KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

PREDICTION OF 3D STRUCTURE OF THE DIMERIC STATE OF HUMAN β_2-ADRENERGIC RECEPTOR

GRADUATE THESIS

AYÇA KOROĞLU

January, 2016

Ayça Koroğlu

M.S. Thesis

2016

PREDICTION OF 3D STRUCTURE OF THE DIMERIC STATE OF HUMAN β_2-ADRENERGIC RECEPTOR

AYÇA KOROĞLU

Submitted to the Graduate School of Science and Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Computational Biology and Bioinformatics

KADIR HAS UNIVERSITY
January, 2016

KADIR HAS UNIVERSITY GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

THE PREDICTION OF 3D STRUCTURE OF THE DIMERIC STATE OF HUMAN β2 ADRENERGIC RECEPTOR

AYÇA KOROĞLU

APPROVED BY:

(Assoc. Prof. Ebru Demet Akdoğan) (Kadir Has University)
(Thesis Supervisor)

(Asst. Prof. Tuğba Arzu Özal İldeniz) (Acıbadem University)

(Prof. Dr. Kemal Yelekçi) (Kadir Has University)

APPROVAL DATE: 27/ Jan/ 2016

“I, Ayça Koroğlu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis”

_____

AYÇA KOROĞLU

THE PREDICTION OF 3D STRUCTURE OF THE DIMERIC STATE OF HUMAN β_2-ADRENERGIC RECEPTOR

Abstract

A significant amount of experimental and computational data points out a possible role of TM6 in the dimerization of β_2-adrenergic receptor (β_2AR). Peptide– and protein–protein docking experiments guided by this assumption were conducted in order to confirm the potential participation of TM6 at the interface region of β_2AR dimers. Firstly, a derived peptide consisting of 23 residues of TM6 was blindly docked to β_2AR monomer using a rigid body approach. The resulting complexes in which the peptide preferred to be near TM5 and TM6 regions have also the highest scores. Secondly, a total of seven peptides were derived from each TM regions and were blindly docked to β_2AR using the same rigid body approach. TM6 was found to be the most preferred binding site region in the receptor for each peptide. As for the protein-protein dockings, a full rigid-body docking returned a total of 16,000 dimer conformations and is followed by a membrane topology filtering. 149 complexes fit the topology requirements. A root mean-squared deviance (RMSD) value of 6 Å was used to cluster all 149 complexes and it is observed that the highest populated cluster of conformers have the highest score value and moreover TM6 presents at the interface region. Finally, an information-driven semi flexible docking approach used by given the interface residue data for seven TM regions. Interface residue data derived from the SASA values of residues in each TM region. The generated complexes were filtered due to membrane topology requirements. The complex with the TM6 interface was the highest scored docking pose among the complexes with a proper membrane topology.

DİMERİK FAZDA İNSAN β_2-ADRENERJİK RESEPTÖRÜNÜN
3D YAPISININ BELİRLENMESİ

Özet

Önemli sayıda deneysel ve bilişsel veri, β_2-adrenerjik reseptörünün (β_2AR) dimerizasyonunda TM6' nın olası bir rolüne işaret ediyor. Bu varsayımın rehberliğinde, β_2AR dimerinin interfaz bölgesinde TM6' nın olası bir katılımını doğrulamak için peptit- ve protein-protein docking deneyleri gerçekleştirildi. İlk olarak, TM6'nın 23 rezidüsüden oluşan bir peptit, β_2AR monomerine rastlantısal olarak rijit kütle yaklaşımıyla dock edildi. Tüm yüksek skorlu konformasyonların reseptörün TM5 ve TM6 bölgelerini tercih ettikleri görüldü. İkinci olarak, herbir TM bölgesinden toplamda yedi tane peptit oluşturuldu ve aynı rijit kütle yaklaşımı kullanılarak β_2AR' a sırayla dock edildi. TM6, her peptit icin reseptörde en çok tercih edilen bağlanma bölgesi olarak bulundu. Protein-protein docking deneylerinde ise, tam rijit kütle docking taraması toplamda 16,000 dimer konformasyonu oluşturdu ve herbir dimer, membran topolojisini esas alan bir seçilim değeri ile filtrelendi. 149 konformasyon 6Å değerinde RMSD ile kümelendi ve en yüksek populasyona sahip olan kümenin TM6 yı interfaz bölgesinde bulunduran ve en yüksek skora sahip olan küme olduğu görüldü. Son olarak, enformasyon güdümlü yarı-esnek bır docking protokolü yedi TM bölgesinden interfazda bulunmaya yatkın rezidüler verilerek kullanıldı. Interfaz rezidü dataları, her TM bölgesindeki rezidülerin SASA değerlerine göre oluşturuldu. Oluşan kompleksler membran topoloji koşullarına göre filtrelendi. TM6' nın interfaz bölgesini oluşturduğu kompleks, membran topolojisine uyan diğerleri içerisinde en yüksek skorlu docking pozu oldu.

Acknowledgements

First and utmost, I would first like to acknowledge my thesis advisor, Assoc. Prof. Ebru Demet AKTEN for giving me the opportunity to work in her project. This thesis would not have been possible without her valuable advice and efforts. However, most importantly, along the way she taught me how to research and analyze. I am grateful to be a student of such a true scientist.

I would like to express my gratitude for graduate scholarship ensured by TUBİTAK for my M.S. study via the project #109M281.

Additionally, I would like to thank Prof. Dr. Kemal YELEKÇİ and Asst. Prof. Tuğba Arzu Özal İLDENİZ for agreeing to serve on my dissertation committee.

I also want to thank my parents for seeing me through all these years and my friend Beril for her love and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

TM	Transmembrane
GPCR	G-Protein Coupled Receptor
ß$_2$AR	Human Beta-2 Adrenergic Receptor
ICL	Intracellular Loop
ECL	Extracellular Loop
Hx8	Helix 8
RMSD	Root-Mean-Square Deviation
PDB	Protein Data Bank
μ-OR	μ-Opioid Receptor
β$_1$AR	Beta-1 Adrenergic Receptor
κOR	κ-Opioid Receptor
CXCR4	C-X-C Chemokine Receptor Type 4

Chapter 1

Introduction

Known as complex molecular machines that are able to moderate many essential physiologic functions, seven transmembrane (TM) helix G-protein coupled receptors (GPCRs) comprise the largest family of integral membrane proteins. These gatekeepers and molecular messengers of the cell are targeted by approximately 40% of the pharmaceutical industry (1, 2). The activation of these receptors is induced by a broad spectrum of molecules ranging from odors to neurotransmitters (3). The formation of a ternary complex between the signalling molecule, the G protein and the active GPCR is crucial for the initiation of any GPCR-mediated signalling pathway (4). In this context, it is suggested that the beta-2 adrenergic receptor ($ß_2$AR) - which is one of the types of GPCRs - is sufficient to activate the signalling pathways while in a monomeric state (5). However, a monomeric receptor's ability to trigger the signaling pathways does not preclude a potential dimer formation from having a functional role as well. As a matter of fact several experimental studies demonstrated that $ß_2$ARs forming functional dimers and even oligomers (6,7,8,9,10).

In this respect, differential epitope tagging and co-immunoprecipitation experiments by Hebert et al. (8) demonstrated that the homodimerization of $ß_2$AR is a reversible process

hence the equilibrium between monomeric and dimeric forms is dynamic. Furthermore, a breakdown in agonist induced stimulation of adenylyl cyclase activity was observed in the presence of a peptide which is consisting from a portion of transmembrane domain 6 (TM6). This finding suggested that the TM6 derived peptide blocked the dimeric interface organization. Assuming that the receptor forms a symmetrical interface, TM6 region was predicted to have a role in dimerization.

A further biophysical approach; bioluminescence resonance energy transfer (BRET) study regarding $ß_2$AR dimerization, unambiguously demonstrated that the human $ß_2$AR forms constitutive homodimers that are expressed at the surface of living mammalian cells, where they interact with agonists (6). In a following study, the same research group took BRET experiments one step further to determine whether the $ß_1$- and $ß_2$AR can form heterodimers in cells and, if so, to investigate the potential functional consequences of such dimerization (11,12). The heterodimerization was occurred owing to their high level of sequence identity. The study highlighted a tendency for heterodimerization between $ß_1$- and $ß_2$AR, and a significantly high amount of constitutive dimeric receptors. Although the adenylyl cyclase activity was not affected by heterodimerization, the agonist-induced internalization of the $ß_2$AR and the activation of the ERK1/2 MAPK signaling pathway was interrupted by the presence of $ß_1$AR/ $ß_2$AR heterodimerization. In the sequel of these findings, the hypothesis that $ß_2$AR homodimerization is required for the trafficking from the endoplasmic reticulum to the plasma membrane was proposed by the same research group (13). Also the study showed that the mutations on the dimerization motif ^{276}GXXXGXXXL^{284} on TM6

confer a dominant negative effect on the transportation of the receptor to the plasma membrane.

For the first time fluorescence recovery after photobleaching studies revealed the formation of stable ß$_2$ARs oligomers (approximately 4- 5 receptor) (14). Following the same year, fluorescence resonance energy transfer (FRET) studies suggested that the ß$_2$AR oligomers consist of more than two monomers, and one probably tetramers in a model lipid bilayer (15). Agonists had little effect on ß$_2$AR oligomerization whereas inverse agonists promoted the higher order oligomerization and stabilize the oligomers by reducing conformational fluctuations in individual monomers.

Computer simulation studies were performed on models of the ß$_2$AR dimer including 5,6 domain swapped dimers -where the dimer interface lies between TM5 and TM6- proposed as the active, high affinity form (16). The correlated mutations amongst the external residues which occurred mainly at the TM5 - TM6 interface were determined in the same study. The presence of correlated mutations which is known to indicate a functionally important interface region, also were observed by a large amount of computational studies (17,18,19,20,21,22).

A computational residue conservation study also proposed TM6 as one of the most potent interface candidate (23). Moreover, a sequence alignment study composed of multiple alignment of 700 GPCRs (24) determined that the external faces of TM5 and TM6 have a high level of conservation and thus assign these regions as potential functional sites for the receptors (25,26). In immunological studies regarding A_1

adenosine receptor (27), intracellular loop 3 (ICL3) which binds TM5 and TM6 had been suggested to becomes less exposed in the dimer, thus again indicating a possible role of TM5 and TM6 in dimerization.

