

The Development and Application of Computational Methods for the Prediction of G Protein-Coupled Receptor Structures

Thesis by
Jenelle K. Bray

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



California Institute of Technology
Pasadena, CA

2010

(Defended August 10, 2009)

© 2010

Jenelle K. Bray

All Rights Reserved

Acknowledgments

I have been fortunate to have had such a supportive advisor as Bill Goddard during my time at Caltech. He has allowed me great freedom in my academic pursuits, and he has also taught me to be confident in my abilities as a scientist. I am also very grateful for the help given to me by many of my group members, and the biogroup as whole. I would especially like to thank Ravi Abrol, Adam Griffith, Victor Kam, John Wendell, and Heather Wiencko.

More than anyone, it was my undergraduate advisor at University of Oregon, Marina Guenza, who encouraged me to pursue a graduate degree. She was extremely generous in her time, and it was in her group that I discovered my passion for research. She was an excellent example of a successful woman in science, and I hope I can follow in her footsteps and encourage other young women to go into chemistry.

I have been lucky to have had outside research opportunities during my time as a graduate student. Both Lubomir Rulisek at the Institute of Organic Chemistry and Biochemistry in Prague and Teresa Head-Gordon at UC Berkeley were great advisors. They were very welcoming and they helped me to learn a great deal in a short amount of time.

The music program at Caltech provided a much needed escape from the hours spent at my computer. Bill and Delores Bing made the concert band and the chamber music program more than I could have ever asked for. It is due to

them that I have had some amazing musical opportunities while at Caltech, from performing at Carnegies to playing Messiaen's Quartet for the End of Time.

I could not have survived my time here without my amazing friends. Kana Takematsu and Adrienne Erickcek were the most caring, understanding, and fun friends possible. I am going to miss the girls' nights with them, and I hope we can have many more in the future. I will also miss my nights watching bad dance movies with Kana and Laurence Yeung. I look forward to the time when we can revive our tiny instrument band. The times spent with my friends at Caltech will be my fondest memories of my graduate career.

Finally, my family has been incredibly loving and supportive. My parents' faith in my ability to succeed, often surpassing my own, has been unwavering. My sister Kiah has been a great friend, and her visits are highlights of the year. My friend Naomi, whom I consider family, has been the best friend anybody could ask for. Without my wonderful family and friends, my success at Caltech would not have been possible.

Abstract

Computational methods for the prediction of G protein-coupled receptor (GPCR) structures were applied to serotonin receptors, and new methods were developed to predict an orphan GPCR structure. First, the MembStruk procedure was used to predict the structures of the serotonin 2b and 2c receptors. Ligand binding sites for agonists and antagonists were predicted for both receptors. In addition, the SAR data for a series of psilocybin analogs bound to serotonin 2c were predicted. There was good agreement with binding and mutagenesis experiments.

A new structure prediction procedure called SuperBiHelix was developed to predict an ensemble of low-lying structures. SuperBiHelix samples the tilt and sweep angles of the transmembrane helices along with the rotation of the helices along the helical axes. The procedure was validated on the β 2-adrenergic receptor and A_{2A} adenosine receptor crystal structures. This procedure was then used to predict the structure of GPR88, an orphan receptor. GPR88 has been identified as a novel target for psychiatric disorders. Three lipids were predicted to bind to GPR88. The head group of a lipid would bind to R113(3) and R116(3) at the extracellular side of the receptor. The lipid tail would bind in an aliphatic pocket in the TM2-TM3-TM6-TM7 region. The predicted bound complexes offer good suggestions for binding and mutagenesis experiments that could help validate the proposed structures.

Table of Contents

List of Figures	iii
List of Tables	v
Introduction	1
Chapter 1: The Structures of Human Serotonin 2b and 2c G Protein-Coupled Receptors Bound to Agonists and Antagonists	3
Abstract	3
Introduction	4
Methodology	7
TM Predictions	7
Packing of the TM Helices into a Seven-Helix Bundle	10
Rigid Body Dynamics of the Bundle in an Explicit Lipid Bilayer	12
Optimization of Individual Helices within the Bundle	13
Prediction of Ligand Binding Sites	21
Molecular Dynamics Simulation	24
Binding Energy Calculations	27
Results and Discussion	29
Serotonin Binding	29
Bound Serotonin Dynamics	31
Comparison to Mutagenesis Experiments	35
Ritanserin Antagonist Binding	38
Metergoline Antagonist Binding	39
Methiothepin Antagonist Binding	41
SAR for Psilocybin Analogs	41
Discussion of Binding Energies	49
Comparison to 5-HT _{2B} Predictions	52
Conclusion	60
Chapter 2: SuperBiHelix: A New Method for Predicting an Ensemble of Low-Lying GPCR Structures	62
Abstract	62
Introduction	62
Methodology	63
Validation	69
SuperBiHelix on Crystal Helices in an Incorrect Template	69
The Effect of SuperBiHelix on Binding Site Predictions	82
SuperBiHelix on Crystal Structures	87
Conclusion	91

Chapter 3: The Structure of the Orphan G Protein-Coupled Receptor GPR88 with Predicted Ligands	93
Abstract	93
Introduction	94
Methodology	95
Results and Discussion	99
BiHelix and ComBiHelix	99
SuperBiHelix and SuperComBiHelix	103
Ligand Docking	108
Conclusion	120
References	121

List of Figures

1	Serotonin 2c hydrophobic profile	9
2	Serotonin 2c rotational profiles	18
3	Final predicted structure of serotonin 2c	20
4	Ligand structures and binding constants for docking to serotonin 2c	21
5	Serotonin 2c antagonist docking spheres	23
6	Energy and volume of serotonin 2c simulation	26
7	Predicted structure of serotonin bound to serotonin 2c	30
8	Structure of serotonin bound to serotonin 2c after 5 ns of dynamics	32
9	Distances between serotonin and residues on serotonin during dynamics	33
10	Predicted structures of antagonists bound to serotonin 2c	40
11	Predicted structures of SAR ligands bound to serotonin 2c	43
12	Calculated vs. experimental binding energies for ligands bound to serotonin 2c	51
13	Alignment of the predicted serotonin 2b and 2c structures	53
14	Predicted structures of serotonin and ritanserin bound to serotonin 2b	55
15	The coordinate system used to described the seven-helix bundle	64
16	Diagram of the SuperBiHelix method	65
17	Diagram of the helical bundle partitioned into quadhelix bundles	67
18	Charged termini residues in the β 2-adrenergic and the A_{2A} adenosine receptors	74
19	B_2 -adrenergic helices in the A_{2A} adenosine template before and after SuperBiHelix	79
20	A_{2A} adenosine helices in the β 2-adrenergic template before and after SuperBiHelix	81
21	B_2 -adrenergic structures before and after SuperBiHelix bound to carazolol	85
22	A_{2A} adenosine structures before and after SuperBiHelix bound to ZM241385	86
23	GPR88 hydrophobic profile	97
24	Lowest energy GPR88 ComBiHelix structure	102
25	Lowest energy GPR88 SuperComBiHelix structure	105
26	Second-lowest energy GPR88 SuperComBiHelix structure	106
27	The three lipid molecules docked into GPR88	109
28	Predicted GPR88 structures bound to FFA	114
29	Predicted GPR88 structures bound to LPA	115
30	Predicted GPR88 structures bound to S1P	116

List of Tables

1	Serotonin 2c TM helix predictions	11
2	Serotonin 2c combinatorial analysis	20
3	Nonbond interactions between each bound ligand and individual residues in the predicted serotonin 2c structure	30
4	Mutations for each ligand that are predicted to decrease binding affinity	49
5	Experimental binding constants and calculated binding energies for ligands bound to serotonin 2c	50
6	Experimental binding constants and calculated binding energies for ligands bound to serotonin 2b	60
7	The differences between the A _{2A} adenosine receptor and β 2-adrenergic receptor templates	69
8	SuperComBiHelix results for β 2-adrenergic receptor with A _{2A} adenosine template ϕ and θ values	71
9	SuperComBiHelix results for the A _{2A} adenosine receptor with β 2-adrenergic template ϕ , θ and η values	72
10	SuperComBiHelix results for the β 2-adrenergic receptor with A _{2A} adenosine template ϕ , θ and η values	73
11	SuperComBiHelix results for the β 2-adrenergic receptor with A _{2A} adenosine template ϕ , θ and η values, with termini residues mutated to alanine during the SuperBiHelix and SuperComBiHelix procedures	75
12	SuperComBiHelix results for the β 2-adrenergic receptor with A _{2A} adenosine template ϕ , θ and η values, with termini residues mutated to alanine during the SuperBiHelix procedures	76
13	SuperComBiHelix results for the β 2-adrenergic receptor with A _{2A} adenosine template ϕ , θ , η , x and y values	78
14	SuperComBiHelix results for the A _{2A} adenosine receptor with β 2-adrenergic template ϕ , θ , η , x and y values	80
15	The ranking of the crystal structure conformation for each helix after the quadhelix protocol that determines which helix conformations to include in the SuperBiHelix bundle energy calculations	88
16	SuperBiHelix and SuperComBiHelix results for the β 2-adrenergic receptor crystal structure	89
17	SuperBiHelix and SuperComBiHelix results for the A _{2A} adenosine receptor crystal structure	90
18	Proteins sequentially related to GPR88	96
19	TM helix predictions for human GPR88	98
20	ComBiHelix results for GPR88 in the β 2-adrenergic receptor template	100
21	SuperComBiHelix results for GPR88	103
22	Neutralized binding energies for three lipids in GPR88	111

23 Nonbond interactions between each bound ligand bound and individual residues in the predicted GPR88 structures	113
--	-----

Introduction

An important class of transmembrane proteins is the G protein-coupled receptor (GPCR) superfamily, which regulates signal transduction. GPCRs are of great interest pharmacologically, for they are the targets of 50% of recently released drugs and 25 of the top 100 best-selling drugs.^[1] Malfunctions in GPCRs play a part in diseases such as ulcers, allergies, migraines, anxiety, psychosis, schizophrenia, hypertension, asthma, congestive heart failure, Parkinson's, and glaucoma.^[2] Despite the great deal of interest in GPCRs, progress in obtaining atomic level experimental structures has been slow, due to challenges involved in GPCR expression, purification, and crystallization. There are currently structures of only four GPCRs: rhodopsin,^[3] the β 1-adrenergic receptor,^[4] the β 2-adrenergic receptor,^[5] and the A_{2A} adenosine receptor.^[6] Consequently, there is a great need for GPCR structure predictions, which computational methods can help fill. Many GPCRs have similar enough structures that an antagonist or agonist targeted at one receptor will bind another type of receptor and lead to unwanted side effects. Thus, atomic-level structures are essential for the rational design of subtype-specific drugs.

The Membstruk procedure^[2, 7] was used to successfully predict serotonin 2b and 2c structures and binding sites,^[8] as described in Chapter 1. The serotonin study led to some important developments in the method, most significantly the neutralization of complexes for binding energy calculations. However, Membstruk required a great deal of user judgment. Additionally, its results were

heavily dependent on the experimental template used because Membstruk sampled very little GPCR conformational space. Thus, a new procedure, BiHelix, was developed that efficiently sampled all of the possible rotations of each helix around their helical axes. This procedure was a big improvement over Membstruk, but it was still dependent on the experimental template used. In order to predict the structures of receptors dissimilar to any available crystal structure, SuperBiHelix was developed. SuperBiHelix samples the tilt and sweep angles of each helix in addition to the rotation around their helical axes. The description and validation of SuperBiHelix is given in Chapter 2. Finally, in Chapter 3, SuperBiHelix was used to predict the structure of the orphan GPCR GPR88. An ensemble of low-lying GPR88 structures was predicted and some possible strong-binding lipids were proposed. Ligand binding and mutagenesis experiments were suggested that could help to validate the proposed structures and binding sites.

Chapter 1

The Structures of Human Serotonin 2b and 2c G Protein-Coupled Receptors Bound to Agonists and Antagonists

Abstract

We used the MembStruk computational procedure to predict the 3-dimensional (3D) structure for the serotonin 5-HT_{2C} G protein-coupled receptor (GPCR). Using this structure, we used the MSCDock computational procedure to predict the 3D structures for bound ligand-protein complexes for agonists such as serotonin and antagonists such as ritanserin, metergoline, and methiothepin. In addition, we predicted the SAR data for a series of psilocybin analogs, both agonists and antagonists. We performed molecular dynamics (MD) on serotonin bound to 5-HT_{2C} and we find the protein and binding site to be stable after 5 ns. We find good agreement with the currently known experimental data and we predict a number of new mutations, which could be used to validate further our predicted structures. This agreement between theory and experiment suggests that our 3D structure is sufficiently accurate for use in drug design. We also compare a preliminary prediction for 5-HT_{2B} with our prediction for 5-HT_{2C} and find a difference in TM5 that contributes to different serotonin binding modes in 5-HT_{2B} and 5-HT_{2C}.

Introduction

The serotonin (5-hydroxytryptamine, or 5-HT) class of G protein-coupled receptors (GPCR) serve as neurotransmitters involved in many processes in the central nervous system, including the regulation of feeding, aggression, mood, perception, pain, and anxiety.^[9] Additionally, 5-HT regulates vascular and nonvascular smooth muscle growth, uterine smooth muscle growth, and gastrointestinal functioning.^[9] Consequently, these receptors are targets for a variety of drug therapies. This family of receptors that consists of at least 15 receptors partitioned into seven main types helps to mediate these physiological functions. This creates a problem for drug design since all 15 likely have similar binding sites, making it difficult to attain selectivity of a drug to just one receptor. This problem is made worse because there are no 3D x-ray structures for any of these receptors. To help fill this gap we have been developing methods (MembStruk) for predicting the 3D structures.

We consider here the 5-HT₂ family, which consists of three structurally related receptors: 5-HT_{2A}, 5-HT_{2B} and 5-HT_{2C}. Drugs that target 5-HT₂ receptors are used to treat such disorders as schizophrenia, depression, and glaucoma. The three 5-HT₂ receptors have high sequence identity within the transmembrane domains, and they play a variety of physiological roles. The majority of current drugs with affinity for 5-HT₂ receptors are not subtype specific,^[10] so it is often unclear which receptor is responsible for the particular physiological effect and which receptor is responsible for unwanted side effects. We expect that detailed knowledge of the ligand-binding environment of the 5-

HT₂ receptors would facilitate the creation of more subtype specific drugs that in turn would lead to a better understanding of the specific physiological role of each receptor.

Since no direct studies have determined how ligands bind to 5-HT₂ receptors,^[9] we will use computational methods to predict the 3D structure of the receptors and of the ligand-receptor complexes. Since there are no experimental structures for comparison, we validate our predicted structures indirectly by comparison with available mutagenesis and ligand binding experiments. Then we predict the results for new mutation and binding studies that could be carried out experimentally. We report here the results for 5-HT_{2C}, and preliminary results for 5-HT_{2B}. These studies should provide the data needed to design sub-type specific drugs and to determine the origin of physiological responses to different agonists and antagonists.

The 5-HT_{2C} receptors are found in many regions of the brain. Some evidence suggests that 5-HT_{2C} may play a role in the mediation of the sleep-wake cycle.^[10] The activation of 5-HT_{2C} appears also to be correlated with penile erections.^[10] Several studies have suggested a prominent role for 5-HT_{2C} receptors in the regulation of food intake and energy balance, and that 5-HT_{2C} agonists act as appetite suppressants.^[11] Additionally, 5-HT_{2C} is believed to be involved in anxiety, feeding behaviors, and anxiety-related disorders like obsessive compulsive disorder (OCD).^[11] In fact, 5-HT_{2C} agonists can intensify the symptoms of schizophrenia and OCD. However, it is difficult to determine whether these symptoms are caused solely by 5-HT_{2C} activation because the

activation of other 5-HT receptors, most notably 5-HT_{2A}, is implicated in similar physiological effects.^[11]

In order to provide the basis for designing subtype specific agonists and antagonists for the 5-HT_{2C} receptor, we set out to obtain the 3D structure of this receptor and for ligands binding to it. We did not use homology modeling because of the two experimentally known 3D structure for GPCRs, 5-HT_{2C} is only 19% homologous to bovine rhodopsin^[3] and 31% homologous to the beta(2)-adrenergic receptor, whose structure has been recently solved.^[5] Additionally, we have found in studies of other GPCRs that systems with even 80% sequence identity can have different binding sites. Instead, we used the MembStruk GPCR structure prediction procedure to predict the 3D structure. MembStruk has been used successfully to predict the structures for CCR1 chemokine receptor, MrgC11 tetrapeptide receptor and DP prostanoid receptor, in each case for which experimental validation was done after the predictions. In addition, predictions of the structures for D2 dopamine receptor, β 2 adrenergic receptor and M1 muscarinic acetylcholine receptor are in good agreement with numerous mutation and ligand binding experiments. Furthermore, predictions of olfactory receptors, the PTC bitter-taste receptor, and rhodopsin agree with somewhat more limited ligand binding experiments.^[2, 7, 12-18]

Here we report the structure prediction for the 5-HT_{2C} receptor, and then we use the MSCDock procedure^[19] to predict the binding sites of serotonin and three strongly binding antagonists. In addition, we predict the binding of ten psilocybin analogs to compare with experimental SAR data. We also compare the

predicted structure for 5-HT_{2C} with a preliminary prediction of 5-HT_{2B}. In all cases, we obtain excellent agreement with currently available data and we predict new experiments to validate further the predicted structures.

Methodology

We used the MembStruk procedure (version 4.30)^[2, 7] to predict the structure of the 5-HT_{2C} receptor. This procedure consists of five main steps:

1. Prediction of the transmembrane (TM) helices from hydrophobic analyses of the amino acid sequence.
2. Assembly of the seven-helix TM bundle and coarse grain optimization of the bundle.
3. Rigid body dynamics of the bundle in an explicit lipid bilayer.
4. Optimization of individual helices.
5. Optimization of the entire protein.

Throughout these steps, MPSim^[20] was used with the Dreiding forcefield^[21] for the molecular mechanics and molecular dynamics steps unless otherwise specified, with a dielectric constant of 2.5, and a 6-12 Lennard-Jones potential for the van der Waals interactions.

TM Predictions

The prediction of the TM helical regions is based on the assumption that the outward facing sections of the TM helices must be hydrophobic because they are in contact with the hydrocarbon tails of the lipid bilayer, and that the hydrophobic center of each helix should be at the center of the membrane.^[22] These ideas are

used in the TM2ndS program to determine the TM helical regions with a hydrophobic profile and helical capping rules. To determine the hydrophobic profile, sequences related to 5-HT_{2C} were found with NCBI BLAST.^[23] Of the ~200 sequences, we selected an ensemble of ~100 sequences having a uniform distribution of sequence identities from 20% to <100%. These sequences (including the target) were aligned with ClustalW to generate pairwise multiple sequence alignments.^[24] From the multiple sequence alignments, we kept only the 40 sequences for mouse, rat and human receptors.

We used the Eisenberg hydrophobicity scale^[25] (which assigns each amino acid a hydrophobicity value ranging from -1.76 to 0.73) to estimate the hydrophobicity for each residue position in the alignment, and we averaged this over all of the sequences in the multiple sequence alignment. The Eisenberg scale was used because we validated our TM predictions on bovine rhodopsin using this hydrophobicity scale.^[2] Then, we calculated the hydrophobic profile by averaging the hydrophobicity over a range of window sizes ranging from 12 to 20, and selected the lowest window size that gives seven peaks with lengths greater than 20 residues. The baseline value is the average hydrophobicity over all of the amino acids in the protein, which for 5-HT_{2C} is 0.007. Because the TM helices can extend past the membrane surface, we use capping rules to refine the helix predictions. Known helix breakers are proline, glycine, arginine, histidine, lysine, aspartic acid and glutamic acid. If any such helix breakers are found within four residues toward the membrane, we keep the initial TM helix predictions. If no helix breaker is found, the helix is extended until a breaker is

found, but only up to six residues. Additionally, the capping rules include the requirement that the intracellular and extracellular loops consist of at least six residues. Then we define the hydrophobic center of each TM helix as the point of maximum hydrophobicity averaged over the range of window sizes. The hydrophobic profile for 5-HT_{2C} is shown in Figure 1.

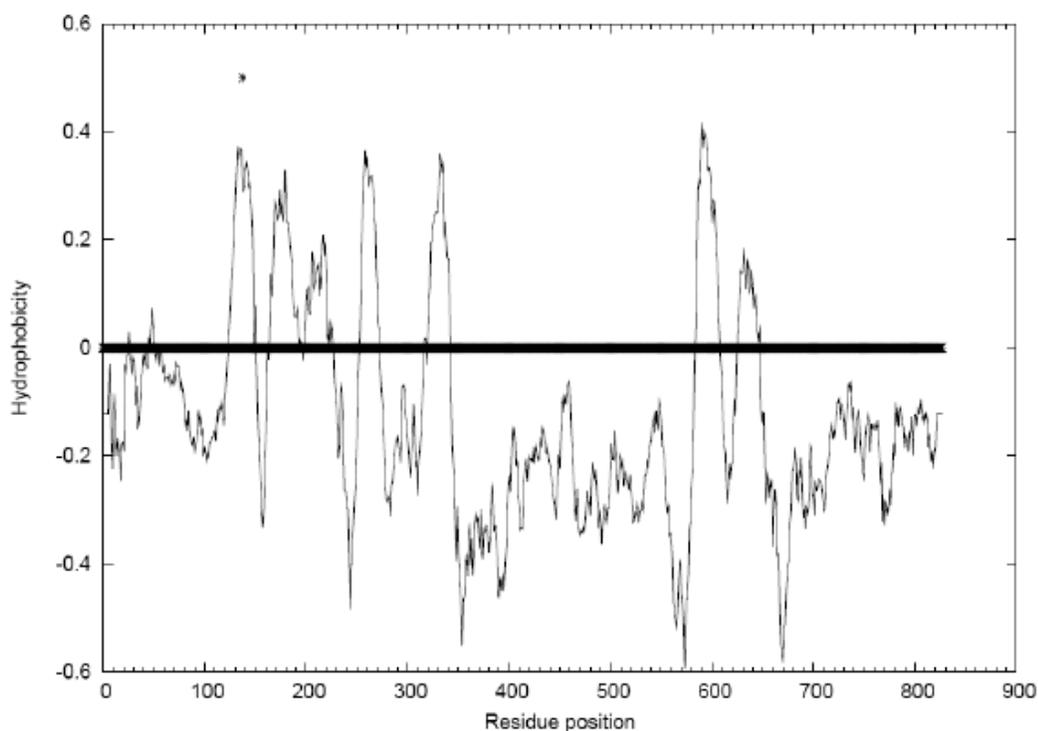


Figure 1. The hydrophobic profiles for the multiple sequence alignment of the 5-HT_{2C} based on aligning the entire sequence. The bold line is the baseline hydrophobicity value.

The edges of some of the TM helices were not well defined, without any residues left to form a loop between TM2 and TM3. Thus, to further refine the predictions, each individual TM helix was BLAST searched separately. For each helix, we consider the ten residues on either side of the hydrophobic center and

make a new BLAST over the GPCR database to obtain a new set of sequences (>20% identity) and then we did a new alignment with these sequences as input to the TM2ndS program. Next, the peak for the specific helix was examined, along with capping rules, to determine the edges of the helix. The hydrophobic centers calculated with the entire protein sequence in the previous step were retained. The predictions, before and after individual helical refinement and helical capping, are given in Table 1. Table 1 shows that the final TM predictions were taken to be the capped individual helix predictions, except for the case of TM5. Here, the capping procedure added six residues to TM5, which we considered to make the helix too long, especially in comparison to TM5 of 5-HT_{2B}. Thus, we adjusted the prediction for TM5 to end at a tyrosine.

Packing of the TM Helices into a Seven-Helix Bundle

The seven TM helices predicted by TM2ndS were next built in to a TM bundle. First, we built each TM helix into a canonical right-handed alpha-helix with extended side chain conformations. Then we placed each so that all seven have their calculated hydrophobic centers in an imaginary plane running through the center of the lipid bilayer.

We orient the most hydrophobic face of each helix to point away from the center of the bundle (based on equal weight for each center) because we expect that it will be energetically favorable for the most hydrophobic part of each helix to be facing outward toward the hydrophobic lipid bilayer. This is achieved by calculating the hydrophobic moment of each helix, considering the fifteen residues on either side of the hydrophobic center, but including just the face of

the helix (considering 180° of the projection on a plane perpendicular to the helical axis). The hydrophobic moment is pointed outward from the center of mass of the protein.

Table 1. TM helix predictions for 5-HT_{2C}

TM	Predicted TM Region	Prediction Method
1	GVQNWPAISIVIIIMTIGGNILVIMA GVQNWPAISIVIIIMTIGGNILVIMAVSME GVQNWPAISIVIIIMTIGGNILVIMA GVQNWPAISIVIIIMTIGGNILVIMAVSME GVQNWPAISIVIIIMTIGGNILVIMAVSME	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
2	YFLMSLAIADMLVGLLVMPSSLAILYD--Y ATNYFLMSLAIADMLVGLLVMPSSLAILYD-- NYFLMSLAIADMLVGLLVMPSSLAILYDY TNYFLMSLAIADMLVGLLVMPSSLAILYD TNYFLMSLAIADMLVGLLVMPSSLAILYD	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
3	PRYLCPVWISLDVLFSTASIMHLCAISLDR RYLCPVWISLDVLFSTASIMHLCAISLDR PRYLCPVWISLDVLFSTASIMHLCAISLDR RYLCPVWISLDVLFSTASIMHLCAISLDR RYLCPVWISLDVLFSTASIMHLCAISLDR	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
4	MKIAIVWAISIGVSVPIPVIG IMKIAIVWAISIGVSVPIPVIGL IMKIAIVWAISIGVSVPIPVIG AIMKIAIVWAISIGVSVPIPVIGL AIMKIAIVWAISIGVSVPIPVIGL	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
5	FVLIGSFVAFFIPLTIMVITYCL FVLIGSFVAFFIPLTIMVITYCLTIYVLR FVLIGSFVAFFIPLTIMVITYCL FVLIGSFVAFFIPLTIMVITYCLTIYVLR FVLIGSFVAFFIPLTIMVITYCLTIY	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
6	LGIVFFVFLIMWCPFFITNILSVLC LGIVFFVFLIMWCPFFITNILSVLCE LGIVFFVFLIMWCPFFITNILSV LGIVFFVFLIMWCPFFITNILSVLCE LGIVFFVFLIMWCPFFITNILSVLCE	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction
7	EKLLNVFVWIGYVCSGINPLVYT EKLLNVFVWIGYVCSGINPLVYT EKLLNVFVWIGYVCSGINPLVYT EKLLNVFVWIGYVCSGINPLVYT EKLLNVFVWIGYVCSGINPLVYT	Full protein, uncapped Full protein, capped Indiv. helix, uncapped Indiv. helix, capped Final Prediction

The predictions are given for both the initial predictions based on the full protein sequence and for the subsequent prediction in which each TM regions was searched and aligned individually. In each case, the predictions are based on hydrophobicity and supplemented with predictions based on capping rules. A gap (-) in a sequence is caused by the multiple sequence alignment. In the final predictions, the calculated hydrophobic center for each helix is indicated by a boxed residue.

The calculation of the center fixes the z coordinate for each helix, while the rotational orientation determines the η angle. There remain four other parameters to uniquely position each helix: the x and y position in the plane through the center of the lipid, the tilt θ of the helix from the z -axis perpendicular to the plane, and the azimuthal angle ϕ for the tilting plane. These four parameters we take from the 7.5 Å electron density map of frog rhodopsin.^[26] This electron density map gives the relative orientations of the helical axes, but no information about translation or rotation of the helices. The original MembStruk was developed using frog rhodopsin because there was no bovine rhodopsin x-ray available at that time. We found that frog rhodopsin worked well and saw no need to change. Only the overall x and y position and tilt (θ , ϕ) are used from the template. No detailed atom information is used. Thus we use only undifferentiated electron density information from experiment. We can take these parameters from any crystal structure or from the structure of a predicted GPCR structure after explicit lipid bilayer dynamics.

The six parameters x , y , z , θ , ϕ , and η for each helix uniquely determine how to pack the seven helices into the bundle. Next, we need to determine the structure within each helix. First, we used SCRWL^[27] to place each side chain appropriate for the seven-helix bundle. Then, we inserted counterions to offset the charged residues not in salt bridges.

Rigid Body Dynamics of the Bundle in an Explicit Lipid Bilayer

In order to allow the helices to deviate from ideal alpha-helix structure, as is experimentally observed in transmembrane helices, we carried out torsional

molecular dynamics (NEIMO MD)^[28] for each individual helix at 300 K for 500 ps. The helices with prolines form bends due to the disruption of the alpha-helix hydrogen bonding network. The lowest energy structure from the last 250 ps was chosen and minimized (using conjugate gradients). Then, all seven minimized helices were bundled together again by CRMS matching the structures back on to the original positions. The side chains were then reassigned using SCRWL, and the entire bundle was minimized. Finally, the helices were again optimized rotationally and translationally as described in the previous step.

Optimization of Individual Helices within the Bundle

Although the hydrophobic nature of the hydrocarbon tails in the membrane are important in determining the rotational orientation of the helices, interhelical hydrogen bonds and salt bridges also help stabilize the structure. Residues conserved across all biogenic amine receptors are expected to be involved in these hydrogen bonding interactions, playing a role in stabilization, binding, and/or activation. Accordingly, it is likely that they will be facing in toward the bundle in order to interact with the ligand or other helices. Moreover, TM3 is near the center of the helical bundle, and hence it is less affected by the membrane environment than the other helices. Consequently, orientation by hydrophobic moment is less appropriate for TM3. Thus, we carried out subsequent optimization of the rotational orientation of the helices by using energy minimization techniques.

Starting with the phobic face orientation, we use energy minimization to optimize the rotation of each helix. Here, each helix is rotated and optimized through a range of angles. For each angle, the helices are rotated in the order 3 2 1 7 6 5 4, starting with the inner TM3 for which the phobic face orientation is least likely to be optimum (because it interacts least with the hydrophobic membrane). The procedure is as follows: The first helix is rotated 5° and -5° , and for these angles, the side chain positions of all seven helices are reassigned using SCRWL and the energy of the active helix is minimized (conjugate gradients) in the field of all of the other helices (whose atoms are kept fixed). The lowest energy conformation for the helix at 0° , 5° or -5° is chosen, and the helix fixed at this angle. Then, in succession, each of the other six helices is put through the same $\pm 5^\circ$ rotation procedure with the other helices fixed. After all seven helices have been rotated $\pm 5^\circ$ and fixed in their lowest energy conformation, the same procedure is repeated for increments of $\pm 10^\circ$. This process is repeated with 5° increments up to $\pm 25^\circ$. Then the process is continued with the angles decreased by 5° increments back down to $\pm 5^\circ$, where at each level all helices are allowed to adjust to the environment of the other helices. The net rotations accomplished by this process for TMs 1 through 7 are

$0^\circ, 5^\circ, -10^\circ, 5^\circ, 55^\circ, 0^\circ, \text{ and } 10^\circ$.

After selecting this lowest energy rotation, the entire helical bundle is energy minimized.

The helical bundle attained by energy-based rotation in the previous step is next immersed in a lipid bilayer made up of 52 explicit DLPC molecules in two

layers around the bundle. The protein is first energy minimized in this bundle, then we carry out 50 ps of rigid-body dynamics at 300 K are (using MPSim).^[20] In this calculation, the helices and lipid bilayer molecules are treated as rigid bodies. This dynamics step helps to optimize the vertical helical translations and rotations in an explicit lipid bilayer and to optimize the positions of the lipid molecules with respect to the helical bundle. After this simulation, we used SCRWL to replace the side chains, now using the lipid molecules as a bump test.

Next, we again considered helical rotations, but this time each helix was rotated a full 360° , in increments of 5° , with the other six helices fixed during the entire range of rotations. For each angle, we used SCRWL to reassign side chains, and then energy minimized all atoms of the seven helices and the explicit lipid bilayer.

For each rotated angle of a helix, the energy of the helix interacting with the six other helices is tabulated, along with the number of interhelical hydrogen bonds and the hydrophobic penalty. The hydrophobic penalty is calculated in order to ensure that it is not favorable for polar and charged residues to point into the lipid bilayer. With the lipids ignored, we consider that any residue with more than 60% of its surface area accessible to a solvent probe of radius 2.0 Å is pointing outward toward the lipid bilayer. Then the percentage of solvent accessible surface for each residue pointing outward within seven amino acids of the hydrophobic center is multiplied by the hydrophobicity value of the amino acid. The hydrophobicity scale used is the average of the augmented Wimley White scale^[29] and the White interface scale.^[30] We use this hydrophobicity scale for the

hydrophobic penalty instead of the Eisenberg scale used to determine the TM regions because it is thermodynamically based, and we are performing energy calculations in this step. We are currently testing our TM prediction algorithm with this thermodynamically based hydrophobicity scale. These calculated hydrophobic penalty values were used to determine which region of angles to choose for each of the six TMs (excluding TM3 which has little direct contact with the lipid bilayer).

The specific residues involved in each interhelical hydrogen bond were identified so that hydrogen bonds involving residues conserved within class A and biogenic amine receptors could be given more importance in choosing angles for combinatorial analysis. The residues focused on were

- ASN71 in TM1, ASP99 in TM2, and TRP179 in TM4 (all highly conserved among class A GPCRs), and
- ASP134 in the center of TM3, TRP324 in TM6, and ASN364 in TM7, (conserved among all mammalian biogenic amine receptors).^[31]

Experiments on various GPCRs suggest that the TM2 aspartic acid forms a hydrogen bonding network with the asparagines in TM1 and TM7.^[32, 33] The TM3 aspartic acid is known to bind the amino functional group in biogenic amines.^[34] The TM4 tryptophan is the most highly conserved residue across the entire GPCR superfamily, and is thought to be integral in structural stabilization.^[34] The TM6 tryptophan is believed to be important in activation.^[34] TM5 is the least conserved helix in GPCRs, suggesting its role in ligand specificity. A proline is highly conserved among class A GPCRs, but its role is most likely to disrupt the

hydrogen bonding within the helix and create flexibility.^[34] Thus, it would not necessarily be pointing in to interact with other helices or the ligand.

