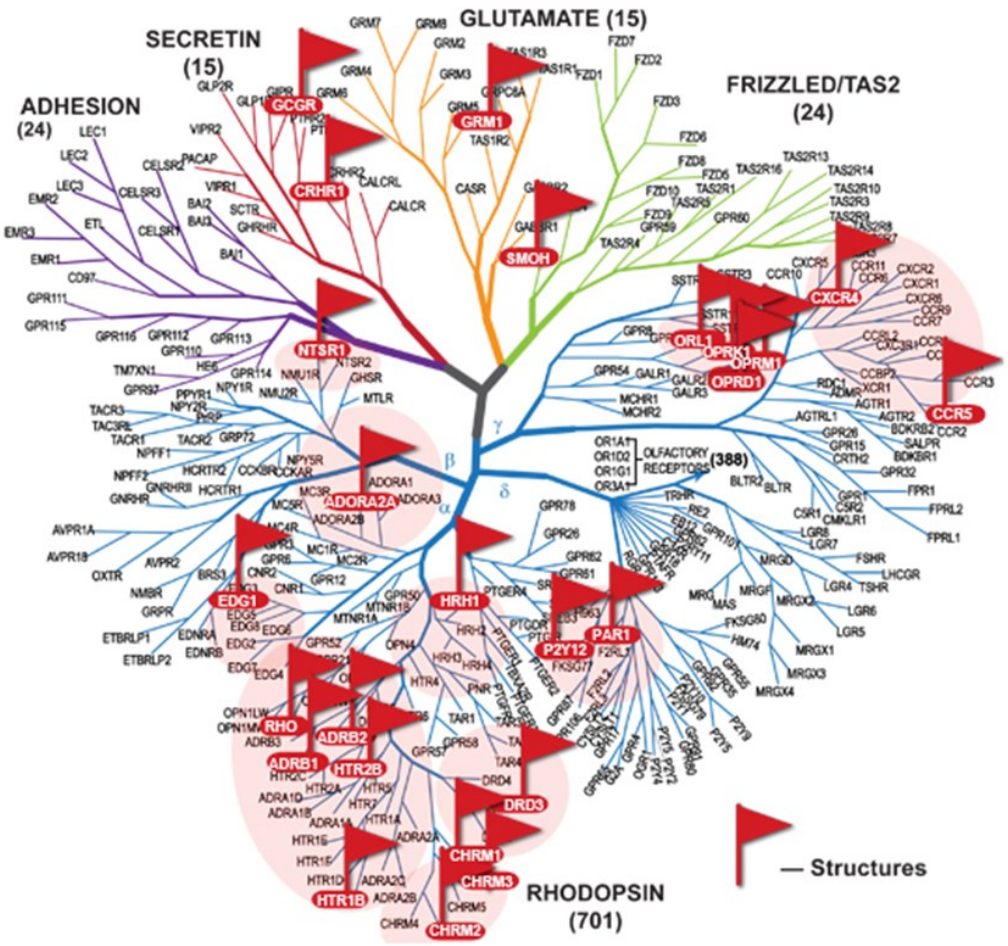


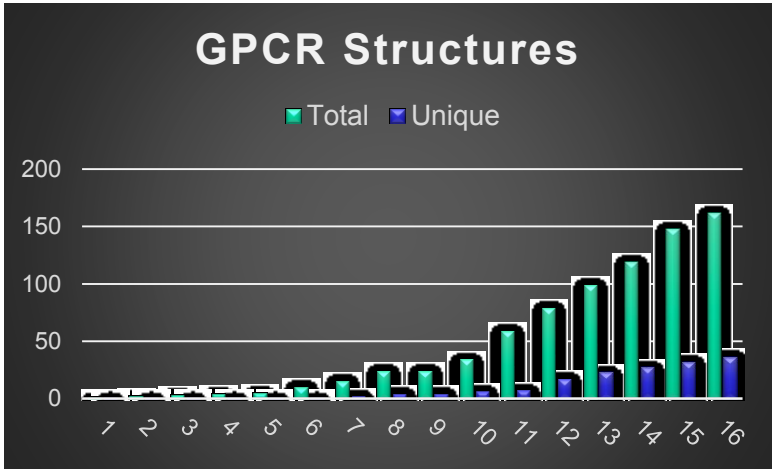
Identifying Template Structures



- **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)**

GPCR Structure Determination Yields Many Structural Templates for Comparative Modeling





GPCR Template Selection

	SMO	GRM5	GRM1	GCGR	CRF1R	FFAR1	P2Y12	P2Y1	PAR1	US28	CXCR4	CCR5	AT1R	NOP	MOR	KOR	DOR	OX2R	OX1R	NTSR1	LPA1	S1P1	M4R	M3R	M2R	M1R	5HT2B	5HT1B	D3R	H1R	A2AAR	B1AR	B2AR	Rho
Rho	12	14	14	15	13	17	18	21	18	20	20	21	19	21	21	19	20	18	16	19	17	19	20	21	23	21	20	21	25	20	20	19	21	100
B2AR	11	13	13	13	13	17	19	20	20	21	23	23	25	25	26	25	26	28	27	20	21	23	30	34	29	33	34	38	36	34	31	63	100	
B1AR	13	13	14	15	15	18	22	19	18	23	22	23	23	25	24	25	26	28	28	22	25	26	29	33	31	32	34	41	38	35	31	100		
A2AAR	12	17	16	14	15	18	18	21	20	18	19	20	24	22	23	24	23	28	29	21	25	27	26	29	25	27	28	30	30	34	100			
H1R	13	12	12	13	15	16	19	20	21	21	23	23	28	27	28	30	27	24	27	19	24	28	39	39	38	40	30	38	34	100				
D3R	13	11	12	14	12	18	22	22	21	22	25	22	23	26	27	25	26	28	26	25	28	26	35	31	33	34	34	39	100					
5HT1B	13	13	13	16	13	15	20	22	18	23	23	21	23	25	26	28	27	27	25	23	26	32	34	38	36	37	32	100						
5HT2B	11	14	15	12	15	22	21	23	21	23	24	21	27	28	27	27	25	25	24	23	20	26	29	31	29	29	100							
M1R	16	13	12	16	14	17	20	21	18	25	21	24	25	24	23	25	23	24	23	22	25	25	67	75	64	100								
M2R	13	11	11	12	12	17	22	21	22	25	22	21	27	26	26	28	26	25	23	22	25	26	84	68	100									
M3R	18	12	11	14	15	17	21	22	20	26	24	23	28	24	25	26	25	26	23	22	25	26	69	100										
M4R	13	12	11	14	13	18	21	22	20	24	23	22	28	28	27	29	27	23	22	22	26	24	100											
S1P1	9	13	16	14	9	15	18	21	17	19	20	20	21	18	18	19	19	23	21	20	41	100												
LPA1	11	13	14	14	12	14	18	18	20	18	17	17	19	24	19	18	21	23	22	16	100													
NTSR1	11	14	15	15	15	17	18	24	20	22	23	21	25	26	22	27	23	22	24	100														
OX1R	10	17	14	16	18	19	18	24	23	19	23	19	22	30	28	29	27	80	100															
OX2R	8	16	15	15	18	19	20	27	22	24	26	21	26	31	31	30	29	100																
DOR	11	15	15	17	17	21	24	28	27	26	26	26	32	61	71	70	100																	
KOR	11	15	12	13	16	18	26	29	26	26	27	28	36	62	70	100																		
MOR	10	16	13	16	16	19	26	28	27	26	29	27	33	60	100																			
NOP	11	15	13	16	17	21	24	27	27	26	30	28	32	100																				
AT1R	13	18	17	15	15	22	29	27	27	29	37	33	100																					
CCR5	11	16	16	12	14	21	27	25	22	31	32	100																						
CXCR4	13	15	16	14	12	22	25	25	23	31	100																							
US28	13	15	14	18	17	21	23	26	25	100																								
PAR1	14	14	13	16	15	25	21	30	100																									
P2Y1	12	14	14	13	13	18	23	100																										
P2Y12	12	11	11	13	13	20	100																											
FFAR1	11	14	13	15	11	100																												
CRF1R	22	14	12	35	100																													
GCGR	19	9	8	100																														
GRM1	10	75	100																															
GRM5	9	100																																
SMO	100																																	

Current available crystal structures share identities often below 30%

Emphasizes need for multiple template modeling

GPCR Template Selection

	SMO	GRM5	GRM1	GCGR	CRF1R	FFAR1	P2Y12	P2Y1	PAR1	US28	CXCR4	CCR5	AT1R	NOP	MOR	KOR	DOR	OX2R	OX1R	NTSR1	LPA1	S1P1	M4R	M3R	M2R	M1R	5HT2B	5HT1B	D3R	H1R	A2AAR	B1AR	B2AR	Rho
Rho	12	14	14	15	13	17	18	21	18	20	20	21	19	21	21	19	20	18	16	19	17	19	20	21	23	21	20	21	25	20	20	19	21	100
B2AR	11	13	13	13	13	17	19	20	20	21	23	23	25	25	26	25	26	28	27	20	21	23	30	34	29	33	34	38	36	34	31	63	100	
B1AR	13	13	14	15	15	18	22	19	18	23	22	23	23	25	24	25	26	28	28	22	25	26	29	33	31	32	34	41	38	35	31	100		
A2AAR	12	17	16	14	15	18	18	21	20	18	19	20	24	22	23	24	23	28	29	21	25	27	26	29	25	27	28	30	30	34	100			
H1R	13	12	12	13	15	16	19	20	21	21	23	23	28	27	28	30	27	24	27	19	24	28	39	39	38	40	30	38	34	100				
D3R	13	11	12	14	12	18	22	22	21	22	25	22	23	26	27	25	26	28	26	25	28	26	35	31	33	34	34	39	100					
5HT1B	13	13	13	16	13	15	20	22	18	23	23	21	23	25	26	28	27	27	25	23	26	32	34	38	36	37	32	100						
5HT2B	11	14	15	12	15	22	21	23	21	23	24	21	27	28	27	27	25	25	24	23	20	26	29	31	29	29	100							
M1R	16	13	12	16	14	17	20	21	18	25	21	24	25	24	23	25	23	24	23	22	25	25	67	75	64	100								
M2R	13	11	11	12	12	17	22	21	22	25	22	21	27	26	26	28	26	25	23	22	25	26	84	68	100									
M3R	18	12	11	14	15	17	21	22	20	26	24	23	28	24	25	26	25	26	23	22	25	26	69	100										
M4R	13	12	11	14	13	18	21	22	20	24	23	22	28	28	27	29	27	23	22	22	26	24	100											
S1P1	9	13	16	14	9	15	18	21	17	19	20	20	21	18	18	19	19	23	21	20	41	100												
LPA1	11	13	14	14	12	14	18	18	20	18	17	17	19	24	19	18	21	23	22	16	100													
NTSR1	11	14	15	15	15	17	18	24	20	22	23	21	25	26	22	27	23	22	24	100														
OX1R	10	17	14	16	18	19	18	24	23	19	23	19	22	30	28	29	27	80	100															
OX2R	8	16	15	15	18	19	20	27	22	24	26	21	26	31	31	30	29	100																
DOR	11	15	15	17	17	21	24	28	27	26	26	26	32	61	71	70	100																	
KOR	11	15	12	13	16	18	26	29	26	26	27	28	36	62	70	100																		
MOR	10	16	13	16	16	19	26	28	27	26	29	27	33	60	100																			
NOP	11	15	13	16	17	21	24	27	27	26	30	28	32	100																				
AT1R	13	18	17	15	15	22	29	27	27	29	37	33	100																					
CCR5	11	16	16	12	14	21	27	25	22	31	32	100																						
CXCR4	13	15	16	14	12	22	25	25	23	31	100																							
US28	13	15	14	18	17	21	23	26	25	100																								
PAR1	14	14	13	16	15	25	21	30	100																									
P2Y1	12	14	14	13	13	18	23	100																										
P2Y12	12	11	11	13	13	20	100																											
FFAR1	11	14	13	15	11	100																												
CRF1R	22	14	12	35	100																													
GCGR	19	9	8	100																														
GRM1	10	75	100																															
GRM5	9	100																																
SMO	100																																	

	D5R	D4R	D3R	D2R	D1R
D1R	58	33	35	37	100
D2R	37	39	57	100	
D3R	34	42	100		
D4R	35	100			
D5R	100				

	GRM8	GRM7	GRM6	GRM5	GRM4	GRM3	GRM2	GRM1
GRM1	41	40	40	64	41	43	44	100
GRM2	45	45	47	45	47	68	100	
GRM3	45	44	46	43	46	100		
GRM4	75	69	69	41	100			
GRM5	42	40	40	100				
GRM6	70	67	100					
GRM7	74	100						
GRM8	100							

Within subfamilies identities increase

List of Templates Ranked by Sequence Identity

Receptor	PDB ID	Identity
5HT-1B	4iar	39
B1AR	4bvn	38
B2AR	2rh1	36
M4R	5dsg	35
M1R	5cxv	34
5HT-2B	4ib4	34
H1R	3rze	34
M2R	3uon	33
M3R	4u15	31
A2AAR	4eiy	30
OX2R	4s0v	28
LPA1	4z35	28
MOR	4dkl	27
NOP	5dhg	26
S1P1R	3v2y	26
DOR	4n6h	26
OX1R	4zjc	26
NTSR1	4xes	25
Rho	1u19	25
CXCR4	3odu	25
KOR	4djh	25
AT1R	4zud	23
US28	4xt1	22
P2Y1	4xnw	22
CCR5	4mbs	22
P2Y12	4pxz	22
PAR1	3vw7	21
FFAR1	4phu	18
GCGR	5ee7	14
SMO	4jkv	13
GRM1	4or2	12
CRF1R	4k5y	12
GRM5	5cgd	11

A,α

⇧

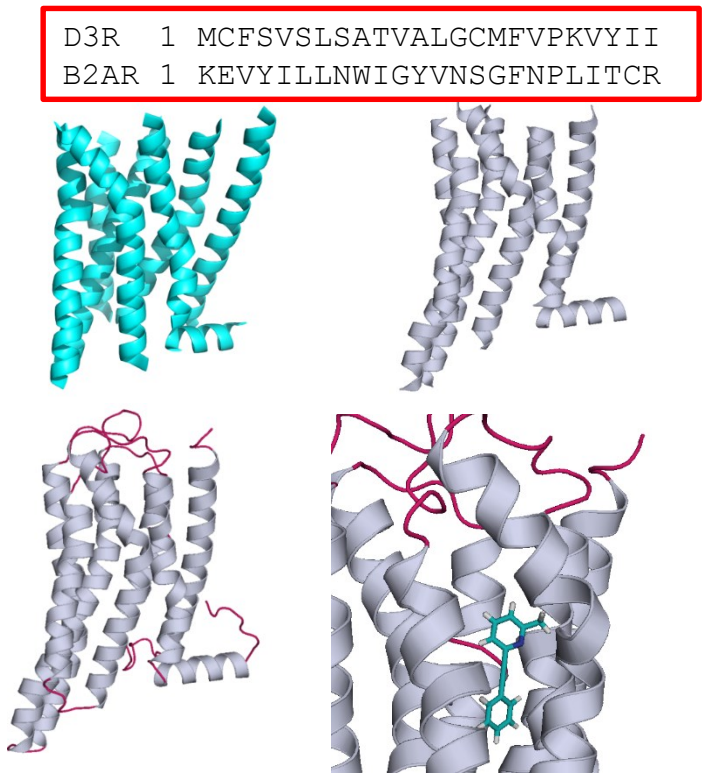
⇩

B/C/F

- Search query in blastp was D3R sequence
- D3R is a Class A Subclass α Receptor
- Top hits include other A α receptors
  - Adrenergic
  - Serotonin
  - Muscarinic
- Top hits are >30% identical
- Template PDBs are easily obtained from RSCB