In spite of various computational and bioinformatics studies that looked for possible dimerization / oligomerization interfaces of GPCRs did not predict exactly the same interfaces for all the GPCR subfamilies, it is noted that some TM segments appear more often than others in the prediction of such interfaces (28). TM4, TM5, TM6 regions lead the way on having the most of the lipid-exposed residues that are predicted more often, this points out a specific role for these three helices. 4.58, 5.48, and 6.42 had been found to have the greatest number of occurrences among the residues identified. 6.30 which is located between TM6 and ICL3 had a similar amount of occurrence as 6.42 suggesting a possible role of ICL3 in dimerization / oligomerization. The atomic force microscopy studies (29) conducted on rhodopsin also indicating that ICL3 facilitates the formation of rows of dimers.

Guided by the experimental and computational findings regarding the location of the interface region, we designed a computational study in which short and long peptides derived from each TM regions of ß$_2$AR that are subjected to a docking procedure. In this procedure, derived peptides were rigidly and blindly docked to the monomeric form of the receptor, in order to determine the TM regions of the receptor for which the docked peptide would have the highest affinity and suggest a plausible interface region. Peptide docking experiments were modeled according to the studies of Hebert and his coworkers (8). In the second part of the study, a protein-protein blind docking tool was used to

predict a dimeric form and an interface by using two identical ß$_2$AR monomers. The receptor structure which was used in all docking runs represents an inactive form with a relatively wrapped up ICL3 that occupies the binding site of G-protein. This position of ICL3 is selected in order to the assumption that it wouldn't interfere any kind of dimeric or oligomeric formation. An inactive state with fewer conformational fluctuation was also preferred because of its favor on the closer packing of dimeric / oligomeric associations, as suggested by FRET experiments (4).

Chapter 2

Beta-2 Adrenergic Receptor

Beta-2 adrenergic receptors (β_2AR) are the members of G-protein coupled receptor (GPCR) superfamily. GPCRs reside on the lipid bilayer of the plasma membrane of almost every eukaryotic cells. In mammalian genomes, they have a highly diversified repertoire from 800 to 1000 genes which corresponds 2-3 percent of the human proteome (30). Their involvement in almost every physiological function makes them essential to human life. In any GPCR malfunction scenario, dozens of diseases, including diabetes, blindness, asthma, depression and some forms of cancer emerge. Therefore their importance nominates them as an enticing target for drug companies.

By the recent classification system GRAFS (**G**lutamate, **R**hodopsin, **A**dhesion, **F**rizzled/Taste2, **S**ecretin) GPCRs are grouped into 6 classes based on sequence homology and functional similarity (31). These groups can be ordered as

- Class A (or 1) Rhodopsin-like (which is by far the largest)
- Class B (or 2) Secretin receptor family
- Class C (or 3) Metabotropic glutamate/pheromone
- Class D (or 4) Fungal mating pheromone receptors

- Class E (or 5) Cyclic AMP receptors
- Class F (or 6) Frizzled/Smoothened

These classes further spread into subfamilies in a hierarchical fashion. β_2AR belongs to the highly abundant Class A-often called the "rhodopsin-like family"-which accounts for about 80 % of all GPCR genes (32). β_2AR is a member of the family biogenic amine receptors and finally subfamily adrenoceptors.

GPCRs have indeed a complex machinery receiving many outputs and compiling many inputs. Upon activation, the extracellular (ligand) and intracellular (G-proteins, arrestins) guests couple with GPCRs, in the transduction the stimuli into biochemical signals (33, 34). β_2AR which is mostly found in muscles interacts with hormone ligands like adrenaline (epinephrine) whose signaling causes dilation of blood vessels, increased contraction speed in cardiac or skeletal muscle and relaxation in smooth muscle.

2.1 Structure

Through 40 years, plenty of varied investigations conducted to elucidate the structure of GPCRs have been provided the most of what we know today about these intriguing macromolecules. Although membrane proteins are notoriously known as difficult to study owing to their partially hydrophobic surfaces, flexibility and lack of stability, by the year 2000 structural biologists managed to obtain the first X-ray crystallography structure of a GPCR (bovine rhodopsin) which has induced the steady increase on publications of the subsequent ones.

GPCRs share a common architecture in which one chain snakes back and forth through the membrane seven times, hence they are also called serpentine receptors. In general these receptors are characterized by an extracellular N-terminal region (N-terminus), followed by seven transmembrane domains (TM1 to TM7) connected by three intracellular (ICL1 to ICL3) and three extracellular loops (ECL1 to ECL3), a cytoplasmic α-helix (Hx8) parallel to the cell membrane, and finally an intracellular C-terminal region (C-terminus) (Figure 2.1 (35) and Table 2.1).

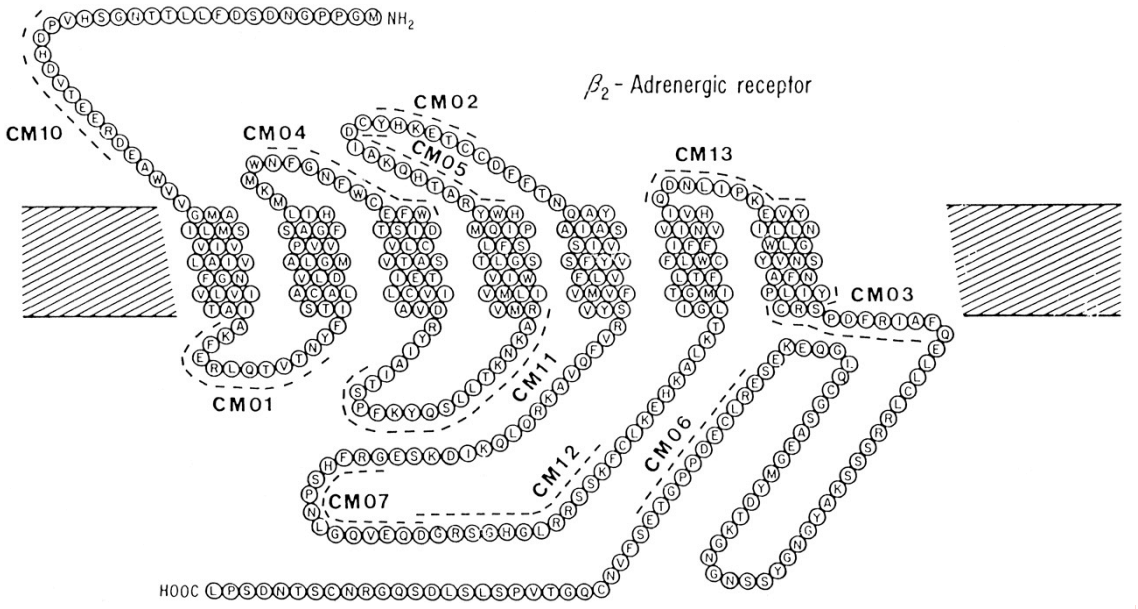


Figure 2.1 2D view and amino acid sequence of human β_2AR (35)

Table 2.1 Topology of GPCR

Feature key	Position(s)	Length	Description
Topological domain	1 – 34	34	Extracellular
Transmembrane	35 – 58	24	Helical; Name=1
Topological domain	59 – 71	13	Cytoplasmic
Transmembrane	72 – 95	24	Helical; Name=2
Topological domain	96 – 106	11	Extracellular
Transmembrane	107 – 129	23	Helical; Name=3
Topological domain	130 – 150	21	Cytoplasmic
Transmembrane	151 – 174	24	Helical; Name=4
Topological domain	175 – 196	22	Extracellular
Transmembrane	197 – 220	24	Helical; Name=5
Topological domain	221 – 274	54	Cytoplasmic
Transmembrane	275 – 298	24	Helical; Name=6
Topological domain	299 – 305	7	Extracellular
Transmembrane	306 – 329	24	Helical; Name=7
Topological domain	330 – 413	84	Cytoplasmic

The X-ray crystallography structure of β_2AR was first published in 2007 (PDB:2RH1). β_2AR arranges itself into a tertiary structure resembling a hourglass with seven helical TM regions. Hourglass figure results from the proline kinks in TM2, TM5, TM6 and TM7 that are in the middle of the membrane span. This structure which is narrower at the waist, forms a cavity at either side of the membrane. The binding domain of the

natural agonist adrenaline is in the cavity formed near the extracellular side of the receptor.

Increased crystallography data, mutagenesis, and computational studies had yielded a residue conservation analysis for GPCRs. According to this analysis, the residues in the intracellular region which affect the coupling selectivity with G proteins are highly conserved (36). In contrast, extracellular region which operates the ligand binding selectivity have a low degree of sequence conservation (Figure 2.2 (37)). This phenomenon can ben explained by the evolutionary pressure of recognizing enormous amount of ligands by the extracellular regions.

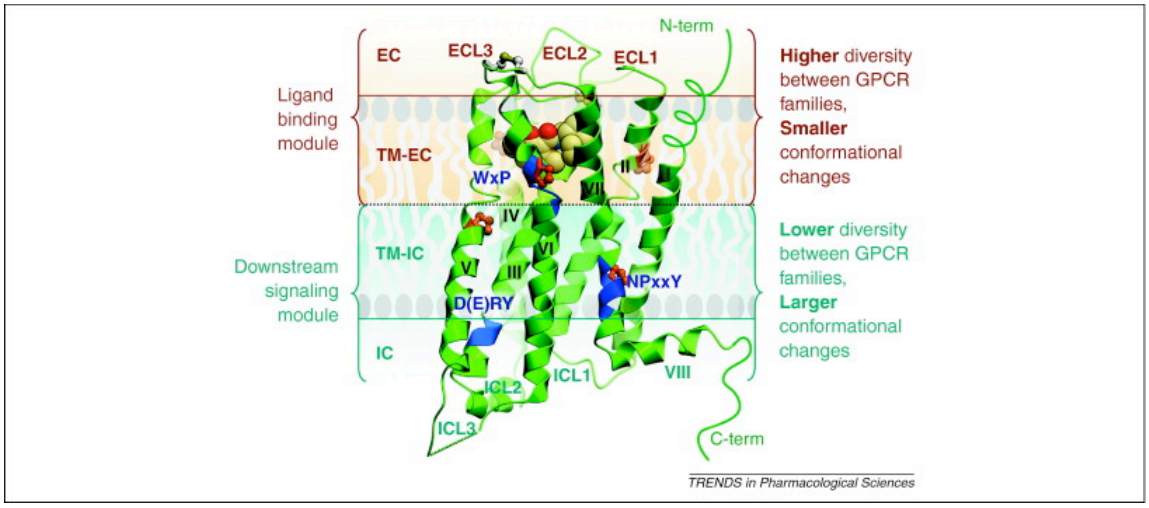


Figure 2.2 3D view of a GPCR with a residue conservation profile (37)

All crystal structures of GPCRs have analogous TM bundle structures despite a low percentage of sequence identity (38). As a matter of fact, it is known that membrane proteins maintain a strong structure conservation for TM bundle even at low sequence

identity (<20 %) (40). Structure conservation in the GPCR family is known to be caused by a highly conserved amino acid in each TM region (38): $N^{1.50}$ in TM1 (98 % of the sequences), $D^{2.50}$ in TM2 (93%), $R^{3.50}$ in TM3 (95%), $W^{4.50}$ in TM4 (96%), $P^{5.50}$ in TM5 (76%), $P^{6.50}$ in TM6 (98%), and $P^{7.50}$ in TM7 (93 %). The superscripts denotes for Ballesteros and Weinstein numbering (40) where the first number (1 through 7) is for the TM regions and the second number is for indicating its position according to the most conserved residue (which is 50) in the TM. Interestingly, the position of these conserved amino acids in TM regions is the same in the superimposition of the current crystal structures.