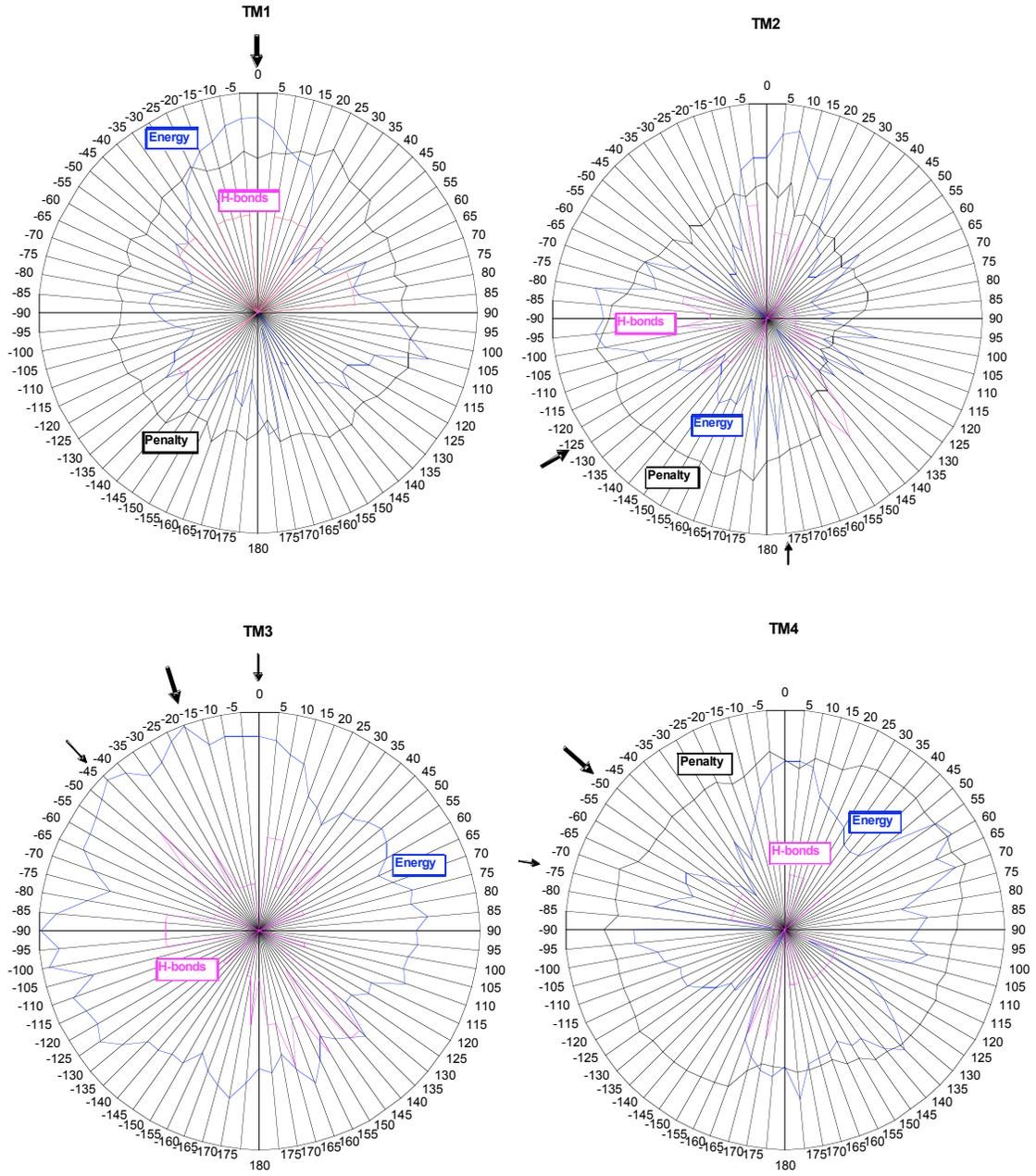
The list of specific hydrogen bonds formed for each helical rotation is used to help choose angles for combinatorial analysis.

- If a hydrogen bond occurs between two conserved residues, the angles for the two helices are retained.
- If the hydrogen bond involves just one conserved residue, we choose the angle or angles that retain the hydrogen bond and have the best combination of energy, hydrogen bonding and hydrophobic penalty.
- If a helix does not have a conserved residue involved in a hydrogen bond, the angle or angles with the best energy, number of hydrogen bonds and hydrophobic penalty are selected.

The rotational profile for each TM is shown in Figure 2. Based on the selection criteria, we selected 0° for TM1; 175° and -125° for TM2; 0° , -20° and -45° for TM3; -50° and -75° for TM4; -35° and -75° for TM5; 90° for TM6; and 170° for TM7. This leads to 24 cases, for each of which we used SCRWL to reassign side chains and then minimized the energy. These structures were ranked by number of hydrogen bonds and then by energy as shown in Table 2. The final rotations for TMs 1 through 7 were

0° , -125° , -20° , -50° , -75° , 90° , and 170° .

The lipid and counter ions were removed from the final structure and it was fully minimized. The final structure is shown in Figure 3.



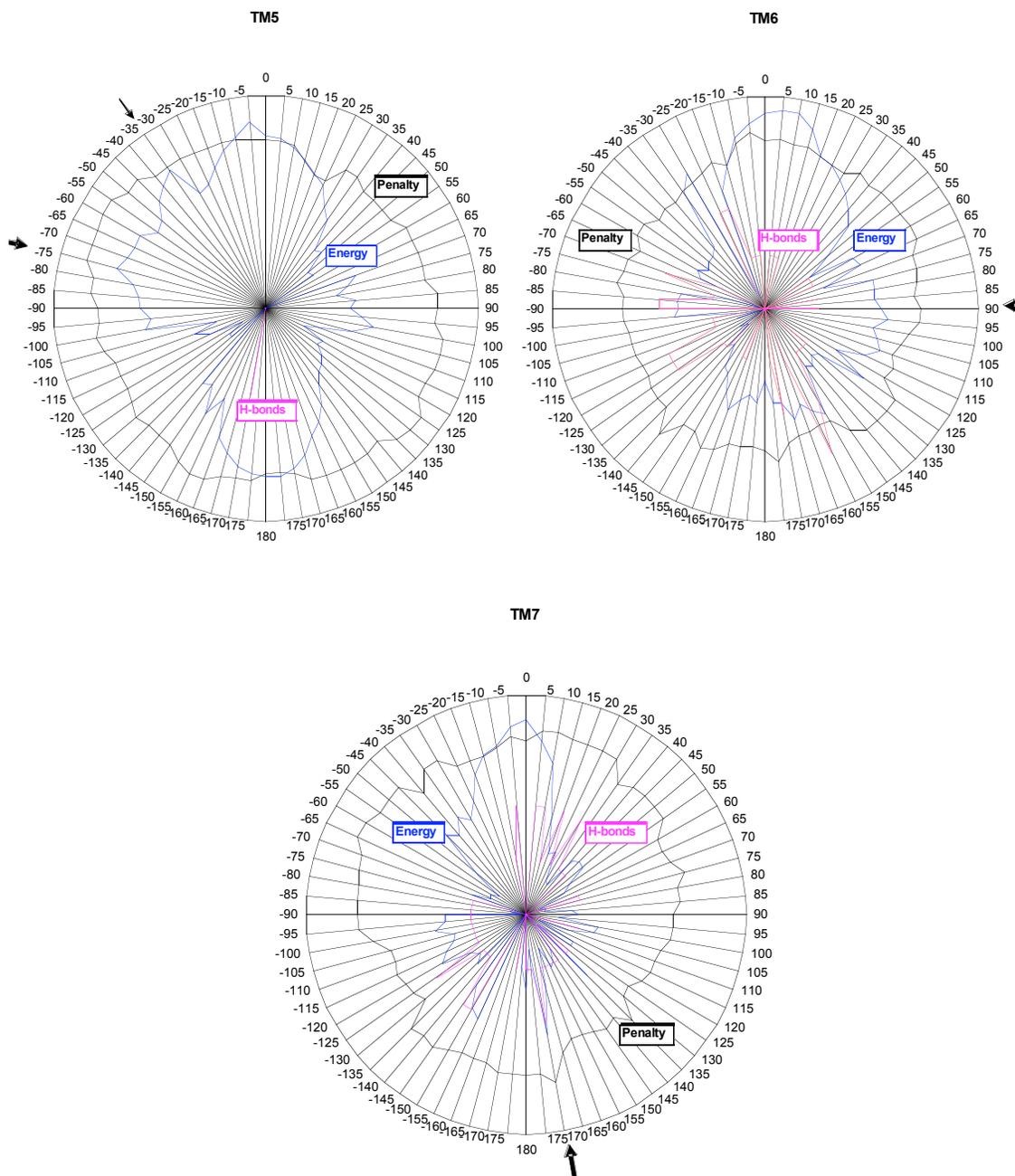


Figure 2. The hydrophobic penalty, energy, and number of hydrogen bonds for each rotation angle for each transmembrane helix. For each property, values toward the outside of the graph are more favorable. The arrows are the angles chosen for each TM, and the bold arrows are the angles from the final structure.

Table 2. Combinatorial analysis

H-bonds	Energy (kcal/mol)	TM Angles
4	597	0 -125 -20 -50 -75 90 170
4	604	0 -125 -20 -50 -35 90 170
3	480	0 175 0 -50 -75 90 170
3	487	0 -125 -45 -75 -35 90 170
3	498	0 -125 -45 -50 -35 90 170
3	512	0 -125 -20 -75 -35 90 170
3	539	0 -125 0 -75 -75 90 170
3	553	0 175 -45 -50 -35 90 170
3	553	0 -125 0 -50 -75 90 170
3	564	0 175 0 -50 -35 90 170
3	576	0 -125 0 -75 -35 90 170
3	590	0 -125 -20 -75 -75 90 170
3	594	0 -125 0 -50 -35 90 170
3	930	0 175 -45 -75 -35 90 170
2	498	0 175 0 -75 -35 90 170
2	522	0 175 -45 -75 -75 90 170
2	524	0 -125 -45 -50 -75 90 170
2	527	0 -125 -45 -75 -75 90 170
2	569	0 175 -20 -50 -35 90 170
2	578	0 175 -45 -50 -75 90 170
2	581	0 175 -20 -75 -75 90 170
2	599	0 175 -20 -50 -75 90 170
2	599	0 175 -20 -75 -35 90 170
1	477	0 175 0 -75 -75 90 170

Calculated energy and number of hydrogen bonds for the combinatorial set of TM angles chosen from the rotational analysis. The structures are ordered by hydrogen bonds and then by energy.

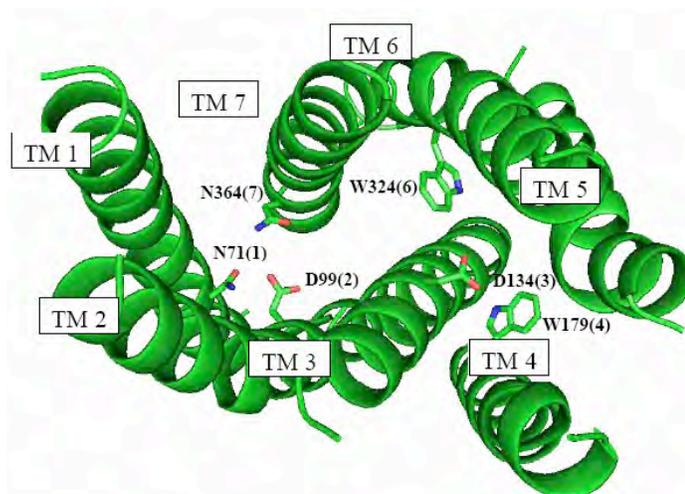


Figure 3. The final predicted structure of apo-5-HT_{2C}. Indicated explicitly are the six amino acids expected to be involved in important interhelical or ligand binding interactions.

Prediction of Ligand Binding Sites

Given the final minimized structure for the apo-protein of 5-HT_{2C}, we then docked various known ligands into the receptor:

- serotonin, the endogenous ligand, an agonist;
- three strongly binding antagonists: ritanserin, metergoline and methiothepin; and
- a series of ten analogs of psilocybin, both agonists and antagonists.

These ligands are shown in Figure 4.

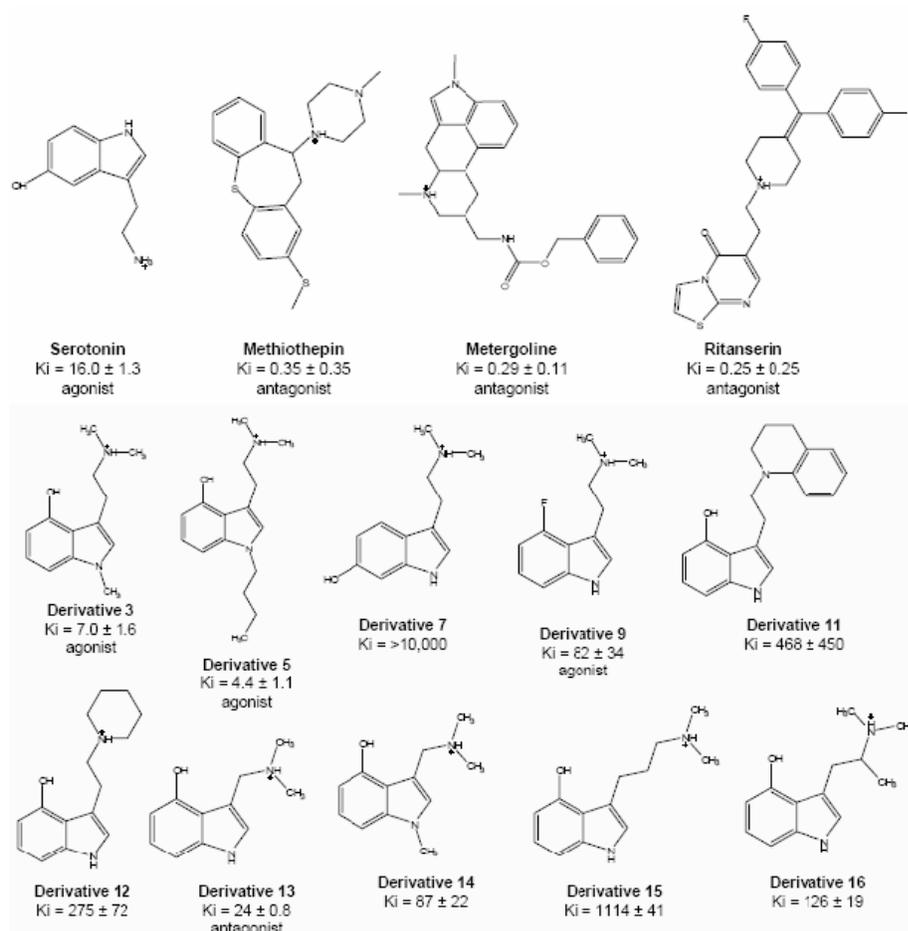


Figure 4. The structures and experimental binding constants to 5-HT_{2C} for serotonin, three strongly binding antagonists, and a set of psilocybin analogs (derivatives 3 to 16). The role of the receptor (whether it is an agonist or antagonist) is given if it has been determined experimentally.

The ligands were built and minimized (Cerius2) using the Dreiding force field with Mulliken charges from quantum mechanics (B3LYP flavor of DFT using the 6-31G** basis set, calculated with Jaguar). We also calculated the pKa values for each ligand using the pKa module of Jaguar to determine if the ligand would be charged, and which atom would become protonated at biological pH 7.4. These pKa values are listed in Table 5. The docking was done with flexible ligands and a rigid protein.

To search for the best binding site for each ligand, we first analyzed the bulky nonpolar residues (leucine, isoleucine, valine, phenylalanine, tryptophan and tyrosine). That is, each was mutated to alanine. Then we used the negative of the molecular surface of the analyzed protein to define the regions in which to sample different ligand conformation. These void regions were mapped with a set of spheres. Then, we defined 36 boxes, each a cube with 10 Å sides, grown from the center of mass of the spheres and covering the entire receptor. Each ligand was docked (using Dock 4.0) into all 36 boxes and ranked by interaction energy and buried surface. Then, we examined the boxes with a combination of the strongest bonding and buried surface, and selected boxes with a continuous range of spheres for subsequent docking with MSCDock. As an example, the sphere set that the antagonists were docked into is shown in Figure 5.

We then used MSCDock to dock each ligand into the selected boxes using the dock with diversity method. This method uses Monte Carlo search techniques to find sites that pass bump tests and groups these into families for which all family heads are at least 0.6 Å CRMS different from the others. This is

carried out until the number of families is saturated. Then we select the 50 families with the best combination of energy and buried surface and continue generating docking configurations until there is an average of six children in each family. Then we calculated the energy for each child of each of these families, and associate the energy with the family head.

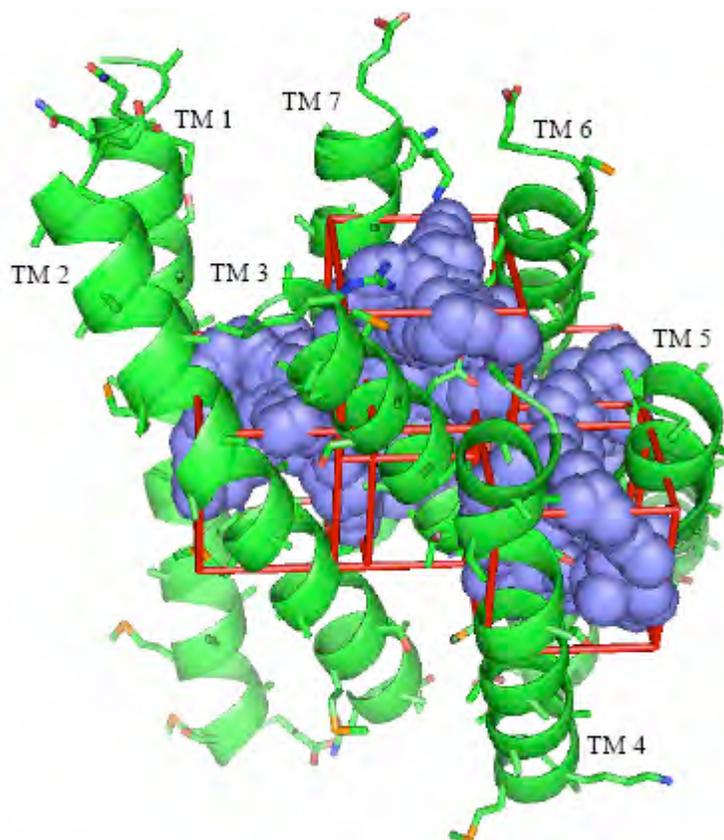


Figure 5. The boxes containing the spheres into which the antagonists ritanserine, methiothepin and metergoline were docked. All of the bulky, nonpolar residues have been alanized.

Next for each of these 50 ligand configurations, we dealnize the bulky nonpolar residues back to their original form and reassign all side chains within 5 Å of the ligand with SCREAM. We minimize (conjugate gradients to 0.2

kcal/molÅ RMS force) the ligand and residues within 5 Å of the ligand with the rest of the protein fixed. Then we minimize the entire complex to 0.2 kcal/molÅ RMS force.

The top ten structures by total energy of the ligand-protein complex were selected as the best docked structures. From these top ten structures, we selected the one with the best contacts between the functional groups of the ligand and the residues in the binding site. Then we used SCREAM to reassign the side chains for the residues interacting most favorably with the functional groups of the ligand to obtain their optimal orientation with respect to the ligand. The ligand and residues within 5 Å were put through one quench-anneal cycle (50 K to 600 K and back over 11.5 ps), selecting the structure with the lowest potential energy structure. This was followed by minimizing the energy of the entire ligand-protein complex.

Serotonin, ritanserin, metergoline and methiothepin were all docked following this procedure. For the 10 SAR ligands, we chose derivative 3 as the template and docked it using the above procedure. Then, we generated a sphere set fitted to a region within 0.8 Å radius around the best docked derivative 3 molecule. Then the entire docking procedure was carried out over this smaller region for the other nine SAR ligands, keeping the bulky residues in place (not mutated to alanine).

Molecular Dynamics Simulation

A simulation was performed on serotonin bound in 5-HT_{2C} in order to determine the stability of the predicted bound structure. NAMD 2.5^[35] was used to perform

5 ns of full solvated lipid bilayer simulation. First, loops were added with a Monte Carlo loop builder. Only the first and last ten residues of the ic3 loop were added. The loops were built so that the disulfide bridge between the two cysteines in ec2 and the top of TM3 was preserved. Ten residues each of the N-terminus and C-terminus were added. Then, the helices and side chains were frozen while loops were put through a quench-anneal cycle. Next, palmitoyloleoyl-phosphatidylethanolamine (POPE) lipid molecules were added, along with a layer of water above and below. Chloride ions were added to neutralize the charge of the system. The membrane and water molecules were minimized with the protein fixed, and then equilibrated for 500 ps in an NPT simulation. Then the loops were minimized along with the membrane and water molecules, and equilibrated for 500 ps in an NPT simulation. Finally, the entire system was minimized, and then 5 ns of NPT simulation was run. All NPT simulations were run using Langevin dynamics with a damping coefficient of 1 ps^{-1} and a bath temperature of 310 K. The pressure was kept constant by Nosé-Hoover Langevin piston pressure control, with a target pressure of 1 atm and barostat oscillation and damping times of 200 fs. The stepsize was 1 fs, with periodic boundary conditions applied. Particle mesh Ewald was used for electrostatic calculations, with a nonbond cutoff of 12.0 Å and the nonbonded interactions of all 1-3 pairs included. The energy and volume of the system equilibrated quickly, as shown in Figure 6.

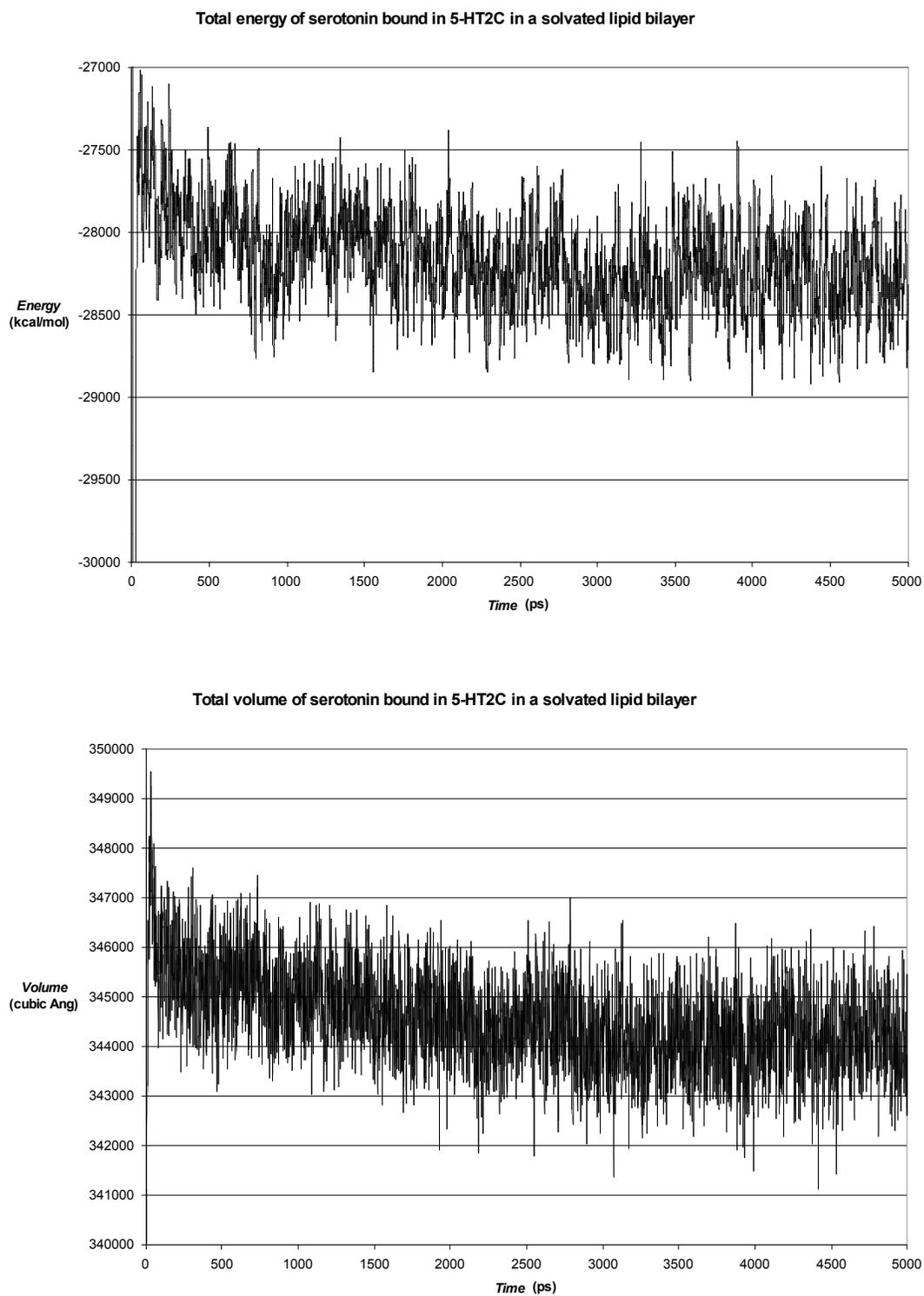


Figure 6. The energy and volume of serotonin bound in 5-HT_{2C} in a solvated lipid bilayer during the 5 ns simulation.

Binding Energy Calculations

For all the docked structures, the binding energies were calculated as

$$E_{\text{bind}} = E_{\text{comp,vac}} - E_{\text{prot,vac}} - E_{\text{lig,vac}} - E_{\text{lig,solv}}$$

Since the binding energy is the change in energy upon binding, more negative binding energies indicate stronger binding. The energies of the separate protein and ligand were not reminimized. That is, they were taken directly from the bound complex (to obtain the snap binding energy). This assumes that there is little change in energy of the separated ligand and protein upon minimization. This also does not take the solvation of the complex or the protein into consideration. However, for small molecules binding in the core of the protein like those we are working with, the solvation energy of the bound complex should not differ greatly from that of the unbound protein. The solvation energy of the ligand was calculated using the Poisson-Boltzmann continuum solvent approximation (dielectric constant of 80 and solvent radius of 1.4 Å) using the Delphi program. The vacuum energies were calculated with MPSim, as in all of the structure prediction steps, so the solute dielectric was 2.5.

Next, we neutralized all charged residues of the system and the ligand by transferring the hydrogen of each salt bridge from the acceptor back to the donor and by adding a proton to each exposed Asp or Glu and removing one from each Lys and Arg. We used a modified Dreiding forcefield that included special hydrogen bonding parameters chosen to reproduce the binding for dimers of analogous residues found from QM. Each full ligand-protein complex was then reminimized and the binding energies for the neutral forcefield were calculated

according to the same equation as for the charged case, but with the addition of a pKa penalty term. This is the energy change required to neutralize the ligand or residue that prefers to be charged in water solvent. For a positively charged ligand this is equal to $1.4 * (pKa - 7.4)$ kcal/mol.

This modified neutral Dreiding forcefield was also used to calculate the contributions of individual residues to binding. As in all of the other Dreiding calculations, the dielectric constant was 2.5, with van der Waals interactions calculated by a 6-12 Lennard-Jones potential. The nonbond interactions were calculated between the bound ligand and all residues within 5 Å of the ligand.

The neutral residue scheme is an improvement over the charged residue scheme for the binding energy calculations because it decreases the large variations between complexes caused by exaggerated long-range Coulombic interactions between charged groups. These exaggerated interactions are due to the fact that the charges are fixed in molecular mechanics, so charge screening is not present to damp the long-range interactions. Additionally, in the hydrophobic membrane environment, many of the residues may already be neutral. The neutralization procedure is carried out for the binding energy calculations and not the docking procedure because the large Coulombic interactions are important to ensure that binding modes with a salt bridge are selected.

Results and Discussion

Serotonin Binding

Serotonin is the endogenous agonist to 5-HT_{2C} with a binding constant of $K_i = 16.0$ nM.^[36] The predicted binding site for serotonin to our predicted structure for 5-HT_{2C} is shown in Figure 7 and the cavity analysis is in Table 3. The most important contacts are

- D134(3) forms a salt bridge with the protonated primary amine site (8.7 kcal/mol), which makes excellent sense. Indeed, experiments suggest that D134(3) creates a salt bridge with the protonated nitrogen of biogenic amines.^[34]
- The protonated primary amine also forms a hydrogen bond with S138(3) (5.7 kcal/mol).
- The aromatic indole group has very good van der Waals interactions with F223(5) (3.0 kcal/mol) and F328(6) (4.9 kcal/mol). The ring is stacked between the two phenylalanines.
- The polar NH of the indole forms a hydrogen bond with S219(5) (6.5 kcal/mol).
- The OH substituent of the indole forms hydrogen bonds both with S141(3) (3.3 kcal/mol) and W324(6) (2.0 kcal/mol).
- There is good van der Waals interaction between the indole and I332(6) (2.0 kcal/mol).

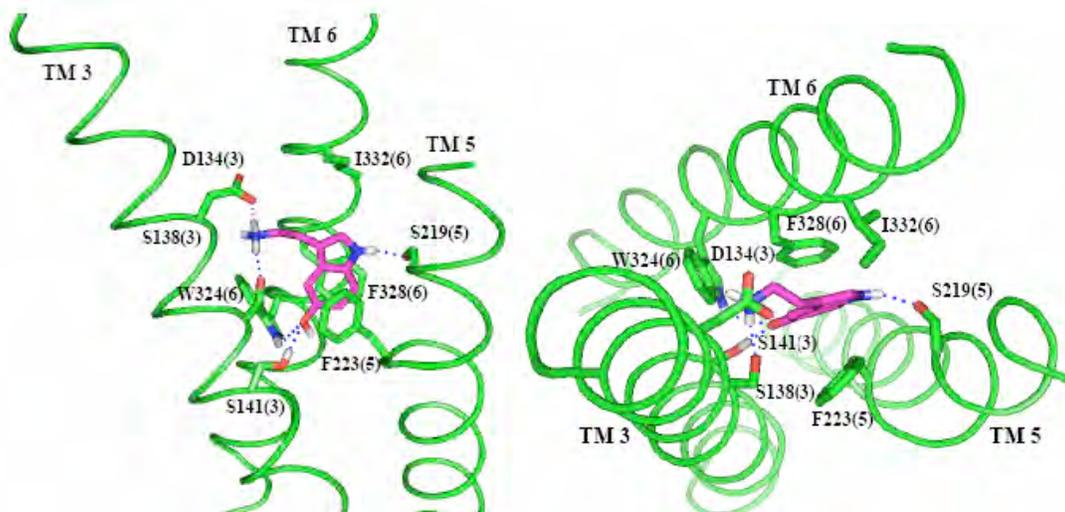


Figure 7. The predicted structure (side and top views) of serotonin bound to 5-HT_{2C}. TMs 1, 2, 4 and 7 are not shown since they do not interact directly with the ligand. The residues shown are those within 5.0 Å of serotonin that have more than 2 kcal of favorable interaction with the ligand. The dotted lines indicate hydrogen bonds with serotonin.

Table 3. Nonbond interaction energies (kcal/mol) between each ligand and individual residues within 5.0 Å of the ligand

	D134	F328	S219	F223	S138	I332	W324	V215	P190	S141
TM helix no.	3	6	5	5	3	6	6	5	4	3
Ritanserin	-9.6	-0.4	-1.5	-1.6	-0.3	-4.1	-8.5	-3.2	-1.9	0.3
Metergoline	-10.9	-0.5	-1.8	-0.9	-7.4	-1.6	-0.2	-5.9	-4.7	-0.2
Methiothepin	-12.2	-1.5	-6.1	-1.8	-1.3	-0.8	-0.3	-2.3	-3.9	0.0
Deriv 5	-9.8	-6.9	-4.9	-4.3	-2.7	-2.9	-2.4	-0.5	-0.2	0.4
Deriv 3	-9.0	-6.1	-4.8	-6.3	-2.4	-2.3	-1.4	-0.7	-0.2	-1.0
Serotonin	-8.7	-4.9	-6.5	-3.0	-5.7	-2.1	-2.0	-0.3	-0.1	-3.3
Deriv 13	-9.0	-6.6	-5.0	-3.2	-1.2	-1.7	-1.9	-0.3	-0.1	-0.8
Deriv 9	-9.0	-5.9	-1.3	-4.8	-2.1	-3.1	-1.7	-0.3	-0.1	-0.2
Deriv 14	-8.4	-6.4	-4.1	-5.5	-1.3	-1.2	-1.0	-0.5	-0.1	-0.2
Deriv 16	-11.2	-4.8	-4.6	-5.8	-1.2	-2.7	-0.9	-0.6	-0.1	-0.1
Deriv 12	-12.3	-6.1	-5.3	-3.2	-1.5	-3.2	-1.8	-1.6	-0.9	-0.7
Deriv 11	-2.7	-6.1	-5.1	-4.8	-1.6	-1.7	-1.7	-2.8	-1.6	-0.6
Deriv 15	-9.5	-4.4	-0.8	-4.0	-1.8	-2.0	-2.2	-0.1	-0.1	-6.1
Deriv 7	-8.2	-6.3	-1.9	-4.2	-1.6	-2.7	-1.4	-0.7	-0.2	-0.2
Average	-9.3	-4.8	-3.8	-3.8	-2.3	-2.3	-2.0	-1.4	-1.0	-0.9

Here we list only the residues that have more than 2 kcal of favorable energy with at least two ligands. The ligands are ordered by decreasing experimental binding strength (see Table 5).

Overall, the bound structure for serotonin looks reasonable, with each polar functional group of serotonin making favorable interactions with a polar or charged residue in the binding pocket and with excellent aromatic interactions.

Bound Serotonin Dynamics

In order to determine the stability of this bound structure, we carried out a 5 ns simulation of this complex in an explicit lipid bilayer. This simulation shows the predicted structure and binding site to be stable for 5 ns. The structure after 5 ns is shown in Figure 8. Some of the hydrogen bond contacts made with serotonin do change throughout the simulation, but the ligand stays in the same binding site, just moving a little downward toward the intracellular region. The hydrogen bond with S219(5) and the van der Waals interactions with I332(6) are lost as serotonin moves downward. The hydrogen bond with S141(3) is lost as the oxygen of the OH substituent of serotonin forms a very strong hydrogen bond with W324 and the hydrogen of the OH group forms a hydrogen bond with the backbone oxygen of S138(3). Both D134(3) and S138(3) maintain their strong hydrogen bonds with the protonated amine group of serotonin. Additionally, both F223(5) and F328(6) preserve their strong van der Waals interactions with the indole serotonin throughout the simulation. The distances between serotonin and key residues during the simulation are shown in Figure 9. Waters enter the binding site and accumulate around the salt bridge between D134(3) and serotonin. There were no additional restraints placed on the helices to prevent them from unraveling, so the C-terminus of TM1 unraveled by two residues, the N-terminus of TM2 by one residue, the N-terminus of TM3 by one residue, and

the N-terminus of TM7 by one residue. The TMs 1-2-7 hydrogen bonding network is retained, although the helices do translate with respect to each other during dynamics.