Comparative Modeling of a GPCR and Ligand Docking using Rosetta

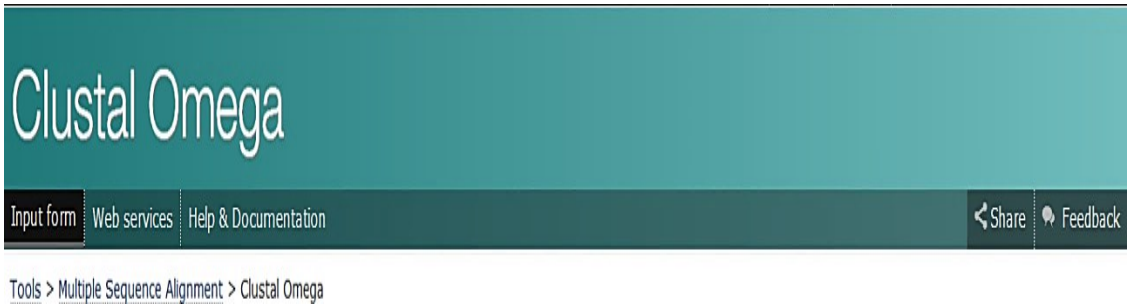
- **Step 1: ALIGN** target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model



Multiple Sequence Alignment of Target with Template Sequences

- Obtain template FASTAs from RSCB or pull from PDB file
- Remove fusion proteins
- Generate multiple sequence alignment (Clustal, MUSCLE) including target and templates

```
1u19        -----PWQFSM--LAAYMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFM
3ODU_A      ANFNKIFL------PTIYSIIFLTGIVGNGLVILVMGYQKKLRSMTDKYRLHLSVADLLF
2RH1_A      ---DEVWVVGMGIVMS---LIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACADLVM
3EML_A      -------IMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAV
                          :    : :  .  * *.  .    ..*:.  :     *: **: .
```



Adjusting Multiple Sequence Alignments to Account for Structural Features

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
```

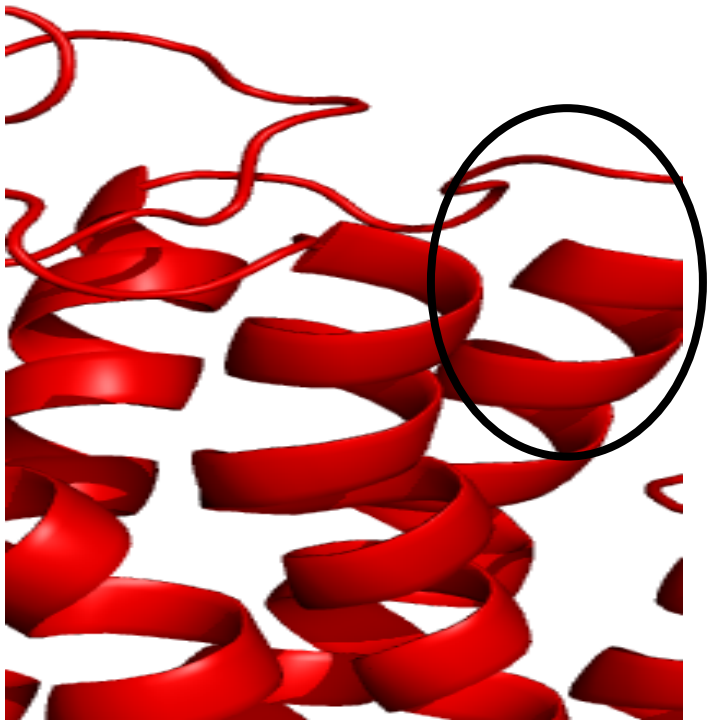
helix regions

highly conserved residues

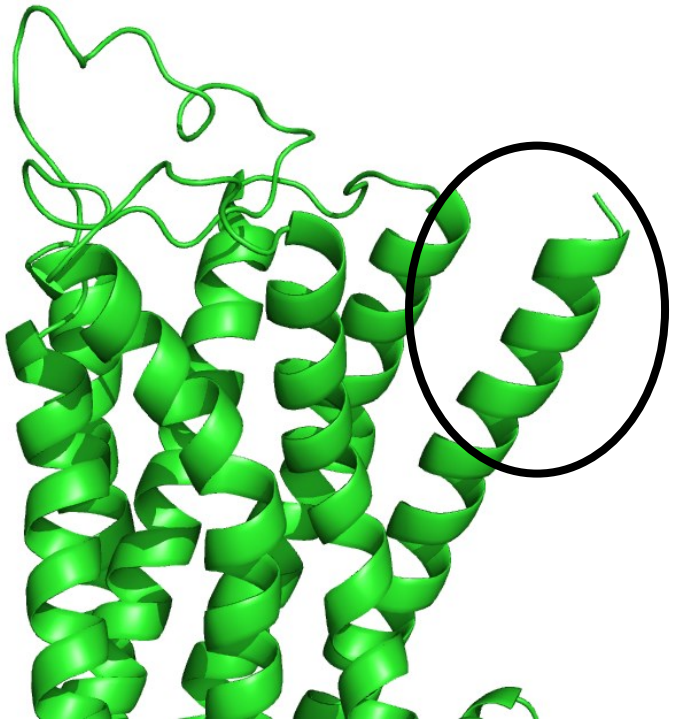
Alignment issues to be resolved

predicted membrane spanning region from OCTOPUS

Optimizing the Alignment is Important for Structural Integrity



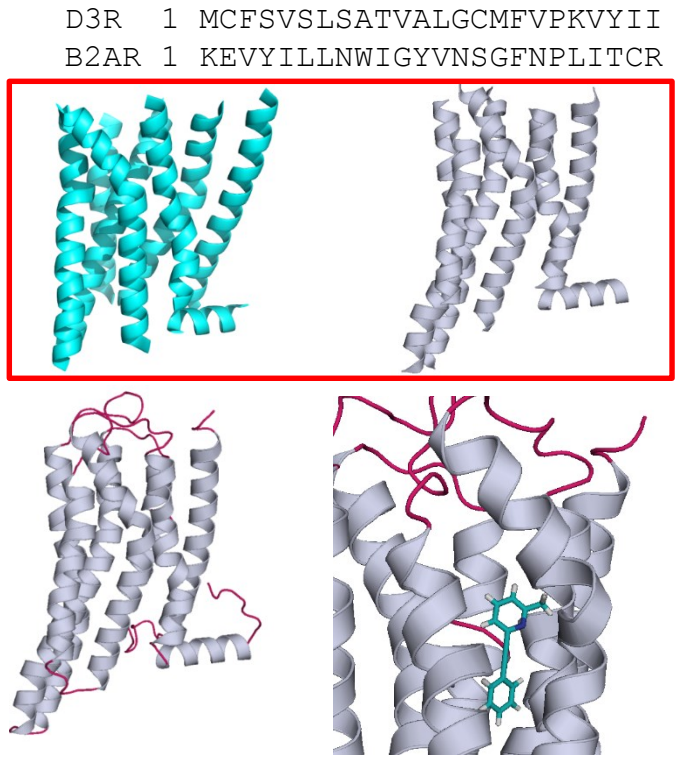
Raw alignment



Adjusted alignment

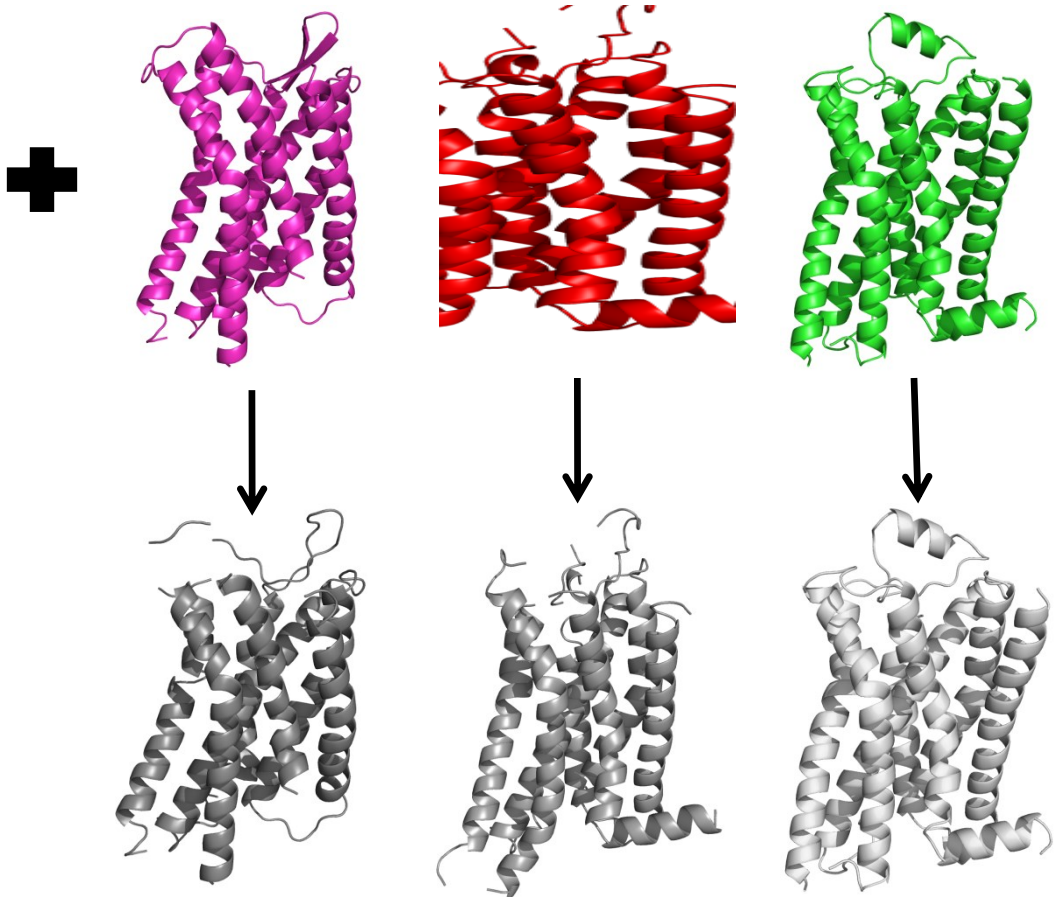
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** **THREAD** the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model



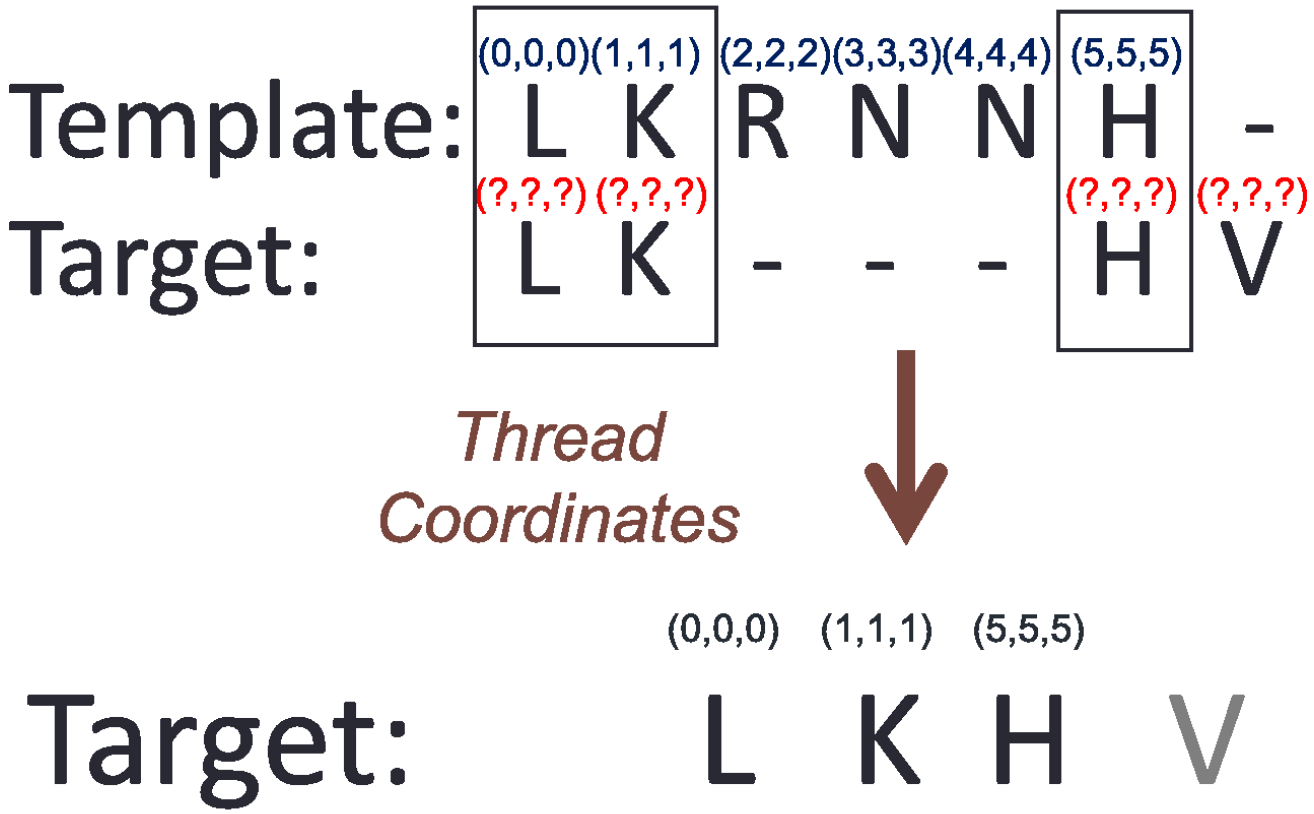
Threading Target Sequence onto Template Structures

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



Using the sequence alignment between target and template, thread the aligned target onto the 3D coordinates of the target structures

Threading



Rosetta Requires Grishin Format for Threading

ClustalO:
- All sequences in one file
- Sequences broken up over several lines

FASTA:
- All sequences in one file
- Each sequence on individual line

Grishin:
- **One file per alignment pair**
- Each sequence on individual line
- Specific header information

```
1u19        -----PWQFSM--LAAYMFL
3ODU_A      ANFNKIFL------PTIYSI
2RH1_A      ---DEVWVVGMGIVMS---L
3EML_A      -------IMGSSVYITVELA
```

```
>2rh1
--------DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKF-ERLQT-VTNYFITSLA
>4bvn
-------LSQQWEAGMSLLMALVVLLIVAGNVLVIAAIGST-QRLQT-LTNLFITSLA
>4iar
-YIYQDSISLPWKVLLVMLLALITLATTLSNAFVIATVYRT-RKLHT-PANYLIASLA
>5cxv
-------KGPWQVAFIGITTGLLSLATVTGNLLVLISFKVN-TELKT-VNNYFLLSLA
```