Both experimental and *in silico* studies have shown that GPCR superfamily share highly conserved sequence motifs (SMs) that inherent functional microdomains (FMs) which are responsible for the activation of the receptor (41-51). The known examples of these SMs/FMs for the Class A receptors are

- "arginine cage" around the (D/E)R(Y/W) motif in TM3 (41),
- the NPxxY motif in TM7 (41-45) and,
- WxP(F/Y) motif in TM6

Additionally, β_2AR contains PSD-95/Discs-large/ZO-1 (PDZ) protein binding sequences (PDZ domain).

A comparative study regarding the conservation of TM regions revealed that the majority of TM regions have kinks and bulges (52). In biogenic amine receptors, TM1 is found to be pointing outside of the bundle. $N^{1.50}$ is suggested to be responsible for the

packing of the TM bundle since its atoms let the residues at positions 1.46 and 7.46 to bind TM1 and TM7. Moreover, $N^{1.50}$ also interacts with $D^{2.50}$ causing a link between TM1 and TM2. The position of TM2 (towards TM1 and away from TM3) at the extracellular part is similar in all structures. TM2 of biogenic amine receptors turns in every ~4.8 residues. Trp residue of ECL1 is found in the (W/F) x (F/L)G motif (53), and leans toward TM2 and TM3. TM3 which has the most conserved residues among all TM regions, is the longest and most tilted helix among the receptor structures. The cytoplasmic side of TM3 contains the (D/E)$R^{3.50}$ (Y/W) motif involved in receptor activation. The central location of TM3 makes the interaction with the ligand at the extracellular part and with the G protein at the intracellular part possible (22). $C^{3.25}$ forms a disulphide bridge with a Cys residue of ECL2 which is important for ligand recognition, since it lines the interhelical binding cavity for many GPCRs. Adrenoceptors (e.g. β_2AR) have a helical ECL2 and this α-helix is not conserved in the other members of the biogenic amine receptor family suggesting that each receptor subfamily has developed a specific N-terminus/ECL2 to adapt to diverse ligands (55-57). TM4 is the shortest helix and is almost perpendicular to the membrane. Important structural diversities at the extracellular part of TM4 are found among structures, which may be related to ECL2 and selectivity of dimerization with other GPCRs, since TM4 has been shown to present at the dimer interface in several GPCRs (58). $P^{5.50}$ makes a kink in TM5, at the 5.43–5.48 turn, which suggests an involvement in the mechanism of ligand-induced receptor activation (59, 60). $Y^{5.58}$ (73 % of the sequences), involves in the stabilization of the active state of the receptor by interacting

with $R^{3.50}$ of the (D/E)RY motif in TM3 which was observed in the crystal structures of β_2AR in complex with G_s protein (PDB:3SN6)(61).

TM6 have the most severe kink which is energetically stabilized through two elements that are $P^{6.50}$ of CWx$P^{6.50}$(Y/F) motif and a structural water molecule between TM6 and TM7. In addition to stabilizing the kink of TM6, this water molecule also binds TM6 and TM7. TM4, TM5 and TM6 are the most diverse helices among the biogenic amine family receptors. TM5 and TM6 give selectivity to the specific ligands. TM7 is kinked at $P^{7.50}$ of the highly conserved NPxxY motif. This region of TM7, involved in important conformational changes associated with GPCR activation (48), is very irregular.

Another analysis on the features of the crystal structures of various biogenic amine receptors (including β_2AR) demonstrated that conserved residues are located inside of the TM bundle, most likely improving the efficiency of packing (39). These inward residues are more conserved than the residues that face to the membrane in the TM bundle. Moreover the extracellular region of TM bundle is more divergent than the intracellular part.

The loops on the other hand differ highly between GPCRs in sequence identity, length, and structure. For biogenic amine receptors, the extracellular loops show less conserved residues than the intracellular loops (ICL1 is highly conserved in all biogenic amine receptors). The extracellular surface of GPCRs is composed of the N-terminus region and ECL1, ECL2 and ECL3. X-ray crystallography studies also highlighted that the N-

terminus and ECL2 have diverse structures that adopt different conformations among the members of the GPCR superfamily (62).

2.2 β_2ARs in Action

2.2.1 Constitutive and ligand determined activity of β_2ARs

β_2ARs have a basal activity which enables them to exist in multiple "active" and "inactive" conformational states even in the absence of ligand (63, 64). Multiscale computational methods showed that the population of the "active" states of β_2AR are found to be much lower than those of the "inactive" states. And ligand binding stabilizes some of the "active" conformations already sampled. Supportively, this was later observed in the NMR study on the dynamics of β_2AR (65).

There are currently 16 crystal structures deposited in the PDB for β_2AR (Table 2.2). The structures of β_2AR bound to the inverse agonists represent the inactive state of the receptor (PDB:2RH1 with carazolol bound and PDB:3D4S with timolol bound). Structures that reflect the active states are bound to the agonist BI-167107 in complex with either a camelid nanobody (PDB:3P0G) or G_s protein (PDB : 3SN6).

However the structures of the active and inactive forms of β_2AR are very similar (RMSD 1.0 Å for residues 29-265 and 299-365) , activation of β_2AR induces a conformational change in TM6 (or vice versa) (66-68), as a result Leu266 in N-terminal end moves by 11 Å. As mentioned before, TM6 is known to be very dynamic with a kink angle among GPCRs. This movement of TM6 consequently alters the shape of the

intracellular face of the molecule and results a splaying of the helices at the cytoplasmic end of the receptor.

This change increases the affinity between β_2AR and G-protein. The crystal structure of the active state bound to the BI-167107 agonist shows hydrogen bonding interactions which stabilize a receptor conformation that includes a 2.1 Å inward movement of TM5 relative to the inactive, carazolol-bound structure (69). Thus, this inward movement causes a rotation and outward movement of TM6 leading a receptor activation (60). The structure of the β_2AR bound to G_s have shown that the C-terminal α5 helix of Gα binds to the intracellular cavity that is opened by the movement of the cytoplasmic end of TM6 away from TM3 and towards TM5. The C-terminal α5 helix of the α-subunit interacts with the short loop connecting TM7 and Hx8, and the inner side of the cytoplasmic TM5 and TM6.

2.2.2 Interacting Partners

The life cycle of β_2ARs, consists of biosynthesis in the endoplasmic reticulum (ER), maturation in the Golgi apparatus, transportation to the membrane, signal transduction and receptor internalization. β_2ARs are coupled by interacting proteins and small molecules through these stages (70, 71). GPCR-interacting proteins (GIPs) are roughly heat-shock proteins, kinases (for example GRKs), GPCR-associated sorting proteins (GASPs), PSD-95/Discs-large/ZO-1 (PDZ) domain-containing proteins, G-proteins, arrestins and many more (70, 72). The heterotrimeric G-protein complex which is vital for GPCR signal transduction is consisting of subunits that are called Gα, Gβ and Gγ (73). Gα comprises two domains: a Ras domain and an alpha-helical domain. β_2AR can

interact with two different G-proteins, G_s and G_i. The structure of the ternary complex of β_2AR, G_s G-protein and agonist BI-167107 (PDB:3SN6) revealed how all the parts of the complex interact. The Ras domain of Gα contacts TM5 and TM6 of β_2AR via its C-terminal helix (α5 helix) such that the G-protein complex binds to the the cytoplasmic face of β_2AR. There are no interactions between β_2AR and the Gβ or Gγ subunits. The binding site in β_2AR into which Gα inserts is only available when TM6 is in its active conformation. Also, mutation of Phe139 to an alanine significantly reduces the binding of Gα to β_2AR (74) suggesting the need for a hydrophobic residue at this position.

When an agonist binds to the extracellular side of β_2AR, a signal to the cytoplasmic side of the GPCR emerges through conformational changes of the receptor. This rearrangement increase the affinity for the GDP-bound form of the G-protein αβγ trimer. Upon binding, the GDP in Gα is exchanged for GTP which causes dissociation of the G-protein complex from the receptor. Afterwards Gα and Gβγ regulate downstream processes separately. Gα subsequently hydrolyses its bound GTP, thus increasing the affinity for Gβγ and stopping the downstream signalling. The trimer is then able to bind to an activated GPCR once more, completing the cycle (75) (Figure 2.3 (76)).

Table 2.2 Structures of β_2AR are grouped based on the efficacy of the ligands: inverse agonists, antagonists, and agonists.

PDB Accession Number	Ligand	Resolution	Organism
	Inverse Agonists		
2RH1	carazolol	2.4Å	*homo sapiens*
2R4R		3.4Å	*homo sapiens*
2R4S		3.4Å	*homo sapiens*
3KJ6		3.4Å	*homo sapiens*
3D4S	timolol	2.8Å	*homo sapiens*
3NY8	ICI-118551	2.84Å	*homo sapiens*
3NY9	compound a[a]	2.84Å	*homo sapiens*
4GBR	CAU	3.99Å	*homo sapiens*
	Antagonists		
3NYA	alprenolol	3.16Å	*homo sapiens*
	Agonists		
3PDS	procaterol	3.5Å	*homo sapiens*
3P0G	BI-167107	3.5Å	*homo sapiens*
3SN6		3.2Å	*homo sapiens*
4LDE	adrenaline	2.79Å	*homo sapiens*
4LDL		3.1Å	*homo sapiens*
4LDO		3.2Å	*homo sapiens*
4QKX	35V	3.3Å	*homo sapiens*

[a] Ethyl 4-[(2S)-2-hydroxy-3-[(propan-2-yl)amino]propoxy]-3-methyl −1- benzofuran −2-carboxylate.