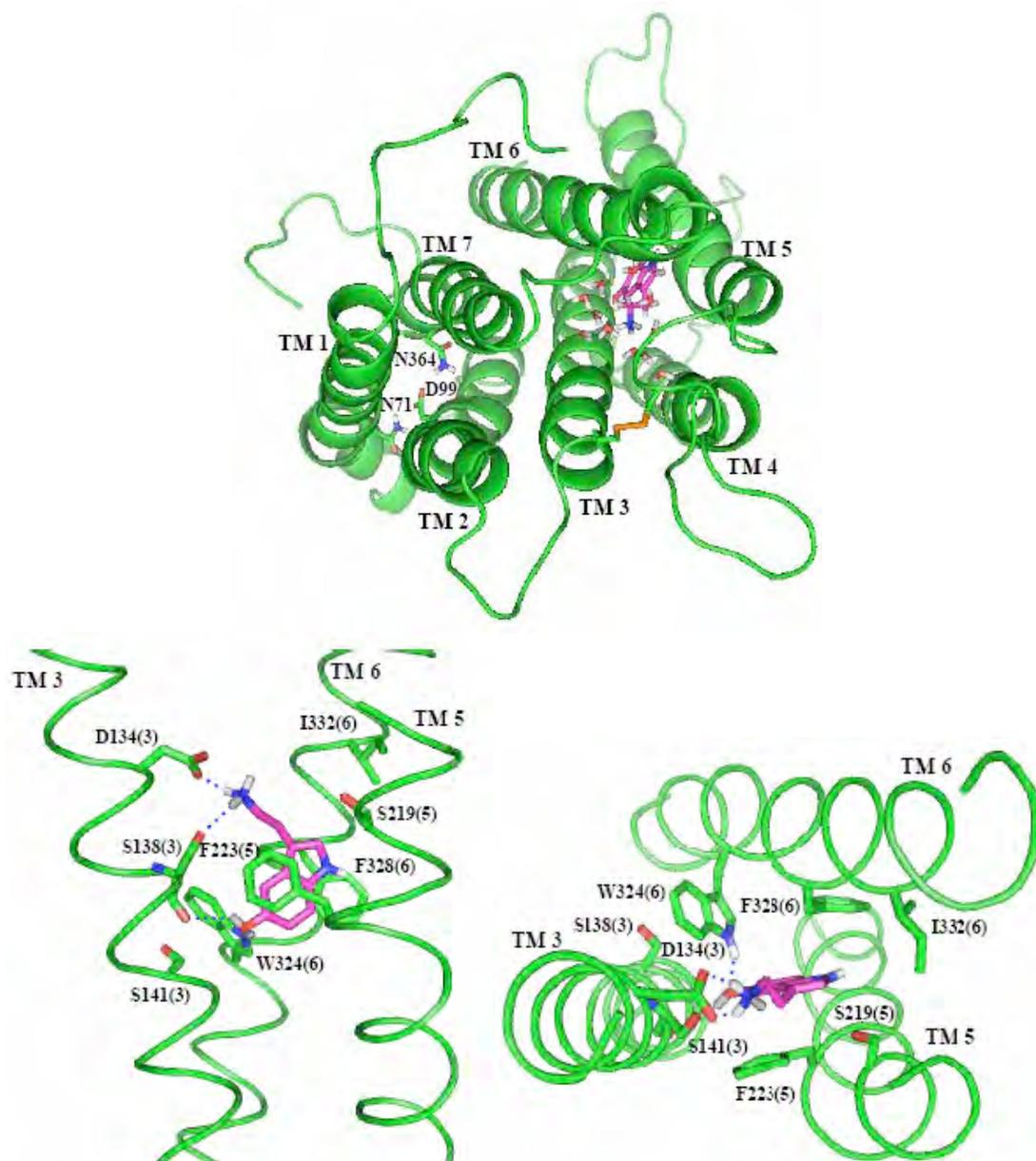
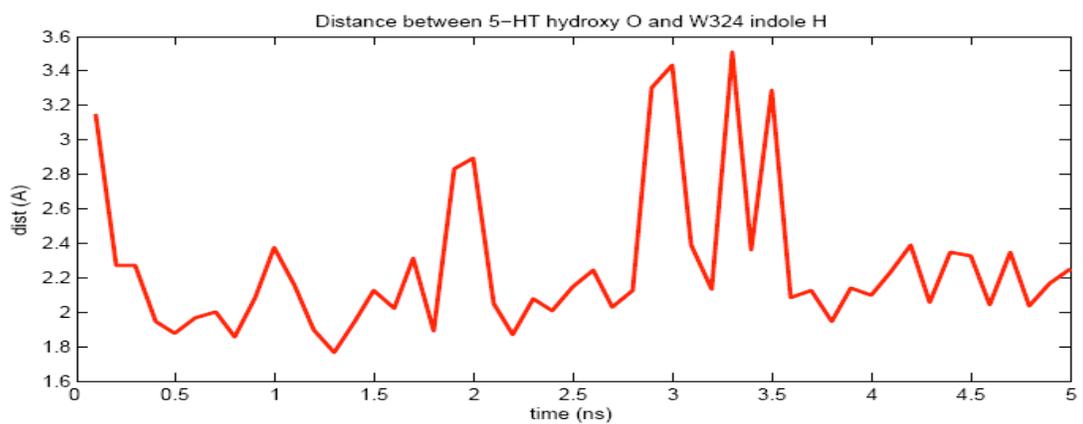
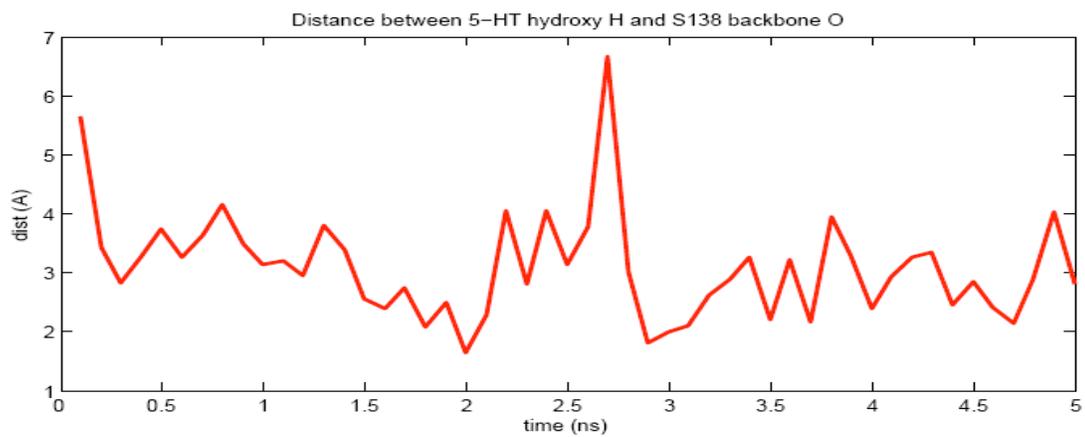
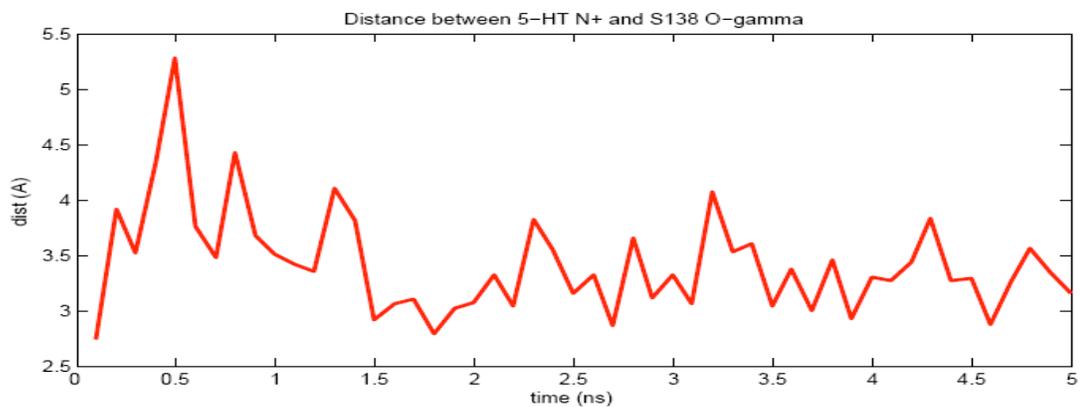
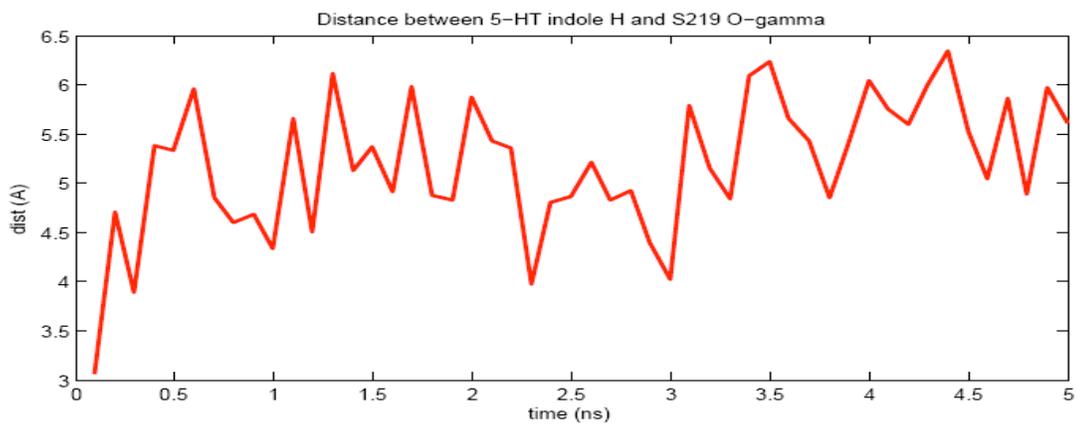
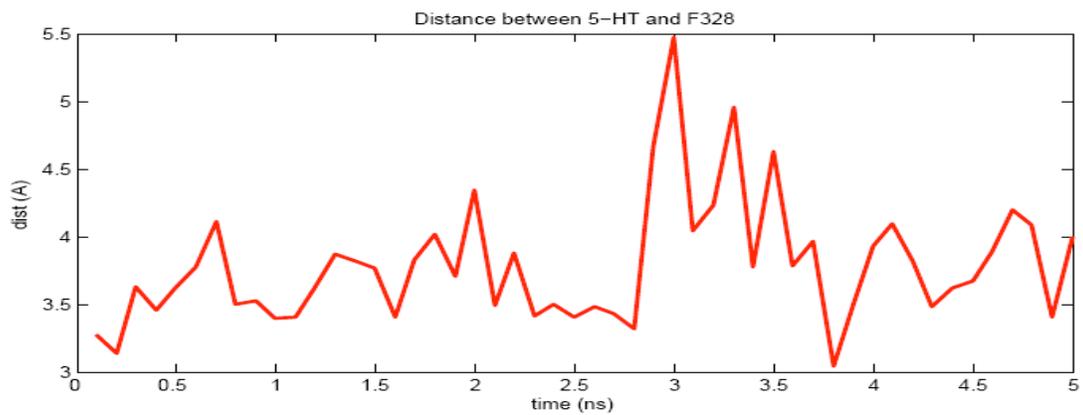
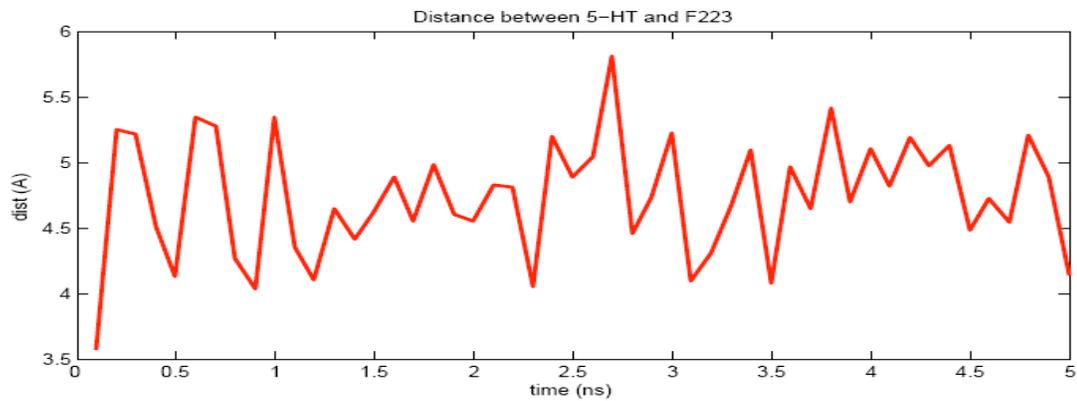


Figure 8. Serotonin bound in 5-HT_{2C} after 5 ns of dynamics. The top shows the TMs 1-2-7 hydrogen bonding network, as well as the disulfide bond between the two cysteines in TM3 and ec2. The water molecules shown are those within 5.0 Å of serotonin. The bottom shows the binding pocket, with the important residues from the structure before minimization or dynamics. The dotted lines indicate hydrogen bonds with serotonin.





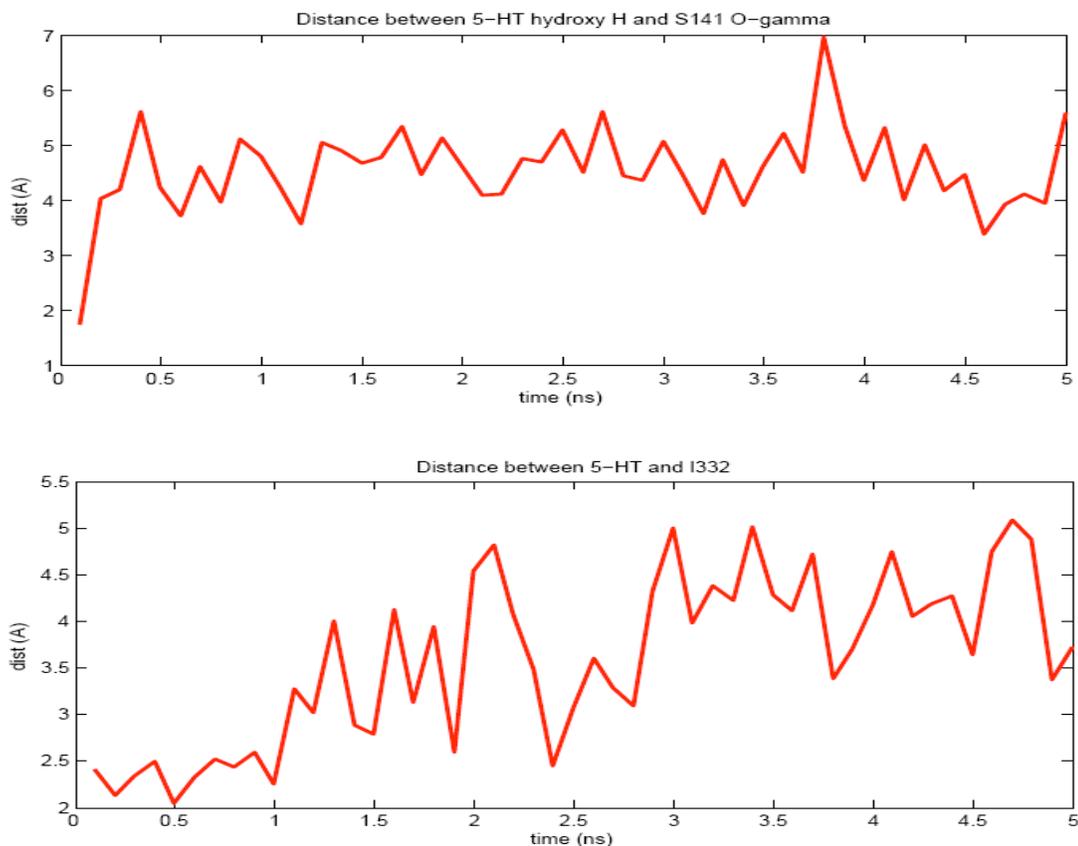


Figure 9. The calculated distances between serotonin and residues predicted to be important in binding at 50 evenly spaced snapshots of the 5 ns simulation.

Comparison to Mutagenesis Experiments

Since no experimentally determined structures for any serotonin receptors are available, experimental validation must come from additional mutagenesis studies, in which residues in the active site are mutated to various amino acids and the effect on ligand binding determined. Many experimental studies have been performed on 5-HT₂ receptors, with the majority of the mutation experiments focused on 5-HT_{2A}. Although there are differences between the receptors, we find 70% sequence identity and 88% sequence similarity between

5-HT_{2A} and 5-HT_{2C} in the predicted TM regions, so it can be useful to compare our predicted structure with experimental findings on 5-HT_{2C}.

- D155(3) in 5-HT_{2A}, which corresponds to D134(3) in 5-HT_{2C} was found to anchor the terminal amine moiety of serotonin,^[37] as we predict in 5-HT_{2C}.
- An alanine substitution of S159(3) in 5-HT_{2A}, corresponding to S138(3) in 5-HT_{2C}, was found to decrease serotonin binding by 18-fold,^[38] leading to the proposal that S159(3) forms a hydrogen bond with the protonated amine group of serotonin, just as we predict for 5-HT_{2C}.
- S239(5) in 5-HT_{2A}, which is an alanine in 5-HT_{2C} that is one turn away from S219(5), is proposed to form a hydrogen bond to the polar NH of the indole in serotonin.^[39] We predict that in 5-HT_{2C}, the NH of the indole in serotonin forms a hydrogen bond with S219(5), but we find this hydrogen bond to disappear after 5 ns of dynamics.
- F243(5) and F244(5) in 5-HT_{2A} are found to have significant interactions with serotonin, with both predicted to point in toward the binding pocket.^[40] F243(5) in 5-HT_{2A} corresponds to F223(5) in 5-HT_{2C}, which we predict to have very good van der Waals interactions with serotonin, but we predict F224(5) not to interact strongly with serotonin in 5-HT_{2C}.
- Mutation of F340(6) to leucine in 5-HT_{2C} significantly decreases serotonin binding, but mutation of F339(6) to leucine did not affect binding.^[41] We predict the same effect in 5-HT_{2C} with serotonin making very good van der Waals interactions with F328(6), but not significantly interacting with F327(6).

- Mutation of W336(6) to alanine was found to cause an almost 1000-fold decrease in serotonin binding in 5-HT_{2A}.^[42] We predict the corresponding residue in 5-HT_{2C}, W324(6), to have good van der Waals and hydrogen bonding interactions with serotonin.

Our predicted serotonin binding site for 5-HT_{2C} has many similarities to experimentally determined features of the 5-HT_{2A} binding site. A homology model of 5-HT_{2C} based on the crystal structure of bovine rhodopsin^[3], produced by the WHAT IF program^[43] does not agree with these mutagenesis experiments as well as our predicted structure. TM5 has S219(5) pointed away from the binding pocket, towards TM6. Additionally, D134(3) is pointed towards the 1-2-3-7 pocket instead of the 3-5-6-7 pocket. Thus, serotonin would not bind in the 3-5-6-7 pocket, as expected for biogenic amine receptors.^[44]

The recently determined experimental structure of the beta(2)-adrenergic receptor^[5] has more similarities to our predicted structure for 5-HT_{2C} than does the structure of bovine rhodopsin, but there are still some key differences in the predicted binding site. The TM5 serine in the beta2 structure that corresponds to S219(5) in 5-HT_{2C} is pointing in toward TM3. However, the well conserved TM3 aspartic acid in beta2 thought to anchor the protonated amine of bound biogenic amines is pointing towards the 1-2-3-7 pocket, as seen in bovine rhodopsin. Thus, in a homology model of 5-HT_{2C} based on the crystal structure of the beta(2)-adrenergic receptor, serotonin cannot make good contacts with both D134(3) and S219(5) as we predict in our structure.

Based on our predicted bound structure and its dynamics, additional experiments that could help validate our predicted binding site are D134A, S138A, F328A, F223A and W324, all of which would lead to significant decreases in serotonin binding affinity. Although the simulation shows the hydrogen bond between serotonin and S219(5) to be absent after 5 ns, it may reappear after a longer simulation, so S219A could also lead to a decrease in serotonin binding affinity.

Ritanserine Antagonist Binding

Ritanserine is an antagonist to 5-HT_{2C} with a binding constant of $K_i = 0.25$ nM.^[45]

It is used for the treatment of many neurological disorders. The predicted binding site for ritanserine to our predicted structure for 5-HT_{2C} is shown in Figure 10 and the cavity analysis is in Table 3. The most important contacts are as follows.

- The protonated nitrogen of the piperidine forms a salt bridge with D134(3) (9.6 kcal/mol).
- The oxygen of the pyrimidine makes a hydrogen bond with W324(6) (8.5 kcal/mol).
- The aromatic thiazolo-pyrimidine group has strong van der Waals interactions with F137(3) (4.9 kcal) and N331(6) (2.2 kcal).
- The fluorenyl groups have strong van der Waals interactions with I332(6) (4.1 kcal/mol) and V215(5) (3.5 kcal/mol).

Note that the sulfur of the thiazole does not make strong interactions with polar or charged residues. This suggests that improved binding might be obtained by

modifying I363(7) or V359(7) to polar residues able to hydrogen bond with the sulfur. Additionally, the fluorines of the antagonists do not interact strongly with polar or charged residues. They may be interacting with the charged head groups of the lipid molecules, water molecules that enter the binding site, or charged or polar residues in the loops.

Metergoline Antagonist Binding

Metergoline is an antagonist to 5-HT_{2C} with a binding constant of $K_i = 0.29$ nM.^[36]

Metergoline is used as an analgesic in migraine headaches. The predicted binding site for metergoline to our predicted structure for 5-HT_{2C} is shown in Figure 10 and the cavity analysis is in Table 3. The most important contacts are as follows.

- D134(3) forms a salt bridge with the protonated nitrogen of the ergoline (10.9 kcal/mol).
- Both the N-H group of the amide and the carboxyl oxygen of the ester makes hydrogen bonds with S138(3) (7.4 kcal/mol).
- The aromatic ergoline group has strong van der Waals interactions with V215(5) (5.9 kcal/mol), P190(4) (4.7 kcal/mol), F214(5) (2.5 kcal/mol), and I189(4) (2.5 kcal/mol).
- The phenyl group has good van der Waals interactions with I142(3) (2.9 kcal/mol) and I182(4) (2.0 kcal/mol).

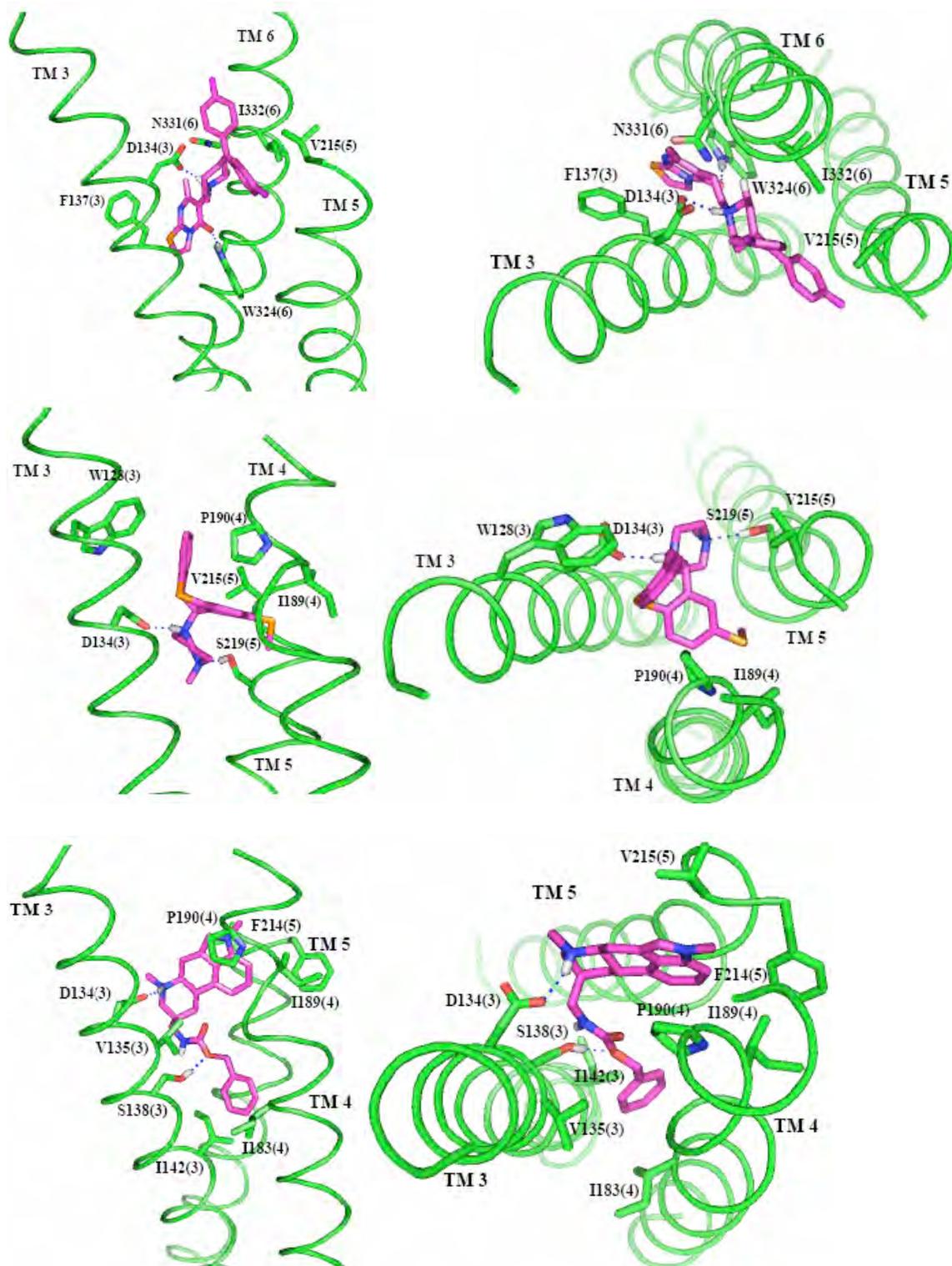


Figure 10. Side and top views of ritanserin (top), methiothepin (middle) and metergoline (bottom) bound in 5-HT_{2C}. The TMs not shown do not directly interact with the ligand. The residues shown are those within 5.0 Å of the ligand that have more than 2 kcal of favorable interaction with the ligand. The hydrogen bonds formed between the protein and the ligand are indicated by dotted lines.

Methiothepin Antagonist Binding

Methiothepin is an antagonist to 5-HT_{2C} with a binding constant of $K_i = 0.35$ nM.^[45] It is used as an antipsychotic. (S)-methiothepin was docked for this study. The predicted binding site for methiothepin to our predicted structure for 5-HT_{2C} is shown in Figure 10 and the cavity analysis is in Table 3. The most important contacts are as follows.

- D134 interacts with the protonated N at the center of methiothepin (12.2 kcal/mol).
- The piperazine NH forms a hydrogen bond with S219(5) (6.1 kcal/mol).
- The two benzene rings have good van der Waals interactions with P190(4) (-3.9 kcal/mol), W128(3) (2.8 kcal/mol) and V225(5) (2.3 kcal/mol).
- I189(4) has good van der Waals interactions with the thiol ether (2.5 kcal/mol).

Note that neither sulfur in methiothepin makes strong interactions with polar or charged residues. This suggests that improved binding might be obtained by modifying I131(3), V135(3), G218(5) or I189(4) to polar residues able to form hydrogen bonds with the sulfurs. If S186(4) was modified to an asparagine or a threonine, it may be long enough to form a hydrogen bond with the thiol ester sulfur.

SAR for Psilocybin Analogs

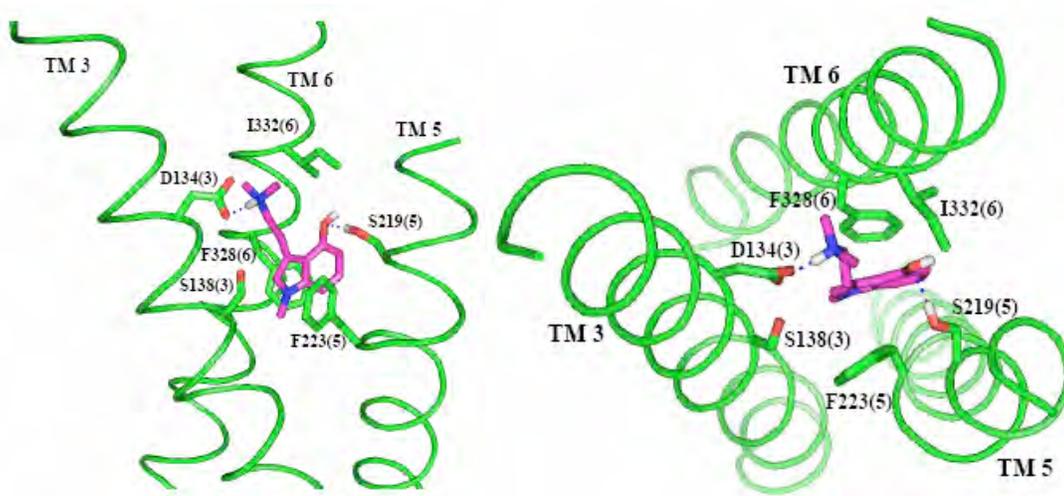
Since there is so little mutation data available for 5-HT_{2C}, we will validate our predicted GPCR structure and binding sites by comparing calculated binding energies with experimental binding affinities. Here we consider a series of

psilocybin analogs in which small structural changes were found to result in large changes in binding affinity.^[46] The binding modes for the SAR ligands are shown in Figure 11. The strength of interaction for each ligand with important residues in the binding cavities is given in Table 3.

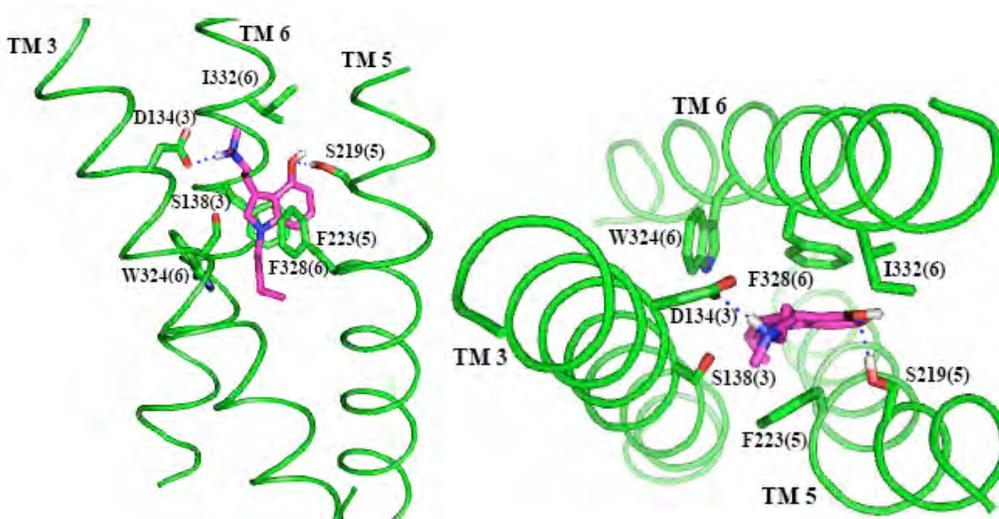
The binding site of the SAR ligands can explain most of the experimental binding data. All of the ligands form a salt bridge with D134 except derivative 11, which pKa calculations show to be neutral in solution. Derivatives 3, 5, 11, 12, 13, 14 and 16 form strong hydrogen bonds between the hydroxyl group of the phenol and S219. The position of the hydroxyl group in derivative 7 prevents any hydrogen bonding with S219, which in part leads to its bad binding. The fluorine in derivative 9 that takes the place of the hydroxyl group does not hydrogen bond with S219, but still has favorable interactions with S219, so the binding to 5-HT_{2C} is not as strong as for derivative 3 or 5, but is still fairly strong. In derivatives 13 and 14, the carbon chain is shorter than in serotonin, but this ligand can still form hydrogen bonds with S219. However, they do not bind as strongly as derivatives 3 and 5, which have the same length carbon chain as serotonin. A cavity analysis shows that there are decreased interactions with I332 and S138 in derivatives 13 and 14 compared to derivatives 3 and 5. Derivative 16 also has decreased interaction with S138, caused by the extra methyl group branching from the carbon chain, causing part of the decrease in binding affinity compared to derivative 3. In derivative 15, the length of the carbon chain connecting the protonated amine group and the aromatic rings does not allow for hydrogen bonding with S219, but instead the N-H group of the indole ring forms a hydrogen

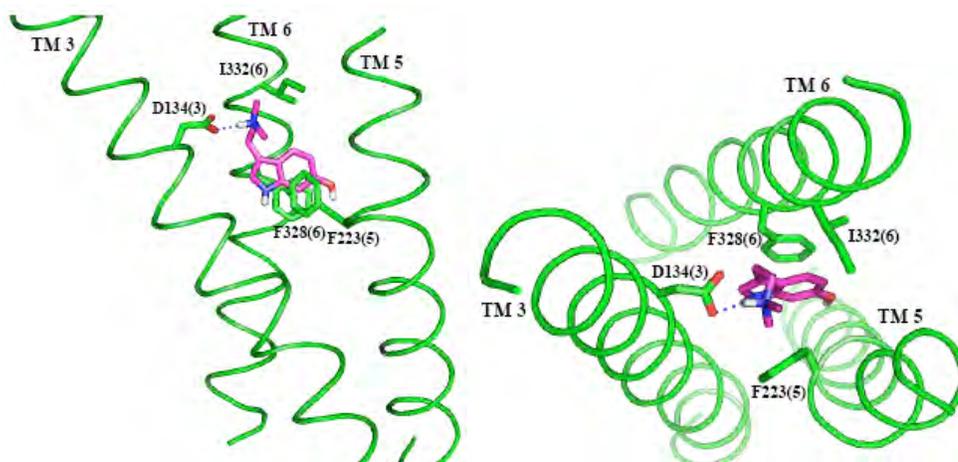
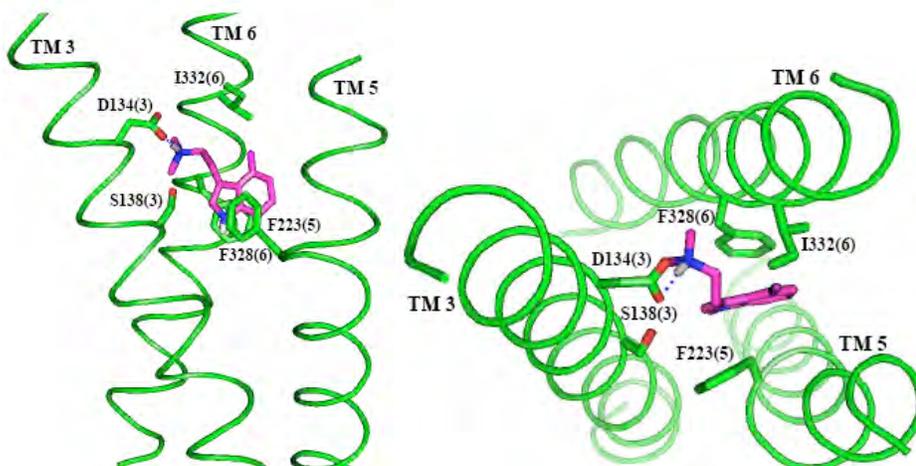
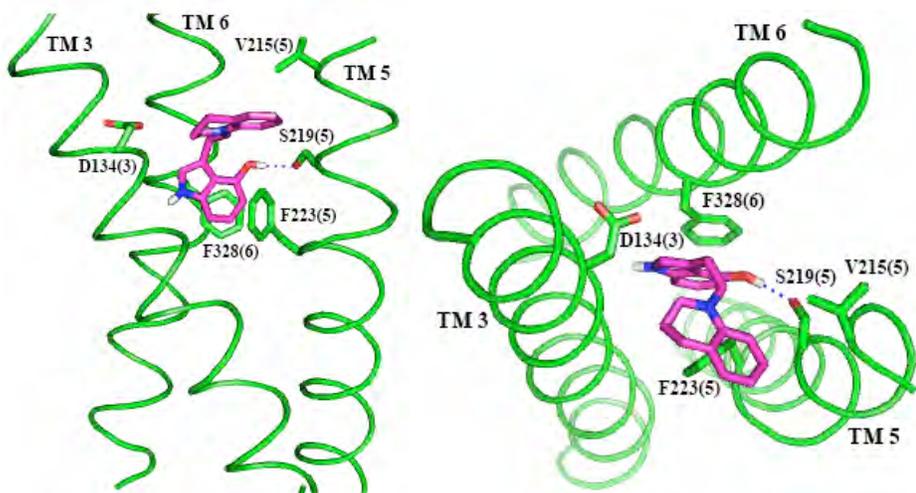
bond to S141. However, experiments show that derivative 15 does not bind well to 5-HT_{2C}, so this hydrogen bond with S141 may not in fact form. S141 could be interacting instead with the backbone of TM3 or forming an interhelical hydrogen bond, maybe with W324, as it does in the dynamics for serotonin bound in 5-HT_{2C}. Similarly, the predicted binding site of derivative 12 cannot explain the experimentally observed low binding affinity.