```
## 3pbl 2rh1_out
#
scores_from_program: 0
0 -------YALSYCALILAIVFGNGLVCMAVLKE-RALQT-TTNYLVVSLAVADLLVATLVMPWVVYLEV-TGGVWNFS-
0 DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKF-ERLQT-VTNYFITSLACADLVMGLAVVPFGAAHIL-MK-MWTFG-
```

Iterative Cycle of Alignment and Threading

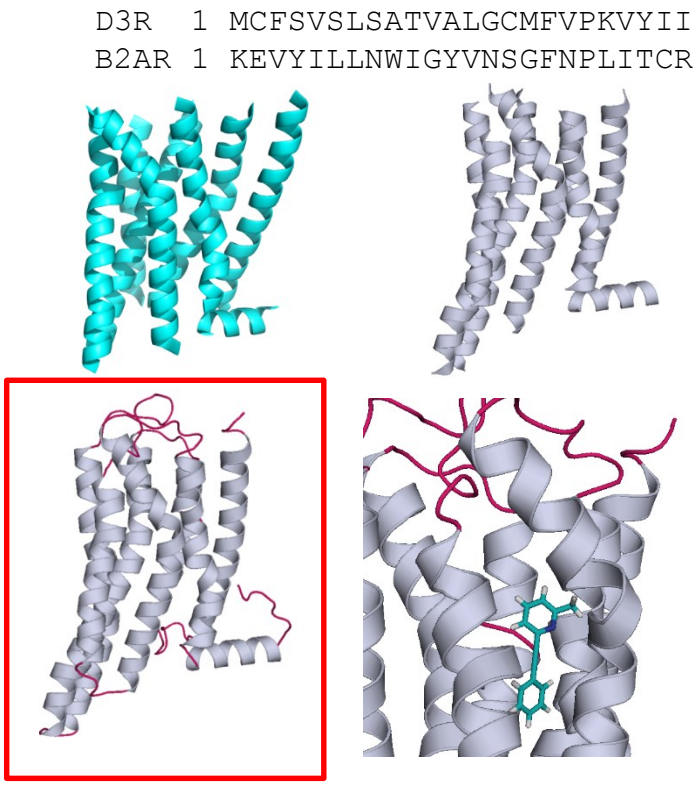
Examining results after threading can give clues about validity of alignment

Sequence Alignment is first step of the “Structural Alignment”

- No major breaks in Secondary Structure Elements
- Proper amino acids are on the right “face” of a helix
- Make sure amino acid placement along the template backbone “makes sense”

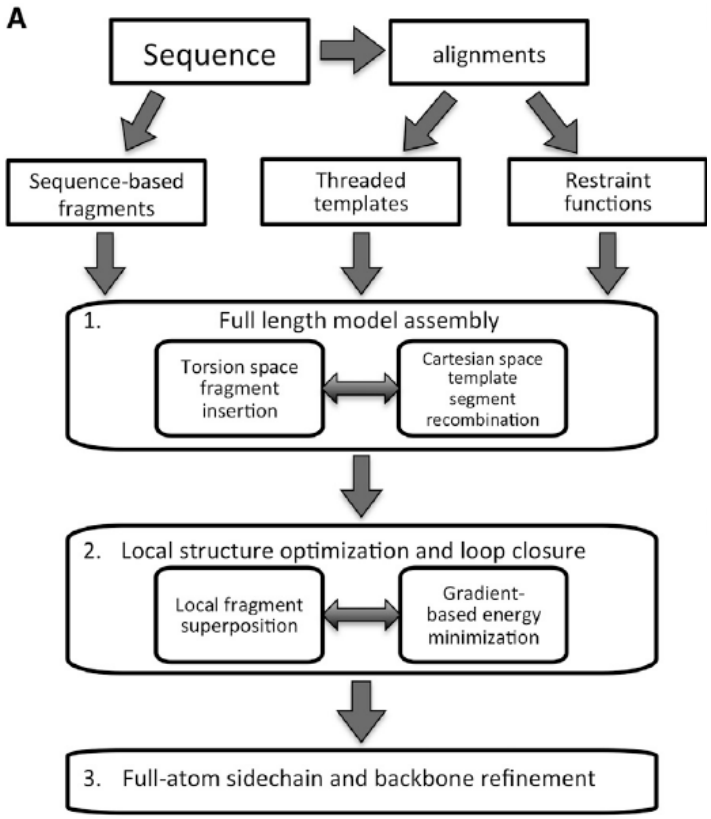
Comparative Modeling of a GPCR and Ligand Docking using Rosetta

- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** **HYBRIDIZE** template segments and fragments from PDB to generate full length models
- **Step 4:** Dock ligand into comparative model

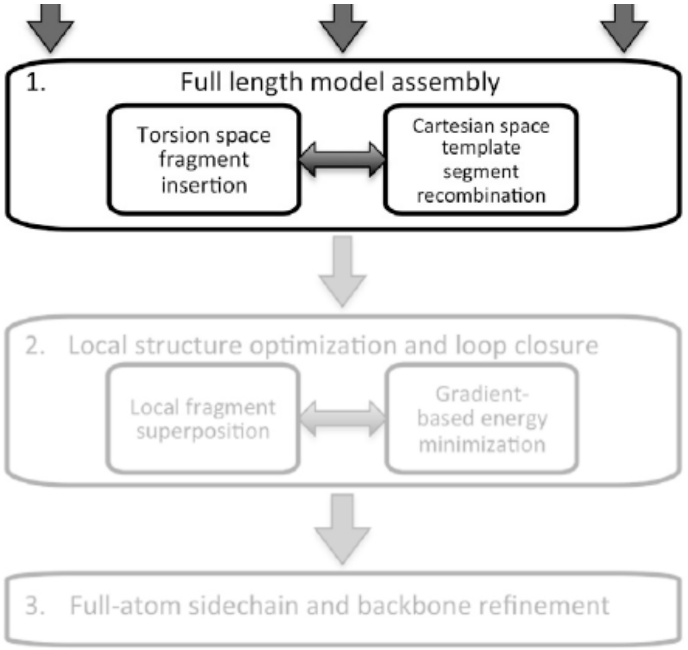


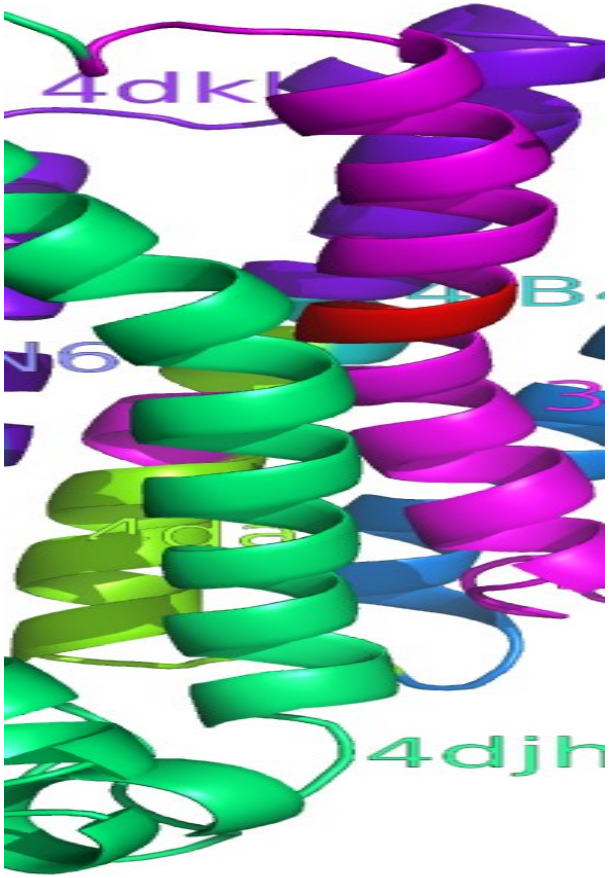
Hybridization

- Combination of template fragments and PDB-derived fragments to generate a full length model
- Broken into a series of smaller steps from coarse sampling to fine sampling to prevent chain breaks
- Resulting model has various combinations of starting template segments

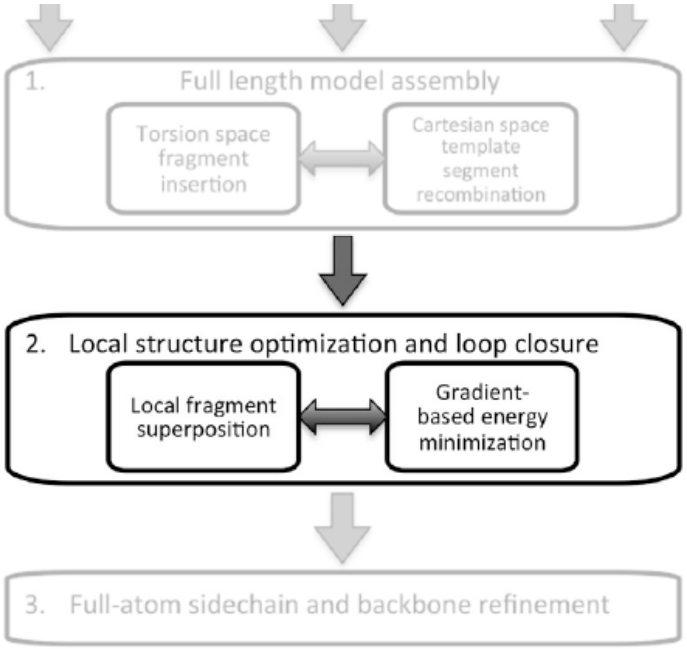


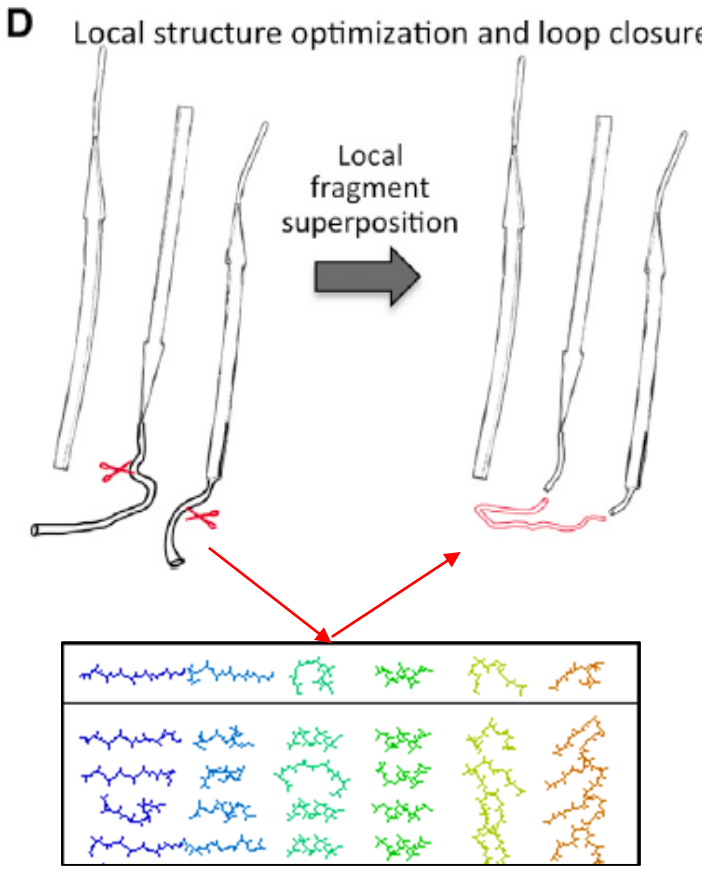
Hybridize: Full Model Assembly



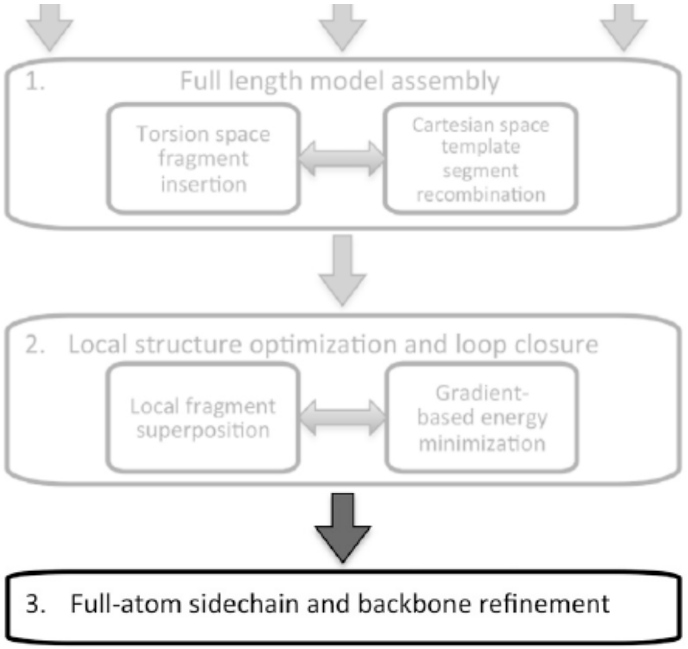


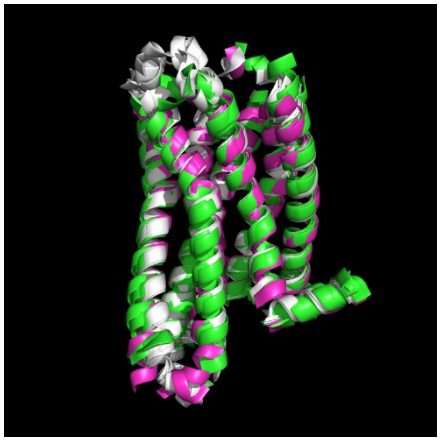
Hybridize: Loop Closure



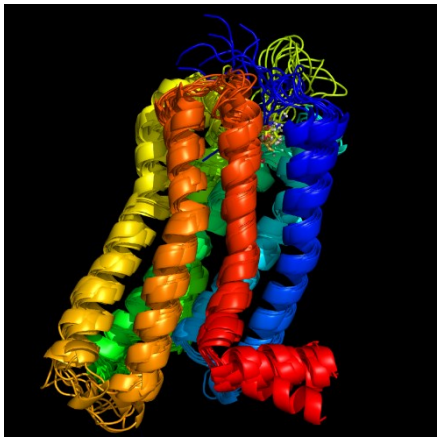


Hybridize: Energy Minimization



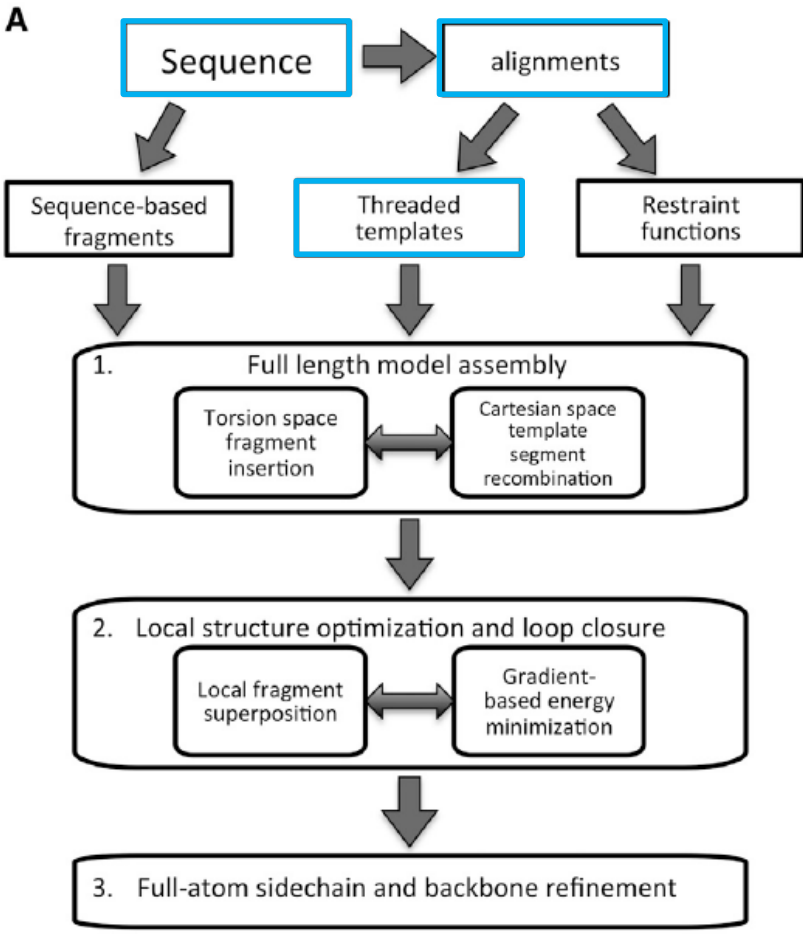


Pre-Relax



Post-Relax

Last Steps Before Hybridization



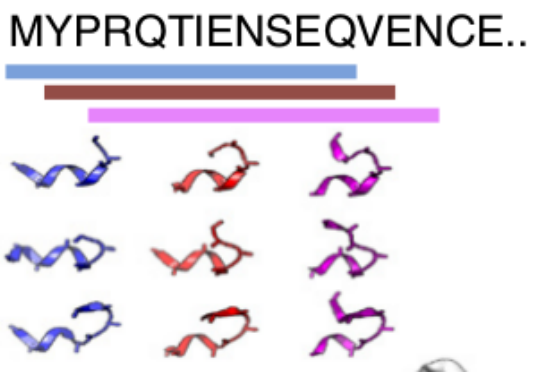
Sequence, alignment, and threading account for initial manual stages of RosettaCM

Fragment files are required

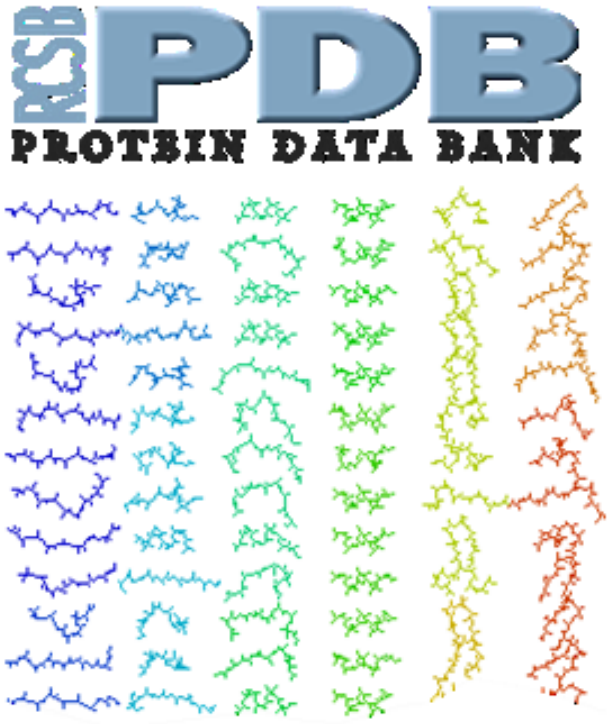
Restraint functions are helpful but not necessary

- Can assist in narrowing the conformational search space
- I.e. disulfide files, experimentally-derived constraints

Fragment Files

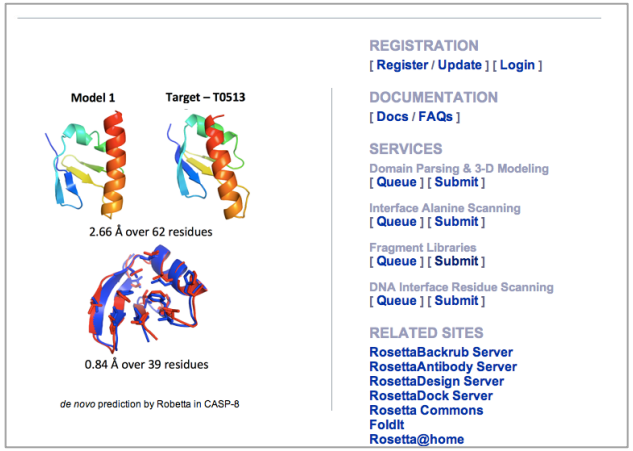


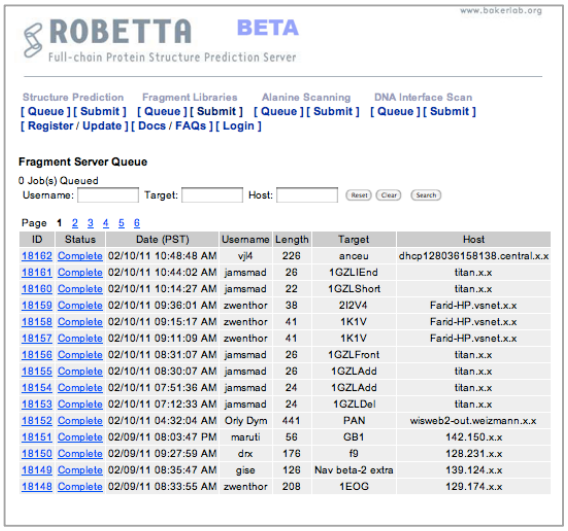
- 3- and 9-mer fragments derived from the PDB
- The target sequence is broken into these fragments and a fragment library is generated based on these segments



Making fragments with Robetta

http://robetta.bakerlab.org/





Restraint Files: Disulfide File

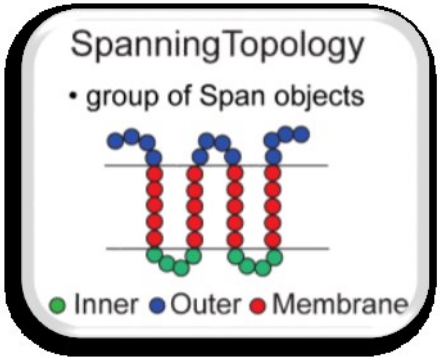
- An important restraint file to use in modeling of GPCRs is a disulfide file
  - Promotes the accurate formation of the conserved disulfide between TM3 and ECL2

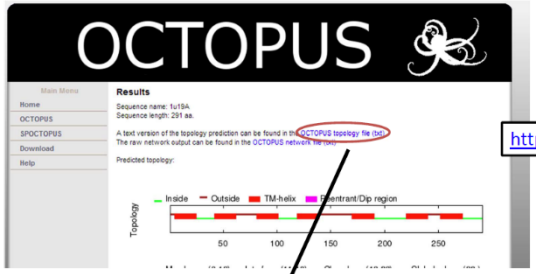
3pbl.disulfide