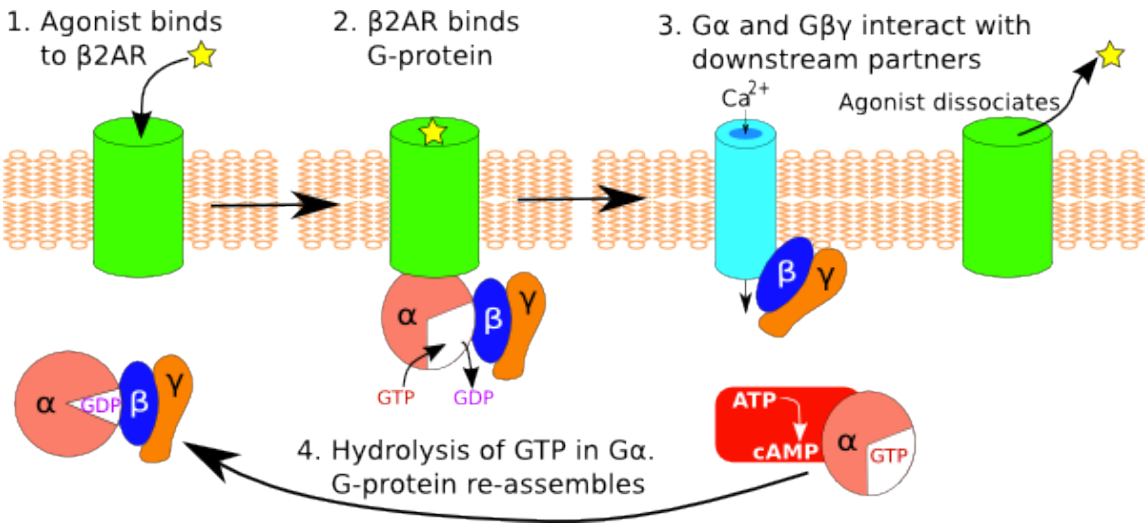


Figure 2.3 An overall mechanism for β_2AR (76)

β_2ARs can bind two different arrestin proteins: arrestin-2 and arrestin-3 (non-visual arrestins) interact and regulate the activity of the receptor (77). Desensitization of β_2ARs has been shown to be mediated by phosphorylation by G-protein-coupled receptor kinase 2 (GRK2) and followed by binding of arrestins to the phosphorylated receptor. Binding of arrestin to the receptor is a prerequisite for subsequent receptor desensitization.

2.2.3 β_2ARs In Their Natural Habitat

Experimental data suggesting that the membrane environment affects oligomerization, stability, and activity of GPCRs. Clearly, the effects of the membrane environment are dependent on the lipid composition which are exist in different combinations in different cell types. Functionally relevant interaction of lipids with GPCRs have two different types: (1)-specific lipid-protein interactions, e.g., the binding of individual cholesterol

molecules to specific residues of the GPCR; and (2)-non-specific interactions, such as hydrophobic mismatch (78-82).

The membrane deforms differently near different parts of the GPCRs. A hybrid Continuum-MD (CTMD) method study concluded that the radial asymmetry of the membrane deformation is due to the radially asymmetric hydrophobic surface of the GPCRs, which have TMs of different hydrophobic lengths (83). This membrane deformation creates a context of adjacent hydrophobic and polar residues at specific sites of the GPCRs where hydrophobic matching may remain incomplete. Moreover, it is due to this radial asymmetry that there exist a few specific sites where specific lipid molecules such as cholesterol bind. The residual mismatch emerged from the hydrophobic mismatch carries a significant energy cost.

Although hydrophobic mismatch is energetically costly, membrane deformations do not achieve a complete hydrophobic matching at all TM segments (83), and residues in the TM-bundle can remain exposed to unfavorable hydrophobic-polar interactions. That adjacencies responsible for the residual exposure are a common feature of GPCRs. This is evident from the recently determined crystal structures and validated structural models of a variety of rhodopsin-like GPCRs, including B_2AR (83-85).

2.2.4 Oligomerization

Dimerization of GPCRs is suggested to have roles in receptor maturation, G-protein coupling, downstream signalling and regulatory processes such as internalization. Figure 2.4 (86) summarize the five stages of the GPCRs that could be mediated by

dimerization. Cell fractionation studies by Salahpour et al. (13) revealed that β_2AR homodimerization is an event occurring as early as the ER.

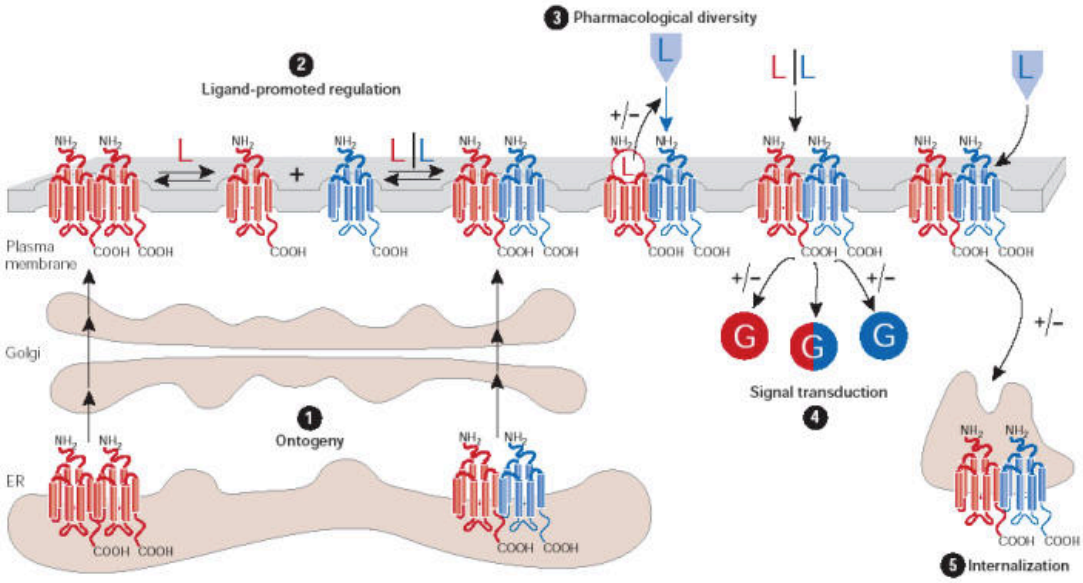


Figure 2.4 Potential roles of G-protein-coupled receptor (GPCR) dimerization (86)

In the membrane environment, the receptor activation is modulated by allosteric communication between protomers of GPCR dimers (87). Oligomerization evokes different activation patterns for the receptors. For instance, the complex consisting of two monomers and a single G protein, is activated by agonist binding to a single monomer and interestingly, inverse agonist binding to the second monomer enhances signaling. On the other hand agonist binding to the second monomer blunts signaling. Additionally, β_2AR oligomerization is regulated by the binding of agonists or the G

protein (15). Therefore it seems like activation and oligomerization have correlative impacts on each other.

Hydrophobic mismatch stemmed from membrane environment, has been shown to promote oligomerization in several studies (79, 88, 89). Clearly, oligomerization reduce the energy cost caused by the residual mismatch. This underlies a stochastic nature for GPCR dimerization in addition to a ligand driven mechanism. A computational study by Periole et al. (90) explored the effects of hydrophobic mismatch on the oligomerization of rhodopsin, using the MARTINI force field. The results demonstrated that oligomerization is driven by frustration of lipid-protein interactions. Rhodopsin significantly altered the local membrane thickness, encouraging oligomerization as a way to reduce unfavorable protein-lipid interactions.

Continuum-Molecular Dynamics (CTMD) software tool yields an energetic quantification for the hydrophobic mismatch deformation in membrane bilayers. This method can predict an optimal monomeric orientation for the oligomers. Results from both CTMD calculations and coarse-grained simulations of the interaction of GPCRs in the membrane provided results consistent with the experimental observations, and brought important information on the mechanism of oligomerization (90, 91). Especially, these calculations suggest that GPCR oligomerization at “specific interfaces” reduces the hydrophobic mismatch and its energy cost (84, 85)

Several crystal structures with homo-oligomeric GPCRs have been published (66, 92-98) (See Table 2.3). The interface via TM1 and TM2 and Hx8 is observed in the

dimeric structures of μ-OR, β_1AR, κOR, opsin, metarhodopsin II. The second interface of μ-OR comprises TM5 and TM6. Moreover the second interface of β_1AR (consisting of TM4 and TM5) resembles the interface previously obtained for rhodopsin using atomic force microscopy (AFM) (99). The crystal structure of the histamine H1 receptor (97) contains a TM4 interface, which is different from the TM4/5 interface of β_1AR due to the absence of TM5 contacts. Similarly, the structures of CXCR4 (94) and squid rhodopsin (98) contain a TM5 interface, which are different from TM5/6 interface of β_1AR and μ-OR.

In a coarse grained study (58), a pair of the DOP (a subtype of opioid receptors) receptor protomers generated by homology modeling, -using the crystal structure of the β_2AR- was placed facing one another at a proposed symmetrical interface, involving residue 4.58, inferred from cysteine cross-linking data on this and other GPCRs (100, 101). This yielded two kinds of interfaces involving either,TM4 or TM4/TM5. The comparison of the interfaces revealed that the TM4 interface was marginally more stable than the interface involving both TM4 and TM5, indicating that the stability of the dimer pair is dependent on the region of contact between the protomers.

A recent Martini MD study which explores the dimerization of β_2AR in lipid bilayers showed that cholesterol binds to TM4, and cholesterol occupancy at this site restricts its involvement at the dimer interface (102). This restriction enables TM1 and TM2 to present in the interface likewise the crystal structure of β_1AR dimer.