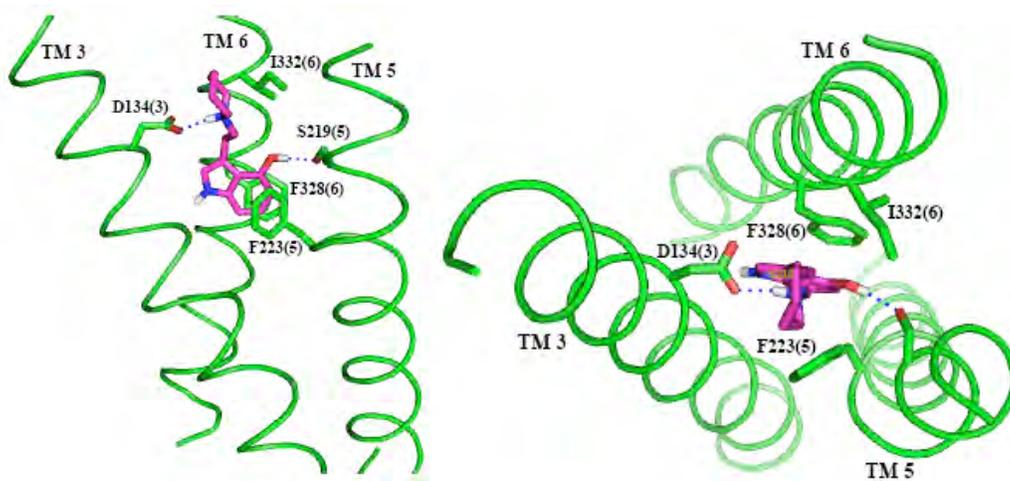
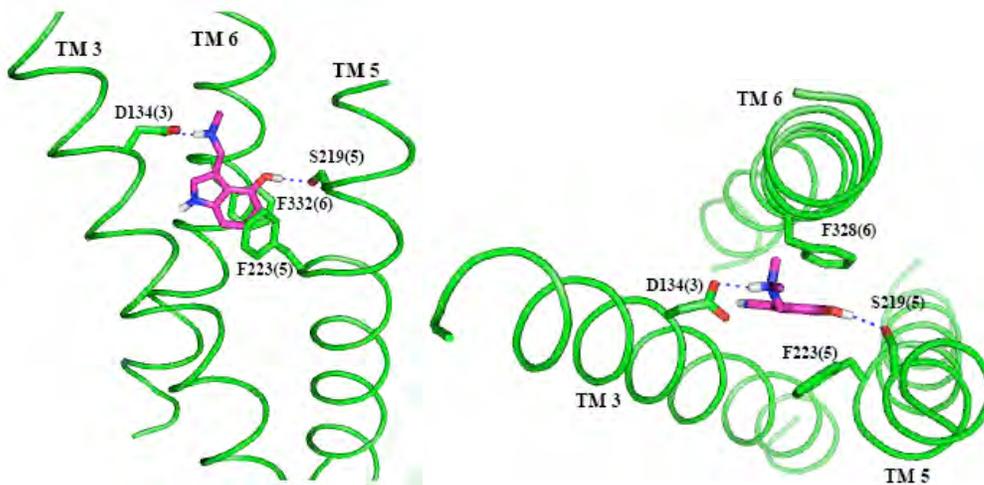
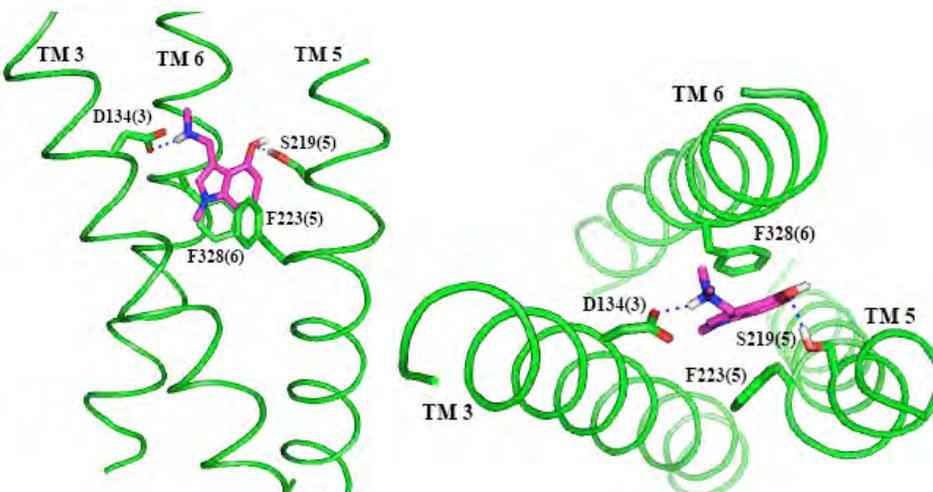
Derivative 3



Derivative 5



Derivative 7**Derivative 9****Derivative 11**

Derivative 12**Derivative 13****Derivative 14**

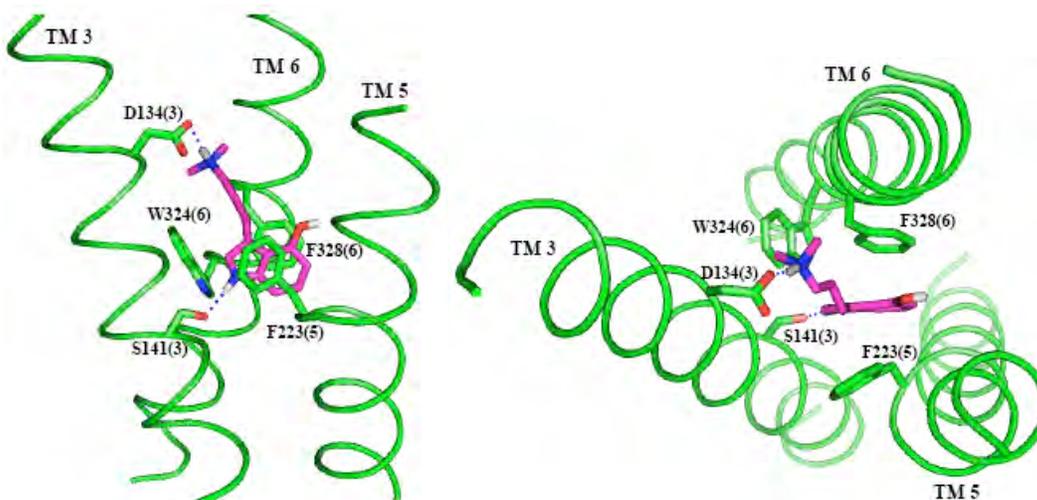
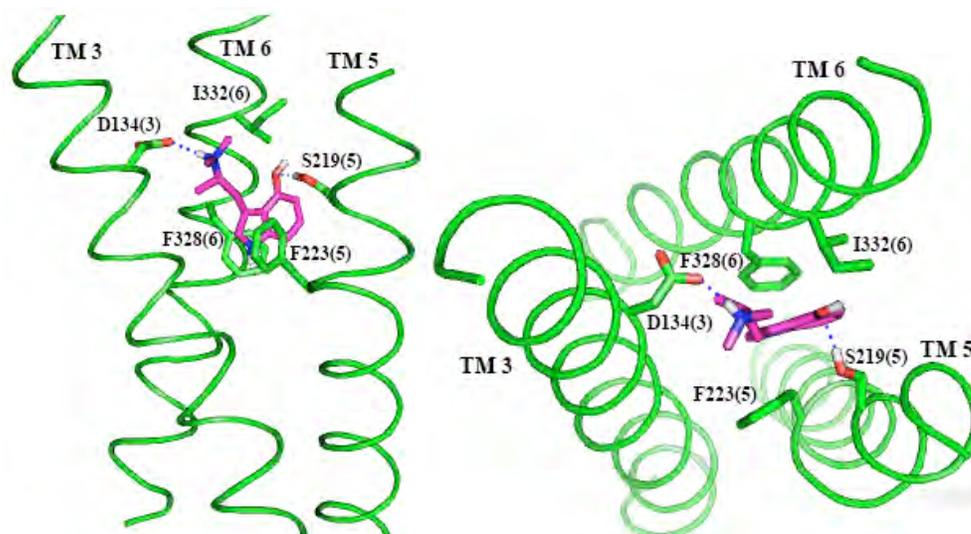
Derivative 15**Derivative 16**

Figure 11. Side and top view of the SAR ligands. The TMs not shown do not directly interact with the ligand. The residues shown are those within 5.0 Å of the ligand that have more than 2 kcal of favorable interaction with the ligand. The hydrogen bonds distances to the ligands, indicated by the dotted lines, are all between 1.85 and 2.00 Å.

First, consider the trend in total binding energy. We predict derivative 5 to be the best and derivative 7 to be worst, which agrees with experiment. Based on

these interactions, proposed mutations that are predicted to decrease significantly the binding affinity of each ligand are listed in Table 4.

- Deriv 3 → 5: This involves replacing the N-Me with N-butyl. Experimentally the binding constant improves from 7.0 to 4.4 nM and we calculate 15% improvement. Our binding site has the butyl in hydrophobic contact with I142(3), I145(3), L227(5) and W324(6) leading to a predicted 4.0 kcal/mol improvement in binding energy over derivative 3. Overall the cavity analysis is quite similar.
- Deriv 3 → 14: This involves removing one CH₂ from the linker connect the indole with the tertiary amine. Experimentally the binding constant changes from 7.0 to 87 nM and we calculate a 13% weakening in binding. Since the tertiary amine needs to keep close to D134(3), the indole is moved slightly away from F223(5), leading to slightly worse binding to both F223(5) and D134(3). Additionally, derivative 3 has better van der Waals interactions with S138(3) and I332(6) than derivative 14. Overall, the cavity analysis is quite similar.
- Deriv 14 → 13: This involves replacing the N-Me of the indole with N-H. Experimentally the binding constant changes from 87 to 24 nM and we calculate a 2% improvement in binding. The difference appears to be from the ligand solvation energy, which is more negative for derivative 14 than derivative 13. Overall, the cavity analysis is quite similar.
- Deriv 13 → 16: This involves adding an extra CHMe to the linker connecting the indole with the tertiary amine. Experimentally the binding

constant changes from 24 to 126 nM but we calculate an 8% improvement in binding. Based on the 3→14 result we would expect that adding a CH₂ would have improved binding (this compound was not studied experimentally). However, we observe that the OH is farther up in the pocket in derivative 16, creating better van der Waals interactions with I332(6) compared to derivative 13. Additionally, derivative 16 seems to have better van der Waals and coulombic interactions with D134(3) than derivative 13.

- Deriv 13 → 15: This involves adding two extra CH₂ groups to the linker connecting the indole with the tertiary amine. Experimentally the binding constant changes from 24 to 1114 nM but we calculate a 13% improvement in binding. The extra CH₂ groups on the linker make it so the OH on derivative 15 cannot form a hydrogen bond with S219(5), but instead the NH makes a hydrogen bond with S141(3). However, since derivative 15 does not bind strongly in experiment, this hydrogen bond may not actually exist, and S141(3) may instead interact strongly with W324(6), as seen in the dynamics of serotonin bound to 5-HT_{2C}.

Table 4. Mutations for each ligand that are predicted to decrease binding affinity

	Proposed Mutations to Decrease Binding Affinity
Ritanserin	D134A, W324A, F137A, I332T, V215T
Metergoline	D134A, S138A, V215T, P190S
Methiothepin	D134A, S219A, P190S
Deriv 5	D134A, F328A, S219A, F223A
Deriv 3	D134A, F223A, F328A, S219A
Serotonin	D134A, W324A, S138A, F328A, F223A
Deriv 13	D134A, F328A, S219A, F223A, I332T
Deriv 9	D134A, F328A, F223A, I332T
Deriv 14	D134A, F328A, F223A, S219A
Deriv 16	D134A, F223A, F328A, S219A
Deriv 12	D134A, F328A, S219A, F223A, I332T
Deriv 11	F328A, S219A, F223A
Deriv 15	D134A, S141A, F328A, F223A
Deriv 7	D134A, F328A, F223A

The mutations for each ligand are listed in order of predicted decreasing effect on binding.

Discussion of Binding Energies

The binding energies were calculated for all of the docked ligands and compared to the experimental binding constants in Table 5. Note that binding constants from different experiments cannot be strictly compared because different hot ligands at different concentrations and different temperatures were used. However, we find that very good binders have consistently strong binding constants across different experiments, even though the values are not always the same.

Table 5. The experimental binding dissociation constants (K_i) and calculated binding energies for ligands bound to 5-HT_{2C}, in order of decreasing experimental binding strength

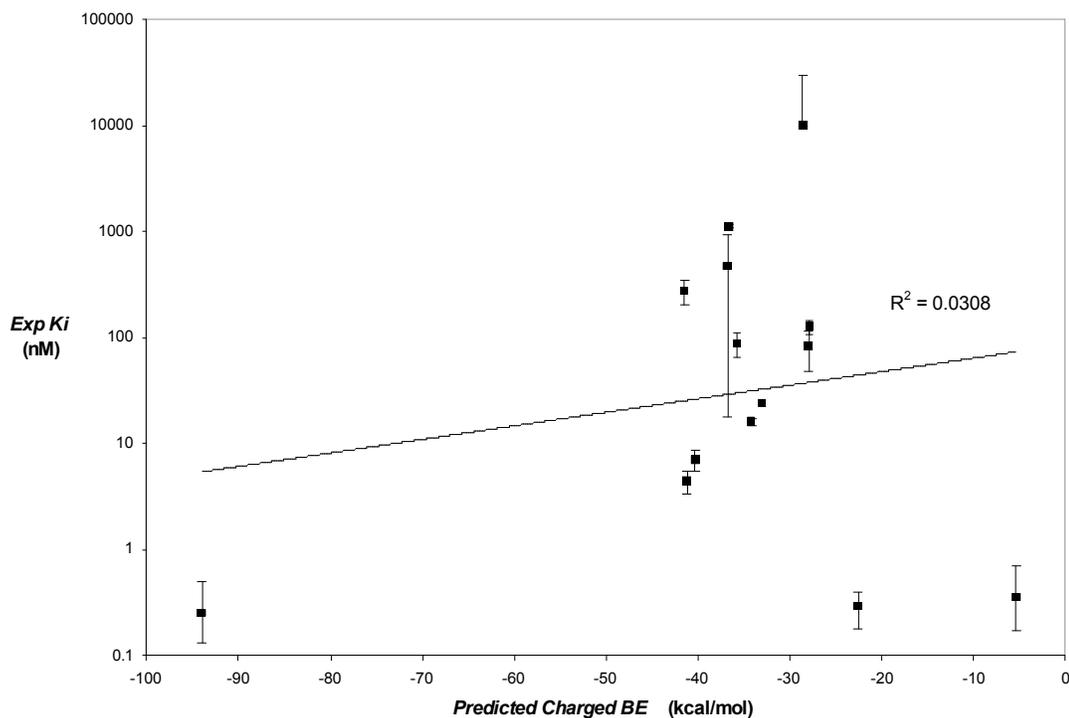
	Calc pKa	pKa pnt	Vac chrg	Vac neut	Lig solv chrg	Lig solv neut	BE chrg	BE neut	Exper K_i	Exper Error
Ritanserin	9.6	3.1	-153.1	-73.4	-59.1	-21.1	-94	-49.2	0.25	±0.25
Metergoline	8.8	2.0	-73.6	-69.7	-51	-18.8	-22.6	-48.9	0.29	± 0.11
Methiothepin	9.5	2.9	-49.9	-59.3	-44.5	-11.5	-5.4	-44.9	0.35	± 0.35
Deriv-5	9.8	3.4	-86.3	-56.7	-45.1	-8.9	-41.2	-44.4	4.4	± 1.1
Deriv-3	9.7	3.2	-86.2	-50.4	-45.9	-8.8	-40.3	-38.4	7	± 1.6
Serotonin	9	2.2	-92.2	-52.5	-58	-12	-34.2	-38.3	16.1	± 1.3
Deriv-13	10.1	3.8	-78.6	-44.9	-45.5	-6.8	-33.1	-34.3	24	± 0.8
Deriv-9	9.6	3.1	-74.9	-42.3	-46.9	-7.6	-28	-31.6	82	± 34
Deriv-14	10.1	3.8	-81.8	-45.7	-46	-8.3	-35.8	-33.6	87	± 22
Deriv-16	9.7	3.2	-72.2	-47.9	-44.3	-7.5	-27.9	-37.2	126	± 19
Deriv-12	10.2	3.9	-83.5	-51.5	-42	-7.1	-41.5	-40.5	275	± 72
Deriv-11	4.3	0.0	-46.4	-42.5	-9.6	-9.5	-36.8	-33.0	468	± 450
Deriv-15	9.8	3.4	-85.6	-49.9	-49	-7.9	-36.6	-38.6	1114	± 41
Deriv-7	9.7	3.2	-78.6	-40.2	-50	-8.8	-28.6	-28.2	10,000	+20,000

More negative binding energies indicate stronger binding. The K_i values and their errors are in units of nM. All energy values are in kcal/mol. The charged binding energy (BE chrg) is defined as the charged vacuum binding energy (vac chrg) minus the solvation energy of the charged ligand (lig solv chrg). The neutral binding energy (BE neut) is the neutral vacuum binding energy (vac neut) minus the solvation energy of the neutralized ligand (lig solv neut) minus the pKa penalty (pKa pnt).

Calculation of binding energies was done for both the charged and neutralized system, and tabulated in Table 5. Using the standard charged model leads to Figure 12, which plots the predicted binding energies against the logarithm of the experimental binding constants. There is no correlation between the charged binding energies and the K_i values ($R^2 = 0.03$). However, the model in which each residue and ligand is treated as neutral leads to a rather good correlation between the binding energies and the K_i values ($R^2 = 0.70$).

Thus, calculating the binding energy with neutral residues and ligands provides a significant improvement over the standard model of charged residues and ligands.

Although the binding energy predictions using neutral residues lead to reasonable correlation with experiment, these calculations are still rudimentary. We consider here minimized structures with continuum solvent approximations. To obtain accurate binding constants we need to carry out molecular dynamics at 300 K and include the entropic contributions to the free energy of binding. Here we must also average the enthalpic parts over the molecular dynamics trajectory while including explicit descriptions of the solvent.



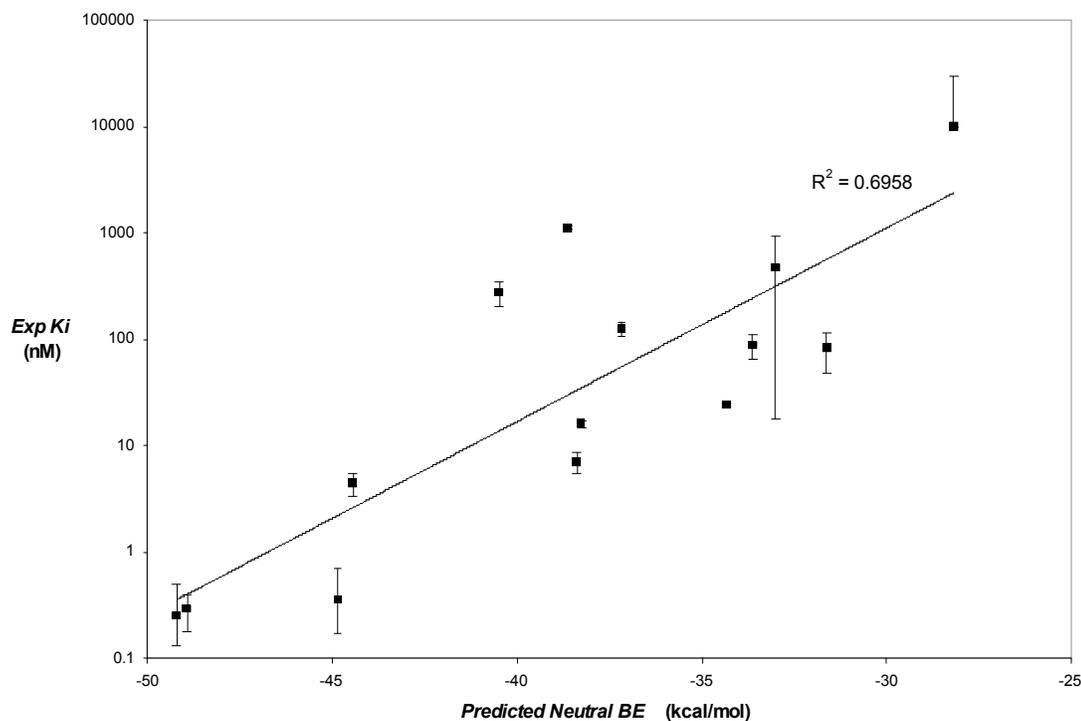


Figure 12. Predicted charged (top) and neutral (bottom) binding energies vs. experimental binding constants with experimental errors.

Comparison to 5-HT_{2B} Predictions

The same procedure used to predict the structure of 5-HT_{2C} was used to predict the structure of 5-HT_{2B}. There is a 3.8 Å RMSD between the two predicted structures, whose alignment is seen in Figure 13. There is close agreement between 5-HT_{2B} and 5-HT_{2C} in the rotations of TM1, TM2, TM4, TM6 and TM7. The conserved aspartic acid in TM3 points between TM5 and TM6 in 5-HT_{2C} while it points between TM4 and TM5 in 5-HT_{2B}. In TM5, the conserved serine points into the binding pocket in 5-HT_{2C}, while it points out toward the membrane in 5-HT_{2B}. Instead, T228 in 5-HT_{2B} TM5, which corresponds to an alanine in 5-HT_{2C}, points toward the binding pocket. TM1 is more bent in 2-HT_{2B} than in 5-HT_{2C}.

HT_{2C} because there is a proline in the middle of the helix in 5-HT_{2B} a turn away from two glycines. TM4 unravels near the end of the helix in both structures because there are two prolines in the helix.

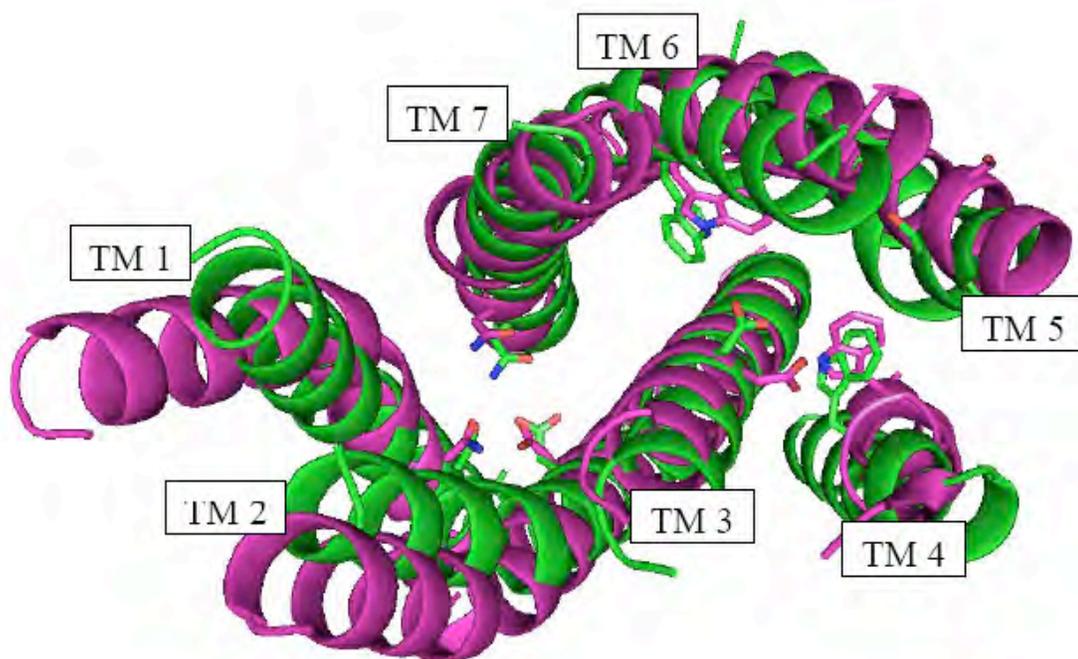


Figure 13. Alignment of the predicted 5-HT_{2B} and 5-HT_{2C} structures. Indicated explicitly are the conserved residues Asn in TM1, Asp in TM2, Asp in TM3, Trp in TM4, Ser in TM5, TRP in TM6, and Asn in TM7.

The bound structures for serotonin and ritanserin to 5-HT_{2B} were predicted, and they are shown in Figure 14. Experimental mutation data for serotonin bound to 5-HT_{2B} can help us assess the accuracy our predicted structure and binding site.^[47] Experimentally, serotonin binds to 5-HT_{2B} with a binding constant of $K_i = 9.0$ nM.^[48]

- The mutation D135A in TM3 was found to create a greater than 800-fold decrease in serotonin binding affinity to 5-HT_{2B}. This agrees with the

prediction that D135(3) creates a salt bridge (16.9 kcal/mol) with the protonated amine of serotonin in 5-HT_{2B}, as it also does in 5-HT_{2C}.

- An alanine substitution of T228(5) was found to cause a 90-fold decrease in serotonin binding affinity to 5-HT_{2B}. Our predicted bound structure sees the OH substituent of the indole in serotonin forming a hydrogen bond with T228(5) (3.0 kcal/mol). The corresponding residue to this TM5 threonine in 5-HT_{2C} is an isoleucine that does not interact with bound serotonin in our predicted structure.
- An alanine substitution of S222(5) did not cause any effect on serotonin binding to 5-HT_{2B}. Our predicted bound structure does not have any interaction with this residue. However, we predict the corresponding residue in 5-HT_{2C} to form a hydrogen bond with the polar NH of the indole group in serotonin.
- The mutation F341A in TM6 caused a 15-fold decrease in serotonin binding affinity to 5-HT_{2B} experimentally. Our predicted structure has good van der Waals interactions with F341(6) (3.4 kcal/mol). We predict that the corresponding residue to this in 5-HT_{2C} also has good van der Waals interactions with serotonin.
- An alanine substitution of W337(6) was found to cause a 11-fold decrease in serotonin binding affinity to 5-HT_{2B}. Our predicted structure has moderate interaction with serotonin (1.1 kcal/mol), but this is not high enough to warrant such a large experimental decrease in binding upon mutation. The corresponding tryptophan in 5-HT_{2C} has a stronger

predicted interaction with serotonin, forming a hydrogen bond to the OH substituent of the indole in serotonin.

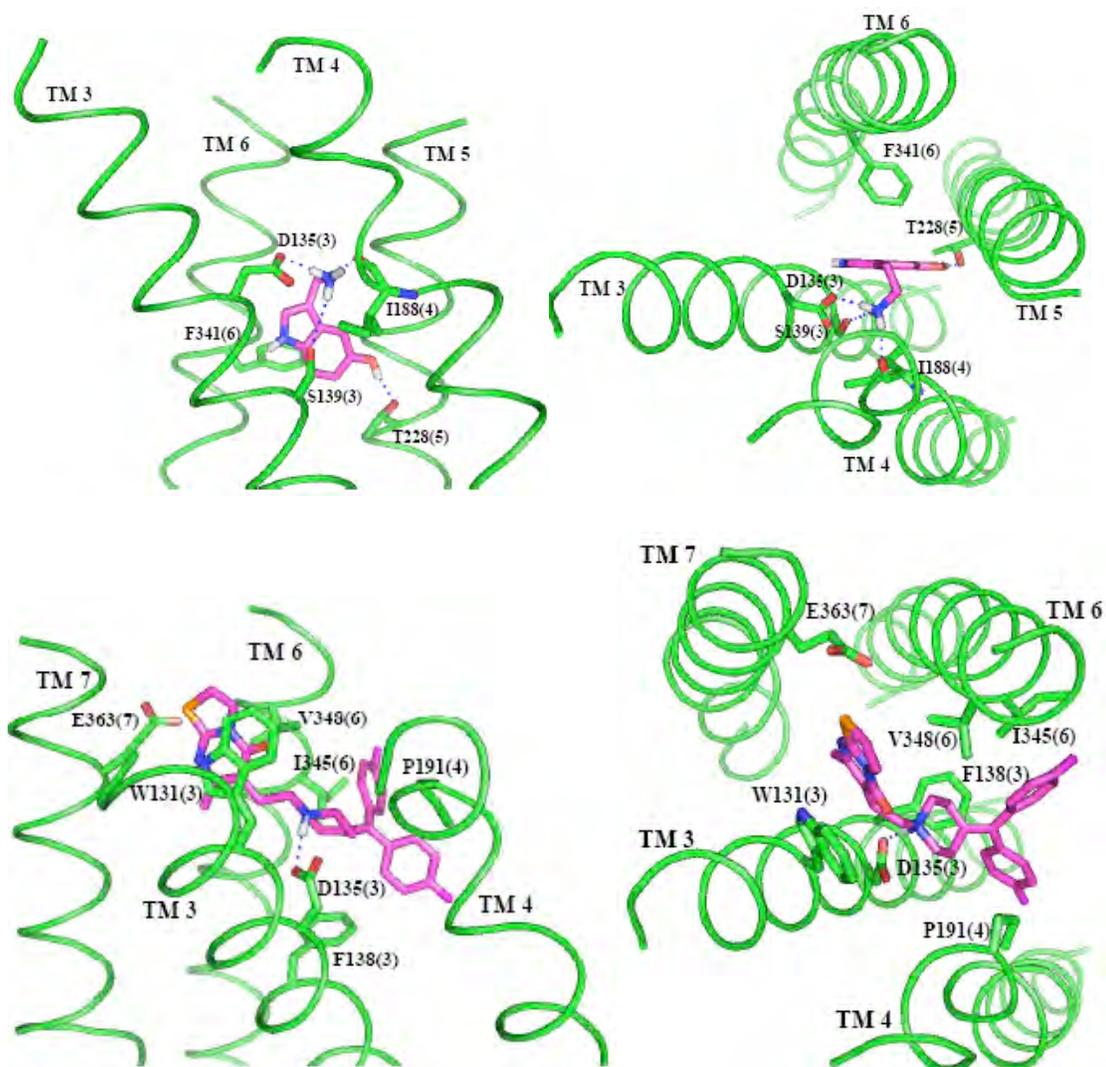


Figure 14. The predicted structure (side and top views) of serotonin (top) and ritanserin (bottom) bound to 5-HT_{2B}. The TMs shown do not interact directly with the ligand. The residues shown are those within 5.0 Å of serotonin that have more than 2 kcal of favorable interaction with the ligand. The dotted lines indicate hydrogen bonds with the ligand.

- The mutation S139A in TM3 was found to cause a 30-fold decrease in serotonin binding affinity to 5-HT_{2B}. The predicted structure has moderate interactions (2.1 kcal/mol) between this residue and serotonin, creating a very weak hydrogen bond with the protonated amine of serotonin. The corresponding residue in 5-HT_{2C} has a stronger interaction with the protonated amine of serotonin.
- The substitution of serine for A187(4) was not found to have any effect on serotonin binding affinity to 5-HT_{2B}. We predict little interaction between this residue and serotonin, but we do predict strong hydrogen bonding between the backbone of I188(4) and the protonated amine of serotonin. We predict no such interaction in 5-HT_{2C}. However, this may be an artefact of the excessive unraveling of TM4 due to prolines.
- The mutation N344A in TM6 was found to cause a 7-fold decrease in serotonin binding affinity to 5-HT_{2B}. We do not see any interaction between this residue in our predicted structure and serotonin. However, N344(6) is forming a hydrogen bond with W367(7), so it may be involved in structural stabilization instead of ligand binding. These residues are also found in 5-HT_{2C}, but they are not predicted to hydrogen bond.
- The mutation D100N in TM2 was found to cause a 12-fold decrease in serotonin binding to 5-HT_{2B}. D100N(2) does not have any predicted interactions with bound serotonin. However, the double mutation D100N in TM2 along with N367D in TM7 has no effect in serotonin binding. Many studies suggest these helices do not directly bind to serotonin, but instead

interact through a hydrogen bonding network to indirectly affect binding through a conformational change in the binding pocket. This hypothesis is supported by the fact that D100(2) and N376(7) are hydrogen bonding in the predicted structure, and experiments show that if the identities of the two amino acids are switched, effectively retaining the hydrogen bond, the original binding constants to serotonin are maintained. These two amino acids are also hydrogen bonded in 5-HT_{2C}, with no interaction with serotonin.

Comparison of the predicted 5-HT_{2B} serotonin bound structure to mutagenesis experiments shows some good agreement with experiment, but there are also some discrepancies. The fact that serotonin makes a hydrogen bond with the backbone of TM4 probably means that TM4 is packed too closely to TM3, or that TM4 has unraveled too much. Also, mutagenesis experiments suggest stronger interactions between serotonin and the residues S139(3) and W337(6). There could be hydrogen bonding to S139(3) and better van der Waals interactions with serotonin if TM3 were rotated more closely to the same angle as it is in 5-HT_{2C}. Despite these inconsistencies, there are strong similarities to the predicted 5-HT_{2C} structure, with the differences in predicted serotonin binding between the two receptors primarily due to the different rotation of TM5.

Our predicted serotonin binding site for 5-HT_{2B} agrees more closely with mutagenesis experiments than do homology models of 5-HT_{2B}. A study comparing homology models of 5-HT based on bacteriorhodopsin and bovine rhodopsin concluded that the bacteriorhodopsin-based model correlated better

with experimental mutagenesis data.^[47] According to this bacteriorhodopsin-based homology model of 5-HT_{2B}, D135(3), S139(3), F341(6) and N344(6) make direct contacts with serotonin. Additionally, W131(3), F138(3), F341(6) and F365(7) form an aromatic box surrounding serotonin in the homology model. This model does not show any significant interaction between serotonin and TM5, while our model shows hydrogen bonding between serotonin and T228(5). Experiments suggest that T228(5) is very important in binding serotonin, as evidenced by the fact that a mutation to alanine causes a 90-fold decrease in serotonin binding. The homology model has serotonin forming a hydrogen bond with N344(6) instead of with T228(5), although experiments only show a 7-fold decrease of binding from the N344A mutation. Additionally, the homology model shows no interaction with W337(6), while our model has moderate interactions with the residue, which agrees with experiment, because the W337A mutation causes a 12-fold decrease in serotonin binding. Thus, overall, our predicted serotonin binding site of 5-HT_{2B} correlates better with the available experimental data than does a bacteriorhodopsin-based homology model.

The binding site for one antagonist, ritanserin, was predicted for 5-HT_{2B}. Experimentally, ritanserin binds to 5-HT_{2B} with a binding constant of $K_i = 1.6 \text{ nM}$.^[48] The most important contacts are as follows.

- The protonated nitrogen of the piperidine forms a salt bridge with D135(3) (16.4 kcal/mol).

- The aromatic thiazolo-pyrimidine group has strong van der Waals interactions with W131(3) (7.8 kcal/mol), V348(6) (5.0 kcal/mol), and E363(7) (2.1 kcal/mol).
- The fluorenyl groups have strong van der Waals interactions with I345(6) (3.2 kcal/mol), P191(4) (2.6 kcal/mol), and F138(3) (2.0 kcal/mol).