```
72 158
```

Restraint Files: Membrane Span Files

- Predict membrane spanning regions using Octopus server
  - http://octopus.cbr.su.se/
- Octopus predictions are converted to a format compatible with Rosetta with the octopus2span.pl script
- Span file list start and stop residues of each TM region





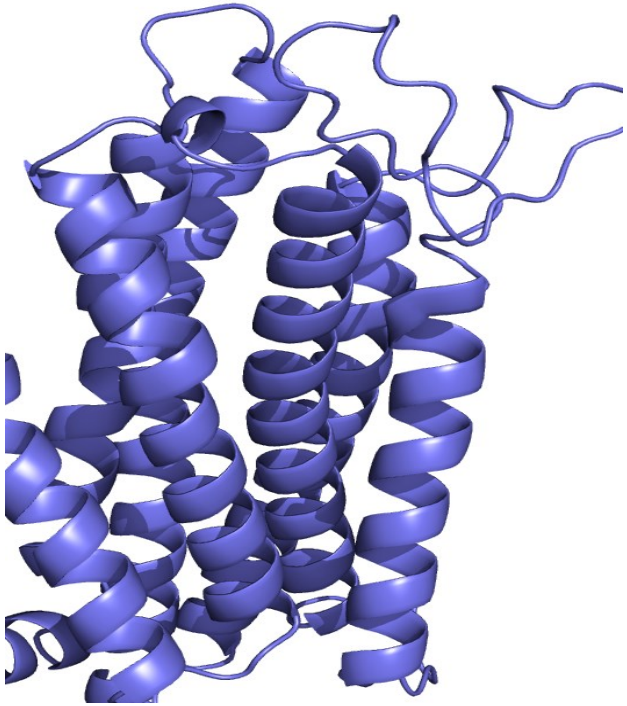
http://octopus.cbr.su.se/

octopus2span.pl 3pbl.octopus

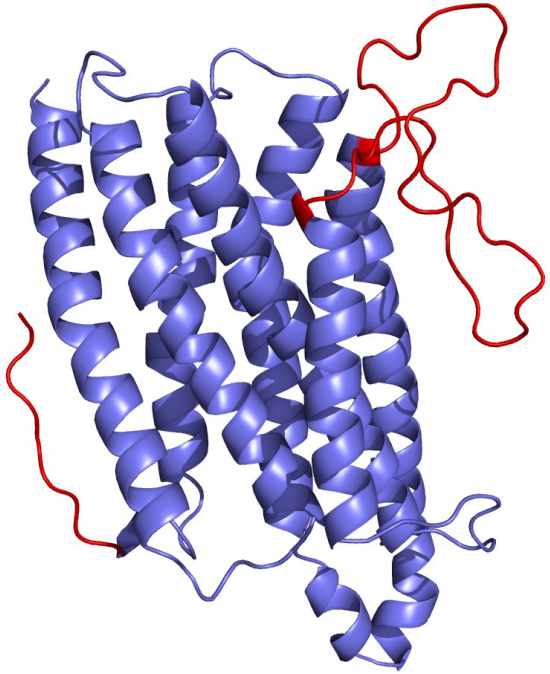
```
TM region prediction for 3pbl.octopus predicted using OCTOPUS
7 280
antiparallel
n2c
   5    22     5    22
  36    56    36    56
  74    96    74    96
 120   134   120   134
 160   177   160   177
 207   227   207   227
 248   264   248   264
```

Restraint Files: Membrane Span Files

With membrane penalties/weights



Without membrane penalties/weights



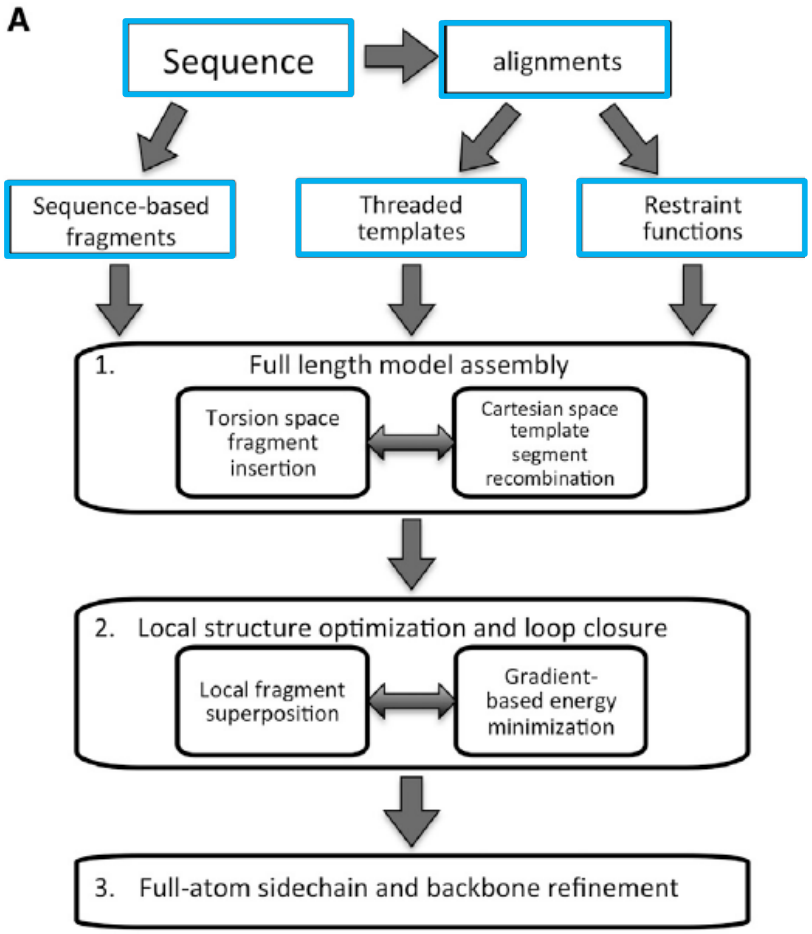
Additional Restraint Files

- Other data that can be helpful in accurately modeling your receptor include
  - NMR
  - EPR
  - EM
  - Mutational analysis
  - H/D Exchange
  - Fluorescence
  - Mass Spec
- All can be incorporated in Rosetta but require specific handling of the constraint type

Example restraint file

```
AtomPair N  256  CA  76  SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AmbiguousConstraint
AtomPair CA 258  CA 207  SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AtomPair CA 258  CA 211  SCALARWEIGHTEDFUNC 10 FLAT_HARMONIC 0 2 5
AtomPair CA 258  CA 227  FLAT_HARMONIC 0 2 5
AtomPair CA 258  CA 230  FLAT_HARMONIC 0 2 5
END
```

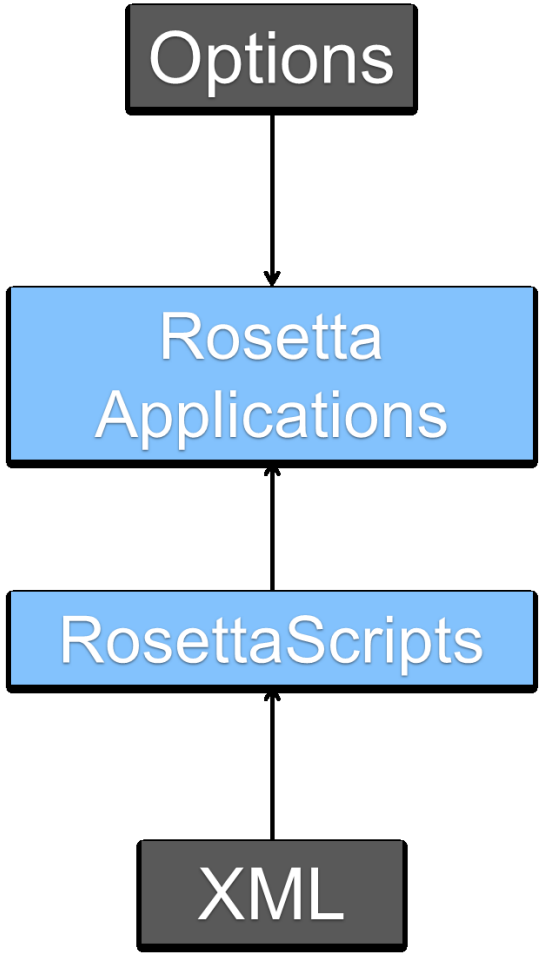
Hybridize: Full Model Assembly



Files needed for RosettaCM

- Partial-threaded structures
- Fragment files (3mer and 9mer lengths)
- Membrane spanning regions (span file)
- Mover definition and options
- Weight patches (pre-generated)

Submitting All Files to RosettaCM Hybridization



- RosettaCM is accessed through the RosettaScripts protocol
  - Relies on input being in XML format
- Additionally controls can be passed in through the command line
  - In general, it is preferred to move all command line options into an option file

RosettaScripts XML

Specifies:

- Score functions to be used
- Methods
- Order of execution

```
<dock_design>
        <TASKOPERATIONS>
        </TASKOPERATIONS>
        <SCOREFXNS>
                <stage1 weights="input_files/stage1_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=1/>
                </stage1>
                <stage2 weights="input_files/stage2_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=0.5/>
                </stage2>
                <fullatom weights="input_files/stage3_rlx_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=0.5/>
                </fullatom>
                <membrane weights="membrane_highres_Menv_smooth" symmetric=0>
                        <Reweight scoretype=cart_bonded weight=0.5/>
                        <Reweight scoretype=pro_close weight=0/>
                </membrane>
        </SCOREFXNS>
        <FILTERS>
        </FILTERS>
        <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">
                        <Fragments 3mers="input_files/3pbl-frags.200.3mers" 9mers="input_files/3pbl-frags.200.9mers"/>
                        <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>
        <APPLY_TO_POSE>
        </APPLY_TO_POSE>
        <PROTOCOLS>
                <Add mover=hybridize/>
                <Add mover=clearconstraints/>
                <Add mover=relax/>
        </PROTOCOLS>
        <OUTPUT scorefxn=membrane/>
</dock_design>
```

Score Function