Table 2.3 Crystal structures of homo-oligomeric GPCRs

PDB	Description	Resolution	Interface
1N3M	Rhodopsin	semi empirical model based on AFM	TM1/TM2/H8, TM4/TM5, TM1/TM2 - ICL3
2Z73	Squid Rhodopsin	2.50Å	TM4/TM5, TM5
3ODU	CXCR4	2.50Å	TM5/TM6
3OE(0, 6, 8, 9)		2.90-3.20Å	
3RZE	Histamine H1 receptor	3.10Å	TM4
3PXO	metarhodopsin II	3.00Å	TM1/TM2/Hx8
4DKL	μ-OR	2.80Å	TM1/TM2/Hx8, TM5/TM6
4DJH	κOR	2.90Å	TM1/TM2/Hx8
4GPO	β_1AR	3.50Å	TM1/TM2/Hx8, TM4/TM5

Chapter 3

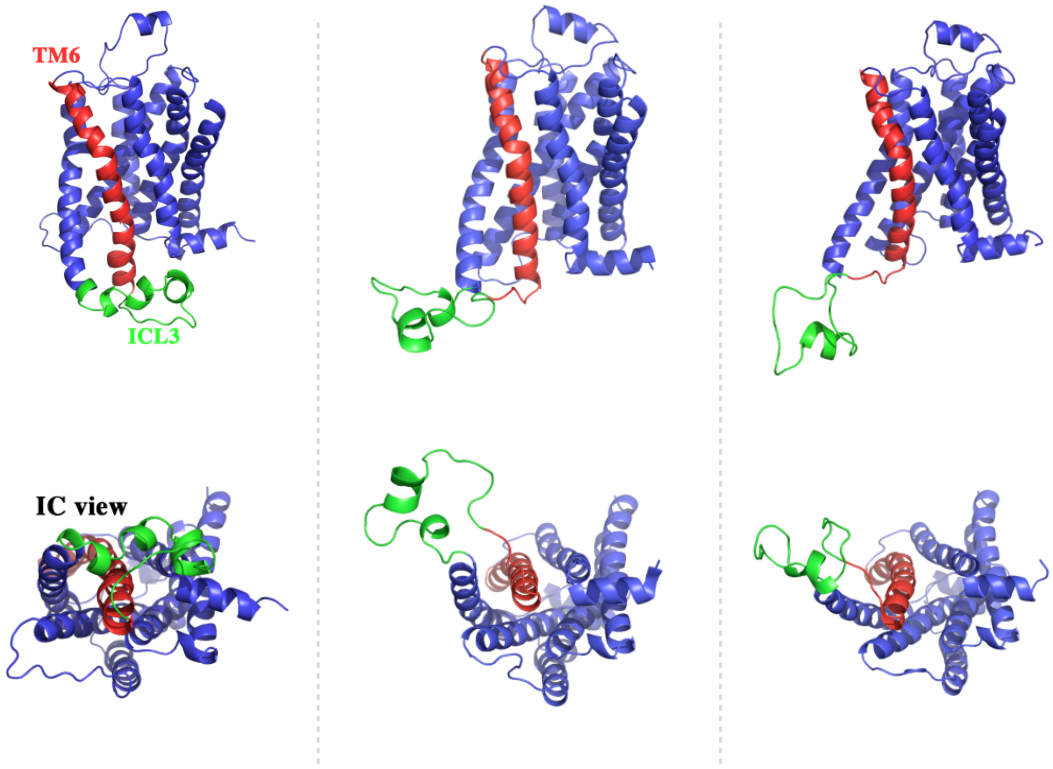
Results and Discussion

3.1 Peptide Docking

To determine the binding affinity of TM6 (transmembrane 6) to monomer ß$_2$AR (beta-adrenergic receptor), a set of docking experiments were conducted. In these docking runs, a peptide molecule consisting of a portion of TM6 was docked to monomer ß$_2$AR using a rigid body approach where both peptide and monomer were taken as rigid entities with no rotatable bonds. The monomeric conformation of the receptor was taken from the X-ray crystallographic structure of ß$_2$AR in complex with partial inverse agonist carazolol (PDB:2RH1, resolution: 2.40 Å). The complex structure includes another protein, T4 lysozyme (T4L), instead of the third intracellular loop (ICL3) region. The removal of highly unstructured ICL3 region had facilitated the crystallization process yielding a conformational stability.

In a previous Molecular Dynamics (MD) study (103), T4L was removed and the structure of ICL3 was generated via the homology modeling tool MODELLER (104). One 800 ns and three 100 ns long MD trajectories produced 5608 snapshots which were then clustered based on the root-mean-square deviation (RMSD) value of ICL3 after

being aligned to the starting conformation with respect to the transmembrane region. The clustering was performed using *K-means* algorithm implemented in the *k-clust* module of the Multiscale Modeling Tools for Structural Biology (MMTSB) Tool Set (105) and yielded a total of 18 clusters. All the snapshots in a cluster were averaged to produce a centroid and the conformation which would be closest to the centroid was selected as a representative of that cluster. The representatives of the 1st , 7th and the 11th clusters out of 18 are illustrated as in Figure 3.1. The side view and the view from the intracellular side (IC view) of these three representatives of the receptor reveals the positioning of ICL3 through the TM bundle. The 1st and 11th cluster representatives have looser conformations comparatively to the conformation of the 7th cluster representative. The position of ICL3 is distant from the TM bundles of the representatives in the 1st and 11th clusters. The representative of the 7th cluster was considered to be the most adequate conformer for the docking study, as it has the most packed ICL3 under the receptor (Figure 3.1a). This type of closely packed conformation would not block the way of another monomer/peptide when the receptor was found in a complex structure.



a b c

Figure 3.1 Schematic illustrations of the the 7th (a) ,1st (b) and 11th (c) clusters' representative conformations which are closest to the centroids. The intracellular view (IC view) of the receptors are located under the side views of each conformations. 7th cluster representative's ICL3 region (green colored) is closely packed under the transmembrane (a). The conformation closest to the 7th cluster's centroid was chosen for the docking studies. Red colored region designated as TM6.

3.1.1 Short Peptide Docking

The short peptide molecule was derived from TM6 region consisting of residues from 276 to 296, GIIMGTFTLCWLPFFIVNIVH as synthesized in the experimental work by Hebert et al. (8). By adopting the same conformation as in the receptor, the peptide was rigidly docked to the receptor using AutoDock v4.0 software tool (106). Conformations were then reevaluated with a knowledge-based scoring function called DSX^{online} v0.88 (107). A rigid body approach was preferred due to computational difficulty of a whole flexibile peptide. Furthermore, a rigid peptide also represent a preserved helical segment likewise the corresponding TM6 preserves its secondary structure while it is embedded in the transmembrane bundle.

AutoDock was employed for the docking studies in order to fact that its capability to predict membrane protein complexes in a high accuracy rate (108-110). its robust and efficient The algorithm implemented in AutoDock was known as highly efficient in the blind docking peptide studies (111). The membrane environment was simply regarded by accounting the constitutive behavior of the GPCR dimers which eventually assign them as ligand independent and the formation of dimeric structure in the endoplasmic reticulum (13). In this way the lipid contribution to dimerization was excluded in this experiment.

AutoGrid v4.0 which is implemented in AutoDock was used to generate a three-dimensional energy grid prior to docking (106). Two different approaches were employed for determining the possible docking site for the peptide. First, a blind docking was performed using a grid box large enough to cover the entire receptor, yet

excluding the intracellular and extracellular loops as illustrated in Figure 3.2a. The grid spacing was increased to 0.45 Å to allow the highest amount of coverage. Taking the number of grid points as 126 (maximum allowed value) in *xy* dimensions and 95 in *z* dimension, the dimensions of the single box became 56.25 Å x 56.25 Å x 42.3 Å. By generating a grid volume that was large enough to cover the entire transmembrane region of the protein, a particular binding site preferred by the peptide over the other sites can be determined.

In the second approach, four separate grid boxes were generated with the default grid spacing of 0.375 Å. As shown in Figure 3.2b and Table 3.1, boxes #2 and #3 sufficiently cover almost all TM6 region whereas boxes #1 and #4 exclude this region completely. The number of grid points in all three *xyz* directions was set to 126 which is the maximum allowed number of grid points and corresponds to a dimension of 46.88 Å x 46.88 Å x 46.88 Å. For each five docking system as illustrated in Figure 3.2, a total of 200 docking runs were performed using Lamarckian genetic algorithm (GA-LS) (114) with default docking parameters (a sample parameter file for docking is given in Appendix A). At the end of each docking, all 200 conformers were clustered based on a 2 Å (RMSD) value. Additionally, all 200 poses were reevaluated with two knowledge-based scoring functions which are found in DSX^{online} v0.88 software tool.

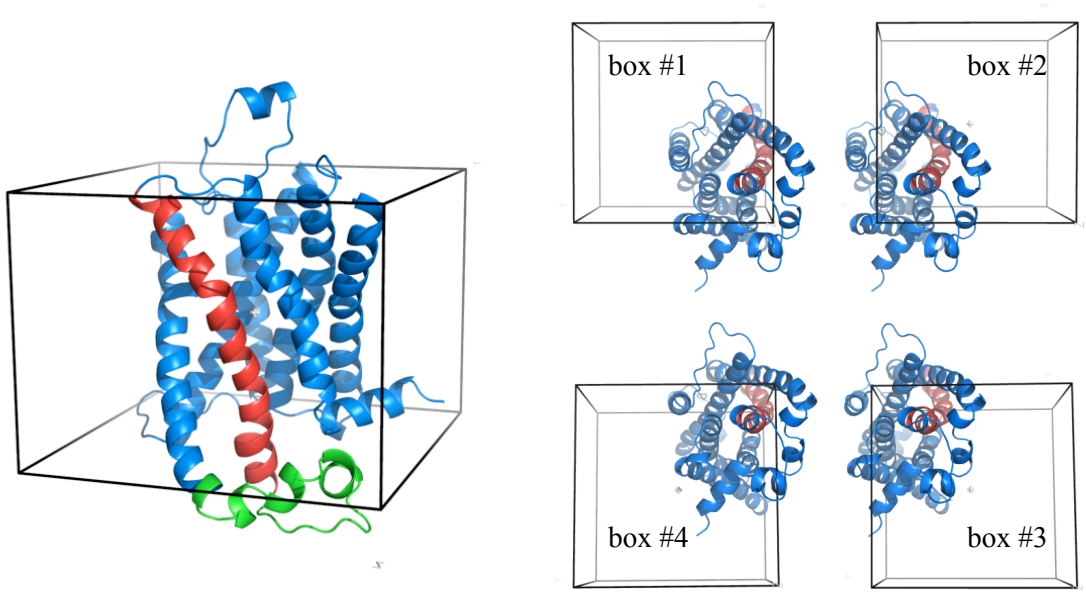


Figure 3.2 β_2AR monomer inside (a) a single box with .45 Å grid spacing and 56.25 x 56.25 x 42.3 Å dimensions, (b) four smaller boxes each with .375 Å grid spacing and 46.88 x 46.88 x 46.88 Å dimensions. The transmembrane 6 is colored in red and the ICL3 is colored in green. (115)

In blind docking using one single box (See Fig. 3.2a), conformers from the first cluster and the highest populated cluster preferred TM5 and TM6 as the binding site. Representative structures taken from the first cluster and the highest populated cluster can be seen in Figure 3.3a and Figure 3.3c. Especially, the pose from the first cluster was found to be well aligned near TM6. Additionally, the profiles in Figure 3.3b and Figure 3.3d illustrates the interacting residues of the receptor within a 5Å distance from the docked peptide. Clearly, TM5 and TM6 dominate as the most preferred binding site.