This predicted binding site for ritanserin in 5-HT_{2B} has little in common with the predicted binding site for ritanserin in 5-HT_{2C}, except that they both interact with TM3 and TM6, and both have salt bridges between ritanserin and the aspartic acid on TM3.

The binding constants were calculated for serotonin and ritanserin bound to 5-HT_{2B}, and the results are shown in Table 6. The calculated binding constants correctly predict that ritanserin binds more strongly than serotonin to 5-HT_{2B}. When the calculated binding constants for 5-HT_{2B} are compared with those for 5-HT_{2C}, the binding constants for 5-HT_{2C} are stronger than those for 5-HT_{2B}. However, the binding constants for the two receptors are much closer together once the complexes are neutralized.

Table 6. The experimental binding dissociation constants (K_i) and calculated binding energies for ligands bound to 5-HT_{2B}, in order of decreasing experimental binding strength

	Calc pKa	pKa pnt	Vac Chrg	Vac neut	Lig solv chrg	Lig solv neut	BE chrg	BE neut	Exper K_i	Exper Error
Ritanserin	9.6	3.1	-127.4	-60.4	-56.7	-20.9	-70.7	-36.4	1.6	±0.7
Serotonin	9	2.2	-108.7	-46.7	-57.5	-11	-51.2	-33.5	9	±1.9

More negative binding energies indicate stronger binding. The K_i values and their errors are in units of nM. All energy values are in kcal/mol. The charged binding energy (BE chrg) is defined as the charged vacuum binding energy (vac chrg) minus the solvation energy of the charged ligand (lig solv chrg). The neutral binding energy (BE neut) is the neutral vacuum binding energy (vac neut) minus the solvation energy of the neutralized ligand (lig solv neut) minus the pKa penalty (pKa pnt).

Conclusion

We find good agreement between the calculated binding sites of the predicted ligand-protein bound complexes of 5-HT_{2C} and experimental binding constants. We also find many similarities between the predicted serotonin binding site for 5-HT_{2C} and mutagenesis experiments for 5-HT_{2A}. Additionally, comparison of the predicted 5-HT_{2C} structure with a preliminary prediction of the 5-HT_{2B} structure along with mutation experiments on serotonin bound in 5-HT_{2B} reveals a great deal of similarity between the two structures. This suggests that the 3D protein structure predicted by the Membstruk method is sufficiently accurate for drug design. Dynamics of serotonin bound in 5-HT_{2C} in an explicit lipid bilayer show the predicted binding site to be stable during 5 ns of simulation. The predicted binding site for serotonin leads to the suggested mutations S138A, S141A,

S219A, F223A, F328A and W324A and I332A, which are predicted to all cause a decrease in binding energies. Mutation experiments for other ligands, including strongly binding antagonists and a series of psilocybin analogs, are also proposed. Thus, a wealth of data is now available to further validate the proposed structure and binding sites of 5-HT_{2C}. The good agreement between experiment and the predicted bound structures indicates that the procedures are useful for studying other systems with less experimental data.

Chapter 2

SuperBiHelix: A New Method for Predicting an Ensemble of Low-Lying GPCR Structures

Abstract

A new procedure called SuperBiHelix has been developed for GPCR structure prediction. It predicts an ensemble of low-lying structures. SuperBiHelix places predicted TM helices in an experimental template, and samples the tilt and sweep angles of the helices along with the rotation of the helices along the helical axes. The procedure was validated on the β 2-adrenergic receptor and A_{2A} adenosine receptor experimental crystal structures. The SuperBiHelix procedure will make it possible to more accurately predict structures of GPCRs that are dissimilar to available experimental structures.

Introduction

The introduction of the BiHelix procedure into the GPCR structure prediction method allowed for more accurate structure prediction. However, the BiHelix procedure relies heavily on the experimental template in which the helices are placed, for the only degree of freedom sampled is the rotation of each helix along its own axis. A new procedure has been developed, SuperBiHelix, which samples the sweep angle and tilt angle of each helix, in addition to the rotation about their own axes. This sampling of additional degrees of freedom will make

the structure prediction less dependent on the template, and therefore less dependent on any experimental input. Thus, it will allow more accurate prediction of structures for receptors dissimilar to any available experimental structure. Finally, it has the additional advantage of predicting more low-lying structures, which could lead to insights on GPCR activation.

The procedure was tested by placing the helices of one experimental structure into another experimental template, and running SuperBiHelix to determine whether the helices recognize their original template. After the procedure was validated, it was used on experimental helices in their original structures to see what other low-lying structure arise.

Methodology

After the helix lengths and shapes are predicted, they are placed in an experimental GPCR bundle template, which is defined by a system of coordinates shown in Figure 15. Each experimental template has 42 degrees of freedom: x , y , z , θ , ϕ and η values for each of the seven TM helices. The helices have kinks and bends, so the helical axis is defined as its moment of least inertia. The hydrophobic center is the residue that crosses $z = 0$, which is defined as the plane that runs through the center of the lipid bilayer, and it is either calculated from the protein's hydrophobic profile or by homology. The x -axis is defined along the axis from the center of TM3, which is in the middle of the bundle, to the center of TM2 in the mid-plane ($z = 0$). The definitions of the x -axis and z -axis implicitly define the y -axis. The x and y values of the helices (where the

hydrophobic centers cross $z = 0$), are defined by the experimental template, and are not sampled by SuperBiHelix. The degrees of freedom that are sampled are θ , the tilt angle of the helix; ϕ , the sweep angle of the helix; and η , the rotation of the helix around the helical axis.

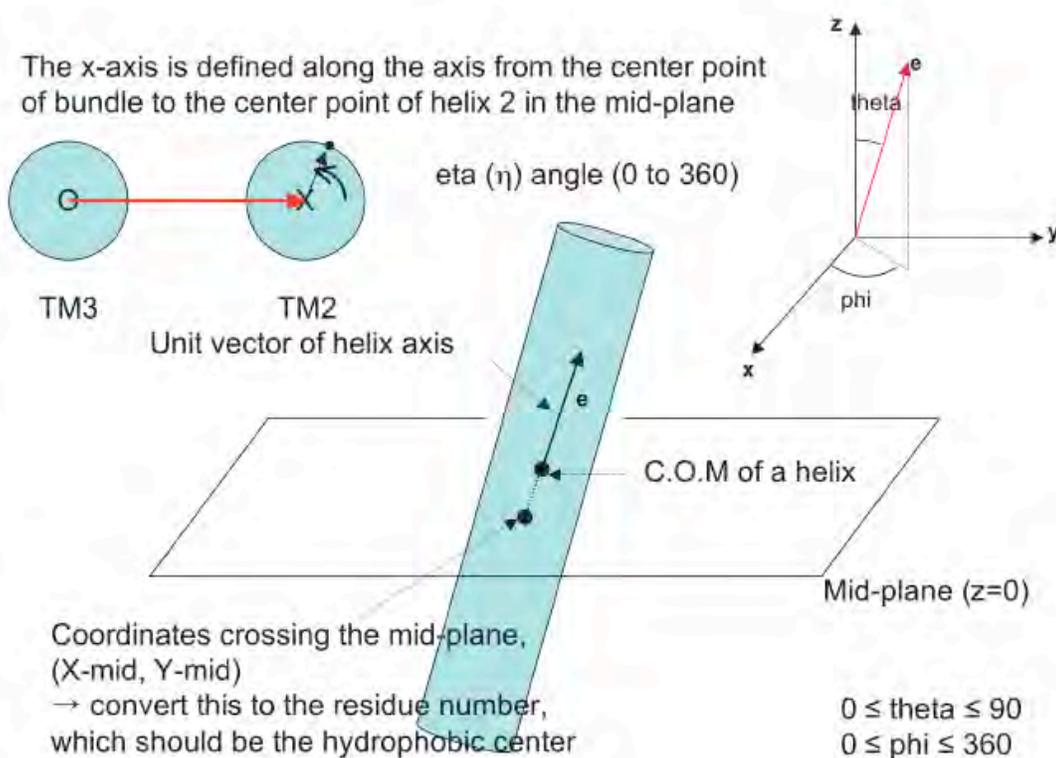


Figure 15. The coordinates used to describe the orientation of the seven helices in a GPCR bundle.

The SuperBiHelix program takes an input GPCR bundle file, and determines its template. It then varies the θ , ϕ and η values. However, even if just sampling a small number of angles for each degree of freedom, for example three, that would lead to $(3 \times 3 \times 3)^7 \approx 10^{10}$ possible configurations for which to predict the side chains and calculate the energy. This is far too computationally expensive.

However, if the energy is approximated to be made up of only interactions between two helices, the calculation is more tractable. A seven-helix GPCR TM bundle has twelve strongly interacting pairs: TM1-TM2, TM1-TM7, TM2-TM3, TM2-TM4, TM2-7, TM3-TM4, TM3-TM5, TM3-TM6, TM3-TM7, TM4-TM5, TM5-TM6 and TM6-TM7. For each of these twelve interacting pairs, θ , ϕ and η are sampled with the other helices absent in order to get the bihelical energies. SCREAM is used to predict the side chain placements, then the side chains are minimized for 10 steps with the backbone fixed. This procedure is illustrated in Figure 16.

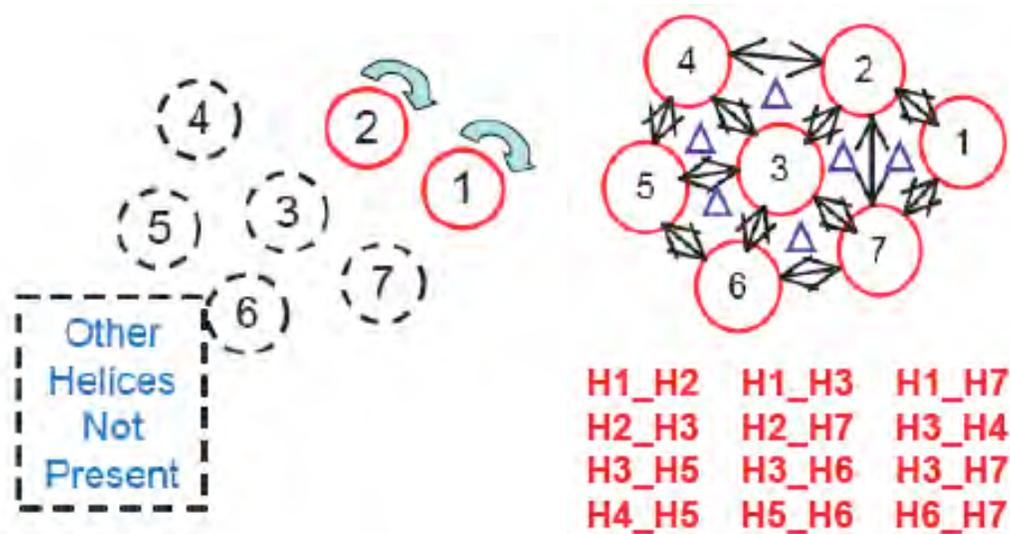


Figure 16. Diagram of the SuperBiHelix method, in which the seven-helix TM bundle is split into twelve helix pairs, and the θ , ϕ and η values for each helix in the pair is sampled with the other helices not present.

Once the bihelical energies have been determined for all possible combinations of θ , ϕ and η , the energy of the entire bundle for each possible conformation must be calculated. The bihelical energies are decomposed into

intrahelical and interhelical components. The energy of the entire complex is then calculated as

$$\begin{aligned}
 E_{intra}(\bar{\eta}, \bar{\theta}, \bar{\phi}) &= \sum_{i=1}^7 \frac{1}{N_i} \sum_{j=J_{i,1}, j>i}^{J_{i,N_i}} E_{i,intra}^{ij}(\eta_i, \eta_j, \theta_i, \theta_j, \phi_i, \phi_j), \\
 E_{inter}(\bar{\eta}, \bar{\theta}, \bar{\phi}) &= \sum_{i=1}^7 \sum_{j=J_{i,1}, j>i}^{J_{i,N_i}} E_{inter}^{ij}(\eta_i, \eta_j, \theta_i, \theta_j, \phi_i, \phi_j), \\
 E_{total}(\bar{\eta}, \bar{\theta}, \bar{\phi}) &= E_{intra}(\bar{\eta}, \bar{\theta}, \bar{\phi}) + E_{inter}(\bar{\eta}, \bar{\theta}, \bar{\phi}), \\
 \bar{\eta} &= (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6, \eta_7), \\
 \bar{\theta} &= (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7), \\
 \bar{\phi} &= (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7),
 \end{aligned}$$

where N_i is the number of helices interacting with helix i , and $J_{i,k}$ is the k th neighbor of helix i . Although the calculation of the energy of a complex based on its bihelical energies is very fast, the calculation of all possible configurations is still too computationally expensive. In practice, the smallest number of conformations sampled would be three values of θ , five values of ϕ and five values of η , which would lead to $(3*5*5)^7 \approx 10^{13}$ total bundle conformations. Thus, a procedure must be developed to determine which conformations for each helix are most favorable, so that fewer total bundle energies have to be calculated.

In order to determine the best conformations for each helix that will lead to the lowest energy bundles, the seven-helix bundle is partitioned into three quadhelix bundles, as shown in Figure 17.

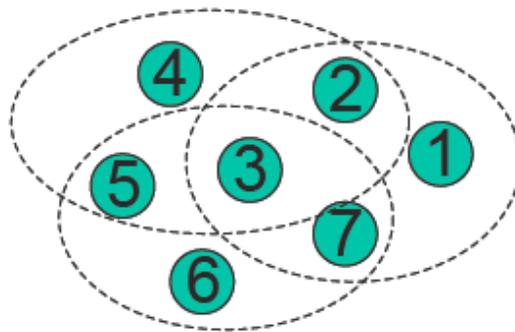


Figure 17. In order to determine best conformations for each helix that will lead to the lowest energy bundles, the seven helix bundle is partitioned into three quadhelix bundles: TM1-TM2-TM3-TM7, TM2-TM3-TM4-TM5 and TM3-TM5-TM6-TM7.

Next, the total bihelical energies of the three quadhelix bundles are calculated.

This is feasible because only $3 \cdot (3 \cdot 5 \cdot 5)^4 \approx 10^8$ bundle energies must be calculated.

The 2000 structures with the lowest energy for each quadhelix are listed by increasing energy. Then, the conformations are ranked for each helix, using the following protocol:

- For TM1, each unique TM1 conformation $(\eta_1, \theta_1, \phi_1)$ in the TM1-TM2-TM3-TM7 bundle list is taken.
- For TM2, each unique TM2 conformation $(\eta_2, \theta_2, \phi_2)$ in the TM1-TM2-TM3-TM7 bundle list is alternated with each unique TM2 conformation in the TM2-TM3-TM4-TM5 bundle list.
- For TM3, each unique TM3 conformation $(\eta_3, \theta_3, \phi_3)$ in the TM1-TM2-TM3-TM7 bundle list is taken in alternation with each unique TM3 conformation in the TM2-TM3-TM4-TM5 bundle list and each unique TM3 conformation in the TM3-TM5-TM6-TM7 bundle list.

- For TM4, each unique TM4 conformation $(\eta_4, \theta_4, \phi_4)$ in the TM2-TM3-TM4-TM5 bundle list is taken.
- For TM5, each unique TM5 conformation $(\eta_5, \theta_5, \phi_5)$ in the TM2-TM3-TM4-TM5 bundle list is alternated with each unique TM5 conformation in the TM3-TM5-TM6-TM7 bundle list.
- For TM6, each unique TM6 conformation $(\eta_6, \theta_6, \phi_6)$ in the TM3-TM5-TM6-TM7 bundle list is taken.
- For TM7, each unique TM7 conformation $(\eta_7, \theta_7, \phi_7)$ in the TM1-TM2-TM3-TM7 bundle list is alternated with each unique TM7 conformation in the TM3-TM5-TM6-TM7 bundle list.

Finally, from each individual helical conformation list, the best 36 conformations for each helix are used to calculate the energy of $36^7 \approx 8 \times 10^{10}$ full bundles, and output the 1000 best energy structures from this procedure.

In a procedure called SuperComBiHelix, these top 1000 helical bundles are built and the side chains are reassigned, given that they will take different conformations than in the bihelical mode. Then the structure is minimized for 10 steps. The energy ranking will be different in SuperComBiHelix than SuperBiHelix because all seven helices are present instead of just two at a time. This procedure results in an ensemble of low-lying structures. Examination of the low-lying structures shows which helices are flexible, and may give insight into activation.

Extensive testing on these methods, shown in the Validation section, lead to several improvements in the procedure. During the side chain prediction steps in

SuperBiHelix and SuperComBiHelix, SCREAM must be used with a 0.5 Å resolution library instead of the 1.0 Å resolution library that is the default for SCREAM. Additionally, the best results arise from alanizing the final two residues of the C- and N-termini for each helix during the SuperBiHelix step. Then, before the SuperComBiHelix step, the alanized residues are mutated back to their original residues for the building of the full bundles. This step reduces artificial long range electrostatic interactions between charged groups that would be located in the polar head group region of the lipid bilayer.

Validation

SuperBiHelix on Crystal Helices in an Incorrect Template

The SuperBiHelix and SuperComBiHelix methods were tested on the β 2-adrenergic receptor^[5] in the A_{2A} adenosine receptor^[6] template, and vice versa. The differences between the x, y, θ , ϕ and η values for the two templates are given in Table 7.

Table 7. The differences between the A_{2A} adenosine receptor and β 2-adrenergic receptor templates

A _{2A} adenosine to β 2-adrenergic template					
TM	x (Å)	y (Å)	θ (°)	ϕ (°)	η (°)
1	2.0	1.7	6.3	17.9	4.8
2	0.2	0.0	7.7	2.4	-18.8
3	0.0	0.0	4.9	2.0	-13.2
4	-1.5	-0.4	-4.1	-3.0	-9.2
5	0.1	-0.6	-1.2	14.9	0.0
6	0.8	0.0	10.8	-24.8	0.9
7	-0.1	1.1	4.1	3.1	2.6

The system of coordinates is described in Figure 15.

Table 7 reveals that x and y have only a small amount of variation between templates, supporting the fact that the SuperBiHelix procedure does not sample x or y . It also shows that ϕ and η vary more among templates than θ , so more ϕ and η values will need to be sampled than θ values.

The first test performed was to take the helices and template of the β_2 -adrenergic receptor, and give the template the ϕ and θ values of the A_{2A} adenosine receptor. For all of the test runs, the helices in the original crystal structures were minimized separately without the other helices present so that the procedure was not biased toward the crystal structure. Then θ was sampled with values of -10° , 0° and 10° , while ϕ was sampled with values of -30° , -15° , 0° , 15° and 30° . The results of this test, using both 1.0 Å and 0.5 Å SCREAM libraries, are shown in Table 8. Table 8 clearly shows that the use of the 0.5 Å SCREAM library instead of the 1.0 Å SCREAM library is necessary for SuperBiHelix and SuperComBiHelix. When the 1.0 Å SCREAM library is used, the structure goes from 1.9 Å backbone RMSD to the crystal structure to a 1.4 Å backbone RMSD. When the 0.5 Å SCREAM library is used, the structure goes to a 0.9 Å backbone RMSD. The closest structure to the crystal structure that the procedure could possibly produce, given the angles sampled, has a 0.8 Å backbone RMSD. Additionally, with the 0.5 Å SCREAM library, the crystal structure is one of the best energy structures, while it is not with the 1.0 Å SCREAM library.

Table 8. SuperComBiHelix results for β 2-adrenergic receptor with A_{2A} adenosine template ϕ and θ values

1.0 Å SCREAM library																	
θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	10	0	0	0	0	10	10		0	-15	0	-15	15	-15	15	150.3	1.4
	10	0	0	0	0	10	10		0	-15	0	0	15	-15	15	155.4	1.4
	10	0	0	0	0	10	10		0	-15	0	-30	15	-15	15	164.8	1.4
	10	0	0	0	0	10	0		-30	0	0	-30	0	-30	15	177.3	2.8
	10	0	0	0	0	10	10		0	0	0	-15	15	-15	15	178.7	1.4
	10	0	0	0	0	10	0		0	0	0	-15	15	-15	15	181.1	1.4
	10	0	0	0	0	10	10		0	0	0	-30	15	-15	15	181.7	1.4
	10	0	0	0	0	10	0		0	0	0	-30	0	-30	15	182.6	1.5
	10	0	0	0	0	10	0		0	0	0	-30	15	-15	15	183	1.4
	10	0	0	0	0	10	0		-15	0	0	-15	15	-30	15	184.8	2
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3	189.7	0
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0	223.4	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0	446.4	1.9

0.5 Å SCREAM library																	
θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	10	10	0	0	0	10	10		15	0	0	-15	0	-15	15	57.2	0.9
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3	60.6	0
	10	0	0	0	0	10	0		0	0	0	-30	15	-30	0	79	1.4
	10	10	0	10	0	10	0		0	-15	-15	-15	0	-30	0	93.4	1.7
	10	10	0	0	0	10	10		15	0	0	-15	15	-15	15	93.6	0.8
	10	0	0	0	0	10	0		0	0	0	-15	15	-30	0	96.1	1.4
	10	10	0	0	0	10	0		15	0	0	0	0	-15	15	96.9	0.9
	10	10	0	0	0	10	0		15	0	0	-15	15	-30	-15	98.4	0.9
	0	0	0	0	0	10	0		15	0	0	-15	15	-30	0	101.1	1.1
	10	10	0	10	0	10	0		15	0	15	-30	15	-30	0	102.3	1.4
	10	10	0	0	0	10	0		15	0	0	-30	0	-15	15	102.3	1
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0	111.6	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0	309.9	1.9

The RMSD is the backbone RMSD to the β 2-adrenergic receptor crystal structure. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

For the same test performed on the A_{2A} adenosine helices with the ϕ and θ values of the β 2-adrenergic template, using the 0.5 Å SCREAM library, the

structure goes from a 2.0 Å backbone RMSD to the crystal structure to a 0.9 Å backbone RMSD. Thus, the 0.5 Å SCREAM library will be used for all SuperBiHelix and SuperComBiHelix runs. The next test was to take the helices and template of the A_{2A} adenosine receptor, and give them the η , ϕ and θ values of the β 2-adrenergic template. The x and y values were still from the A_{2A} adenosine template. SuperBiHelix and SuperComBiHelix were run, sampling θ with values of -10° , 0° and 10° , ϕ with values of -30° , -15° , 0° , 15° and 30° , and η with values of -30° , -15° , 0° , 15° and 30° . The results are given in Table 9.

Table 9. SuperComBiHelix results for the A_{2A} adenosine receptor with β 2-adrenergic template ϕ , θ and η values

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	-6	-8	-5	4	1	-11	-4		-18	-2	-2	3	-15	25	-3		-5	19	13	9	0	-1	-3	88.2	0.2
	-10	-10	0	0	10	-10	0		-30	0	0	-15	-15	-30	0		-15	15	15	15	0	0	0	112.1	1.1
	-10	-10	-10	10	0	-10	0		-30	-15	0	-30	-15	0	0		-15	15	15	15	0	0	-15	116.4	1
	-10	-10	0	0	10	-10	0		-30	0	0	0	-15	-30	0		-15	15	15	15	0	0	0	118.2	1.1
	-10	-10	-10	10	0	-10	0		-30	-15	0	-30	-15	-15	0		-15	15	15	15	0	0	0	125.8	1
	-10	-10	-10	10	0	-10	0		-30	-15	-15	-30	-15	15	0		-15	15	15	15	0	0	-15	132.6	1
	0	-10	-10	10	0	-10	0		-30	-15	0	-30	-15	0	0		0	15	15	15	0	0	-15	134.9	1
	0	-10	-10	0	0	-10	0		-30	-30	0	-30	-15	15	-15		0	15	15	0	0	0	0	134.9	1.1
	-10	-10	-10	10	0	-10	0		-30	-15	0	-30	-15	15	0		-15	15	15	15	0	0	-15	136.8	0.9
	0	-10	-10	0	0	-10	0		-30	-30	0	-30	-15	0	-15		0	15	15	0	0	0	0	138.5	1.1
	-10	-10	0	0	10	-10	0		-30	0	0	-15	-15	-30	-15		-15	15	15	15	0	0	0	138.5	1.1
	-10	-10	0	0	0	-10	0		-15	0	0	0	-15	30	0		0	15	15	15	0	0	0	257.6	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	794.6	2

The RMSD is the backbone RMSD to the A_{2A} adenosine receptor crystal structure. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

Table 9 shows that SuperBiHelix works on the A_{2A} adenosine receptor with β 2-adrenergic template ϕ , θ and η values: the structure goes from a 2.0 Å backbone RMSD to the crystal structure to a 1.1 Å backbone RMSD, with the lowest possible RMSD being 0.8 Å.

For the same test performed on the helices and template of the β 2-adrenergic receptor with the η , ϕ and θ values of the A_{2A} adenosine template, the results are not as promising, as shown in Table 10.

Table 10. SuperComBiHelix results for the β 2-adrenergic receptor with A_{2A} adenosine template ϕ , θ and η values

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3		5	-19	-13	-9	0	1	3	55.6	0.1
	10	0	0	0	0	10	0		-15	0	0	-30	15	-30	0		30	-15	-15	0	0	0	0	89.5	2.1
	10	0	0	0	0	10	10		0	-15	0	-30	-15	0	15		0	-15	-15	0	-15	0	0	92.5	1.7
	10	0	0	0	0	10	0		0	-15	0	-30	15	-30	-15		0	-15	-15	0	0	0	0	95	1.5
	0	0	0	0	0	10	0		-30	0	0	-30	-15	-30	0		-30	-15	-15	0	-15	0	0	95.4	2.7
	10	0	0	0	0	10	10		-30	-15	0	-30	-15	-30	15		-15	-15	-15	0	-15	0	-15	99.8	2.9
	10	0	0	0	0	10	0		-15	0	0	-30	15	-30	0		15	-15	-15	0	0	0	0	100.7	2.1
	0	0	0	0	0	10	0		0	-15	0	-30	15	-30	-15		0	-15	-15	0	0	0	0	103.3	1.4
	10	0	0	0	0	10	0		0	0	0	-30	-15	-30	0		15	-15	-15	0	-15	0	0	104.4	1.7
	10	0	-10	10	0	10	10		-30	-15	0	-30	0	-30	15		0	-30	-15	0	0	0	0	106.4	3
	10	0	0	0	0	10	10		-30	-15	0	-30	-15	0	15		0	-15	-15	0	-15	0	0	106.5	2.9
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0		0	-15	-15	-15	0	0	0	108.3	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	437.2	1.9

The RMSD is the backbone RMSD to the β 2-adrenergic receptor crystal structure. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

As Table 10 reveals, SuperBiHelix and SuperComBiHelix fail when run on the β 2-adrenergic receptor helices with A_{2A} adenosine template ϕ , θ and η values: the structure goes from a 1.9 Å backbone RMSD to the crystal structure to a 2.1

Å backbone RMSD. There must be an explanation for why the SuperBiHelix procedure fails with β 2-adrenergic receptor helices in the A_{2A} adenosine template, but succeeds with A_{2A} adenosine receptor helices in the β 2-adrenergic. Inspection of the helices of the two receptors reveals that the β 2-adrenergic helices have more charged residues near the ends of the helices than the A_{2A} adenosine helices, as shown in Figure 18.

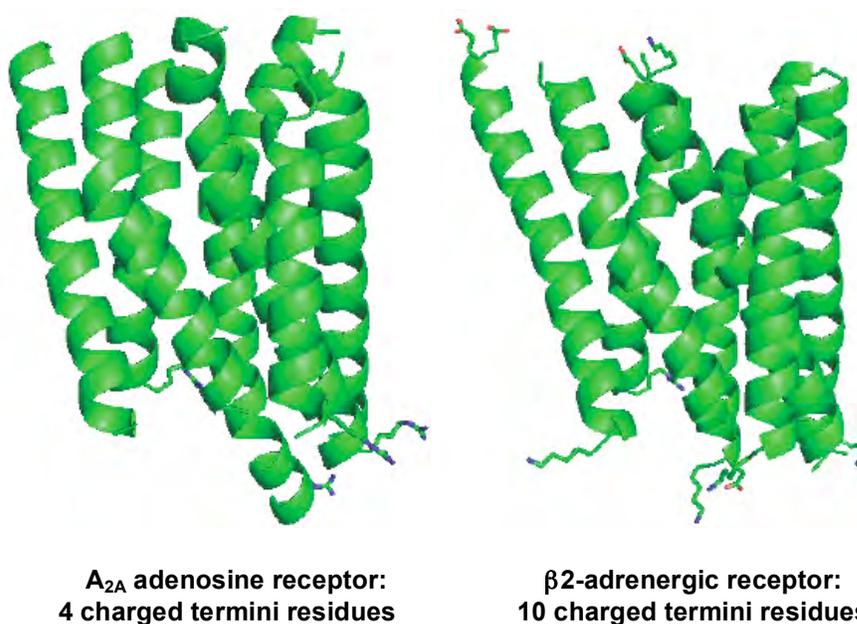


Figure 18. A comparison of charged termini residues in the A_{2A} adenosine and β 2-adrenergic receptors. A terminus residue is defined to be any residue within two residues of the end of a TM helix. Charged residues are ARG, GLU, ASP and LYS.

The A_{2A} adenosine TM helices have four charged termini residues, where a terminus residue is defined to be any residue within two residues of the end of a helix, while the β 2-adrenergic TM helices have ten charged termini residues. Coulombic interactions between charged residues at the termini are sensitive to

small changes in helix orientation. Plus, these termini residues are located in the polar head group region of the lipid bilayer, which would provide significant charge screening. Based on this difference in the number of charged residues, a SuperBiHelix run was again performed on the β 2-adrenergic receptor helices with A_{2A} adenosine template ϕ , θ and η values, but with the final two residues, excepting GLY or PRO, of each helix mutated to alanine. The results of this test are shown in Table 11.

Table 11. SuperComBiHelix results for the β 2-adrenergic receptor with A_{2A} adenosine template ϕ , θ and η values, with any residue within two residues of the end of a TM helix mutated to alanine during the SuperBiHelix and SuperComBiHelix procedures

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3		5	-19	-13	-9	0	1	3	369.9	0
	-10	0	0	0	0	10	0		15	0	0	-30	-15	-30	0		-15	-15	-15	0	-15	0	0	388.6	2
	10	10	0	10	0	10	0		15	0	0	-30	-15	-30	0		0	-15	-15	0	0	0	0	402.3	1.5
	0	0	0	0	0	10	0		0	0	0	-30	-15	-30	0		15	-15	-15	0	-15	0	0	403	1.7
	0	0	0	0	0	10	0		0	0	0	-30	-15	-30	0		0	-15	-15	0	-15	0	0	404	1.6
	10	10	0	0	0	10	0		15	0	0	-15	-15	-30	0		0	-15	-15	-15	0	0	0	409.2	1.2
	10	10	0	0	0	10	0		15	0	0	-30	-15	-30	0		0	-15	-15	-15	0	0	0	419.4	1.2
	10	10	0	0	0	10	0		15	0	0	-15	0	-15	15		0	-15	-15	0	0	0	0	423.4	0.9
	10	10	0	0	0	10	0		15	0	0	-30	0	-30	0		0	-15	-15	-15	-15	0	0	424.3	1
	10	10	0	0	0	10	0		0	0	0	-15	-15	-30	0		0	-15	-15	-15	0	0	0	425.8	1.5
	10	10	0	0	0	10	0		15	0	0	-30	-15	-30	0		0	-15	-15	-15	-15	0	0	426.1	1.2
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0		0	-15	-15	-15	0	0	0	450.9	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	768.7	1.9

The RMSD is the backbone RMSD to the β 2-adrenergic receptor crystal structure. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

The alanization of the termini residues improves the results a little bit for the β 2-adrenergic receptor in the A_{2A} adenosine template, in that the backbone

RMSD values of the top ten SuperComBiHelix structures to the β 2-adrenergic receptor crystal structure are less than or equal to 2.0 Å. However, the best energy structure still has a 2.0 Å RMSD from the crystal structure, which is worse than the original structure in the incorrect template. So, the procedure was once again changed so that the termini are alanized for the SuperBiHelix procedure, but then mutated back to the original residues for the SuperComBiHelix procedure. The results for this test are shown in Table 12.

Table 12. SuperComBiHelix results for the β 2-adrenergic receptor with A_{2A} adenosine template ϕ , θ and η values, with any residue within two residues within the end of a TM helix mutated to alanine during the SuperBiHelix procedure

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3		5	-19	-13	-9	0	1	3	55.6	0
	10	10	0	0	0	10	0		15	0	0	-15	0	-30	0		0	-15	-15	-15	0	0	0	78	0.9
	0	0	0	0	0	10	0		0	0	0	-30	-15	-30	0		0	-15	-15	0	-15	0	0	81.7	1.6
	10	10	0	0	0	10	0		15	0	0	-15	15	-30	0		15	-15	-15	-15	0	0	0	92.6	0.9
	10	10	0	0	0	10	10		0	0	0	-15	0	-15	15		0	-15	-15	0	0	0	0	94	1.3
	10	10	0	10	0	10	0		15	0	0	-30	15	-30	0		15	-15	0	-15	-15	0	0	95.6	1.3
	10	10	0	0	0	10	10		0	-15	0	-15	0	-15	15		0	-15	-15	15	0	0	0	95.6	1.4
	0	0	0	0	0	10	0		0	0	0	-30	-15	-30	0		0	-15	-15	-15	0	0	0	96.8	1.6
	10	10	0	0	0	10	10		0	-15	0	-30	0	-15	15		0	-15	-15	0	0	0	0	96.9	1.4
	10	10	0	0	0	10	10		15	0	0	-30	-30	-15	15		0	-15	-15	0	-15	0	0	96.9	1.4
	10	10	0	0	0	10	10		0	0	0	-30	-15	-15	15		0	-15	-15	-15	-15	0	0	99	1.5
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0		0	-15	-15	-15	0	0	0	108.3	0.8
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	437.2	1.9

The original helices, with all charged termini residues present, are then used for the SuperComBiHelix procedure. The RMSD is the backbone RMSD to the β 2-adrenergic receptor crystal structure. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

Table 12 shows that the procedure in which the termini are alanized for SuperBiHelix and dealanized for the SuperComBiHelix works well. With

β 2-adrenergic receptor with A_{2A} adenosine template ϕ , θ and η values, the backbone RMSD to the β 2-adrenergic receptor crystal structure goes from 1.9 Å to 0.9 Å, with the closest possible RMSD given the angles sampled being 0.8 Å. Thus, the procedure will be permanently changed to include the alanization and dealanization steps.

All of the tests so far have kept the x, y and z values of the original template while giving the template the incorrect ϕ , θ and η values. However, while we have methods to calculate z based on the primary sequence, we do not have any way of calculating x or y for a receptor with an unknown structure. Thus, for the next test, β 2-adrenergic helices were given the x, y, ϕ , θ and η values the A_{2A} adenosine template, and SuperBiHelix and SuperComBiHelix were run, with the results seen in Table 13. As seen in Table 13, when the x and y values are incorrect, the SuperBiHelix and SuperComBiHelix procedures causes the backbone RMSD to the β 2-adrenergic receptor crystal structure to go from 2.0 Å to 1.6 Å, a modest improvement. The lowest possible RMSD, given the angles sampled, is 1.2 Å. The procedure does not do a good job at predicting the conformation of TM1, as seen in an alignment between the crystal structure and the structures before and after SuperBiHelix and SuperComBiHelix, shown in Figure 19. This is due to the fact that TM1 has the fewest interactions with other helices, because it does not have direct interaction with TM3 like the other TM helices. Additionally, Table 7 shows that in the x and y values, which are not sampled by SuperBiHelix, TM1 has larger deviations between the β 2-adrenergic and A_{2A} adenosine templates than any other helix.