```
<dock_design>
        <TASKOPERATIONS>
        </TASKOPERATIONS>
        <SCOREFXNS>
                <stage1 weights="input_files/stage1_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=1/>
                </stage1>
                <stage2 weights="input_files/stage2_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=0.5/>
                </stage2>
                <fullatom weights="input_files/stage3_rlx_membrane.wts" symmetric=0>
                        <Reweight scoretype=atom_pair_constraint weight=0.5/>
                </fullatom>
                <membrane weights="membrane_highres_Menv_smooth" symmetric=0>
                        <Reweight scoretype=cart_bonded weight=0.5/>
                        <Reweight scoretype=pro_close weight=0/>
                </membrane>
        </SCOREFXNS>
```

- Score functions are set
- The first three weights are files that need to be copied into a directory for running
- The last score function “membrane_highres_Menv_smooth” is a default score function in Rosetta and does not to be moved

Hybridize Mover

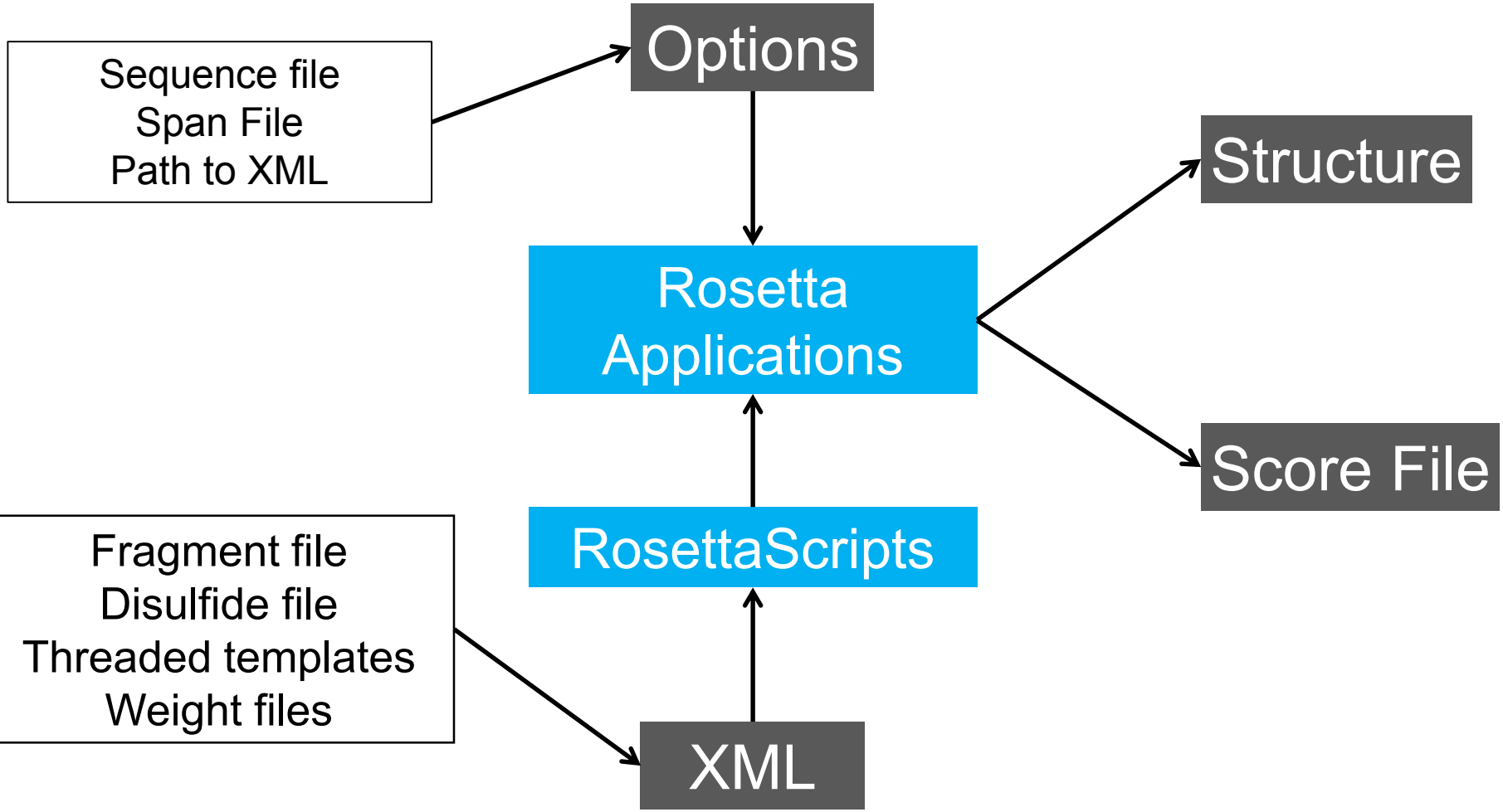
```
<MOVERS>
        <Hybridize name=hybridize stage1_scorefxn=stage1 stage2_scorefxn=stage2 fa_scorefxn=fullatom
        batch=1 stage1_increase_cycles=1.0 stage2_increase_cycles=1.0 linmin_only=1 realign_domains=0
        disulf_file="input_files/3pbl.disulfide">
                <Fragments 3mers="input_files/3pbl-frags.200.3mers" 9mers="input_files/3pbl-frags.200.9mers"/>
                <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>
```

- Hybridize mover specifies which score function to use at different stages of modeling
- The disulfide file is pointed to in the main mover
- Within the mover we point to the fragment files and threaded template files
- The ClearConstraintsMover and FastRelax movers are additional step to assist in energy minimization

Options File Directs Rosetta to XML and Additional Inputs

```
-database /home/benderb/Rosetta/main/database/          #### path to Rosetta database

# i/o
-in:file:fasta input_files/3pbl.fasta                   #### fasta of final sequence to be modeled
-parser:protocol input_files/rosetta_cm.xml             #### path to XML script
-out:path:all output_files/                             #### designates where to put pdbs/silent files/scorefiles/etc
# output styles
-out:pdb                                                #### specifiies output format as pdbs
-out:file:scorefile 3pbl_scores.out                     #### gives specific name for scorefile (default is scores.sc)
-nstruct 1                                              #### specifies number of models to be created


# membrane options
-in:file:spanfile input_files/3pbl.span                 #### lists transmembrane spanning regions for membrane scoring
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-rg_reweight .1

# 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      #### path to membrane weights file
-use_bicubic_interpolation
-hybridize:stage1_probability 1.0
-sog_upper_bound 15

# reduce memory footprint
-chemical:exclude_patches LowerDNA  UpperDNA Cterm_amidation SpecialRotamer VirtualBB ShoveBB VirtualDNAPhosphate
VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated
thr_phosphorylated  tyr_phosphorylated tyr_sulfated lys_dimethylated lys_monomethylated  lys_trimethylated
lys_acetylated glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm
```

What goes in, What comes out



Running RosettaCM

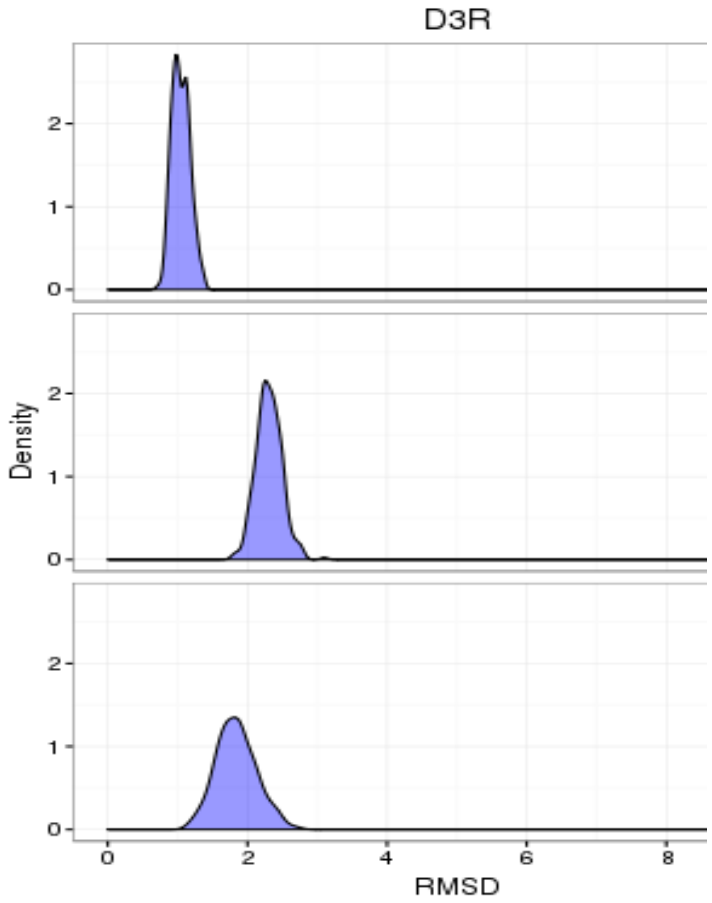
- If all files are set correctly the command line for running RosettaCM is:

```
/path/to/Rosetta/main/source/bin/rosetta_scripts.linuxgccrelease \
@rosetta_cm.options
```

- By using the options file all command line options are contained in one, manageable file
- The options file points to the XML file with directs RosettaCM
- This protocol directly outputs pdbs which are ready for analysis/docking studies

- A single GPCR model typically takes 20-40 minutes depending on length
- It is advised to run 1000-5000 models
- Model selection is specific to goals of modeling but is oten based on energy and structure-based clustering

Modeling Results of D3R



- The sampling of RosettaCM places important regions within 1-2 Å RMSD
- The overall RMSD including all loop regions averages just over 2 Å

- This level of accuracy is critical for having reliable models for ligand docking

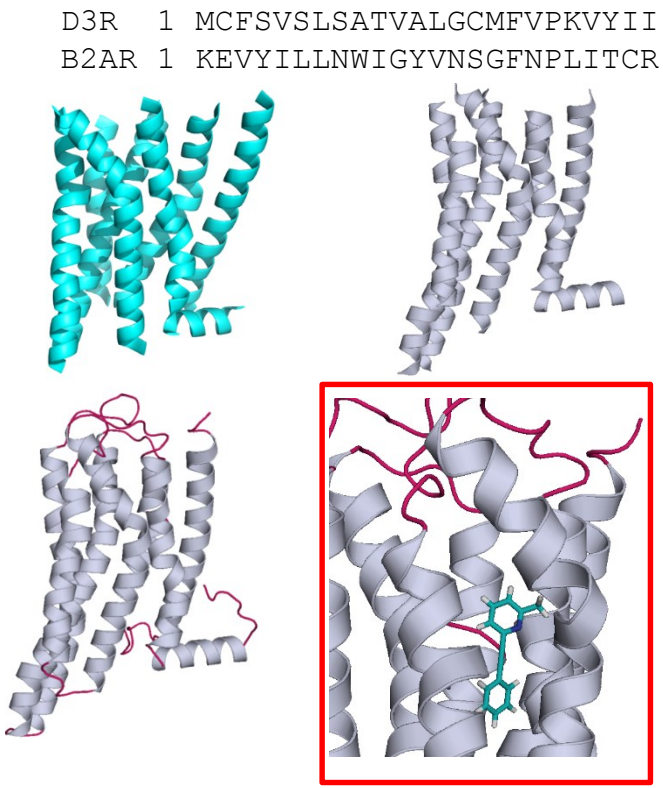
Using RosettaCM online

- ROBETTA web-based application uses RosettaCM to generate 3D models from sequence.
- Completely automated.
- Generates alignments with Hhsearch, SPARKS-X, and RaptorX
- Does not include final loop building and membrane relax

http://robetta.bakerlab.org/submit.jsp

Comparative Modeling of a GPCR and Ligand Docking using Rosetta

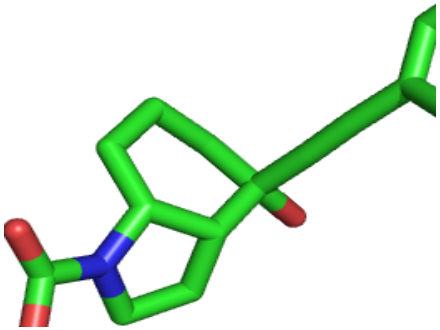
- **Step 1:** Align target sequence with sequence of template structure
- **Step 2:** Thread the target sequence onto the backbone of the template structure
- **Step 3:** Hybridize template segments and fragments from PDB to generate full length models
- **Step 4:** **DOCK** ligand into comparative model



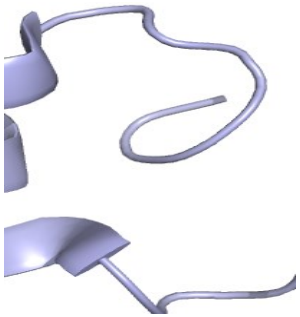
Ligand Docking

Center for Structural Biology and Institute of Chemical Biology
Departments of Chemistry, Pharmacology, and Biomedical Informatics

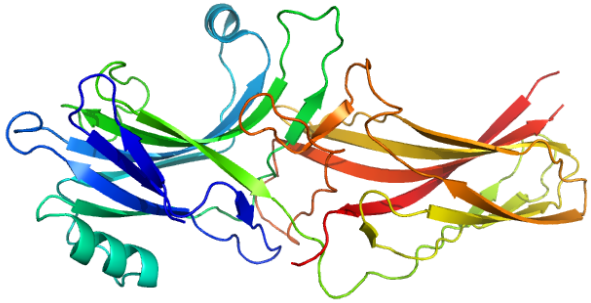
Ligands = Small molecules



Ligand Docking
(RosettaLigand)



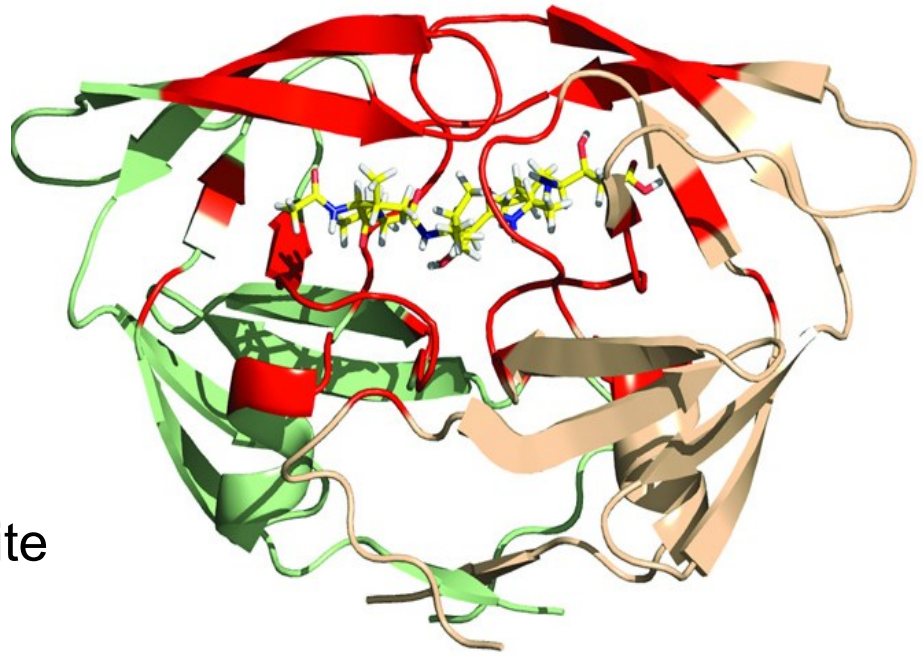
Peptide Docking
(FlexPepDock)



Protein-Protein
Docking
(RosettaDock)

RosettaLigand Past

- Meiler and Baker 2006
  - Protein and ligand ensembles
  - Side chain flexibility
- Kaufmann et. al. 2008
  - Rotatable bond angle sampling
- Davis and Baker 2009
  - Backbone flexibility near binding site
- Lemmon and Meiler 2012
  - XML Scripts

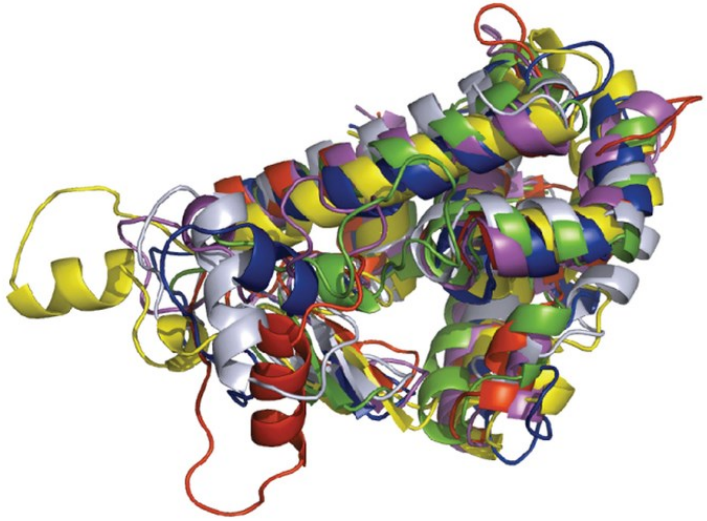


HIV-1 PR homodimer (green/wheat) with acetylpepstatin (yellow) in binding site (red)

G. Lemmon et. al. *Chemical biology & drug design* (2012).

RosettaLigand Now

- **Combs et. al. 2013**
  - **Docking into comparative models**
- Allison et. al. 2013
  - Dock/design of interface
- Lemmon and Meiler 2013
  - Docking with interface waters
- **Deluca and Meiler 2015**
  - **High throughput screening and improved low resolution sampling**

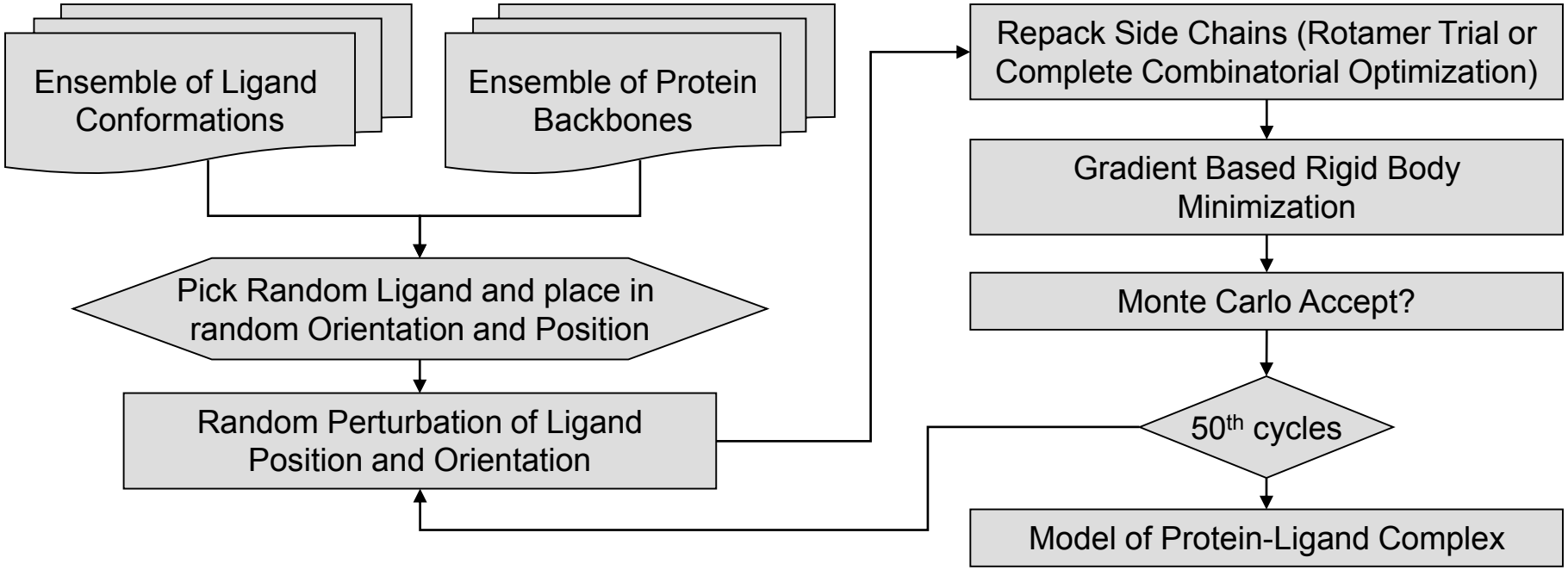


T4-Lysozyme Comparative Models