Moreover, DSX/CDS and DSX/PDB score values of the first cluster were found to be the highest among the first clusters of five docking results as listed in Table 3.1. For docking runs inside box #1 and box #4 which do not include TM6 region, the most preferred binding site was determined as TM2, TM3 and TM4, as the peptide had the

highest AutoDock score value (See Figures 3.4a,b and 3.5a,b). The clustering analysis showed that the lowest energy clusters for box #1 and box #4, designated as the "first cluster" in Table 3.1, are the least populated among the others, with 38 and 43 conformations respectively. Also, the highest-populated cluster contains 109 and 73 conformers located mostly near TM3 and TM4 (See Figures 3.4c,d and 3.5c,d). Moreover, the DSX score value of their first cluster are the lowest ones among other first clusters, -89.9 and -93.3 for DSX/CDS and -100.2 and -98.4 for DSX/PDB.

For docking runs inside box #2 and box #3 which include the whole TM6 region, TM5 and TM6 were determined as the most preferred binding site regions (See Figures 3.6a-d). Interestingly, the number of conformations in the first cluster is found as 199 and 197 for boxes #2 and #3, respectively (See Table 3.1). This is a significant amount of conformers observed in the first cluster, considering the total number of conformers as 200. Moreover their DSX score values were found to be relatively high among all five docking runs. All these docking results are in the favor of TM5 and TM6 which can be suggested as the most preferred binding site with the highest number of docked conformers in the first cluster and the highest score values.

Table 3.1 Docking results for the short peptide consisting of residues from 276 to 296 in TMVI. (115)

		Blind Docking	**Docking with a Binding Site**			
		Covered TM Regions	Box Number and Covered TM Regions			
		All TM helices	Box #1 TM:(2-5)	Box #2 TM:(3,5-7)	Box #3 TM:(1,5-8)	Box #4 TM:(1-4,7,8)
First Cluster	Size[1]	66	38	199	197	43
	TM Position[2]	5,6	2,3,4	5,6	1,5,6,7	2,3,4
	AD score[3]	**-10.5**	**-10.6**	**-10.6**	**-10.4**	**-10.6**
	DSX / CSD score[4]	**-102.6**	-89.9	**-101.0**	-93.3	-88.2
	DSX / PDB score[5]	**-114.2**	-100.2	**-111.5**	**-112.3**	-98.4
Highest-Populated Cluster	Size	84	109	_	_	73
	TM Position	1,5,6,7	3,4,5	_	_	1,2,3,4
	AD score	-10.2	-9.1	_	_	**-10.6**
	DSX / CSD score	-97.7	-77.1	_	_	**-114.2**
	DSX / PDB score	**-115.9**	-98.1	_	_	**-127.9**

[1] Number of docked poses in the cluster
[2] Transmembrane region to which the docked pose is within a distance of 5 Å at most.
[3] Highest negative AutoDock score in the first cluster (kcal/mol)
[4] DSX / CSD score of the best pose of AutoDock (unitless)
[5] DSX / PDB score of the best pose of AutoDock (unitless)

In docking results of box #1 and box #4, the difference between the mean energies of clusters are less than 2.5 kcal/mol which is the standard deviation of the AutoDock force field (106). Therefore, it is not easy to predict which cluster is the correct one. The failure of AutoDock for these docking cases that do not include TM6 (box #1 and #4) suggests that there is no transmembrane helix strongly preferred in these confined regions. On the other hand, for docking runs where TM6 was incorporated, the docking results strongly suggest TM6 as the most preferred binding site based on both score value and the number of docked conformers in the first cluster.

On the other hand, it is important to mention about the orientations of these peptides with respect to the target monomer. Except the docking result of the highest populated cluster for box #4, in all docking runs the peptide was found to be antiparallel to $ß_2$AR monomer. In other words, the amino acid sequence in the peptide is the reverse of the corresponding amino acid sequence in TM6 on the receptor. Although $ß_2$AR dimers are known to exist in parallel orientations, previous studies have shown that in rare occasions they also form antiparallel dimers (49, 50, 74, 113, 114). However, despite this unfavorable orientation, the best pose of the peptide provide a valuable and satisfactory information about the preference of TM5 and TM6 over the other helices.

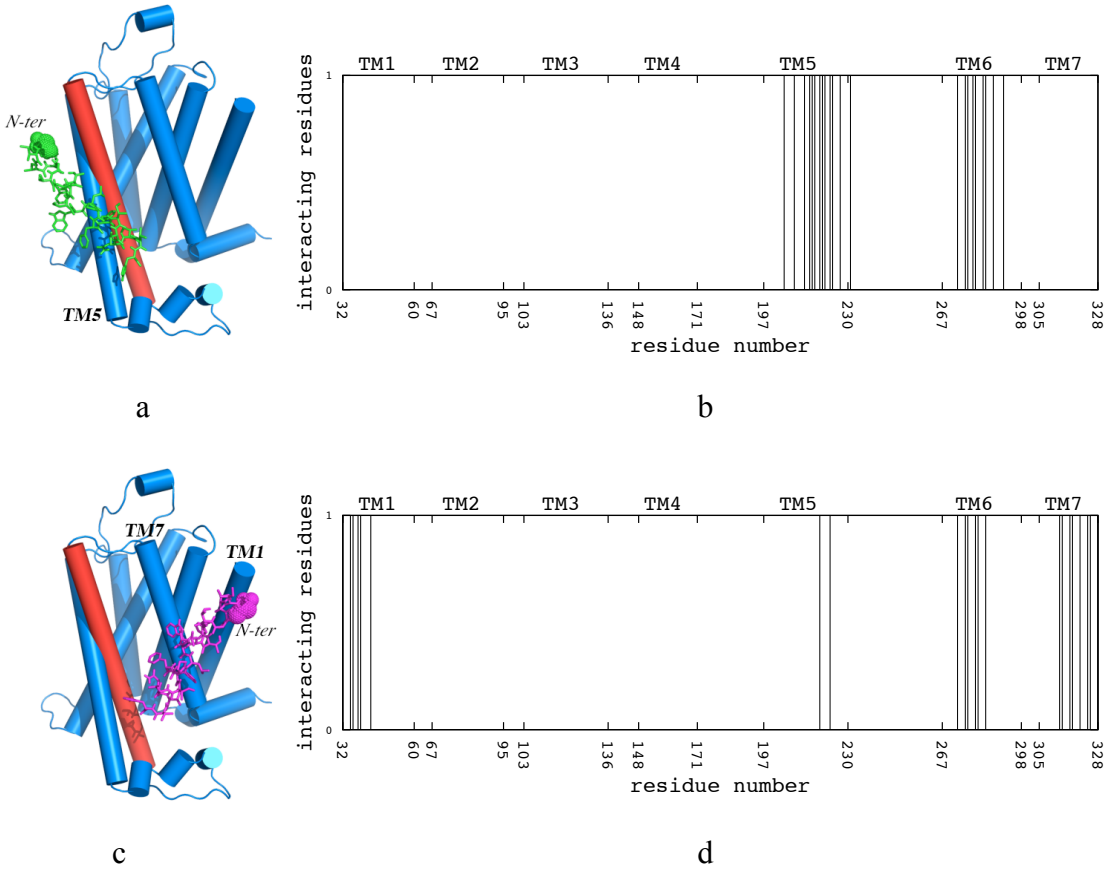
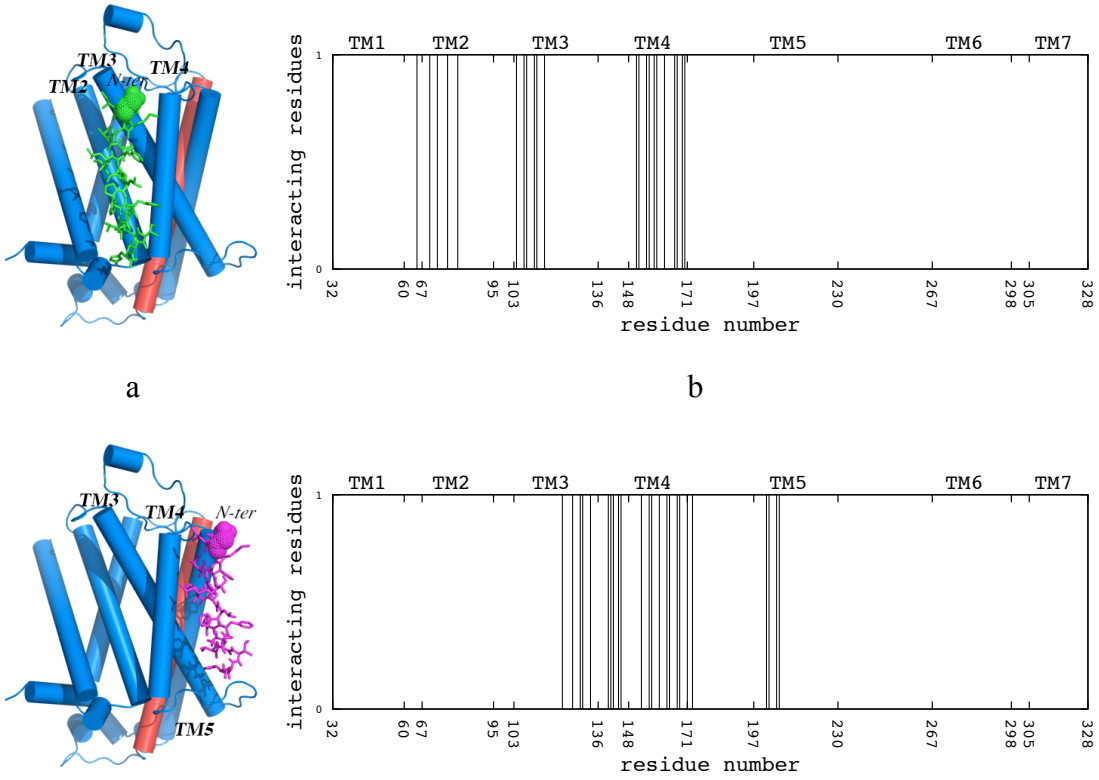