Table 13. SuperComBiHelix results for the β 2-adrenergic receptor with A_{2A} adenosine template ϕ , θ , η , x and y values

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	0	0	0	0	10	0		0	0	0	-30	0	-30	-15		-30	-15	-15	-30	-15	0	15	164.6	1.6
	0	0	0	0	0	10	0		0	0	0	-30	0	-30	-30		-30	-15	-15	-30	-15	0	0	176.6	1.6
	0	0	0	0	0	10	0		0	0	0	-30	0	-30	15		-30	-15	-15	-30	-15	-15	15	177.7	1.7
	0	0	0	0	0	10	0		-15	0	0	-30	0	-30	-30		-30	-15	-15	-30	-15	-15	0	187.2	2
	0	0	0	0	0	10	0		0	15	0	-30	0	-30	-15		-30	-15	-15	-30	-15	0	15	188.2	1.7
	0	0	0	0	0	10	0		-15	0	0	-30	0	-30	30		-30	-15	-15	-30	-15	-15	15	196.7	2.1
	0	0	0	0	0	10	0		0	0	0	-30	0	-30	-30		-30	-15	-15	-30	-15	-15	0	198.9	1.6
	0	0	0	0	0	10	0		-15	0	0	-30	-15	-30	-30		-30	-15	-15	-30	-15	0	0	203.9	2.1
	0	0	0	0	0	10	0		-15	0	0	-30	0	-30	30		-30	-30	-15	-30	-15	-15	15	205	2.1
	0	0	0	0	0	10	0		0	0	0	-30	0	-15	-15		-30	-30	-15	-30	-15	0	15	206.4	1.7
	10	10	0	0	0	10	0		15	0	0	0	15	-30	0		0	-15	-15	-15	0	0	0	626.5	1.2
	6	8	5	-4	-1	11	4		18	2	2	-3	15	-25	3		5	-19	-13	-9	0	1	3	696	1
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	1032.7	2

The RMSD is the backbone RMSD to the β 2-adrenergic receptor crystal structure. The structure highlighted in orange is the structure closest to the crystal structure, given that the x and y values are from the A_{2A} adenosine template. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

For A_{2A} adenosine helices with the x , y , ϕ , θ and η values of the β 2-adrenergic template, SuperBiHelix and SuperComBiHelix causes the RMSD to the A_{2A} adenosine receptor crystal structure to go from 2.1 Å to 1.4 Å. This is a good improvement, given that the best RMSD possible, given the angles sampled, is 1.2 Å. These results are shown in Table 14, and the alignments between the crystal structure and the structures before and after SuperBiHelix and SuperComBiHelix are shown in Figure 20.

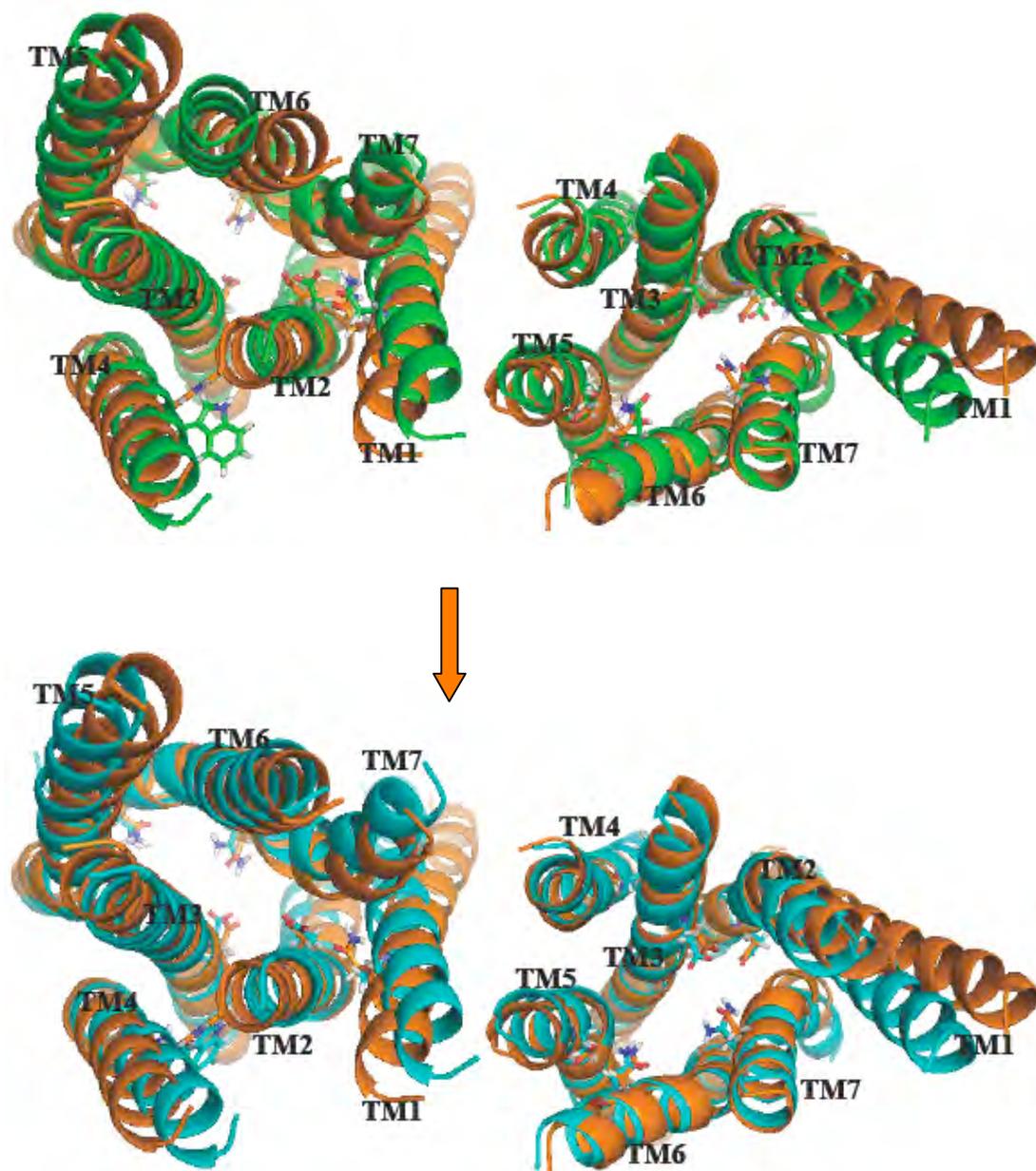


Figure 19. Intracellular (left) and extracellular (right) views of the β_2 -adrenergic receptor helices in the A_{2A} adenosine template before (top, green) and after (bottom, blue) SuperBiHelix and SuperComBiHelix. The β_2 -adrenergic crystal structure (orange) is present for comparison. The residues shown are those that are important for ligand binding or highly conserved in class A GPCRs. They are N51(1), D79(2), D113(3), W158(4), S204(5), N293(6) and N312(7).

Table 14. SuperComBiHelix results for the A_{2A} adenosine receptor with β 2-adrenergic template ϕ , θ , η , x and y values

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	-10	-10	0	0	-10	0		-15	-15	0	30	0	0	0		0	30	15	-30	0	0	0	224.2	1.4
	0	-10	-10	0	0	-10	0		-30	-30	0	30	0	0	0		0	30	15	-30	0	0	0	235.9	1.7
	0	-10	-10	0	0	-10	0		-15	-15	0	30	-15	15	0		0	30	15	-30	-15	0	0	238.7	1.4
	0	-10	-10	0	0	-10	0		-30	-15	0	30	0	0	0		15	30	15	-30	0	0	0	240	1.6
	0	-10	-10	0	10	-10	0		-30	-15	0	30	15	0	0		15	30	15	-30	30	0	0	241.1	2
	0	-10	-10	0	0	-10	0		-15	-15	0	30	0	30	0		15	30	15	-30	0	0	0	243.5	1.4
	0	-10	-10	0	0	-10	0		-15	-15	0	30	0	0	0		15	30	15	-30	0	0	0	243.7	1.4
	0	-10	-10	0	10	-10	0		-15	-15	0	30	15	-15	0		0	30	15	-30	30	0	0	244.3	1.8
	0	-10	-10	0	10	-10	0		-15	-15	0	30	15	0	0		0	30	15	-30	30	0	0	245.2	1.8
	0	-10	-10	0	0	-10	0		-30	-15	0	30	0	0	0		0	30	15	-30	0	0	0	249.4	1.6
	-6	-8	-5	4	1	-11	-4		-18	-2	-2	3	-15	25	-3		-5	19	13	9	0	-1	-3	280.1	1
	-10	-10	0	0	0	-10	0		-15	0	0	0	-15	30	0		0	15	15	15	0	0	0	354.7	1.2
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	849.7	2.1

The RMSD is the backbone RMSD to the A_{2A} adenosine receptor crystal structure. The structure highlighted in orange is the structure closest to the crystal structure, given that the x and y values are from the β 2-adrenergic template. The structure highlighted in orange is the crystal structure. The structure highlighted in green is the original structure in the incorrect template. The structure highlighted in yellow is the closest structure to the crystal structure given the angles sampled.

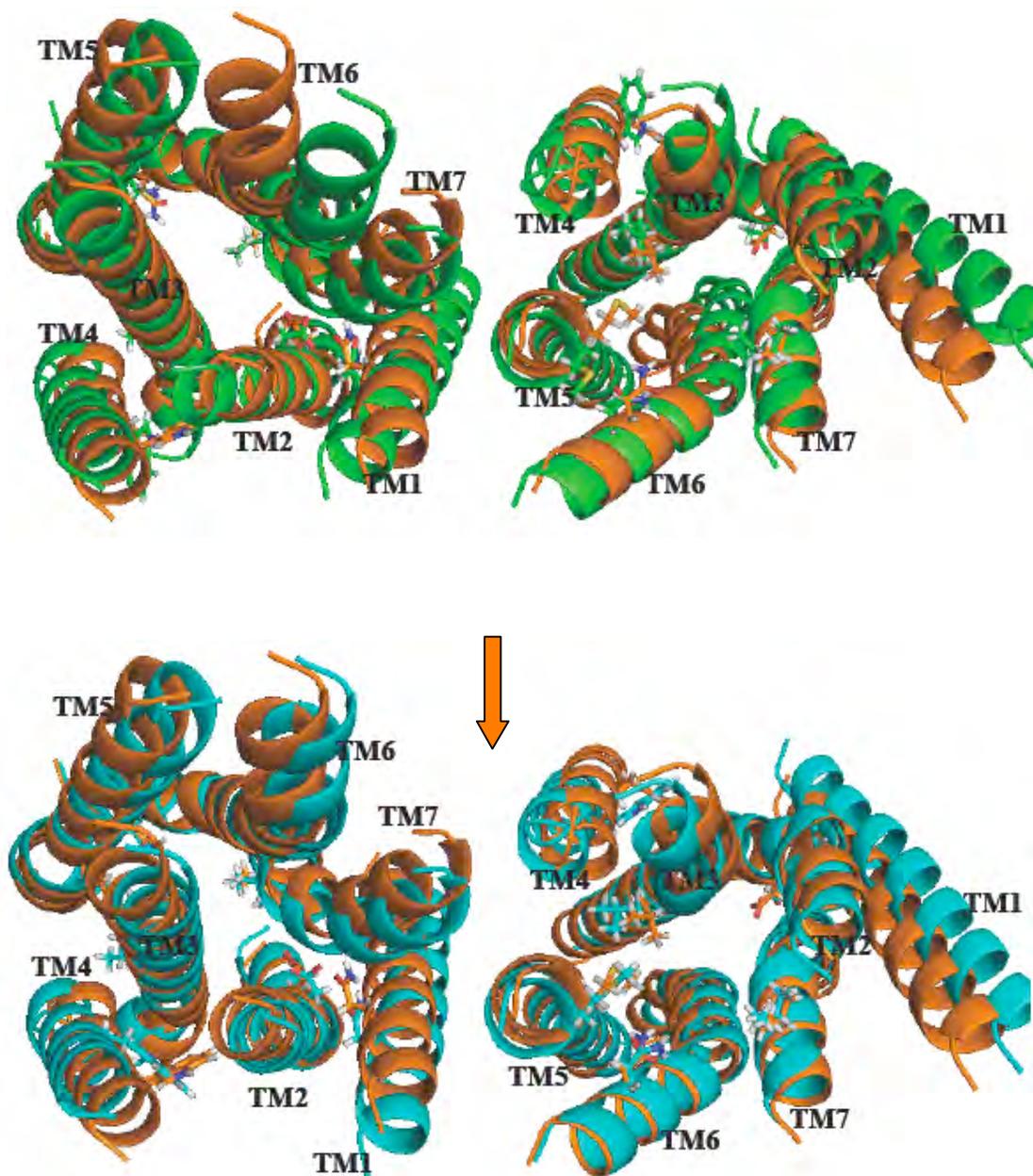


Figure 20. Intracellular (left) and extracellular (right) views of the A_{2A} adenosine receptor helices in the β 2-adrenergic template before (top, green) and after (bottom, blue) SuperBiHelix and SuperComBiHelix. The A_{2A} adenosine crystal structure (orange) is present for comparison. The residues shown are those that are important for ligand binding or highly conserved in class A GPCRs. They are N24(1), D52(2), L85(3), W129(4), M177(5), N253(6) and I274(7).

The Effect of SuperBiHelix on Binding Site Predictions

Although RMSD is a good metric for testing SuperBiHelix, it does not take ligand binding into account. One of the main purposes of predicting GPCR structures is for drug design, so it is important to measure how well ligand binding can be predicted in structures predicted by SuperBiHelix. Thus, carazolol was docked into the β 2-adrenergic structures before and after SuperBiHelix and SuperComBiHelix, and ZM241385 was docked into the A_{2A} adenosine structures before and after SuperBiHelix and SuperComBiHelix. Then these docked results were compared to the ligand-bound crystal structures, to see whether SuperBiHelix improved docking. The ligands were also docked into the ligand-free crystal structure for purposes of comparison.

Each structure was prepared for docking by mutating all the bulky, nonpolar residues to alanine. Then a sphere set representing the binding region was formed by taking a 2.0 Å radius around the coordinates of the crystal-bound ligand. The ligands were assigned Mulliken charges from quantum mechanics (B3LYP flavor of DFT using the 6-31G** basis set, calculated with Jaguar). GenDock, which is based on Dock6,^[49] was used to dock the ligand with the DarwinDock method. DarwinDock generates a large number of poses using Dock6, then clusters using Voronoi clustering. The program then adds 5000 new ligand poses (again from Dock6). If the fraction of new families is less than 1/20, then completeness is achieved. The program then scores the family heads of the Voronoi families. Then, 1/10 of the families are scored completely based on

the family head energies. Finally, the top 100 structures are passed on to the next step.

For each of these 100 ligand configurations, the bulky nonpolar residues are dealanized back to their original form, and all of the side chains in a 4 Å unified binding site, defined as all residues within a 4 Å radius of the docked ligand in all of the complexes, are reassigned with SCREAM. The complexes are ranked by total energy, and the best 50% are kept for neutralization. The complexes are neutralized, and then ranked by total energy, and the best 50% are kept for minimization. The 4.0 Å unified binding site is minimized for 50 steps. The complexes are ranked by total energy, and the best 50% are kept for the final step, resulting in 13 final structures. All of the full complexes are minimized for either 500 steps or to a 0.25 kcal/molÅ RMS force threshold. The final 13 structures are put through a quench-anneal cycle (50 K to 600 K and back over 11.5 ps) with the charged forcefield, selecting the configuration with the lowest potential energy structure. Then the structures are reneutralized and reminimized.

The 13 final docked structures for each structure were all compared to the ligand-bound crystal structure. The similarity to the ligand-bound crystal structure was measured with the contact RMSD. This is calculated by first determining the closest contacts for the ligand on each residue for the crystal structure, and finding their contact distances. Then the same contact distances are determined for the predicted ligand-bound complex. Finally, the RMSD is calculated between the two sets of distances.

For carazolol docked into the ligand-free β 2-adrenergic crystal structure, the contact RMSD is 2.4 Å. For the β 2-adrenergic helices in the A_{2A} adenosine template before SuperBiHelix, the lowest contact RMSD in the final 13 docked structures is 4.5 Å, and after SuperComBiHelix it is 4.4 Å. These bound structures are seen in Figure 21. Although there is very little improvement in the contact RMSD from SuperBiHelix, inspection of the bound structures shows marked improvement. In the crystal structure, carazolol forms strong hydrogen bonds with D113(3) and N312(7). In the structure before SuperBiHelix, carazolol only has a hydrogen bond with D113(3). However, after SuperComBiHelix, carazolol makes strong hydrogen bonds with both D113(3) and N312(7). So, SuperBiHelix does make the binding site more like that of the crystal structure.

For ZM241385 docked into the ligand-free A_{2A} adenosine crystal structure, the contact RMSD is 2.4 Å. For the A_{2A} adenosine helices in the β 2-adrenergic template before SuperBiHelix, the lowest contact RMSD in the final 13 docked structures is 4.6 Å, and after SuperComBiHelix it is 3.4. Thus, SuperBiHelix makes the binding site much more like that of the crystal structure. As seen in Figure 22, which shows the docked structures, the docked ligand in the best energy SuperComBiHelix structure is very similar to the pose in the crystal structure. They both make strong hydrogen bonds with N253(6). The docked ligand in the structure before SuperBiHelix takes a different pose, and does not form any hydrogen bond with N253(6).

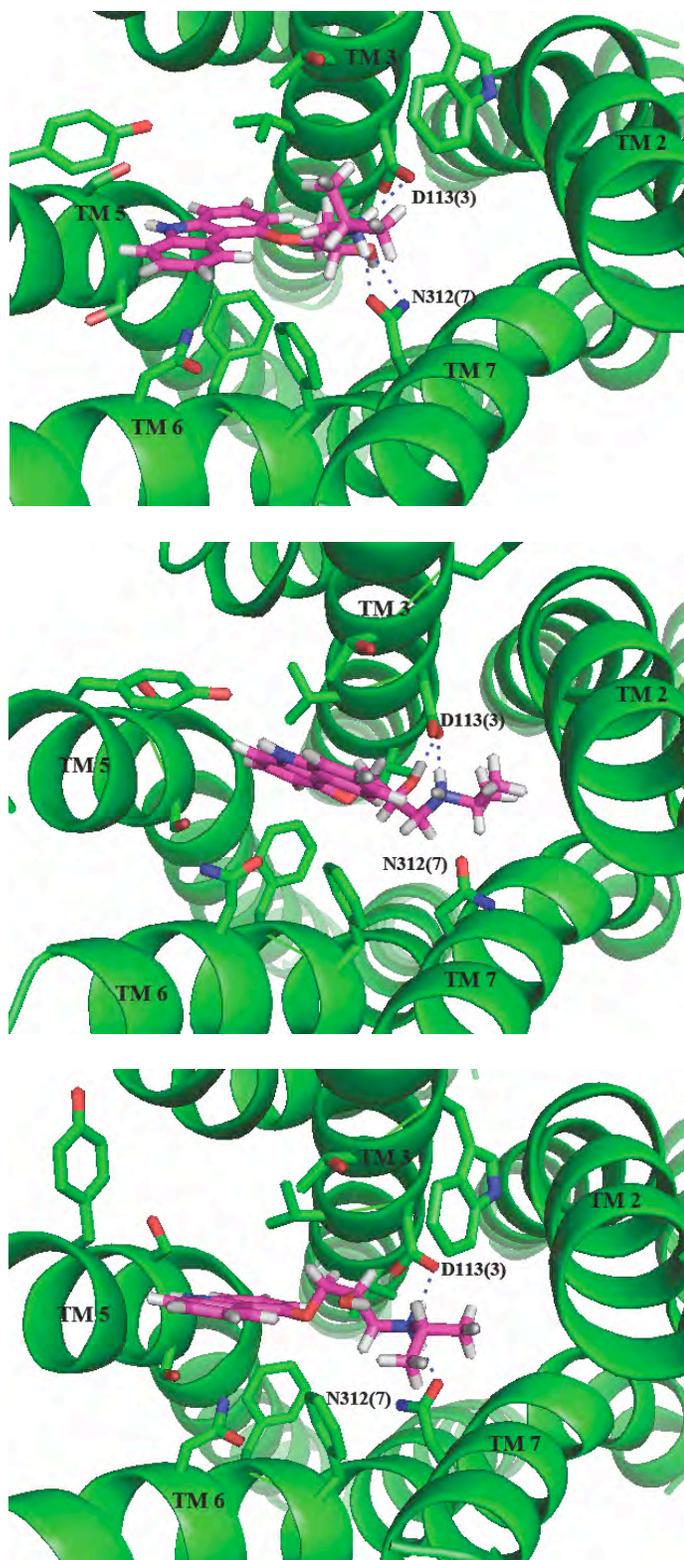


Figure 21. The carazolol-bound β 2-adrenergic crystal structure (top), along with carazolol docked into the β 2-adrenergic helices in the A2A adenosine template before SuperBiHelix (middle) and after SuperComBiHelix (bottom).

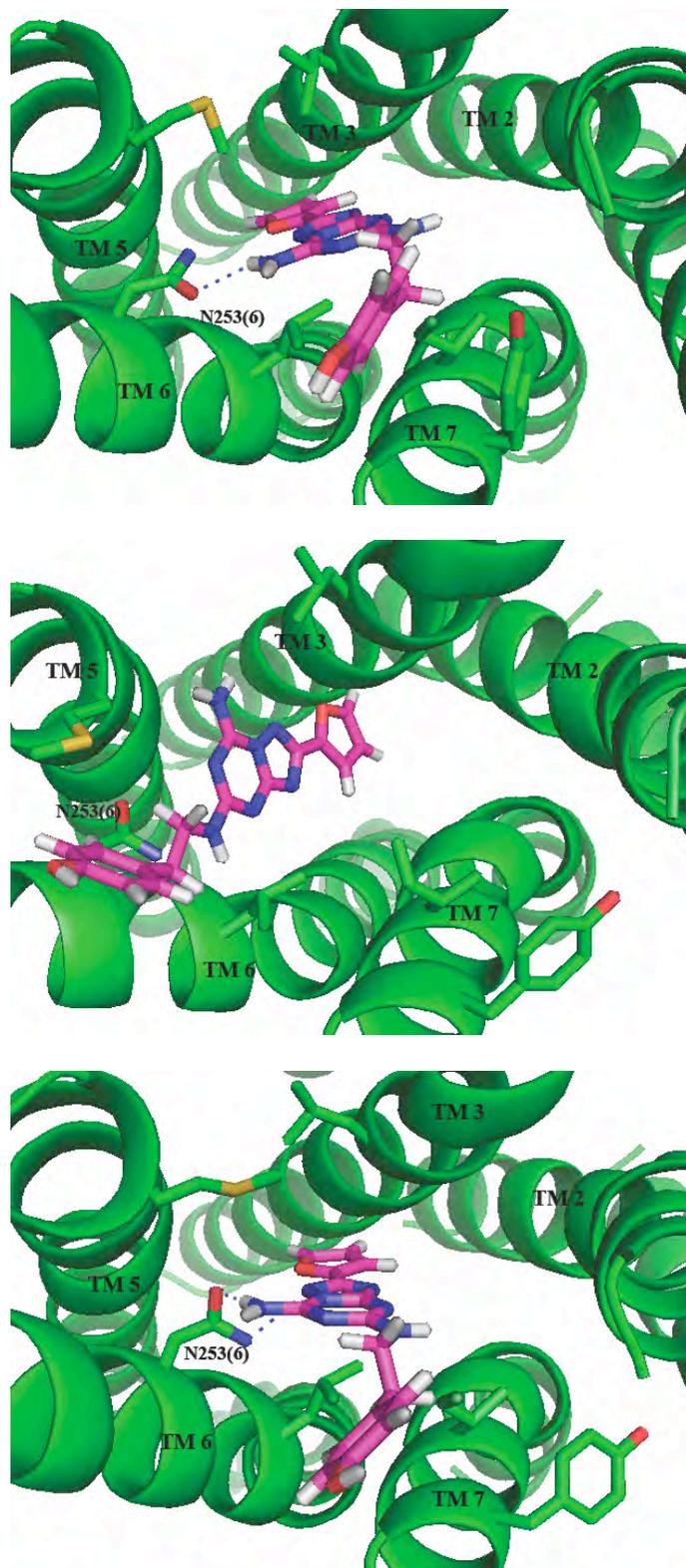


Figure 22. The ZM241385-bound A_{2A} adenosine crystal structure (top), along with ZM241385 docked into the A_{2A} adenosine helices in the β₂-adrenergic template before SuperBiHelix (middle) and after SuperComBiHelix (bottom).

While SuperBiHelix improves the binding site predictions for both β 2-adrenergic helices in the A_{2A} adenosine template and A_{2A} adenosine helices in the β 2-adrenergic template, it has more effect on the A_{2A} adenosine helices in the β 2-adrenergic template. This agrees with the RMSD calculations for SuperBiHelix, in which there is a larger effect on the A_{2A} adenosine helices in the β 2-adrenergic template than the β 2-adrenergic helices in the A_{2A} adenosine template.

SuperBiHelix on Crystal Structures

The SuperBiHelix and SuperComBiHelix procedures were developed by testing them on helices from one crystal structure in the template of another crystal structure. It is also important to determine whether performing SuperBiHelix and SuperComBiHelix on a crystal structure itself returns the original structure. Therefore SuperBiHelix and SuperComBiHelix were run on the A_{2A} adenosine receptor and the β 2-adrenergic crystal structure, sampling θ with values of -10° , 0° and 10° , ϕ with values of -30° , -15° , 0° , 15° and 30° , and η with values of -30° , -15° , 0° , 15° and 30° . The first test was to see how well the quadhelix protocol that determines which conformations to include in the SuperBiHelix bundle energy calculations worked. In order for the original crystal structure to show up in the best energy SuperBiHelix and SuperComBiHelix structures, the crystal conformation for each helix must be in that helix's top 36 conformations. The ranking of the crystal conformation for each helix for the A_{2A} adenosine receptor and the β 2-adrenergic receptor is shown in Table 15.

Table 15. The ranking of the crystal structure conformation for each helix after the quadhelix protocol that determines which helix conformations to include in the SuperBiHelix bundle energy calculations

TM	$\beta 2$	A _{2A}
1	1	1
2	1	1
3	1	4
4	2	7
5	3	4
6	12	2
7	1	1

Results are for both the A_{2A} adenosine receptor and the $\beta 2$ -adrenergic receptor.

The crystal structure conformation for each helix is indeed in the top 36 helical conformations. In fact, for many helices, the crystal structure is the best conformation. Even the worst ranking for a helix is for TM6 in the $\beta 2$ -adrenergic receptor, ranked number 12. So, the quadhelix protocol works well for the A_{2A} adenosine receptor and the $\beta 2$ -adrenergic receptor crystal structures.

The next test was to determine whether the ranking of the top 1000 SuperBiHelix structures is improved by SuperComBiHelix. The SuperBiHelix and SuperComBiHelix results are shown for the $\beta 2$ -adrenergic receptor crystal structure in Table 16. In the SuperBiHelix structures, the crystal structure is rank 78, and after SuperComBiHelix, it is rank 25. Not only does SuperComBiHelix cause significant improvement in the rank of the crystal structure, it also slightly improves the backbone RMSD values of the 10 structures.

Table 16. SuperBiHelix and SuperComBiHelix results for the β 2-adrenergic receptor crystal structure

SuperBiHelix

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	0	0	0	0	0	0		0	0	0	-15	0	-30	-15		0	0	0	0	0	0	0	338.9	0.7
	0	0	0	0	0	0	0		0	0	0	-15	0	-30	-15		0	0	0	15	0	0	0	341.2	0.7
	0	0	0	0	0	0	0		0	0	0	-15	0	-15	-30		0	0	0	0	0	0	0	343	0.6
	0	0	0	0	0	0	0		0	0	0	-15	0	15	0		0	0	0	0	0	0	0	343.3	0.5
	0	0	0	0	0	0	0		0	0	0	-15	0	-15	-15		0	0	0	0	0	0	0	343.8	0.5
	0	0	0	0	0	0	0		0	0	0	-15	-15	15	0		0	0	0	0	-15	0	0	344.4	0.6
	0	0	0	0	0	0	0		0	0	0	-15	0	-15	-30		0	0	0	15	0	0	0	345.4	0.7
	0	0	0	0	0	0	0		0	0	0	-15	0	15	15		0	0	0	0	0	0	0	345.5	0.5
	0	0	0	0	0	0	0		0	0	0	-15	0	15	0		0	0	0	15	0	0	0	345.7	0.5
	0	0	0	0	0	0	0		0	0	0	0	0	-30	-15		0	0	0	15	0	0	0	346.3	0.7
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	357.6	0

SuperComBiHelix

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	0	0	0	0	0	0		0	0	0	15	0	0	-15		0	0	0	0	0	0	0	12.7	0.4
	0	0	0	0	0	0	0		0	0	0	0	0	0	-15		0	0	0	0	0	0	0	13.5	0.4
	0	0	0	0	0	0	0		0	0	0	-15	0	-15	0		0	0	0	0	0	0	0	18	0.4
	0	0	0	0	0	0	0		0	0	0	-15	-15	-15	0		0	0	0	0	-15	0	0	25.3	0.6
	0	0	0	0	0	0	0		0	0	0	-15	-15	0	-15		0	0	0	0	-15	0	0	27.2	0.6
	0	0	0	0	0	0	0		0	0	0	0	0	-15	0		0	0	0	0	0	0	0	28.9	0.4
	0	0	0	0	0	0	0		0	0	0	-15	0	0	-15		0	0	0	0	0	0	0	29.3	0.4
	0	0	0	0	0	0	0		0	0	0	15	-15	15	15		0	0	0	0	-15	0	0	30.6	0.6
	0	0	0	0	0	0	0		0	0	0	15	0	0	0		0	0	0	0	0	0	0	31.3	0.3
	0	0	0	0	0	0	0		0	0	0	-15	0	0	15		0	0	0	0	0	0	0	31.3	0.4
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	46.4	0

The top ten structures for both SuperBiHelix and SuperComBiHelix are shown, along with the crystal structure, which is highlighted in orange. The crystal structure shows up as rank 78 for SuperBiHelix and rank 25 for SuperComBiHelix.

The SuperBiHelix and SuperComBiHelix results for the A_{2A} adenosine receptors crystal structure are in Table 17. The rank of the crystal structure goes from 2nd in SuperBiHelix to 6th in SuperComBiHelix. Although SuperComBiHelix makes the crystal structure rank slightly worse than in SuperBiHelix, SuperComBiHelix significantly improves the backbone RMSD values of the top

ten structures. Thus, SuperBiHelix and SuperComBiHelix are successful for both the β 2-adrenergic and A_{2A} adenosine crystal structures.

Table 17. SuperBiHelix and SuperComBiHelix results for the A_{2A} adenosine receptor crystal structure

SuperBiHelix																									
θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	0	0	-10	10	0	0		0	0	15	-30	0	0	0		0	0	0	-30	30	0	0	392.6	1.3
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	396.9	0
	0	0	0	-10	0	0	0		0	0	15	-30	0	-15	0		0	0	0	-30	-15	0	0	398.3	1
	0	0	0	-10	10	0	0		0	0	15	-30	0	0	0		0	0	15	-30	30	0	0	400.6	1.3
	0	0	0	-10	0	0	0		0	0	15	-30	0	0	0		0	0	0	-30	0	0	0	400.7	0.9
	0	0	0	-10	0	0	0		0	0	15	-30	0	0	0		0	0	0	-30	-15	0	0	401.2	0.9
	0	0	0	-10	10	0	0		0	0	15	-30	-15	0	0		0	0	0	-30	15	0	0	401.3	1.3
	0	0	0	0	0	0	0		0	0	0	15	0	0	0		0	0	0	0	0	0	0	402.4	0.4
	0	0	0	-10	10	0	0		0	-15	15	-30	0	0	0		0	0	0	-30	30	0	0	402.4	1.3
	0	0	0	-10	0	0	0		0	0	15	-30	-15	0	0		0	0	0	-30	0	0	0	402.8	1

SuperComBiHelix																									
θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)	RMSD (Å)
	0	0	0	0	0	0	0		0	0	0	0	-15	0	0		0	0	0	0	0	0	0	59.1	0.4
	0	0	0	0	0	0	0		0	0	-15	0	0	0	0		0	0	0	0	0	0	0	71.2	0.4
	0	0	0	0	0	0	0		0	0	-15	0	0	15	0		0	0	0	0	0	0	0	74.7	0.6
	0	0	0	0	0	0	0		0	0	-15	0	0	-15	0		0	0	0	0	0	0	0	76.3	0.7
	0	0	0	0	0	0	0		0	0	0	-15	0	-15	0		0	0	0	0	0	0	0	77	0.6
	0	0	0	0	0	0	0		0	0	0	0	0	0	0		0	0	0	0	0	0	0	78	0
	0	0	0	0	10	0	0		0	0	0	0	0	-15	0		0	0	0	0	15	0	0	79.1	1
	0	0	0	0	0	0	0		0	0	0	0	0	-15	0		0	0	0	0	0	0	0	85.8	0.5
	0	0	0	0	0	0	0		0	0	0	0	0	-15	-15		0	0	0	0	0	0	0	86.2	0.6
	0	0	0	0	0	0	0		0	0	-15	0	0	15	15		0	0	0	0	0	0	0	87.8	0.6

The top ten structures for both SuperBiHelix and SuperComBiHelix are shown, which include the crystal structure, highlighted in orange.

The SuperComBiHelix results of both the β 2-adrenergic and A_{2A} adenosine crystal structures show that there is the most variation in the sweep angles of the helices. Neither receptor shows any variation in TM1 and TM2. The sweep

angle of TM3 differs from the crystal structure in the A_{2A} adenosine receptor, but not in the β 2-adrenergic receptor. The sweep angles of TM4, TM5, TM6 and TM7 vary for both receptors. Finally, TM5 is the only helix whose η value changes from the crystal structure, for both receptors. Thus, it seems that TM5 is most flexible in both receptor.

It must be taken into account that the ligand is not present during the SuperBiHelix and SuperComBiHelix procedures. The crystal structures are determined with the ligand bound, so the ligand-free crystal structure may still not be the lowest energy structure. The presence of the ligand would change the order of the structures. Additionally, the presence of the loops may change the ordering of the structures. However, even taking these factors into account, the SuperBiHelix procedure is a very promising method for the prediction of new GPCR structures.

Conclusion

SuperBiHelix and SuperComBiHelix allow for the efficient sampling of GPCR conformational space. This makes it possible to predict structures of receptors that are dissimilar to any experimental crystal structure. It also predicts an ensemble of low-lying structures, mirroring the flexibility of GPCR structures. When helices from one crystal structure are placed into the template of another structure, SuperBiHelix and SuperComBiHelix successfully move the experimental helices closer to their original template. The procedure also improves binding site predictions and makes ligand binding calculations more

accurate. The success of SuperBiHelix and SuperComBiHelix on experimental crystal structures will now lead to better predictions of GPCR structures and binding sites, and therefore more successful rational drug design.