Combs et. al. (2013)

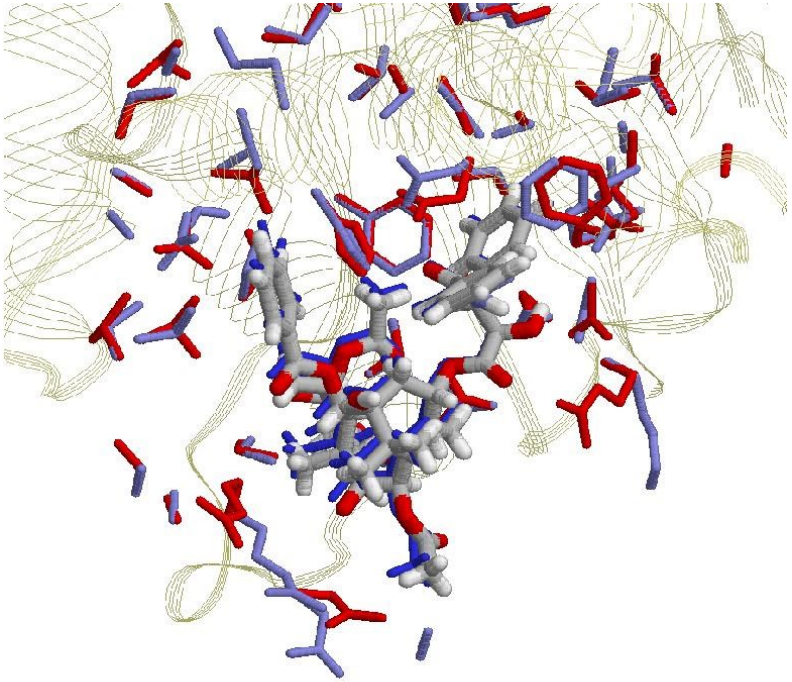
ROSIE Ligand Docking Server: rosie.rosettacommons.org

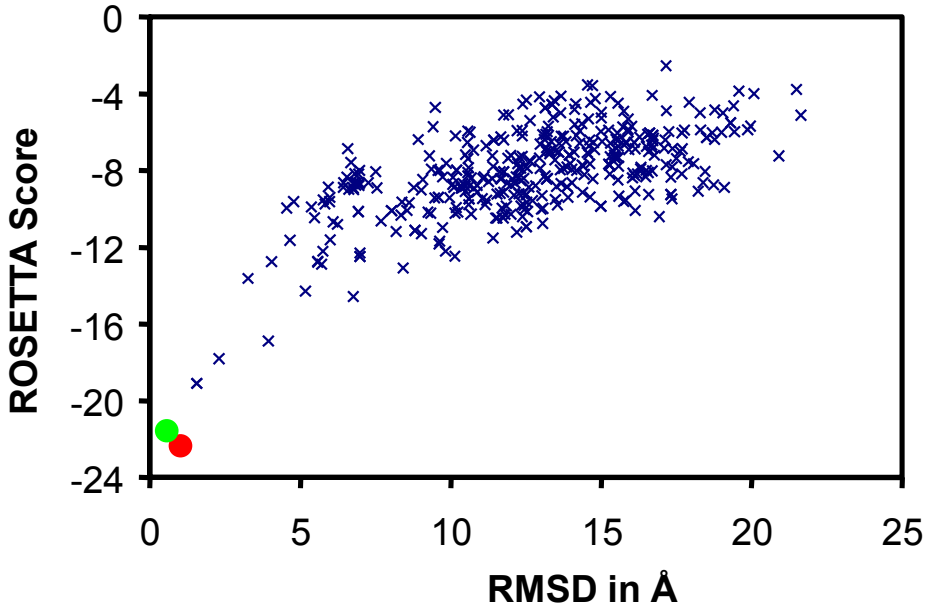
Rosetta Protein – Ligand – Docking

- Protocol



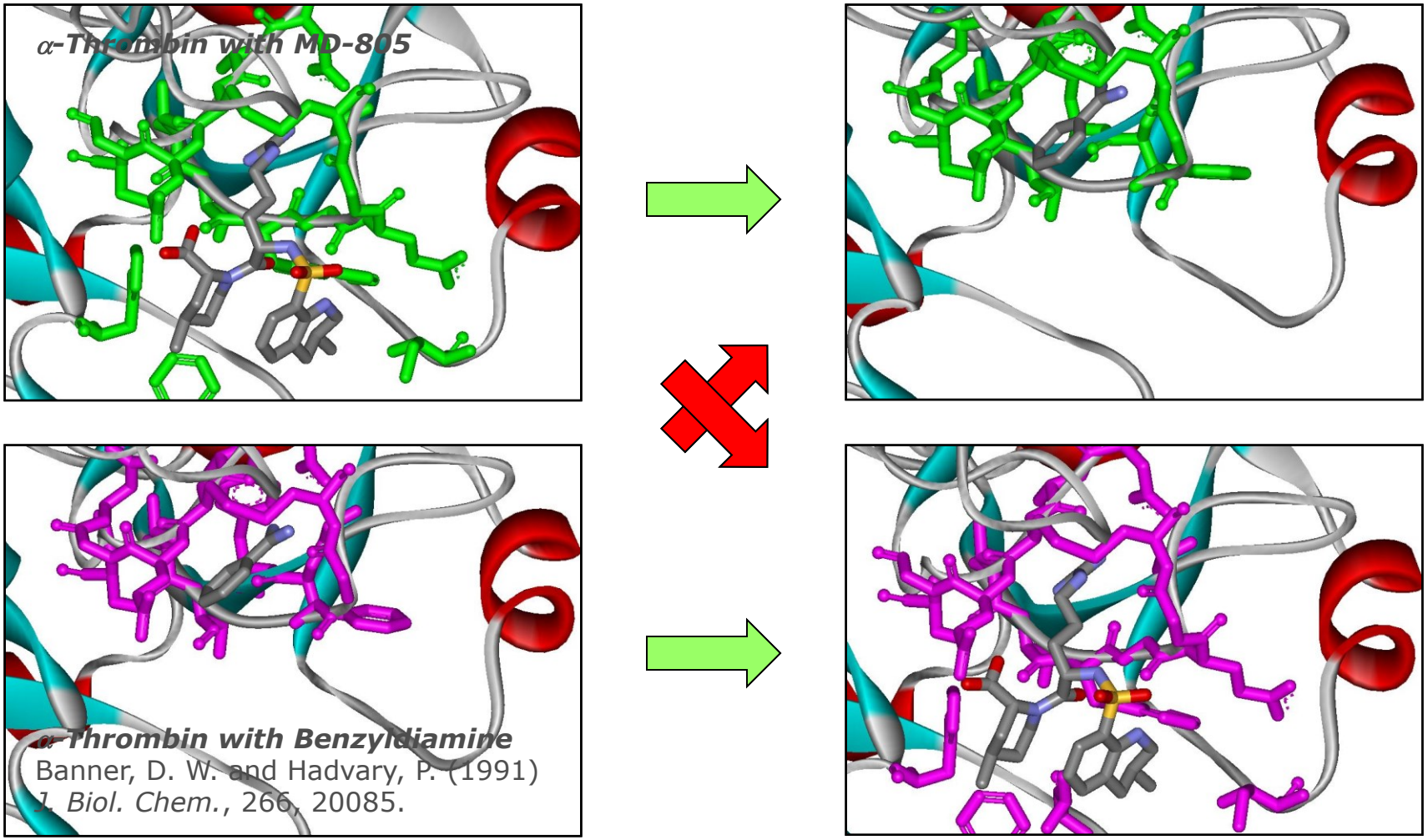
Taxol-Tubulin Complex





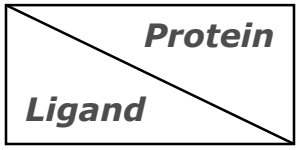
Lowe, J., Li, H., Downing, K.H., Nogales, E. (2001) *J. Mol. Biol.* 313, 1045.

“Cross-Docking” Experiment



“Cross-Docking” Benchmark of 20 Protein-Ligand Complexes

- 10 proteins in complex with 2 ligands each
- 70 (80)% best RMSD below 2.0Å
- 75 (90)% below 2.0Å in first percentile



2

	1aq1	1dm2
1aq1	0.42	0.57
1dm2	0.44	0.35

	1dbj	2dbl
1dbj	1.12	0.94
2dbl	0.72	0.47

	1dwc	1dwd
1dwc	0.55	0.56
1dwd	0.84	0.36

	1fm9	2prg
1fm9	0.30	1.52
2prg	1.03	0.38

	1p8d	1pq6
1p8d	0.54	1.00
1pq6	0.44	0.18

	1p8d	1pqc
1p8d	0.54	0.70
1pqc	1.82	0.44

	1ppc	1pph
1ppc	0.26	0.61
1pph	0.55	0.62

	1pq6	1pqc
1pq6	0.18	0.65
1pqc	1.72	0.44

	2ctc	7cpa
2ctc	0.84	0.37
7cpa	1.00	0.76

	4tim	6tim
4tim	0.61	0.32
6tim	0.42	0.20

3

	1aq1	1dm2
1aq1	0.28	6.61
1dm2	0.48	0.34

	1dbj	2dbl
1dbj	0.99	0.78
2dbl	0.72	0.83

	1dwc	1dwd
1dwc	7.57	0.59
1dwd	0.88	0.30

	1fm9	2prg
1fm9	0.27	1.56
2prg	1.10	0.43

	1p8d	1pq6
1p8d	0.45	1.64
1pq6	0.67	0.39

	1p8d	1pqc
1p8d	0.45	1.39
1pqc	2.11	0.66

	1ppc	1pph
1ppc	0.25	0.54
1pph	0.61	0.60

	1pq6	1pqc
1pq6	0.39	3.43
1pqc	2.44	0.66

	2ctc	7cpa
2ctc	0.68	0.69
7cpa	0.86	0.76

	4tim	6tim
4tim	0.57	0.26
6tim	2.67	0.39

4

	1aq1	1dm2
1aq1	1	3
1dm2	1	1

	1dbj	2dbl
1dbj	1	1
2dbl	1	1

	1dwc	1dwd
1dwc	2	1
1dwd	1	1

	1fm9	2prg
1fm9	1	1
2prg	1	1

	1p8d	1pq6
1p8d	1	1
1pq6	1	1

	1p8d	1pqc
1p8d	1	1
1pqc	7	1

	1ppc	1pph
1ppc	1	1
1pph	1	1

	1pq6	1pqc
1pq6	1	7
1pqc	12	1

	2ctc	7cpa
2ctc	1	1
7cpa	1	1

	4tim	6tim
4tim	1	1
6tim	6	1

5

	1aq1	1dm2
1aq1	0.49	4.59
1dm2	0.49	0.57

	1dbj	2dbl
1dbj	0.00	0.00
2dbl	1.14	1.37

	1dwc	1dwd
1dwc	6.28	6.67
1dwd	0.79	6.66

	1fm9	2prg
1fm9	0.55	1.62
2prg	1.60	1.71

	1p8d	1pq6
1p8d	0.64	1.75
1pq6	1.11	0.87

	1p8d	1pqc
1p8d	0.64	1.98
1pqc	3.19	0.56

	1ppc	1pph
1ppc	2.44	2.45
1pph	2.33	0.97

	1pq6	1pqc
1pq6	0.87	1.66
1pqc	1.51	0.56

	2ctc	7cpa
2ctc	2.80	2.34
7cpa	1.02	0.48

	4tim	6tim
4tim	0.81	1.29
6tim	1.54	1.48

6

	1aq1	1dm2
1aq1	1	27
1dm2	1	1

	1dbj	2dbl
1dbj	0	0
2dbl	1	1

	1dwc	1dwd
1dwc	58	244
1dwd	1	2

	1fm9	2prg
1fm9	1	1
2prg	1	1

	1p8d	1pq6
1p8d	1	1
1pq6	1	1

	1p8d	1pqc
1p8d	1	1
1pqc	162	1

	1ppc	1pph
1ppc	15	27
1pph	2	1

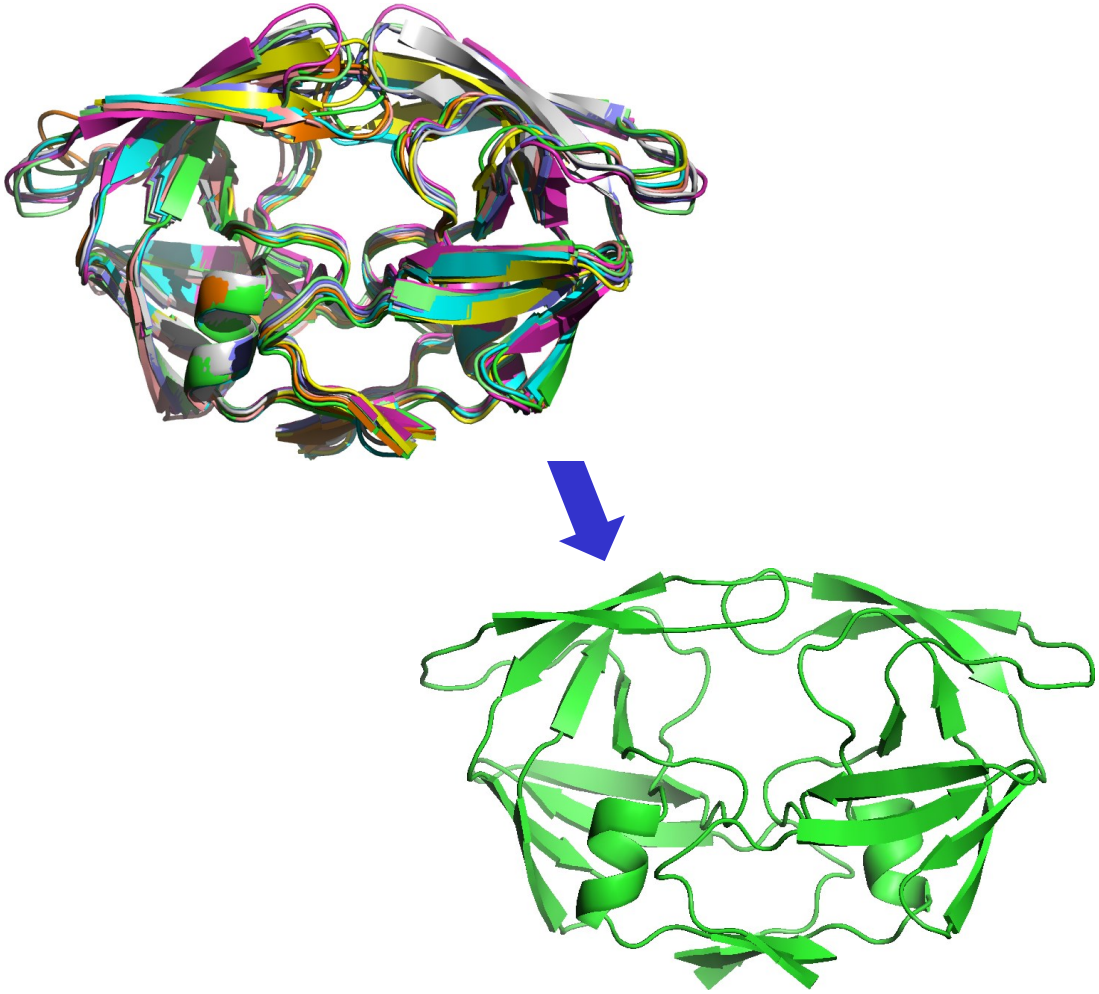
	1pq6	1pqc
1pq6	1	1
1pqc	1	1

	2ctc	7cpa
2ctc	2	16
7cpa	1	1

	4tim	6tim
4tim	1	1
6tim	1	1

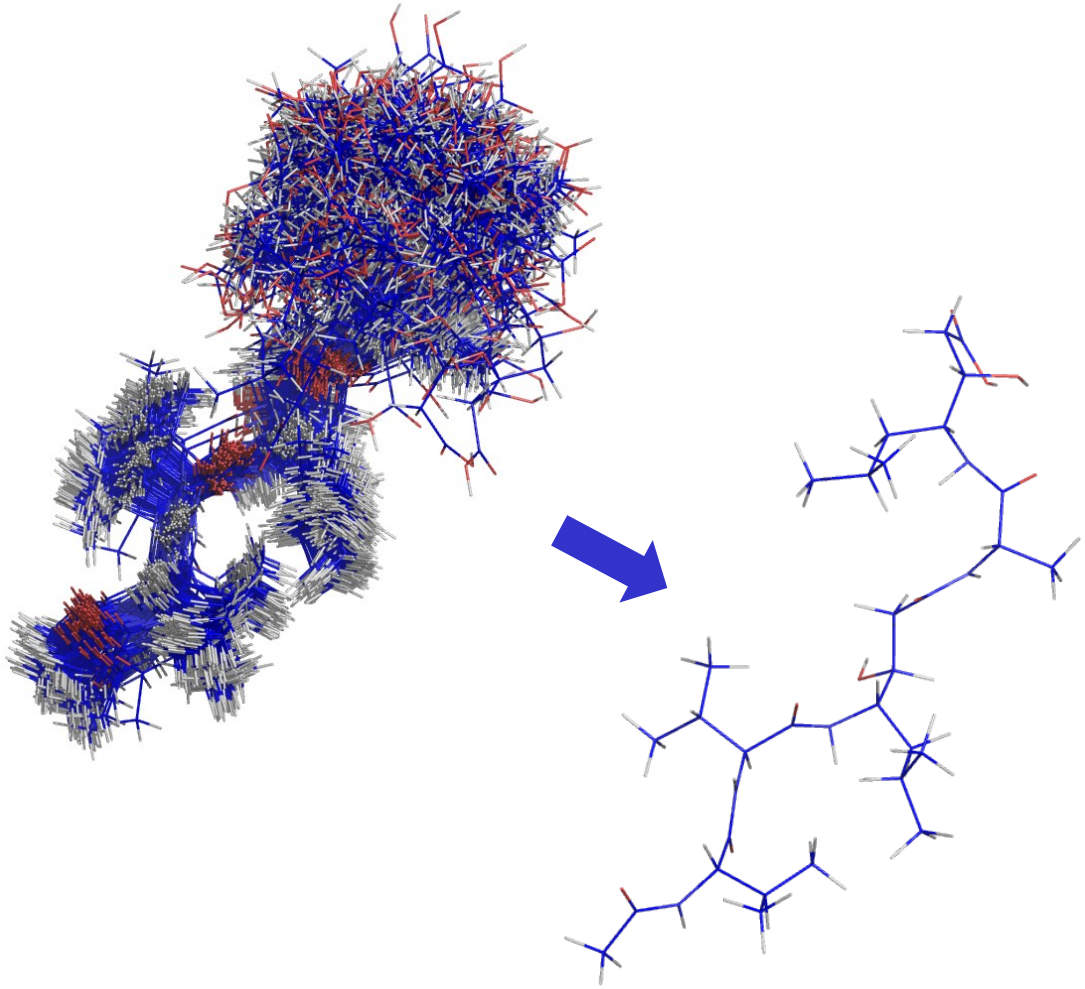
RosettaLigand Algorithm

- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement



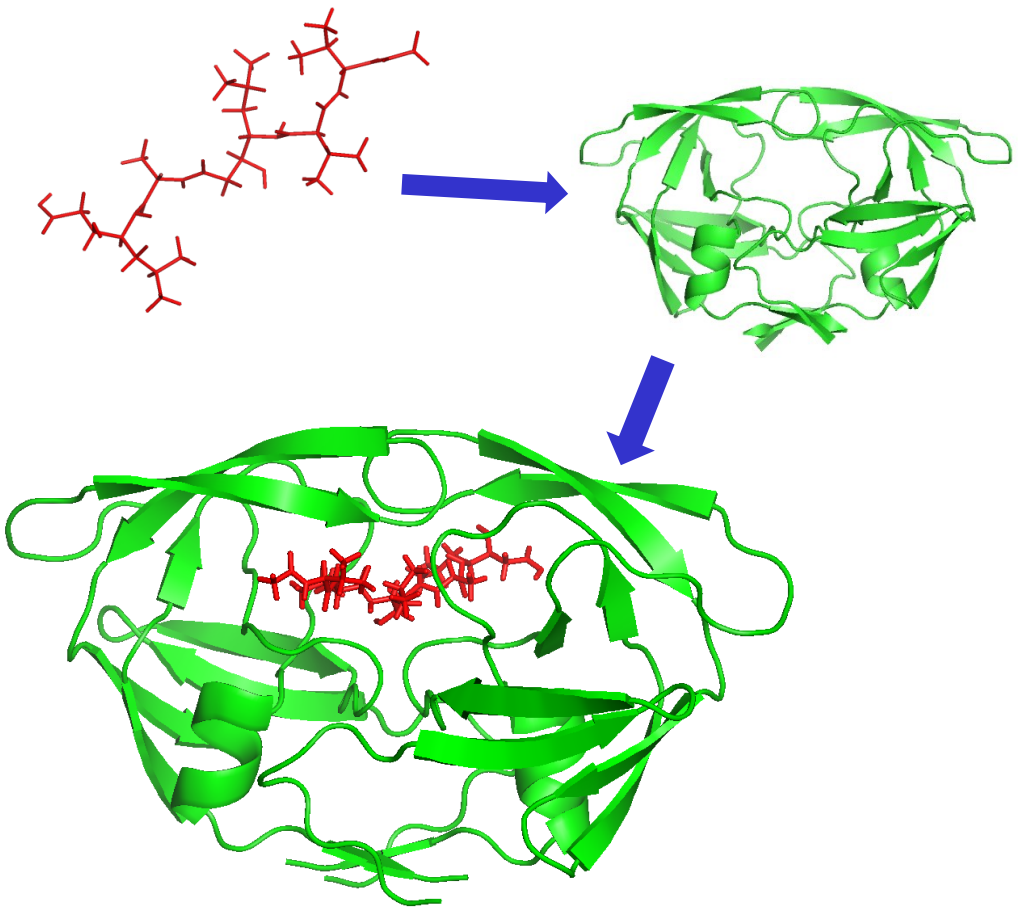
RosettaLigand Algorithm

- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement



RosettaLigand Algorithm