Figure 3.3 (a) Snapshot of the short peptide's best pose (green) using one single box, (b) the corresponding profile of the neighboring residues, (c) snapshot of the short peptide's pose from highest-populated cluster (magenta) and (d) the corresponding profile of the neighboring residues. The dots on the peptides correspond to amino terminal end. TM6 region of the receptor illustrated in red. Interacting residues for the first cluster are Ile205, Tyr209, Val213, Met215, Val216, Phe217, Tyr219, Ser220, Arg221, Phe223, Gln224, Lys227, Gln231, Lys273, Gly276, Ile277, Met279, Gly280, Thr283, Leu284, Leu287, Ile291. Interacting residues for the highest populated cluster are Trp32, Gly35, Met36, Ile38, Val39, Ile43, Tyr219, Phe223, Lys273, Gly276, Ile277, Gly280, Thr281, Leu284, Trp313, Ile314, Val317, Asn318, Phe321, Leu324, Ile325, Arg328. (115)



a b

c d

Figure 3.4 (a) Snapshots of the short peptide's best pose (green) in box #1, (b) the corresponding profile of the neighboring residues, (c) snapshot of the short peptide's pose from highest-populated cluster (magenta) in box #1 and (d) the corresponding profile of the neighboring residues. Interacting residues for the first cluster are Gln65, Tyr70, Thr73, Cys77, Val81, Phe104, Glu107, Phe108, Ser111, Ile112, Leu115, Arg151, Val152, Leu155, Met156, Trp158, Ile159, Gly162, Phe166, Leu167, Ile169, Gln170. Interacting residues for the highest populated cluster are Glu122, Val126, Val129, Asp130, Phe133, Lys140, Tyr141, Gln142, Leu144, Leu145, Ile153, Met156, Val157, Val160, Leu163, Thr164, Leu167, Pro168, Met171, Trp173, Ala202, Ser203, Val206, Ser207.

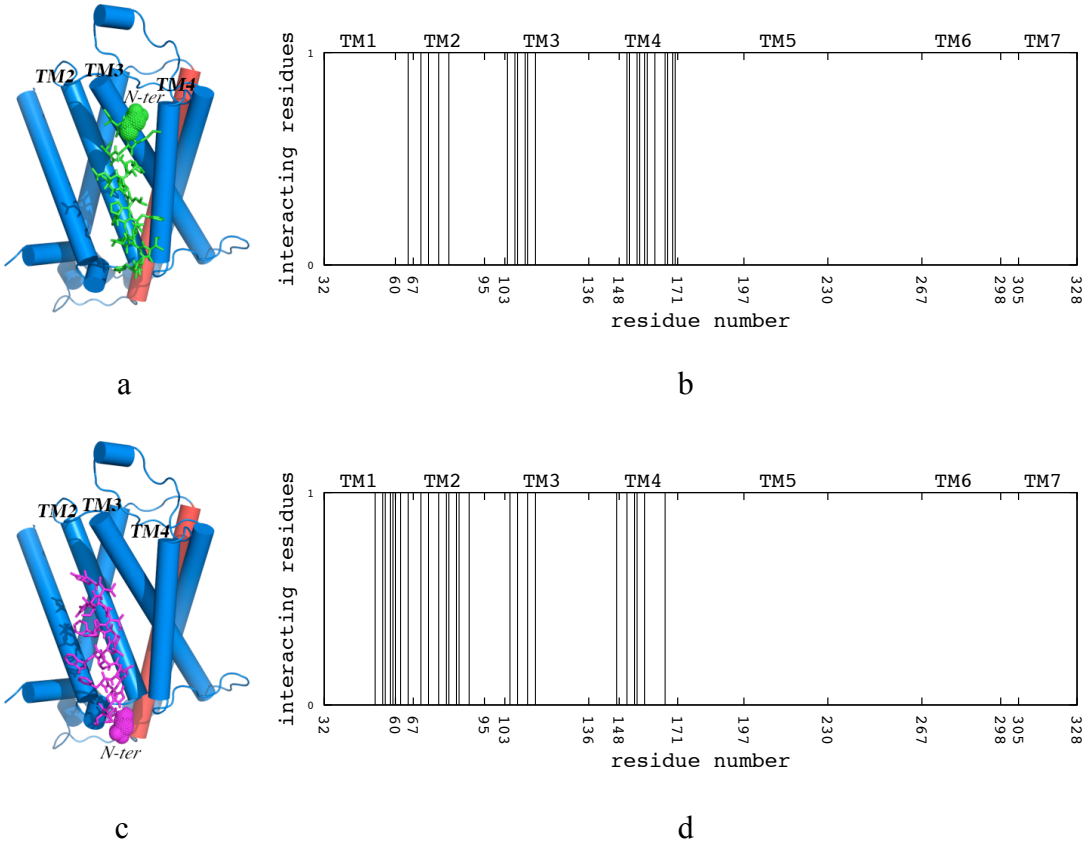


Figure 3.5 (a) Snapshot of the short peptide's best pose (green) in box #4, (b) the corresponding profile of the neighboring residues, (c) snapshot of the short peptide's pose from highest-populated cluster (magenta) in box #4 and (d) the corresponding profile of the neighboring residues. Interacting residues for the first cluster are Gln65, Tyr70, Thr73, Cys77, Val81, Phe104, Glu107, Phe108, Ser111, Ile112, Leu115, Arg151, Val152, Leu155, Met156, Trp158, Ile159, Gly162, Phe166, Leu167, Ile169, Gln170. Interacting residues for the highest populated cluster are Val52, Ile55, Thr56, Ile58, Ala59, Lys60, Glu62, Gln65, Tyr70, Thr73, Cys77, Leu80, Val81, Leu84, Ala85, Phe89, Trp105, Phe108, Ile112, Leu115, Lys147, Arg151, Ile154, Leu155, Trp158, Phe166.

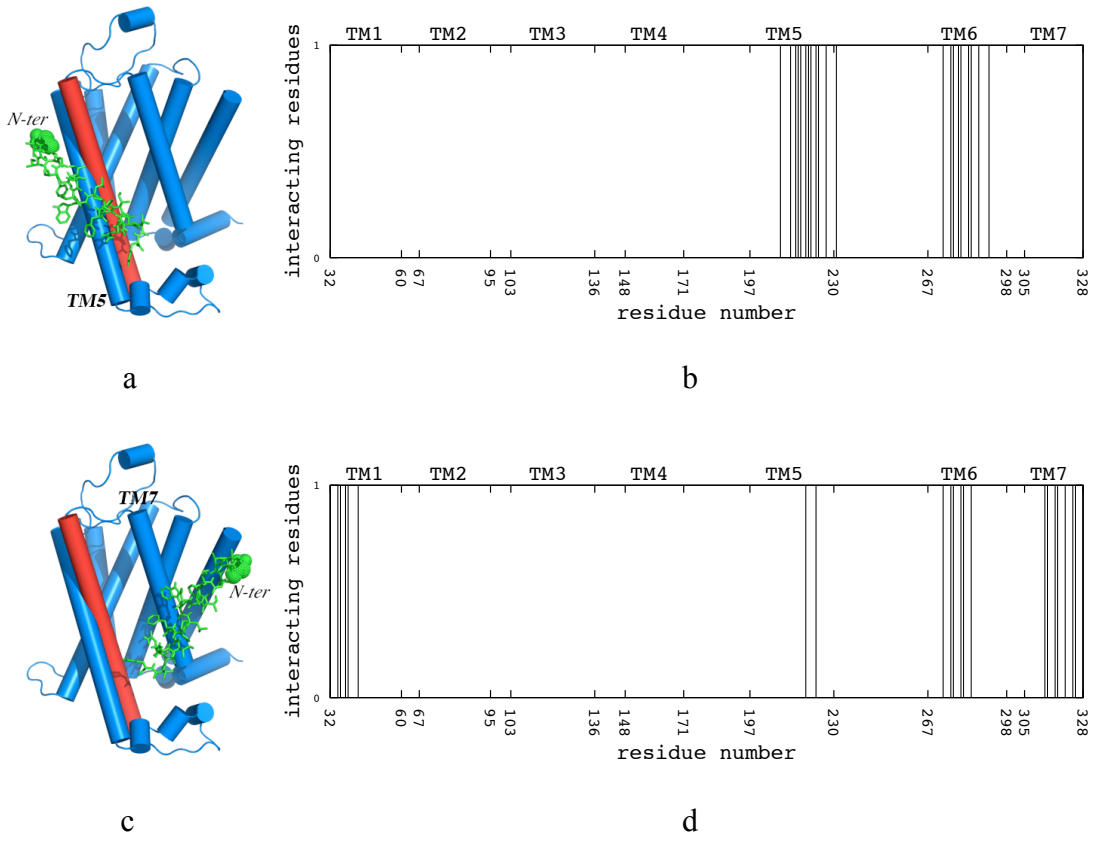


Figure 3.6 (a) Snapshot of the short peptide's best pose (green) in box #2, (b) the corresponding profile of the neighboring residues within a distance of 5Å, (c) Snapshot of the short peptide's best pose in box #3, (d) the corresponding profile of the neighboring residues within a distance of 5Å. Interacting residues for box #2 are Tyr209, Val213, Met215, Val216, Phe217, Tyr219, Ser220, Arg221, Phe223, Gln224, Lys227, Gln231, Lys273, Gly276, Ile277, Met279, Gly280, Thr283, Leu284, Leu287, Ile291. Interacting residues for box#3 are Trp32, Gly35, Met36, Ile38, Val39, Ile43, Tyr219, Phe223, Lys273, Gly276, Ile277, Gly280, Thr281, Leu284, Trp313, Ile314, Val317, Asn318, Phe321, Leu324, Ile325, Arg328.

3.1.2 Long Peptide Docking

Seven distinct peptide molecules were derived from each helical transmembrane regions of ß$_2$AR. Each peptide incorporates a large portion of the helix between residues 32-60, 67-95, 103-136, 148-171, 197-230, 267-298 and 305-328 corresponding to TM1 through TM7, respectively. Thus, for seven TM helices, seven independent blind docking experiments with 200 runs each, were performed using a single box with 0.45 Å grid spacing, similar to the blind docking of the short peptide. This approach enabled to compare the binding of TM6-derived peptide with those of other TM-derived peptides, to find out a possible significance of TM6 region to be at a certain site of the receptor, and consequently at the interface of a dimeric structure.