Chapter 3

The Structure of the Orphan G Protein-Coupled Receptor GPR88 with Predicted Ligands

Abstract

GPR88 is an orphan GPCR, with no known ligands that bind to it. It has been identified as a novel target for psychiatric disorders. In order to determine which types of ligands that could bind to it, three low energy structures were predicted for GPR88. Based on inspection of the structures, lipids were predicted to bind to GPR88. The head group of a lipid would bind to R113(3) and R116(3) at the extracellular side of the receptor. The lipid tail would bind in an aliphatic pocket in the TM2-TM3-TM6-TM7 region. Three lipids were docked to the predicted GPR88 structures: S1P, LPA and FFA. The predicted bound structures show that all three lipids would likely bind to GPR88, with the head group forming hydrogen bonds to R113(3) and R116(3). The residues contributing favorably to the binding of the lipid tails would be L120(3), L124(3), W84(2), W322(7), C325(7), L305(6), S321(7), L328(7), V301(6), and S127(3), in order of decreasing predicted strength of interaction. The predicted bound complexes offer good suggestions for binding and mutagenesis experiments that could help validate the predicted structures.

Introduction

A GPCR for which the endogenous ligand is unknown is called an orphan receptor. One such orphan receptor that was discovered in 2000 is the GPCR GPR88.^[50] In fact, not only is the endogenous ligand unknown for GPR88, no ligand is known to bind the GPR88. Additionally, its function is unknown. It is found in both human and rodent striatum,^[50, 51] which implies that its function may be psychiatric. Supporting this is the fact that schizophrenia-like phenotypes are displayed by GPR88 knock-out mouse.^[52] In addition, antidepressant treatments affect levels of GPR88 expression.^[53] Finally, GPR88 has been identified as a novel target for movement disorders as well as psychiatric disorders.^[54]

The primary sequence of GPR88 shows the highest homology to biogenic amine receptors. However, some highly conserved residues in biogenic amine receptors, including the aspartic acid in TM3 that makes a salt bridge with the protonated amine of biogenic amine ligands, are not present in GPR88. This implies that GPR88 may represent a novel subtype of GPCR.^[50] A model 3D structure of GPR88 could help to determine the type of ligands that will bind to GPR88, and perhaps even help classify the receptor. However, the lack of any experimental structural, ligand binding or mutation data will make any predictions difficult to verify even indirectly. Hopefully, predictions will lead to direct binding and mutation experiments with GPR88 that could lead to rational drug design for GPR88.

Methodology

The Membstruk methodology for determining the serotonin receptors in Chapter 1 gave reasonable structures, but it was fairly ad hoc and it required knowledge of the system, including important residues in the system for binding and stabilization. This method would not work for GPR88 because there is no experimental knowledge of its structure or binding. Instead, the SuperBiHelix method described in Chapter 2 was used. The starting structure for SuperBiHelix was the best structure from the BiHelix procedure.

In order to use the BiHelix procedure, the helices must first be predicted that will be put into each experimental template. A new procedure, PredicTM, was developed in the group to predict the TM helices of the protein. In this procedure, like the old TM prediction method, sequences are found with BLAST that are related to the target sequence. For GPR88, BLAST was run to get structures up to an e-value of 0.2, which gave 290 structures. The top 20 structures closest to human GPR88 are shown in Table 18. Then MAFFT^[55] is used to align these sequences, and any segments of the alignment corresponding to gaps in the target sequence are removed. The White octanol hydrophobicity scale^[56], which is a thermodynamic scale based on the partitioning of amino acids from water to octanol, is used to assign a hydrophobicity value to each amino acid in the alignment. These values are then averaged over windows of residue length 7–21, then averaged over these averages to decrease the noise in the hydrophobic profile. The hydrophobic

profile for GPR88, in which seven hydrophobic peaks that represent the seven TM helices, is shown in Figure 23.

Table 18. Proteins sequentially related to GPR88

Accession	E-value	% Identity	Receptor
Q9GZN0	0	100	GPR88 human
Q9EPB7	0	95	GPR88 mouse
Q9ESP4	0	94	GPR88 rat
P25962	2.564e-14	23	Beta-3 adrenergic mouse
P79148	6.3153e-13	21	Beta-1 adrenergic dog
Q9TT96	2.3998e-12	20	Beta-1 adrenergic bovine
P13945	5.3462e-12	22	Beta-3 adrenergic human
P28221	5.3462e-12	19	5-HT1D human
Q9TST6	6.9824e-12	21	Beta-1 adrenergic cat
O02666	9.1193e-12	22	Alpha-1D adrenergic rabbit
Q28524	1.191e-11	21	Beta-3 adrenergic rhesus
P08588	1.191e-11	19	Beta-1 adrenergic human
P23944	1.191e-11	19	Alpha-1D adrenergic rat
Q28927	1.5555e-11	19	Beta-1 adrenergic sheep
O02662	2.0316e-11	23	Beta-3 adrenergic dog
Q95252	2.6533e-11	21	Beta-3 adrenergic pig
P97714	4.5259e-11	19	Alpha-1D adrenergic mouse
P34971	1.0083e-10	20	Beta-1 adrenergic mouse
P47899	2.2462e-10	19	Beta-1 rhesus
Q7TQP2	3.8314e-10	19	GP135 mouse

The top 20 sequences were found by BLAST, ordered by increasing e-value.

The hydrophobic profile does a good job of predicting the transmembrane regions of the seven helices. However, GPCR crystal structures show that these helices often extend past the hydrophobic membrane region of the lipid bilayer. The hydrophobic profile cannot detect this helix extension, but secondary structure prediction can. Thus, secondary structure prediction can be used in conjunction with the hydrophobic profile. The secondary structure servers Porter^[57] and APSSP2^[58] were used to do the predictions. The secondary structures along with the raw predictions are shown in Table 19.

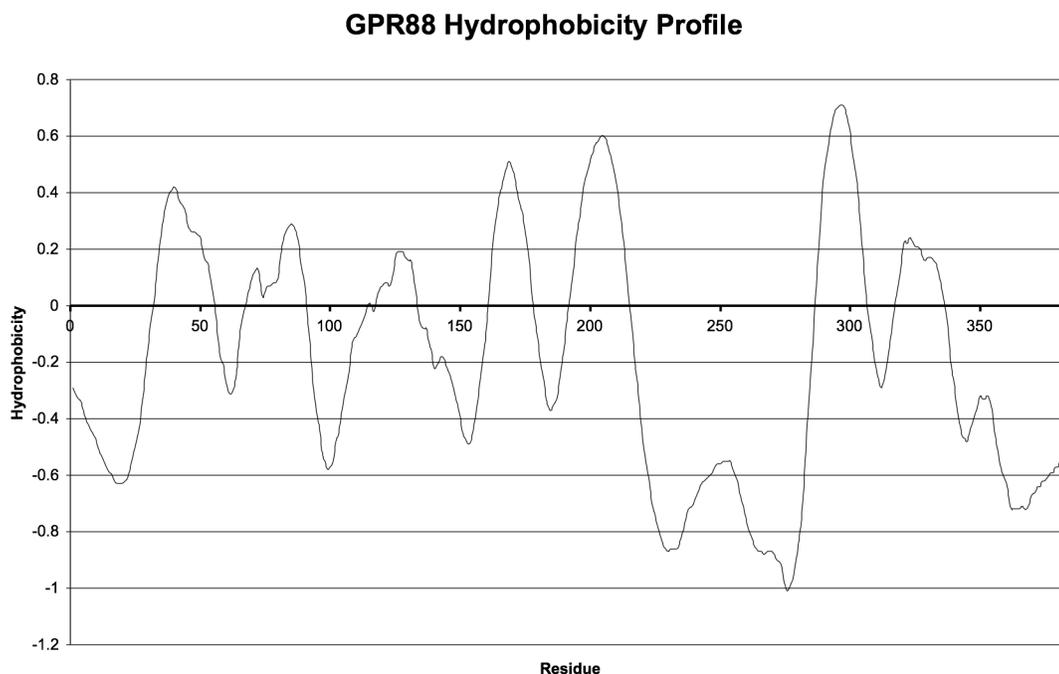


Figure 23. The hydrophobic profile for human GPR88, based on a MAFFT alignment of related sequences, with the seven peaks representing the transmembrane helices. The hydrophobicity scale is thermodynamic, so the cutoff for the TM part of the helices is a hydrophobicity value 0.

The entire sequence of GPR88 was inputted into the secondary structure servers to see the length of the transmembrane helix predictions. The longer of the two secondary structure predictions was taken as the final prediction for each helix, except for TM7. In TM7, the Porter predict predicts TM7 to extend out toward the C-terminus, but examination of the hydrophobic profile reveals that the extension is probably the eighth helix, which exists in some GPCRs and runs parallel to the membrane.

Table 19. TM helix predictions for human GPR88

TM	Predicted TM Region	Prediction Method
1	RIPVSLLYSGLAIGGTLANGMVIY VSLLYSGLAIGGTL WAGRRIPVSLLYSGLAIGGTLANGMVIYLVSS WAGRRIPVSLLYSGLAIGGTLANGMVIYLVSS	Raw Porter APSSP2 Final Prediction
2	NAFIVNGCAADLSVCALWMPQEAVLG AADLSVCALWMPQEAVL TSNAFIVNGCAADLSVCALWMPQEAVLGLL TSNAFIVNGCAADLSVCALWMPQEAVLGLL	Raw Porter APSSP2 Final Prediction
3	GLLGLGLTVSLLSHCL LTVSLLSHCLVALNRYLLI GSYRLLRGLLGLGLTVSLLSHCLVALNRYLL GSYRLLRGLLGLGLTVSLLSHCLVALNRYLLI	Raw Porter APSSP2 Final Prediction
4	MLALSWALALGLVLLLP ATYQALYQRRHTAGMLALSWALALGLVLL RRHTAGMLALSWALALGLVLLLP ATYQALYQRRHTAGMLALSWALALGLVLLLP	Raw Porter APSSP2 Final Prediction
5	PALLAAAALLAQTALLLHCYLG YPALLAAAALLAQTALLLHCYLGIVRR ALLAAAALLAQTALLLHCYLGIVRRVRSVKR YPALLAAAALLAQTALLLHCYLGIVRRVRSVKR	Raw Porter APSSP2 Final Prediction
6	LLCCVFLLATQPLVWVSLAS PRRAQRRLSGLSVLLCCVFLLA HPRRAQRRLSGLSVLLCCVFLLATQPLVWVSLASGF HPRRAQRRLSGLSVLLCCVFLLATQPLVWVSLASGF	Raw Porter APSSP2 Final Prediction
7	VHAASWLLCCALSALNPLLY GVHAASWLLCCALSALNPLLYTWRNEEFRRSVRSV GVHAASWLLCCALSALNPLLYTW GVHAASWLLCCALSALNPLLYTW	Raw Porter APSSP2 Final Prediction

The raw predictions are from the hydrophobic profile. The final predictions take the longer of the two secondary structure predictions, except for TM7, in which the longer prediction also includes TM8.

With the residues making up each TM helix having been predicted, the helix shapes are then predicted. Methods have been developed to perform molecular dynamics on canonical helices in order to predict the kinks and bends caused by helix breakers like proline. However, these methods do not work in the BiHelix procedure as well as using homology helices. Homology helices are obtained by

using one of the available GPCR crystal structures: the β 1-adrenergic receptor,^[4] β 2-adrenergic receptor,^[5] A_{2A} adenosine receptor,^[6] or rhodopsin.^[59] The target protein is aligned with each of these receptors, and then the residues in the crystal helices are mutated to the corresponding residues of the target. The helices are allowed to minimize to accommodate the new side chains. If the predicted helices for the target protein are longer than the crystal helices, then the crystal helix lengths are used.

Each set of homology helices based on an experimental crystal structure is then placed in its respective templates. When using homology helices, the hydrophobic centers of the helices are those of the corresponding experimental helices, as defined by the OPM database.^[60] The hydrophobic centers are aligned along the z-axis, while the template defines the x and y values in the plane as well as the tilt and sweep angles, θ and ϕ . In order to optimize the rotation of each helix on its own axis, BiHelix and ComBiHelix are run for each set of homology helices in their respective templates. Then the template that gives the lowest ComBiHelix energy is used for SuperBiHelix and SuperComBiHelix.

Results and Discussion

BiHelix and ComBiHelix

After running ComBiHelix for GPR88 in the available crystal templates, the β 2-adrenergic receptor template and homology helices gave the lowest energy for

GPR88. The ComBiHelix results for the top 10 structures of GPR88 in the β 2-adrenergic receptor template are given in Table 20.

Table 20. ComBiHelix results for GPR88 in the β 2-adrenergic receptor template

H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)
0	0	0	180	0	90	60	-376.7
0	0	0	180	0	60	60	-373.4
0	0	0	0	30	60	60	-371.5
0	0	0	0	0	90	60	-370.8
0	0	0	0	0	60	60	-369.0
0	0	0	210	0	90	60	-367.2
0	0	0	120	30	60	60	-367.1
0	0	0	150	30	60	60	-366.8
0	0	0	210	0	60	60	-365.5
0	0	0	120	0	60	60	-360.9

The zeros represent the same rotation for the helix as in the β 2-adrenergic receptor crystal structure. TM4 and TM5 are the most flexible rotationwise, while TM1, TM2, TM3 and TM7 are fixed.

The ComBiHelix results show that TM1, TM2 and TM3 have the same rotations as the β 2-adrenergic receptor. TM7 is rotated 60° clockwise from the β 2-adrenergic receptor, but is fixed in that position. TM6 prefers either a rotation of 60° or 90° from the β 2-adrenergic receptor. TM4 and TM5 are the most flexible rotationwise. Interestingly, the ComBiHelix results of the β 2-adrenergic receptor itself show that TM1, TM4 and TM5 are the most flexible rotationwise. Thus, it is not surprising that TM4 and TM5 are also flexible in GPR88. Additionally, among the proteins included in the BLAST search to create the hydrophobic profile for GPR88, TM5 has the least similarity to the other proteins in the alignment.

The lowest energy structure from the ComBiHelix procedure for GPR88 is shown in Figure 24. The predicted structure seems quite stable in that there are several strong interhelical hydrogen bonds. TM2 and TM3 are strongly coupled by a salt bridge near the extracellular side between R116(3) and E88(2) in addition to a strong hydrogen bond between R113(3) and Q87(2). The TM1-TM2-TM7 hydrogen bonding network present in class A GPCRs is found in the GPR88 structure, with hydrogen bonds between N49(1), D77(2) and N332(7). There is a hydrogen bonding network between the intracellular sides of TM2, TM3 and TM4, connecting T65(2), N137(3) and R156(4). However, the residues involved in that network are very near the termini of their helices, so they may interact with the head groups of the lipid bilayer instead of each other. Finally, there is a weak hydrogen bond between T205(3) and Q298(6). Given that the strongest hydrogen bonding networks involve TM1, TM2, TM3 and TM7, it is not surprising that the ComBiHelix analysis shows that those helices are nonflexible rotationwise, in order to preserve the hydrogen bonding networks.

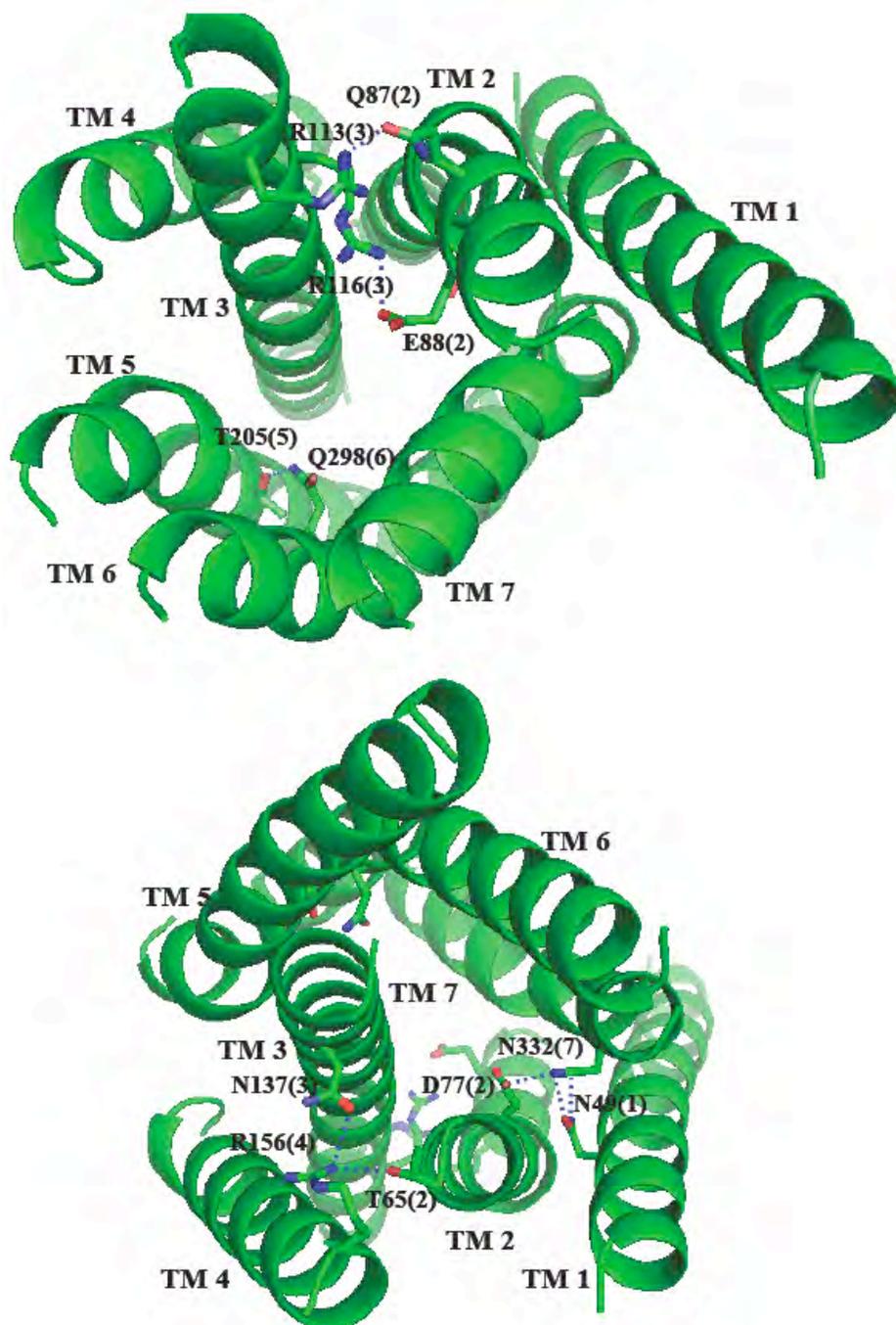


Figure 24. The lowest energy structure from ComBiHelix analysis on GPR88, in the β_2 -adrenergic receptor template. The top is the extracellular view and the bottom the intracellular view. The residues shown are those involved in interhelical hydrogen bonding, with the dotted lines showing hydrogen bonds.

SuperBiHelix and SuperComBiHelix

SuperBiHelix and SuperComBiHelix were run on the best BiHelix structure of GPR88 in the β 2-adrenergic receptor template. The angles sampled were -10° , 0° and 10° for the tilt angle, θ ; -30° , -15° , 0° , 15° and 30° for the sweep angle, ϕ ; and -30° , -15° , 0° , 15° and 30° for the rotation around the helical axis, η . These angles were chosen because an examination of experimental crystal structure templates shows that the templates do not vary by much more than these values. Additionally, it is assumed that the starting structure, the lowest energy BiHelix structure, is somewhat close to the global lowest energy structure. The results from SuperBiHelix and SuperComBiHelix are shown in Table 21.

Table 21. SuperComBiHelix results for GPR88

θ	H1	H2	H3	H4	H5	H6	H7	ϕ	H1	H2	H3	H4	H5	H6	H7	η	H1	H2	H3	H4	H5	H6	H7	Energy (kcal/mol)
	-10	0	0	0	10	10	10		-30	15	15	15	15	-30	0		15	-15	0	-15	-15	0	0	-468.8
	0	0	-10	0	10	0	0		-15	-15	15	30	0	-30	15		15	0	-15	-15	-15	0	15	-466.4
	0	0	-10	0	10	0	0		-15	-15	15	30	0	-15	15		15	0	-15	-15	-15	0	15	-465
	0	0	-10	0	10	0	0		-15	-15	15	15	0	-15	15		15	0	-15	-15	-15	0	15	-462.6
	-10	0	0	0	10	10	10		-30	15	15	-30	15	-30	0		15	-15	0	30	-15	0	0	-459.2
	-10	0	-10	10	-10	0	10		-15	-15	-15	0	15	0	30		0	-30	15	-15	-30	30	-30	-457.4
	0	0	-10	10	10	0	0		-15	-15	30	-30	0	-15	15		0	30	0	-15	-30	-30	-30	-454.2
	0	0	-10	0	10	0	0		-15	-15	15	15	0	-30	15		15	0	-15	-15	-15	0	15	-453.5
	-10	0	0	0	10	10	10		-30	15	15	0	15	-30	0		15	-15	0	0	-15	0	0	-450.5
	-10	0	0	0	10	10	10		-30	15	15	-15	15	-30	0		15	-15	0	0	-15	0	0	-450

The starting structure for the SuperBiHelix and SuperComBiHelix procedure, which would be 0° for all angles, is the lowest energy structure from BiHelix in the β 2-adrenergic receptor template.

The original structure, the lowest energy structure from ComBiHelix, does not show up in the 2000 best structures from SuperBiHelix. However, its energy (-461.3 kcal/mol) actually makes it the fifth lowest energy of the SuperComBiHelix structures. This is due to the fact that procedure that determines the best SuperBiHelix structures based on the energies of the TM1-TM2-TM3-TM7, TM2-TM3-TM4-TM5 and TM3-TM5-TM6-TM7 quadehelix bundles can overlook good energy structures. Thus, the original ComBiHelix structure was kept for further analysis and docking. An inspection of the SuperComBiHelix results shows that the sweep angles show the most variation, especially in TM4. The lowest energy structure, shown in Figure 25, was kept for docking and further analysis. The next three structures by energy only differ by 15° in the sweep angles of TM4 and TM6, so they are probably important structures. Thus, the second lowest energy structure, shown in Figure 26, was also kept for docking and further analysis.

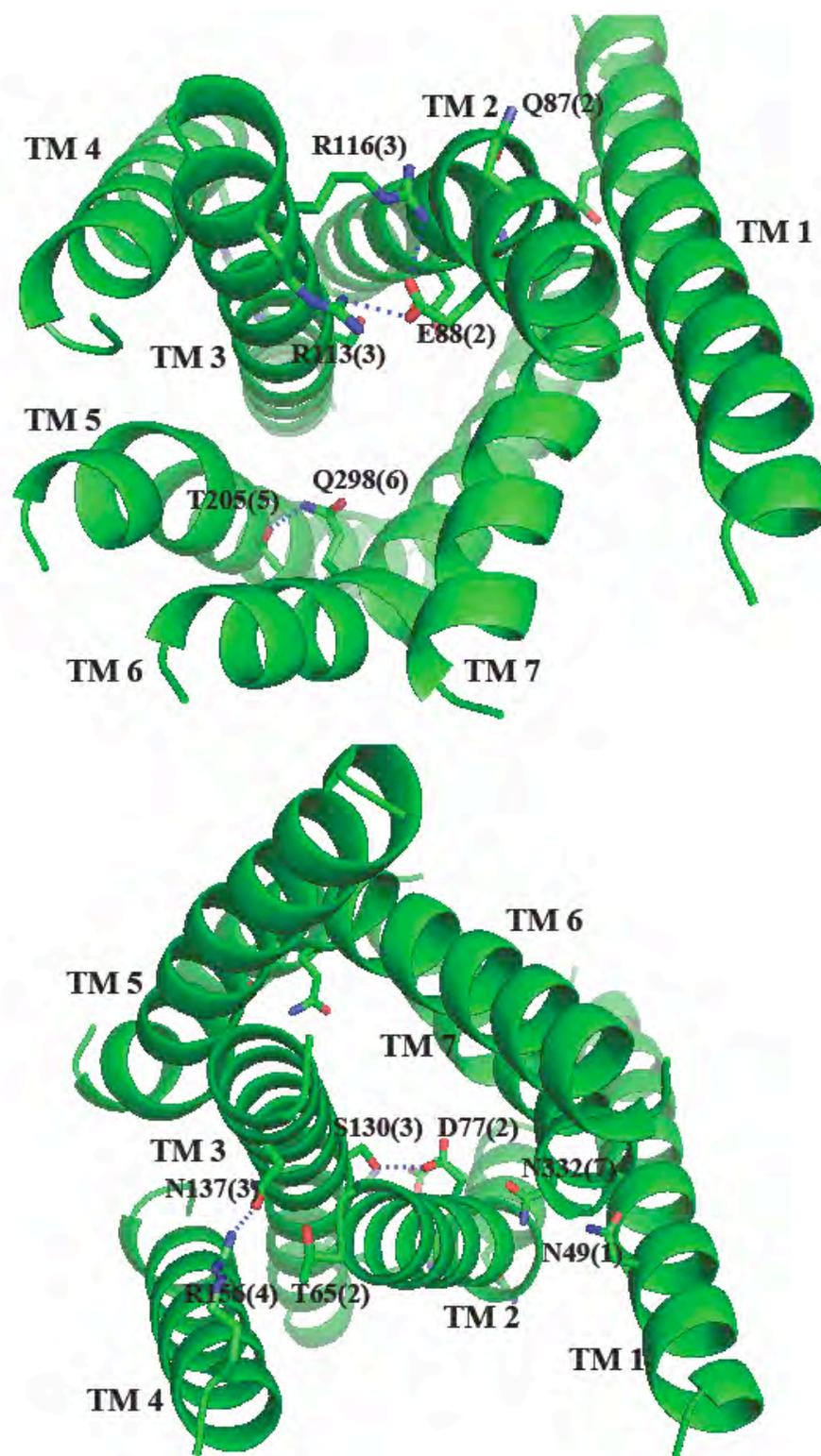


Figure 25. The lowest energy structure from SuperComBiHelix analysis on GPR88. The top is the extracellular view and the bottom the intracellular view. The residues shown are those involved in interhelical hydrogen bonding, or those that were in the BiHelix structure, with the dotted lines showing hydrogen bonds.

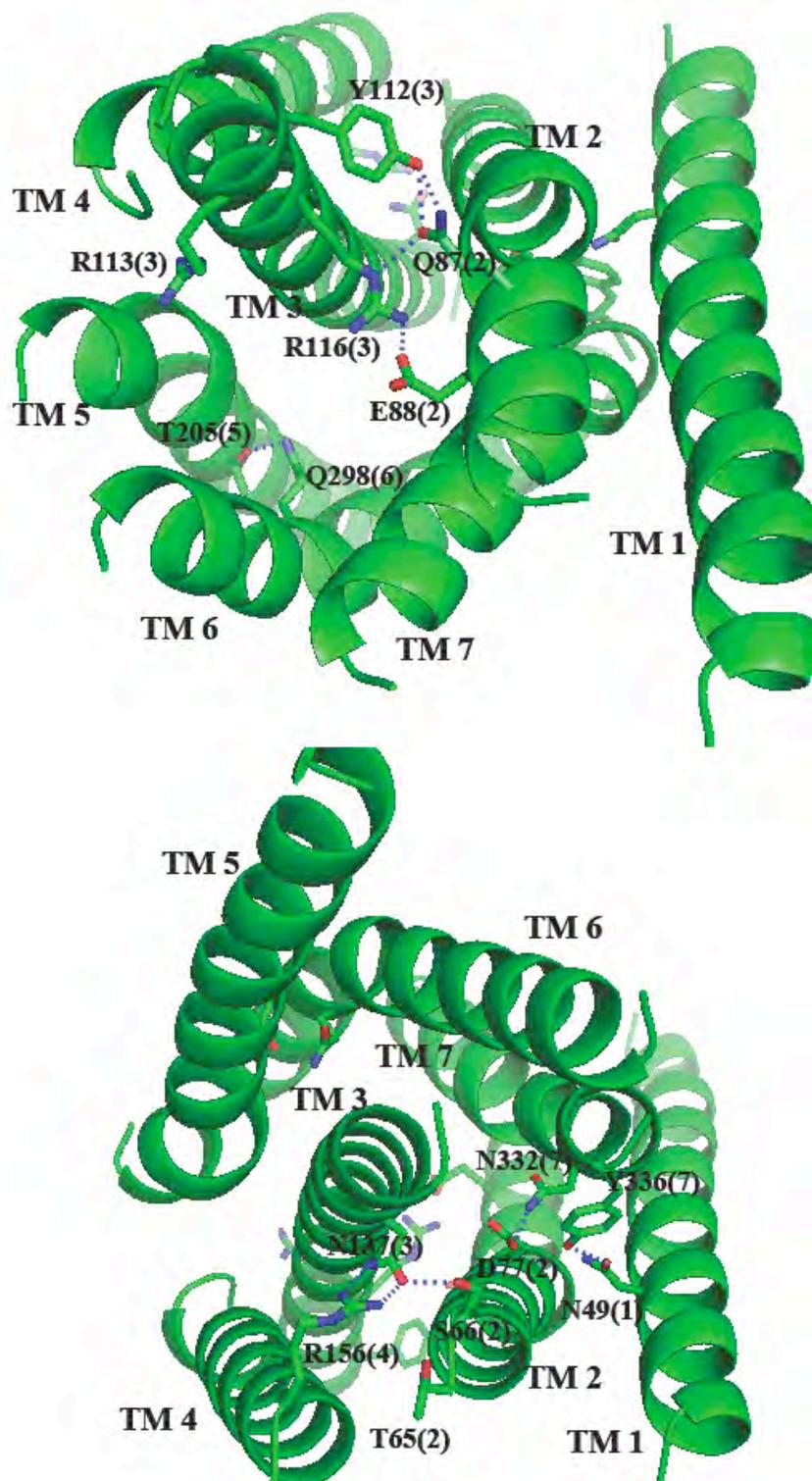


Figure 26. The second-lowest energy structure from SuperComBiHelix analysis on GPR88. The top is the extracellular view and the bottom the intracellular view. The residues shown are those involved in interhelical hydrogen bonding, or those that were in the BiHelix structures, with the dotted lines showing hydrogen bonds.

The two lowest energy structures from SuperComBiHelix have similar hydrogen bonding networks to the BiHelix structure.

- The lowest energy structure has a salt bridge between R113(3) and E88(2). However, there are no hydrogen bonds involving R116(3) or Q87(2). On the second-lowest energy structure, R116(3) is making a salt bridge with E88(2) and forming a hydrogen bond with Q87(2). Q87(2) is also in a hydrogen bond with Y112(3). R116(3) is not involved in any hydrogen bonding in the second-lowest energy structure.
- Both SuperComBiHelix structures have a hydrogen bond between T205(5) and Q298(6).
- Like the BiHelix structure, both structures have a hydrogen bond between R156(4) and N137(3). There is no interaction with T65(2) like in the BiHelix structure. The second-lowest energy structure does have a hydrogen bond between S66(2) and N137(3).
- The lowest energy structure does not have the TM1-TM2-TM7 hydrogen bonding network. Instead, D77(2) is making a hydrogen bond with S103(3). N49(1) and N332(7) are not involved in any interhelical hydrogen bonding. In the second-lowest energy structure, D77(2) is in a hydrogen bond with N337(2). Additionally, N49(1) is forming a hydrogen bond with Y336(7).

Based on interhelical hydrogen bonding, the second lowest energy ComBiHelix structure seems to be more stable than the lowest energy structure. Additionally, the fact that the second-, third- and fourth-lowest energy structures are almost

identical points to the fact that the structure is very important. However, the lowest energy structure was also kept for docking because it does have the best energy.

Ligand Docking

The docking of ligands to GPR88 is difficult because there are no known ligands that bind to the receptor. Based on a visual inspection of the three predicted structures of GPR88, there is an open, aliphatic binding pocket running vertically in the TM3-TM6-TM7-TM7 region. It looks like a lipid could dock in GPR88 with a negatively charged head group forming a salt bridge with either R113 or R116 on TM3 and the aliphatic tail running down into the aliphatic binding pocket. Three lipids that commonly bind to GPCRs were docked: lysophosphatidic acid (LPA), sphingosene-1-phosphate (S1P) and a free fatty acid (FFA). These lipids are shown in Figure 27. The ligands were built and minimized using the Dreiding force field with Mulliken charges from quantum mechanics (B3LYP flavor of DFT using the 6-31G** basis set, calculated with Jaguar). The carboxylic acid head group of FFA was given a -1 charge, and one hydrogen was added in the neutralization step. The phosphoric acid head group of LPA and S1P will have a -2 charge at biological pH. However, in the docking algorithm, a -2 charge could lead to artificially large long-distance Coulombic interactions. Thus, a -1 charge was given to the phosphoric acid head group, treating each of the three oxygens alike. In S1P, the nitrogen was protonated. In the neutralization step, one hydrogen was added to the -1 charged phosphate group to make it neutral in

LPA and S1P, and one hydrogen was taken away from the protonated nitrogen in S1P.

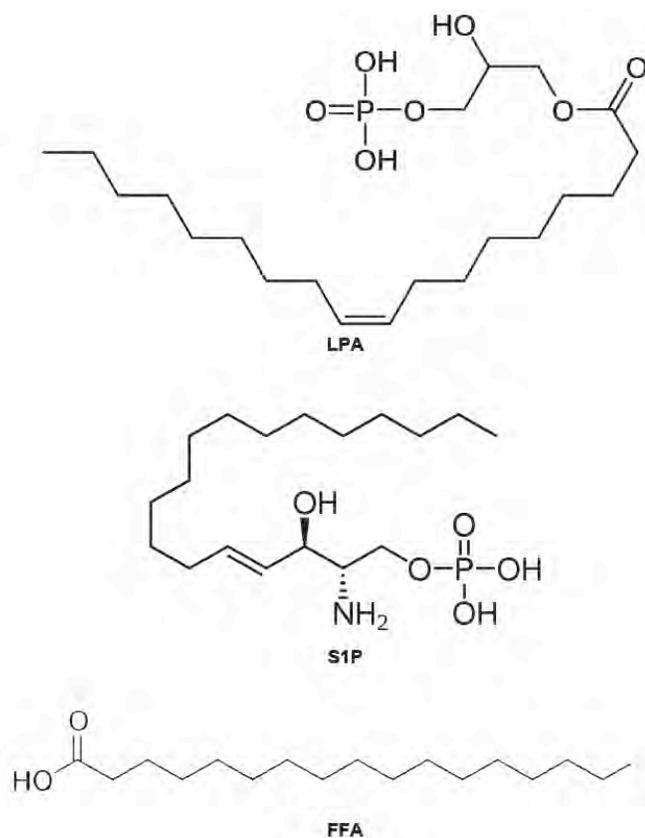


Figure 27. The three lipid molecules that were docked into GPR88.

Given the large number of rotatable bonds in a lipid, it would be very difficult to directly dock the entire lipid in the protein. Instead, just the head of each lipid was docked. First, all of the bulky nonpolar residues (tryptophan, tyrosine, phenylalanine, valine, isoleucine and leucine) were mutated to alanine in order to allow the ligand to find the best conformation. Then, a sphere region around R113 and R116 on TM3 was created as the binding region for the lipid head groups. GenDock, an in-group docking program based on Dock6^[49], was then

used to dock the ligand (or head group in this case) with the DarwinDock method. DarwinDock generates a large number of poses using Dock6, then clusters using Voronoi clustering. The program then adds 5000 new ligand poses (again from Dock6). If the fraction of new families is less than 1/20, then completeness is achieved. The program then scores the family heads of the Voronoi families. Then, 1/10 of the families are scored completely based on the family head energies. Finally, the top 100 structures are passed on to the next step.