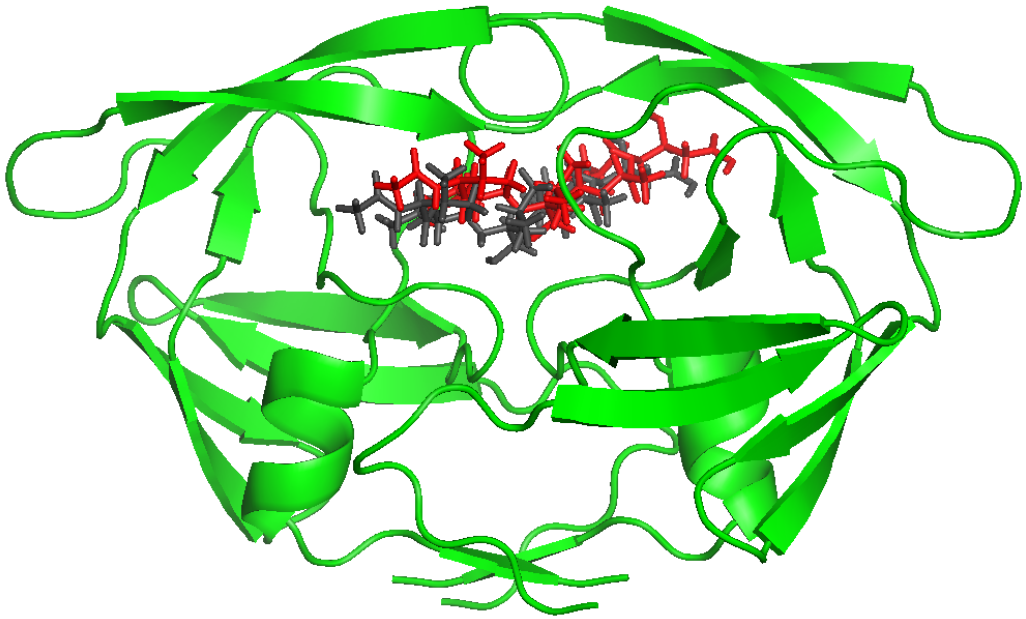
- Ensemble selection
- Initial placement
- Transform
- Docking Cycles
- High res refinement



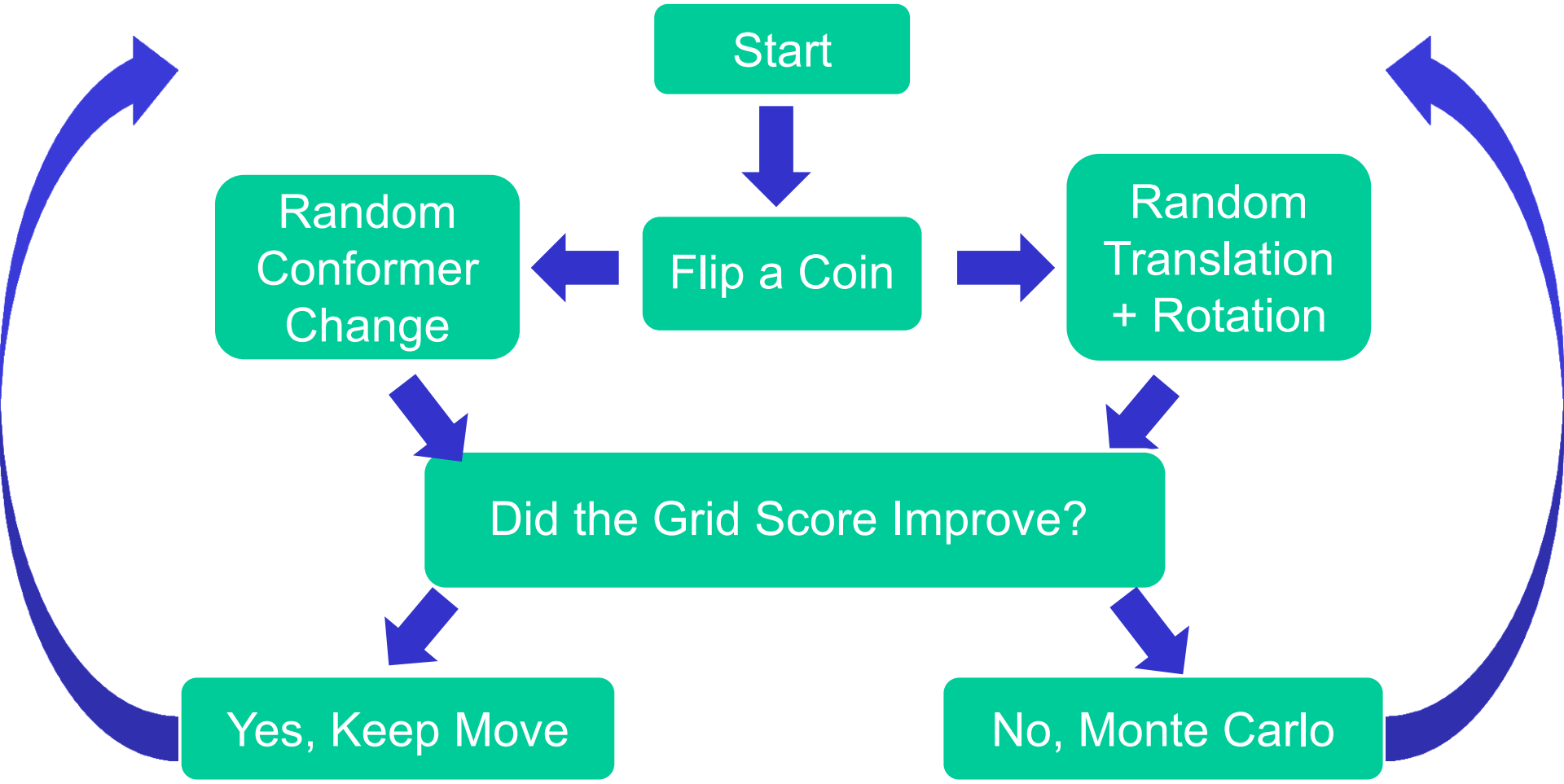
RosettaLigand Algorithm

- Ensemble selection
- Initial placement
- **Transform**
- Docking Cycles
- High res refinement

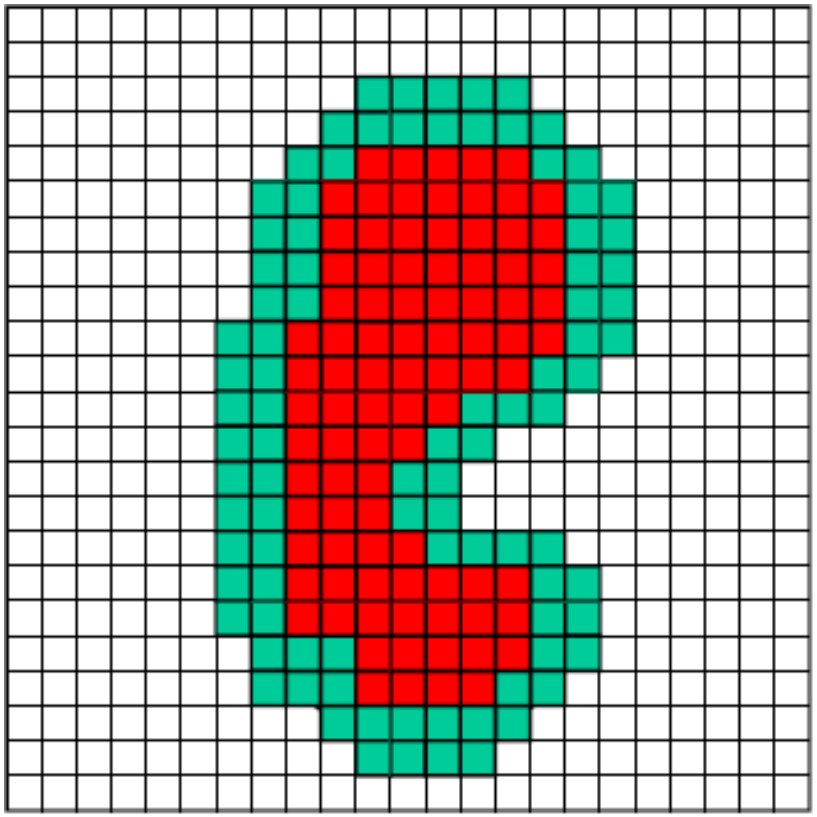
Grid based Monte-Carlo Translation, Rotation, and Conformation Sampling

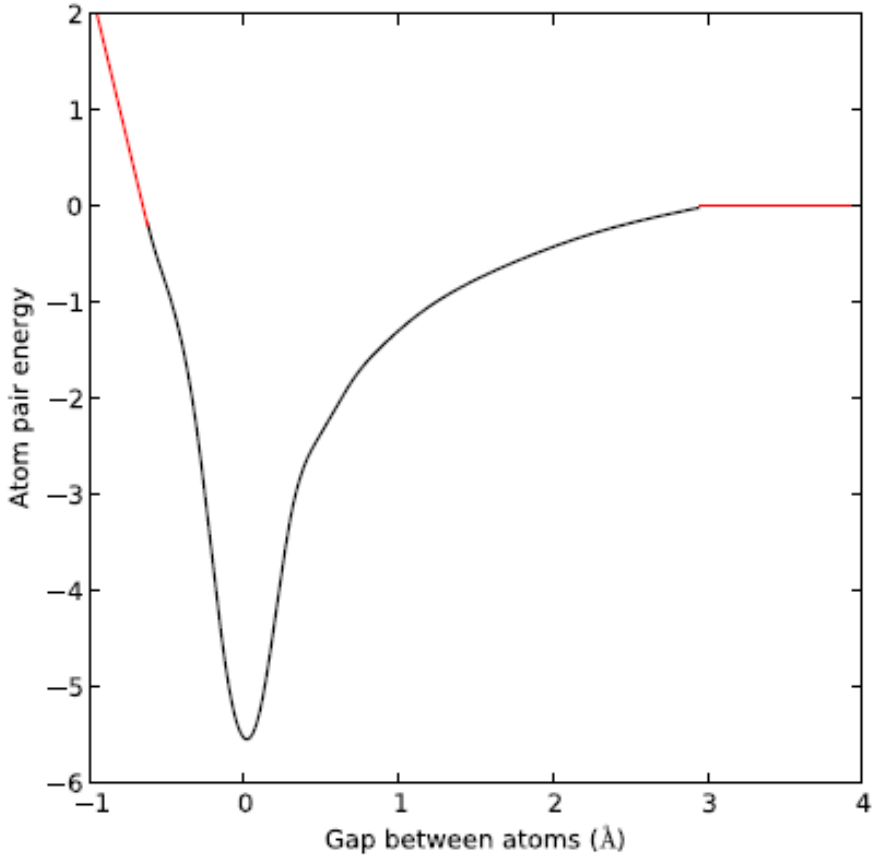


Transform Algorithm



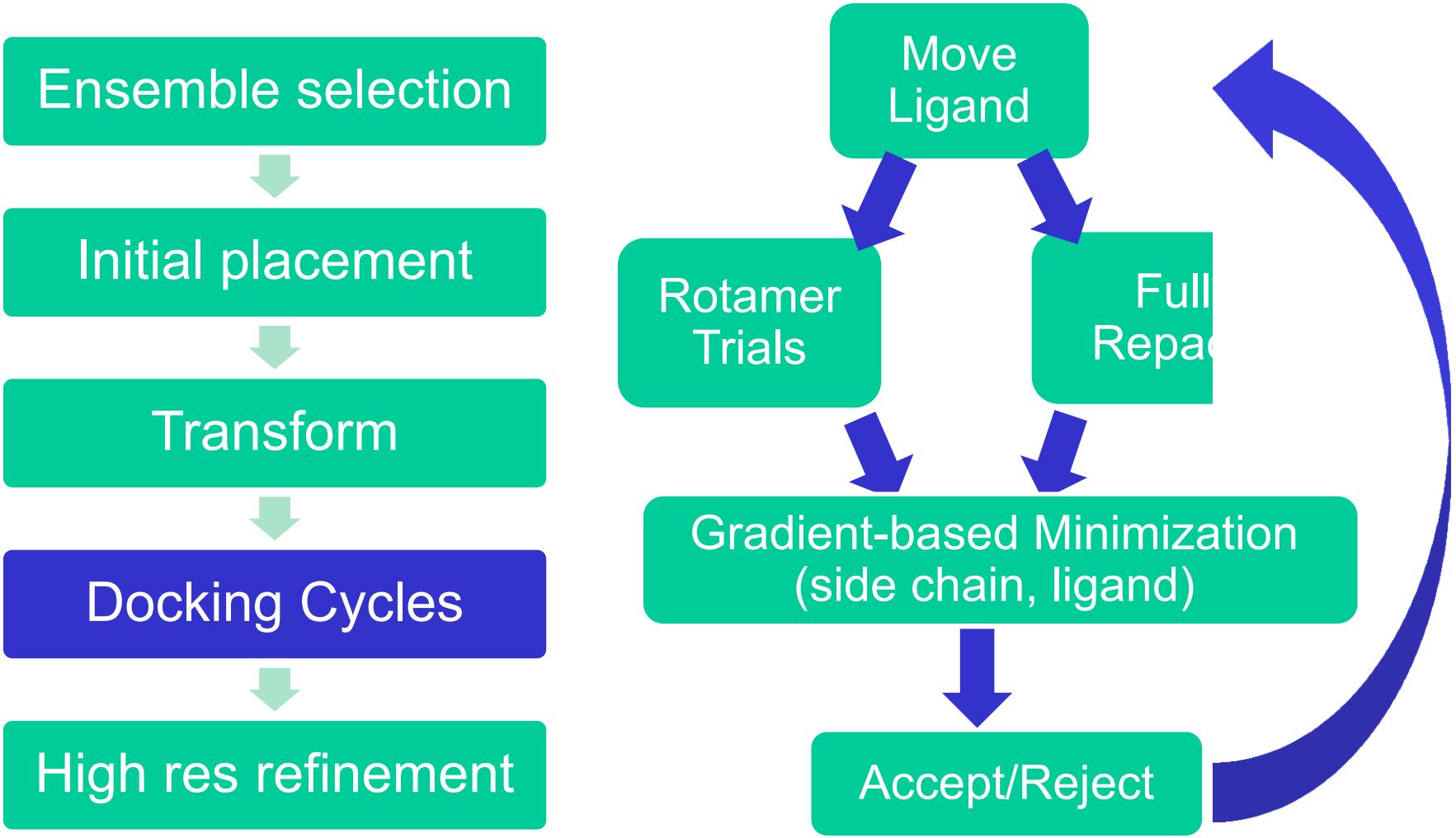
Simple Shape Complementarity



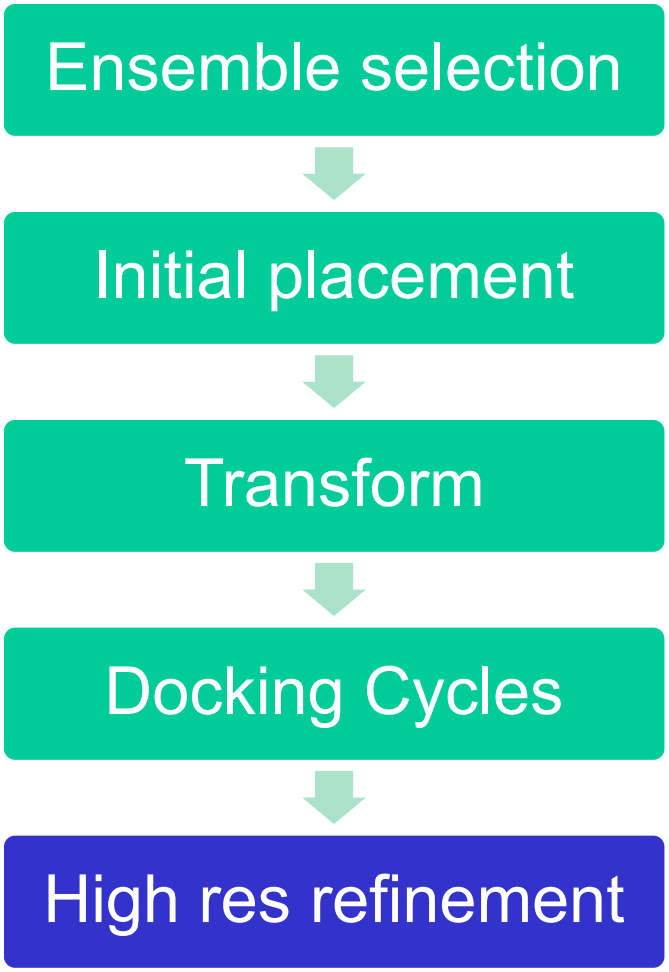


Red = Repulsion Teal = Attraction

RosettaLigand Algorithm



RosettaLigand Algorithm



Gradient Based Minimization of side-chain and backbone torsion angles

References

Methodology

- **DeLuca, S., Khar, K. & Meiler, J. Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One* 10, e0132508 (2015).**
- **S. Combs *et al.*, Small-molecule ligand docking into comparative models with Rosetta, *Nature Protocols* 8, 1277–1298 (2013).**
- G. Lemmon, J. Meiler, R. Baron, Ed. RosettaLigand docking with flexible XML protocols, *Methods in Mol Biol* **819**, 143–155 (2012).
- I. W. Davis, D. Baker, RosettaLigand docking with full ligand and receptor flexibility., *Journal of molecular biology* **385**, 381–92 (2009).
- K.W. Kaufmann *et al.,* Small Molecule Rotamers Enable Simultaneous Optimization of Small Molecule and Protein Degrees of Freedom in ROSETTALIGAND Docking, German Conference on Bioinformatics, pp. 148–157 (2008)
- J. Meiler, D. Baker, ROSETTALIGAND : Protein – Small Molecule Docking with Full Side-Chain Flexibility, *Proteins* **548**, 538–548 (2006).

References

Applications

- K. J. Gregory *et al.*, Probing the metabotropic glutamate receptor 5 (mGlu5) positive allosteric modulator (PAM) binding pocket: discovery of point mutations that engender a “molecular switch” in PAM pharmacology., *Molecular pharmacology* **83**, 991–1006 (2013).
- B. Allison *et al.*, Computational design of protein-small molecule interfaces., *Journal of structural biology* (2013), doi:10.1016/j.jsb.2013.08.003.
- G. Lemmon, J. Meiler, Towards ligand docking including explicit interface water molecules., *PloS one* **8**, e67536 (2013).
- G. Lemmon, K. Kaufmann, J. Meiler, Prediction of HIV-1 protease/inhibitor affinity using RosettaLigand., *Chemical biology & drug design* **79**, 888–96 (2012).
- K. W. Kaufmann, J. Meiler, Using RosettaLigand for small molecule docking into comparative models., *PloS one* **7** (2012), doi:10.1371/journal.pone.0050769.
- K. W. Kaufmann *et al.*, Structural determinants of species-selective substrate recognition in human and Drosophila serotonin transporters revealed through computational docking studies., *Proteins* **74**, 630–42 (2009).