For each of these 100 ligand configurations, the bulky nonpolar residues are dealanized back to their original form, and all of the side chains in a 4 Å unified binding site, defined as all residues within a 4 Å radius of the docked ligand in all of the complexes, are reassigned with SCREAM. The complexes are ranked by total energy, and the best 50% are kept for neutralization. The complexes are neutralized, and then ranked by total energy, and the best 50% are kept for minimization. The 4.0 Å unified binding site is minimized for 50 steps. The complexes are ranked by total energy, and the best 50% are kept for the final step, resulting in 13 final structures. All of the full complexes are minimized for either 500 steps or to a 0.25 kcal/molÅ RMS force threshold.

From the top structures, three structures were inspected: the structure with the best total complex energy, the structure with the best unified cavity energy, and the structure with the best ligand interaction energy. The best head group position was selected from these three structures as the one with the best contacts between the functional groups of the ligand and the residues in the binding site. This entire procedure was carried out for the head groups of LPA,

S1P and FFA in the three GPR88 predicted structures, resulting in nine complexes. Then, for each structure, the lipid's tail group was manually placed vertically in the TM3-TM6-TM7-TM2 binding site. Then the tail group and all residues within a 4.0 Å radius, except the head group which was kept frozen, were minimized and then put through a quench-anneal cycle (50 K to 600 K and back over 11.5 ps), selecting the configuration with the lowest potential energy structure. The quench-anneal cycle was carried out with the charged force field. Next, the head and tail group were connected, minimized, and the entire lipid and residues within a 4.0 Å radius were put through a quench-anneal cycle. Finally, the complex was neutralized, and the binding site, then the entire complex, was minimized. The neutralized binding energies for the three lipids in the three GPR88 predicted structures are given in Table 22. The binding energy is calculated as described in Chapter 1.

Table 22. Neutralized binding energies for three lipids in GPR88

	BiHelix_1	SuperBiHelix_1	SuperBiHelix_2
LPA	-63 kcal/mol	-35 kcal/mol	-61 kcal/mol
S1P	-55 kcal/mol	-51 kcal/mol	-68 kcal/mol
FFA	-58 kcal/mol	-52 kcal/mol	-52 kcal/mol

The binding energies were calculated for three GPR88 predicted structures: the best energy ComBiHelix structure (BiHelix_1), the best energy SuperComBiHelix structure (SuperBiHelix_1), and the second-best energy SuperComBiHelix structure (SuperBiHelix_2).

The binding energy calculations show there are similar binding energies for all of the complexes, except for LPA bound in the best energy SuperComBiHelix structure. An analysis of the interactions between the lipid and specific residues

in the binding pocket, shown in Table 23, should give further insight into why this complex has a lower binding energy than the others. Given that the docking method for the lipids was very crude, and the binding energies for the three lipids are fairly similar, the binding energies probably do not reveal which lipid binds the most weakly or strongly. The predicted bound structures for FFA are in Figure 28, the predicted bound structures for LPA are in Figure 29, and the predicted bound structures for S1P are in Figure 30.

Inspection of the bound structures in Figure 28 and the cavity analysis in Table 23 shows many important features of FFA binding to GPR88.

- For FFA binding, the carboxylic head group forms hydrogen bonds and has strong Coulombic interactions with R113(3) and R113(6) in all three predicted structures. There are also strong van der Waals interactions in all three predicted structures between FFA and W84(2), W322(7) and S321(7).
- In the two SuperComBiHelix structures, there are good van der Waals interactions between FFA and L124(3), C325(7) and L120(3).
- Both the lowest energy ComBiHelix structure and the second lowest energy SuperComBiHelix structures have good van der Waals interactions between FFA and L305(6) and V301(6).
- E88(2) has good interactions with the FFA in best ComBiHelix structure, and moderate interactions with FFA in the SuperComBiHelix structures.
- Only the second best SuperComBiHelix structure has good interactions between FFA and S127(3).

Thus, based on the three predicted structures for GPR88 bound to FFA, the carboxylic acid head group will bind strongly to R113(3) and R116(3), and the tail will have strong van der Waals interactions with W84(2), W322(7) and S321(7). Additionally, there will be possible good van der Waals interactions with L120(3), L124(3), V301(6), L305(6), C325(7), S127(3) and L328(7).

Table 23. Nonbond interaction energies (kcal/mol) between each ligand and individual residues within 5.0 Å of the ligand

Residue	TM	BiHelix1 FFA	Super1 FFA	Super2 FFA	BiHelix1 LPA	Super1 LPA	Super2 LPA	BiHelix1 S1P	Super1 S1P	Super2 S1P	Mean Energy
R113	3	-2.2	-3.5	-2.7	-2.7	-6.5	-6.7	-6.7	-7.6	-4.4	-4.8
R116	3	-4.5	-3.5	-2.7	-8.7	-1.0	-2.3	-3.7	-0.9	0.2	-3.0
L120	3	-2.5	-1.9	-0.8	-2.9	-3.7	-2.5	-0.9	-3.7	-2.5	-2.4
L124	3	-0.3	-2.4	-1.9	-3.1	-3.4	-2.3	-2.0	-2.7	-2.2	-2.2
W84	2	-3.3	-1.9	-2.7	-0.7	-0.7	-2.1	-2.6	-0.3	-3.1	-1.9
W322	7	-1.4	-1.9	-2.1	-0.5	-1.4	-4.5	-1.8	-4.2	0.8	-1.9
C325	7	-0.7	-1.9	-1.7	-2.3	-1.8	-0.8	-0.9	-2.2	-2.3	-1.6
L305	6	-1.3	-0.3	-1.8	-2.7	0.0	-3.3	-1.1	0.0	-2.7	-1.5
S321	7	-1.3	-1.5	-1.2	-1.3	0.0	-1.5	-1.0	0.0	-5.3	-1.5
L328	7	-0.6	-1.5	-1.4	-2.4	1.0	-0.3	-1.7	-1.7	-2.4	-1.2
E88	2	-1.5	-0.6	-0.9	-0.3	-1.4	-0.3	-0.6	-3.3	-1.0	-1.1
V301	6	-1.4	-0.9	-1.3	0.4	0.0	-2.2	-1.0	0.0	-0.8	-0.8
S127	3	0.1	0.1	-1.3	-0.9	-1.3	0.6	-0.9	-1.3	-2.0	-0.8
S304	6	0.0	0.0	-0.8	0.0	0.0	-2.2	0.0	0.0	-0.1	-0.3

The energies were calculated for three GPR88 predicted structures: the best energy ComBiHelix structure (BiHelix1), the best energy SuperComBiHelix structure (Super1), and the second-best energy SuperComBiHelix structure (Super2). The residues listed are those that have more than 2 kcal/mol of favorable energy with at least one ligand. Those residues that are more than 5.0 Å from the ligand in the respective structure are given 0.0 kcal/mol interaction energy.

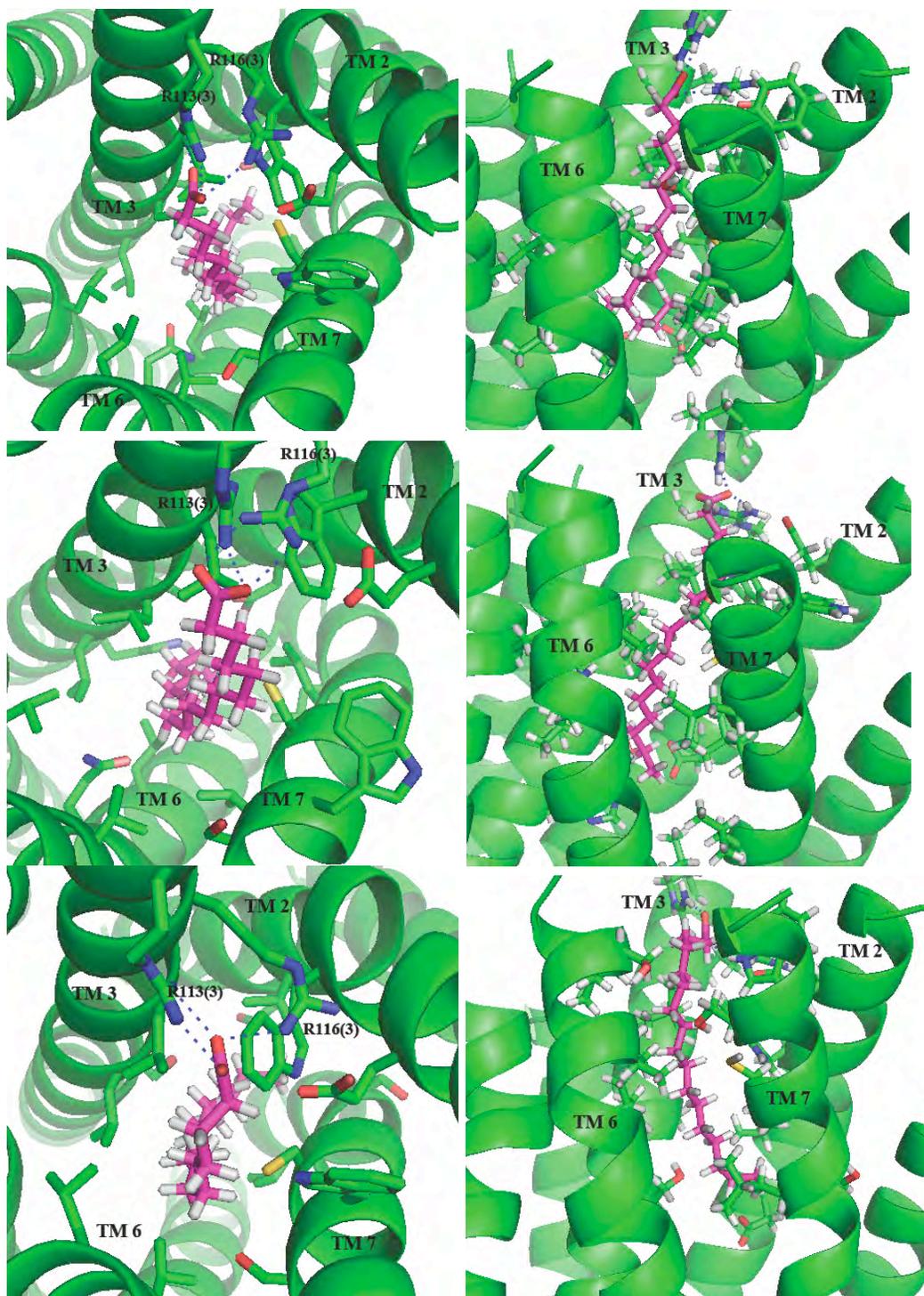


Figure 28. Extracellular (left) and side (right) views of the predicted bound structure of FFA in the lowest energy ComBiHelix structure (top), the lowest energy SuperComBiHelix structure (middle), and the second-lowest energy SuperComBiHelix structure (bottom). The dotted lines represent hydrogen bonds to the labeled residues.

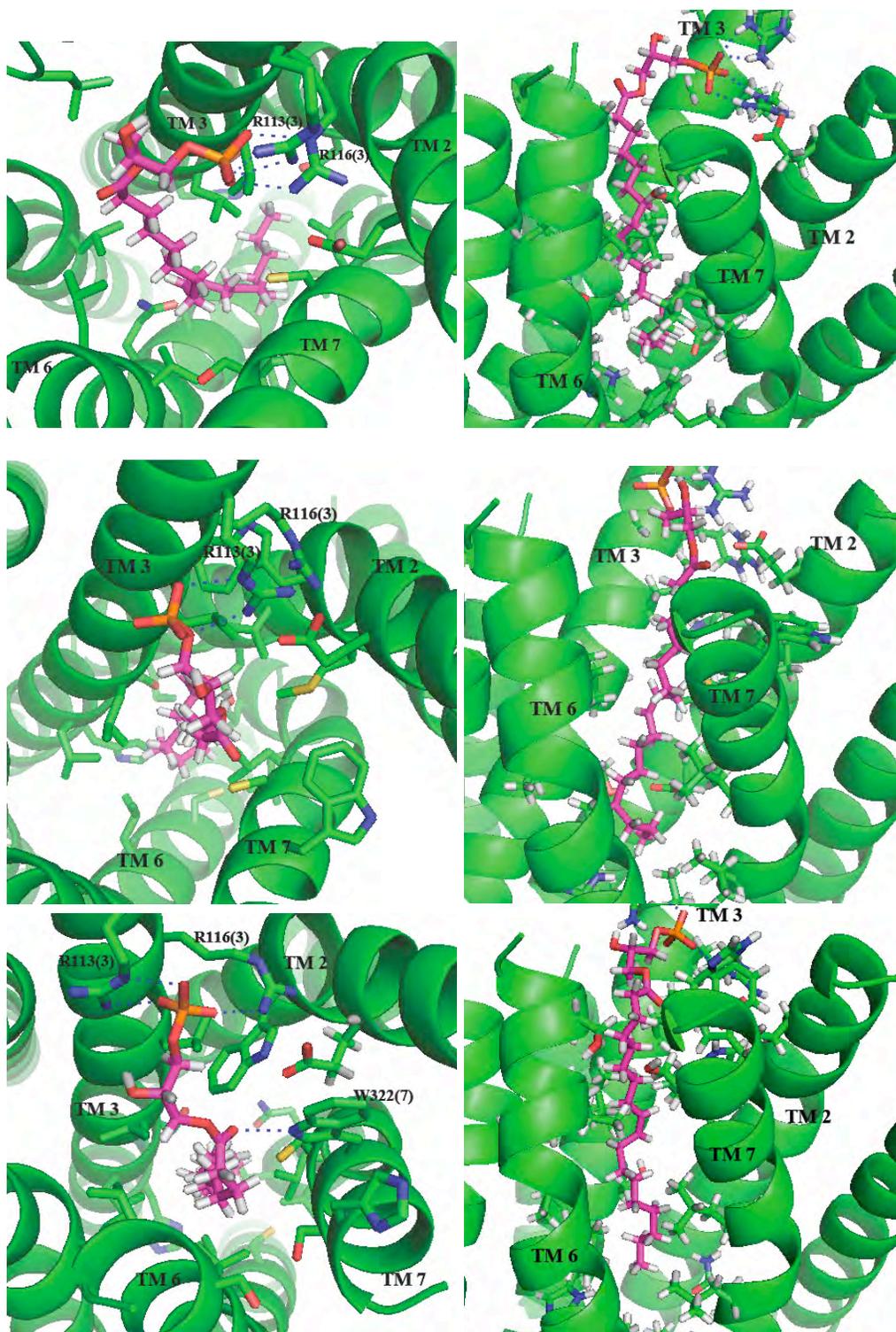


Figure 29. Extracellular (left) and side (right) views of the predicted bound structure of LPA in the lowest energy ComBiHelix structure (top), the lowest energy SuperComBiHelix structure (middle), and the second-lowest energy SuperComBiHelix structure (bottom). The dotted lines represent hydrogen bonds to the labeled residues.

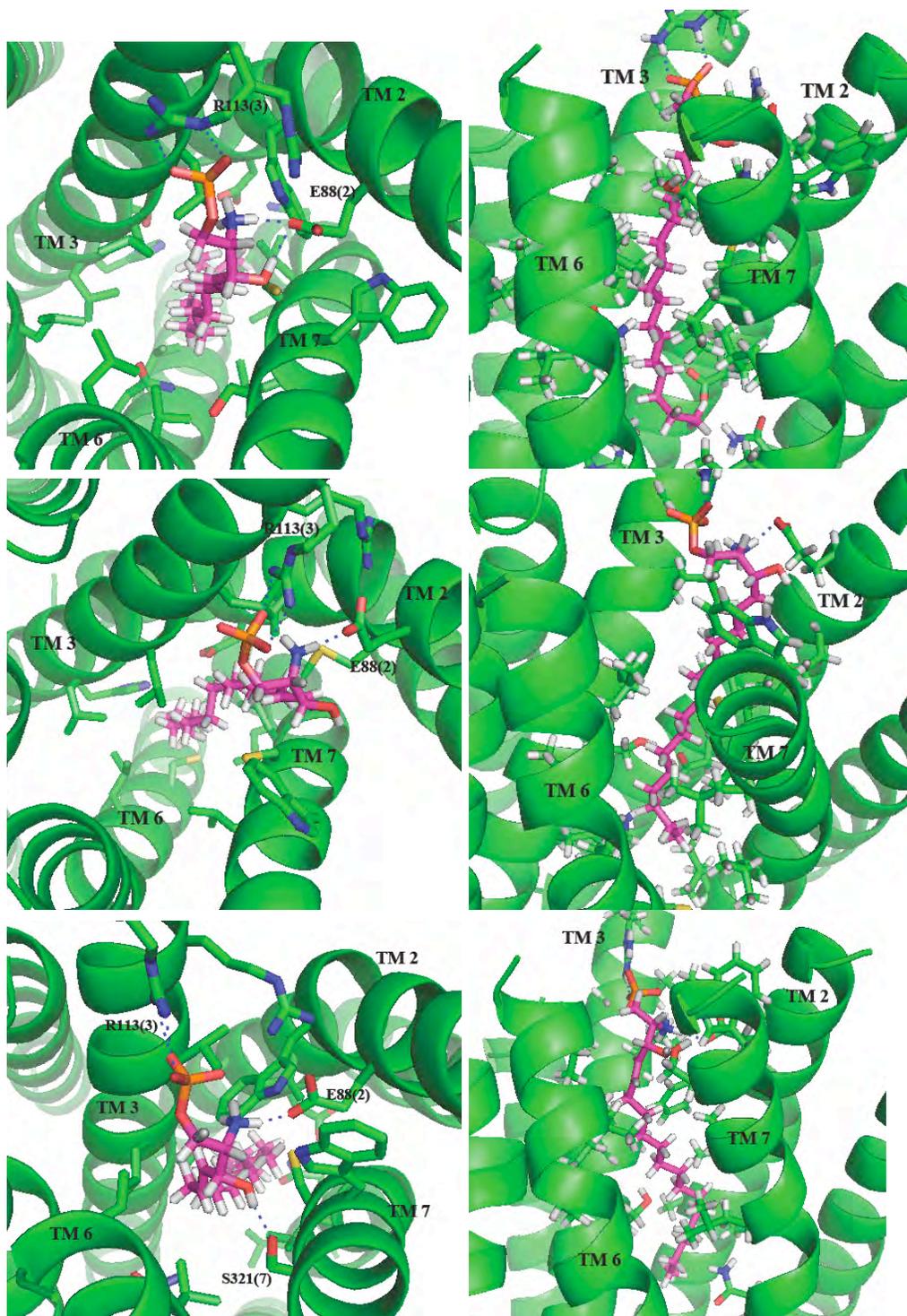


Figure 30. Extracellular (left) and side (right) views of the predicted bound structure of S1P in the lowest energy ComBiHelix structure (top), the lowest energy SuperComBiHelix structure (middle), and the second-lowest energy SuperComBiHelix structure (bottom). The dotted lines represent hydrogen bonds to the labeled residues.

Inspection of the bound structures in Figure 29 and the cavity analysis in Table 23 shows many important features of LPA binding to GPR88.

- The LPA head group forms a hydrogen bond with R113(3) in all the predicted structures. The head group also forms a hydrogen bond with R116(3) in the lowest energy ComBiHelix structure and the second lowest energy SuperComBiHelix structure. However, the lowest energy SuperComBiHelix structure does not have strong interaction with R116(3), which in part explains its low LPA binding energy.
- All three predicted structures have strong van der Waals interactions with L120(3) and L124(3).
- The carbonyl group of LPA forms a hydrogen bond with the W322(7) in the second best SuperComBiHelix structure, but not in the other two structures.
- The best energy ComBiHelix structure and the second best SuperComBiHelix structure have very good van der Waals interactions with L305(6) and S321(7). Those residues in the best SuperComBiHelix structure are not even within 5 Å of the ligand, contributing to the low binding energy of LPA in this structure.
- The second-lowest energy SuperComBiHelix has good van der Waals interactions with W84(2), V301(6) and S304(6).
- The lowest energy ComBiHelix and SuperComBiHelix structures have good van der Waals interactions with C325(7).

- The lowest energy ComBiHelix structure has good van der Waals interactions with L328(7).

Based on the three predicted structures of LPA bound to GPR88, the phosphate head group will bind strongly to R113(3) and R116(3), and the tail will have strong van der Waals interactions with L120(3) and L124(3). The carbonyl group may form a hydrogen bond to W322(7). There will be possible good van der Waals interactions with L305(6), L321(7), C325(7), W84(2), V301(6), S304(6) and L328(7).

Inspection of the bound structures in Figure 30 and the cavity analysis in Table 23 shows many important features of S1P binding to GPR88:

- All three structures show strong hydrogen bonds between the phosphate head group of S1P and R113(3). The head group does not form hydrogen bonds to R116(3) as seen in with the other two lipids. However, the lowest energy ComBiHelix structure does have good van der Waals interactions with R116(3).
- All three structures show hydrogen bonds between the protonated nitrogen group of S1P and E88(2). However, there are unfavorable van der Waals interactions that make it so the overall interaction energy is moderately favorable. The lowest energy SuperComBiHelix structure also has a hydrogen bond between E88(2) and the hydroxy group of S1P.
- The second best SuperComBiHelix structure forms a very strong hydrogen bond between the hydroxy group on S1P and S321(7).

- All three predicted structures have good van der Waals interactions between S1P and L124(3) and L328(7).
- The best two SuperComBiHelix structures have good van der Waals interactions between S1P and L120(3), C325(7) and S127(3).
- There are good van der Waals interactions between W84(2) and L305(6) and the S1P in the lowest energy ComBiHelix structure and second lowest energy SuperComBiHelix structure.
- W322(7) has good van der Waals interactions with S1P in the lowest energy ComBiHelix structure and the lowest energy SuperComBiHelix structure.

Based on the three predicted structures of S1P bound to GPR88, the phosphate head group will bind strongly to R113(3), with the protonated nitrogen group forming a hydrogen bond with E88(2). The tail will have strong van der Waals interactions with L328(7) and L124(3). The hydroxy group on S1P may form a hydrogen bond with either S321(7) or E88(2). Additionally, there will be possible good van der Waals interactions with L120(3), C325(7), S127(3), W84(2), L305(6) and W322(7).

In summary, the docking of the LPA, S1P and FFA to the three predicted GPCR structures leads to the prediction that GPR88 will bind lipids. The negatively charged head group will bind to R113(3) and R116(3). In S1P, the protonated nitrogen group will form a hydrogen bond with E88(2). S1P may form a hydrogen bond between its hydroxy group and S321(7). LPA may form a hydrogen bond between its carbonyl group and W322(7). For all three lipids, the

important der Waals interactions, in order of increasing strength, are predicted to be L120(3), L124(3), W84(2), W322(7), C325(7), L305(6), S321(7), L328(7), V301(6), and S127(3).

Conclusion

Three low energy structures were predicted for the orphan GPCR GPR88. These predicted structures had an aliphatic binding pocket in the TM2-TM3-TM6-TM7 region that would likely bind to an aliphatic lipid tail. The head group of a lipid could bind to R113(3) and R116(3) at the extracellular side of the receptor. Three lipids that commonly bind to GPCRs were docked: S1P, LPA and FFA. The predicted bound structures show that all three lipids would likely bind to GPR88, with the head group forming hydrogen bonds to R113(3) and R116(3). The residues contributing favorably to the binding of the lipid tails are L120(3), L124(3), W84(2), W322(7), C325(7), L305(6), S321(7), L328(7), V301(6), and S127(3), in order of decreasing predicted strength of interaction. Mutagenesis experiments on these residues could help to indirectly confirm the validity of one of the predicted GPR88 structures.

References

1. Chamberlain, A.K. and J.U. Bowie, *Analysis of side-chain rotamers in transmembrane proteins*. Biophysical Journal, 2004. **87**(5): p. 3460-3469.
2. Trabanino, R.J., et al., *First principles predictions of the structure and function of G-protein-coupled receptors: Validation for bovine rhodopsin*. Biophysical Journal, 2004. **86**(4): p. 1904-1921.
3. Palczewski, K., et al., *Crystal structure of rhodopsin: A G protein-coupled receptor*. Science, 2000. **289**(5480): p. 739-745.
4. Warne, T., et al., *Structure of a beta(1)-adrenergic G-protein-coupled receptor*. Nature, 2008. **454**(7203): p. 486-491.
5. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta(2)-adrenergic G protein-coupled receptor*. Science, 2007. **318**(5854): p. 1258-1265.
6. Jaakola, V.P., et al., *The 2.6 Angstrom Crystal Structure of a Human A(2A) Adenosine Receptor Bound to an Antagonist*. Science, 2008. **322**(5905): p. 1211-1217.
7. Vaidehi, N., et al., *Prediction of structure and function of G protein-coupled receptors*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12622-12627.
8. Bray, J.K. and W.A. Goddard, *The structure of human serotonin 2c G-protein-coupled receptor bound to agonists and antagonists*. Journal of Molecular Graphics & Modelling, 2008. **27**(1): p. 66-81.
9. Roth, B.L., et al., *5-hydroxytryptamine(2)-family receptors (5-hydroxytryptamine(2A), 5-hydroxytryptamine(2B), 5-hydroxytryptamine(2C)): Where structure meets function*. Pharmacology & Therapeutics, 1998. **79**(3): p. 231-257.
10. Leysen, J.E., *5-HT₂ Receptors*. Current Drug Targets - CNS & Neurological Disorders, 2004. **3**(1): p. 11-26.
11. Roth, B.L., et al., *The multiplicity of serotonin receptors: Uselessly diverse molecules or an embarrassment of riches?* Neuroscientist, 2000. **6**(4): p. 252-262.
12. Floriano, W.B., et al., *Modeling the human PTC bitter-taste receptor interactions with bitter tastants*. Journal of Molecular Modeling, 2006. **12**(6): p. 931-941.

13. Vaidehi, N., et al., *Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation*. Journal of Biological Chemistry, 2006. **281**(37): p. 27613-27620.
14. Hummel, P., et al., *Test of the Binding Threshold Hypothesis for olfactory receptors: Explanation of the differential binding of ketones to the mouse and human orthologs of olfactory receptor 912-93*. Protein Science, 2005. **14**(3): p. 703-710.
15. Hall, S.E., et al., *Predicted 3-D structures for mouse 17 and rat 17 olfactory receptors and comparison of predicted odor recognition profiles with experiment*. Chemical Senses, 2004. **29**(7): p. 595-616.
16. Floriano, W.B., N. Vaidehi, and W.A. Goddard, *Making sense of olfaction through predictions of the 3-D structure and function of olfactory receptors*. Chemical Senses, 2004. **29**(4): p. 269-290.
17. Kalani, M.Y.S., et al., *The predicted 3D structure of the human D2 dopamine receptor and the binding site and binding affinities for agonists and antagonists*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(11): p. 3815-3820.
18. Freddolino, P.L., et al., *Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(9): p. 2736-2741.
19. Cho, A.E., et al., *The MPSim-Dock hierarchical docking algorithm: Application to the eight trypsin inhibitor cocrystals*. Journal of Computational Chemistry, 2005. **26**(1): p. 48-71.
20. Lim, K.T., et al., *Molecular dynamics for very large systems on massively parallel computers: The MPSim program*. Journal of Computational Chemistry, 1997. **18**(4): p. 501-521.
21. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field for Molecular Simulations*. Journal of Physical Chemistry, 1990. **94**(26): p. 8897-8909.
22. Donnelly, D., J.P. Overington, and T.L. Blundell, *The Prediction and Orientation of Alpha-Helices from Sequence Alignments - the Combined Use of Environment-Dependent Substitution Tables, Fourier-Transform Methods and Helix Capping Rules*. Protein Engineering, 1994. **7**(5): p. 645-653.
23. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.

24. Chenna, R., et al., *Multiple sequence alignment with the Clustal series of programs*. Nucleic Acids Research, 2003. **31**(13): p. 3497-3500.
25. Eisenberg, D., R.M. Weiss, and T.C. Terwilliger, *The Helical Hydrophobic Moment - a Measure of the Amphiphilicity of a Helix*. Nature, 1982. **299**(5881): p. 371-374.
26. Schertler, G.F.X., *Structure of rhodopsin*. Eye, 1998. **12**: p. 504-510.
27. Bower, M.J., F.E. Cohen, and R.L. Dunbrack, *Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool*. Journal of Molecular Biology, 1997. **267**(5): p. 1268-1282.
28. Mathiowetz, A.M., et al., *Protein Simulations Using Techniques Suitable for Very Large Systems - the Cell Multipole Method for Nonbond Interactions and the Newton-Euler Inverse Mass Operator Method for Internal Coordinate Dynamics*. Proteins-Structure Function and Genetics, 1994. **20**(3): p. 227-247.
29. Jayasinghe, S., K. Hristova, and S.H. White, *Energetics, stability, and prediction of transmembrane helices*. Journal of Molecular Biology, 2001. **312**(5): p. 927-934.
30. Wimley, W.C. and S.H. White, *Experimentally determined hydrophobicity scale for proteins at membrane interfaces*. Nature Structural Biology, 1996. **3**(10): p. 842-848.
31. Kroeze, W.K., K. Kristiansen, and B.L. Roth, *Molecular Biology of Serotonin Receptors - Structure and Function at the Molecular Level*. Current Topics in Medicinal Chemistry, 2002. **2**(6): p. 507-528.
32. Ballesteros, J.A., L. Shi, and J.A. Javitch, *Structural mimicry in G protein-coupled receptors: Implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors*. Molecular Pharmacology, 2001. **60**(1): p. 1-19.
33. Eilers, M., et al., *Comparison of Class A and D G protein-coupled receptors: Common features in structure and activation*. Biochemistry, 2005. **44**(25): p. 8959-8975.
34. Bywater, R.P., *Location and nature of the residues important for ligand recognition in G-protein coupled receptors*. Journal of Molecular Recognition, 2005. **18**(1): p. 60-72.
35. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
36. Xie, J.L., S. Dernovici, and P. Ribeiro, *Mutagenesis analysis of the serotonin 5-HT_{2C} receptor and a Caenorhabditis elegans 5-HT₂ homologue: conserved*

- residues of helix 4 and helix 7 contribute to agonist-dependent activation of 5-HT₂ receptors.* Journal of Neurochemistry, 2005. **92**(2): p. 375-387.
37. Kristiansen, K., et al., *A highly conserved aspartic acid (Asp-155) anchors the terminal amine moiety of tryptamines and is involved in membrane targeting of the 5-HT_{2A} serotonin receptor but does not participate in activation via a "salt-bridge disruption" mechanism.* Journal of Pharmacology and Experimental Therapeutics, 2000. **293**(3): p. 735-746.
 38. Almaula, N., et al., *Mapping the binding site pocket of the serotonin 5-hydroxytryptamine(2A) receptor - Ser(3.36(159)) provides a second interaction site for the protonated amine of serotonin but not of lysergic acid diethylamide or bufotenin.* Journal of Biological Chemistry, 1996. **271**(25): p. 14672-14675.
 39. Ebersole, B.J., et al., *Molecular basis of partial agonism: Orientation of indoleamine ligands in the binding pocket of the human serotonin 5-HT_{2A} receptor determines relative efficacy.* Molecular Pharmacology, 2003. **63**(1): p. 36-43.
 40. Shapiro, D.A., et al., *Differential modes of agonist binding to 5-Hydroxytryptamine(2A) serotonin receptors revealed by mutation and molecular modeling of conserved residues in transmembrane region 5.* Molecular Pharmacology, 2000. **58**(5): p. 877-886.
 41. Choudhary, M.S., S. Craigo, and B.L. Roth, *A Single Point Mutation (Phe-340-→Leu-340) of a Conserved Phenylalanine Abolishes 4-[I-125]Iodo-(2,5-Dimethoxy)Phenylisopropylamine and [H-3] Mesulergine but Not [H-3] Ketanserin Binding to 5-Hydroxytryptamine(2) Receptors.* Molecular Pharmacology, 1993. **43**(5): p. 755-761.
 42. Roth, B.L., et al., *Identification of conserved aromatic residues essential for agonist binding and second messenger production at 5-hydroxytryptamine(2A) receptors.* Molecular Pharmacology, 1997. **52**(2): p. 259-266.
 43. Vriend, G., *What If - a Molecular Modeling and Drug Design Program.* Journal of Molecular Graphics, 1990. **8**(1): p. 52-56.
 44. vanRhee, A.M. and K.A. Jacobson, *Molecular architecture of G protein-coupled receptors.* Drug Development Research, 1996. **37**(1): p. 1-38.
 45. Bonhaus, D.W., et al., *RS-102221: A novel high affinity and selective, 5-HT_{2C} receptor antagonist.* Neuropharmacology, 1997. **36**(4-5): p. 621-629.
 46. Sard, H., et al., *SAR of psilocybin analogs: Discovery of a selective 5-HT_{2C} agonist.* Bioorganic & Medicinal Chemistry Letters, 2005. **15**(20): p. 4555-4559.

47. Manivet, P., et al., *The serotonin binding site of human and murine 5-HT_{2B}-receptors - Molecular modeling and site-directed mutagenesis*. Journal of Biological Chemistry, 2002. **277**(19): p. 17170-17178.
48. Wainscott, D.B., et al., *[³H]Rauwolscine: an antagonist radioligand for the cloned human 5-hydroxytryptamine_{2B} (5-HT_{2B}) receptor*. Naunyn-Schmiedeberg's Archives of Pharmacology 1998. **357**(1): p. 17-24.
49. Lang, P.T., et al., *DOCK 6: Combining techniques to model RNA-small molecule complexes*. Rna-a Publication of the Rna Society, 2009. **15**(6): p. 1219-1230.
50. Mizushima, K., et al., *A novel G-protein-coupled receptor gene expressed in striatum*. Genomics, 2000. **69**(3): p. 314-321.
51. Ghate, A., et al., *Identification of novel striatal genes by expression profiling in adult mouse brain*. Neuroscience, 2007. **146**(3): p. 1182-1192.
52. Pausch, M., et al. *Schizophrenia-like phenotypes displayed by GPR88 knock-out mice: Behavioral and neurochemical profile*. 2007: Oxford Univ Press.
53. Conti, B., et al., *Region-specific transcriptional changes following the three antidepressant treatments electro convulsive therapy, sleep deprivation and fluoxetine*. Molecular Psychiatry, 2007. **12**(2): p. 167-189.
54. Massart, R., et al. *Striatal synaptic and cortical nuclear orphan G protein-coupled receptor GPR88, a promising target for psychiatric and movement disorders*. 2008: Elsevier Science Bv.
55. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform*. Nucleic Acids Research, 2002. **30**(14): p. 3059-3066.
56. Wimley, W.C., T.P. Creamer, and S.H. White, *Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides*. Biochemistry, 1996. **35**(16): p. 5109-5124.
57. Pollastri, G. and A. McLysaght, *Porter: a new, accurate server for protein secondary structure prediction*. Bioinformatics, 2005. **21**(8): p. 1719-1720.
58. Raghava, G.P.S., *APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning*. CASP5, 2002. **A-132**.
59. Okada, T., et al., *The retinal conformation and its environment in rhodopsin in light of a new 2.2 angstrom crystal structure*. Journal of Molecular Biology, 2004. **342**(2): p. 571-583.

60. Lomize, M.A., et al., *OPM: Orientations of proteins in membranes database*. *Bioinformatics*, 2006. **22**(5): p. 623-625.