For each docking, all 200 docked poses were clustered with an RMSD value of 2 Å. The results of the first cluster and the highest-populated cluster are listed in Table 3.2. Remarkably, for all seven peptides, the best pose in the first cluster and the pose in the highest-populated cluster were always found near TM5, TM6 and TM7 regions. The score values ranked in the top three for all 10 clusters (7 *first* and 3 *highest populated*) are typed in bold and underlined as in the Table 3.1. Accordingly, the TM1-, TM2-, and TM6-derived peptides were found to dock with three high score values, which are all ranked in the top three and their best poses are located near TM5, TM6, and TM7 regions of the receptor. Moreover, for peptides derived from TM1 through TM4, the first cluster was observed to be the highest-populated cluster with a number of peptide conformations varying between 104 and 178, and additionally, their corresponding best

Table 3.2. Docking results for the seven long peptides derived from different transmembrane regions of the receptor.

		Docked Peptide						
		TM1	TM2	TM3	TM4	TM5	TM6	TM7
First Cluster	Size[1]	104	178	136	162	14	62	9
	TM Position[2]	5,6	5,6,7	5,6,7	5,6,7	1,5,6,7	6,7	6,7
	AD score[3]	**-9.5**	**-9.8**	**-10.6**	**-9.8**	-8.3	-9.1	-9.2
	DSX / CSD score[4]	-95.7	**-103.6**	-86.3	-82.0	-79.7	**-101.6**	-78.2
	DSX / PDB score[5]	**-115.2**	-71.8	-106.7	-99.2	-95.8	**-121.5**	-77.9
Highest-Populated Cluster	Size	–	–	–	–	52	73	58
	TM Position	–	–	–	–	5,6,7	5,6,7	5,6,7
	AD score	–	–	–	–	-7.9	-8.8	-8.8
	DSX / CSD score	–	–	–	–	**-100.3**	-90.7	-97.4
	DSX / PDB score	–	–	–	–	**-114.7**	-111.0	-110.2

[1] Number of docked poses in the cluster
[2] Transmembrane region to which the docked pose is within a distance of 5 Å at most.
[3] Highest negative AutoDock score in the first cluster (kcal/mol)
[4] DSX / CSD score of the best pose of AutoDock (unitless)
[5] DSX / PDB score of the best pose of AutoDock (unitless)

pose has a high negative AutoDock score varying between -10.6 and -9.5 kcal/mol. As listed in Table 3.2, the highest negative AutoDock score value of -10.64 belongs to TM3-derived peptide that was bound to the monomer near TM5, TM6, and TM7 regions. Moreover the highest negative DSX/PDB score value of -121.5 belongs TM6-derived peptide.

The position of the peptide conformation from the first cluster and the highest populated cluster were illustrated in Figure 3.7 alongside their interacting residues within a radius of 5 Å (See Figure 3.7b and d). Figure 3.7a and c illustrate the position of the best pose and the pose from the highest populated cluster of TM6-derived peptide interacting with the monomer's TM6 region. Similar profiles for the other six peptides are given in Appendix B. While TM6-derived peptide's best pose docked in an antiparallel orientation to the receptor, the peptide from the highest populated cluster's orientation with respect to the receptor is almost 90 degrees. Even if it is known that antiparallel dimeric membrane proteins do exist, a 90° degrees orientation in one monomer in a dimeric membrane protein is not plausible. Although the spatial positions of the peptides were not favorable, it is remarkable that TM6-derived peptide preferred to docked to TM6 region of the receptor with a very high score among the other dockings.

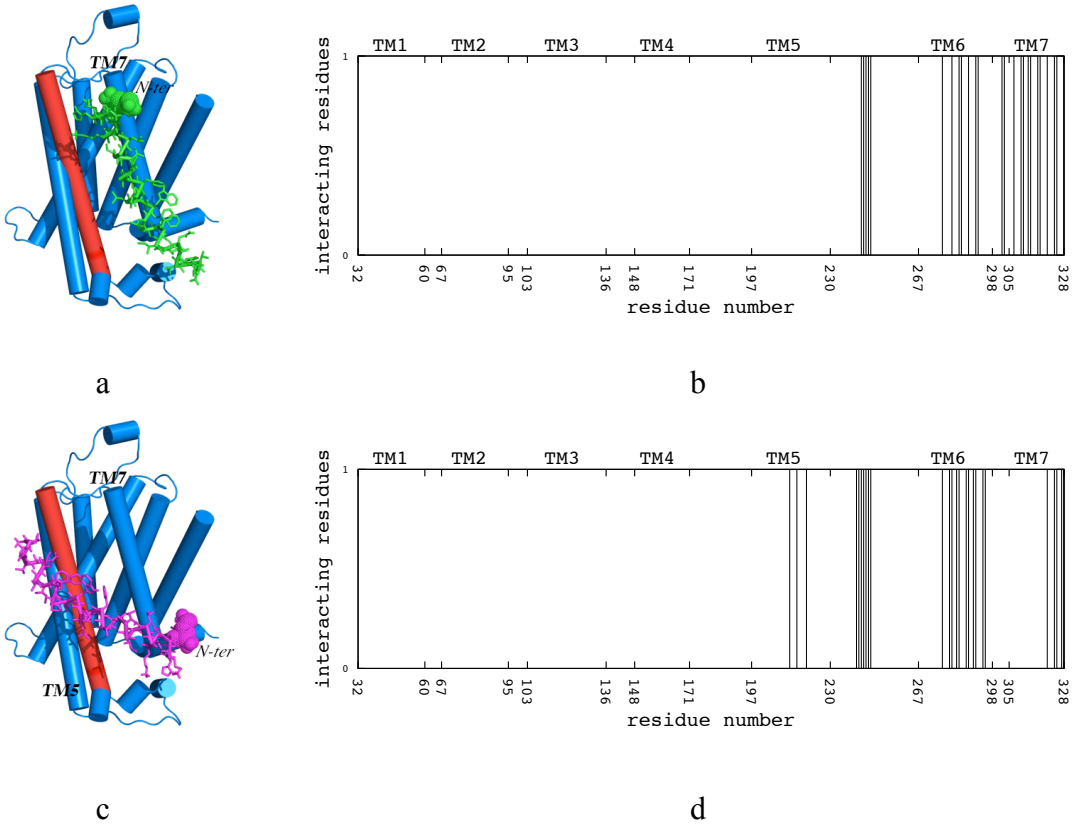


Figure 3.7 (a) Snapshot of the TM6-derived peptide's best pose (green), (b) the corresponding profile of the neighboring residues within a distance of 5Å, (c) Snapshot of the TM6-derived peptide from the highest populated cluster (magenta), (d) the corresponding profile of the neighboring residues within a distance of 5Å. Interacting residues for the first cluster are Gln243, Asn244, Leu245, Ser246, Gln247, Ile277, Thr281, Leu284, Cys285, Pro288, Ile291, Val292, Leu302, Ile303, Val307, Leu310, Leu311, Trp313, Ile314, Val317, Asn318, Phe321, Leu324, Ile325, Arg328. Interacting residues for the highest populated cluster are Val213, Val216, Ser220, Hse241, Val242, Gln243, Asn244, Leu245, Ser246, Gln247, Ile277, Gly280, Thr281, Thr283, Leu284, Leu287, Pro288, Phe290, Ile291, Ile294, Val295, Phe321, Leu324, Ile325, Cys327, Arg328, Arg333. (115)

3.2 Protein-Protein Docking

In this part of the thesis, the protein-protein docking tool ZDOCK v3.0.2 (116) and HADDOCK2.2 (117, 118) was used to estimate the dimer structure of ß$_2$AR and consequently the interface region. Two identical copies of the monomeric conformation of ß$_2$AR, that were also used in the previous peptide dockings, were docked to each other to generate a set of dimer forms of the receptor, which are further analyzed with the membrane topology filtering, cluster analysis and rescoring. The existence of TM6 region at the interface of these dimeric structures had been investigated.

3.2.1 Setting up and analyzing the ZDOCK runs

ZDOCK program which is based on a FAST Fourier Transform takes two monomer coordinates as the input data and searches all possible binding modes in the translational and rotational space in order to estimate their relative orientation to generate the final complex structure. Moreover, it takes into account surface complementarity, electrostatics, and desolvation to find the optimal fit between two proteins.

The input parameters were provided via a web-based interface. Default parameters cannot be modified, however, one can select certain residues that are not likely to appear in the protein-protein interface and these residues will simply be blocked from the conformational search space.

The computing time scales as $1/\Delta 3$, where Δ is the rotational sampling interval. A value of 6˚ was used since it provides more accurate predictions by sampling 54,000 poses.

The average computing time with a 6˚ rotational sampling interval and a total of 128 x 128 x 128 grid points with a spacing of 1.2 Å is 30 min on a 16-processor IBM SP3.

By default, the program can yield up to 2000 complexes that have the highest scores among 54,000 poses in a single run. To obtain the largest possible amount of distinct complexes, the docking runs were separated into two groups. In group A, two identical conformations of the receptor that include ICL3 were used as the initial conformations subjected to the docking runs. In the case of group B, two identical ß$_2$AR structures excluding ICL3 were used as the input conformations. Furthermore, certain residues such as intra- and extracellular loop regions were blocked from being in the binding site during conformational search. Accepting the fact that the intracellular and extracellular loops couldn't be located at the interface region, blocking the loops prevented the generation of unrealistic dimer conformers for a membrane protein.

A total of eight different ZDOCK runs were performed. Table 3.3 summarizes eight different layouts in protein-protein docking experiments, each generating 2000 complexes, which amounts to a total of 16,000 conformers to be analyzed.

All 16,000 dimer conformations were filtered based on their "membrane topology" feature consisting of the tilt angle about the *z*-axis that is perpendicular to the surface of the membrane and the *z*-offset taken as the displacement of the geometrical center of the monomer along the *z*-axis. The tilt angle is calculated as the angle between the first principle coordinate axis and the *z*-axis. 0.4 radians and 6 Å were employed as maximum allowable values for tilt angle and z-offset, respectively.

Consequently, a total of 149 conformations have passed the membrane topology filtering. Further examination of these 149 conformations revealed that 47 of them were parallel dimers with monomers oriented in the same direction, whereas the remaining 102 conformers had the two monomers arranged with an antiparallel orientation. Although the antiparallel association was considered as a rare and nonphysiological state, it was previously observed in several crystal packing of GPCRs (49, 60, 74, 113, 114) .

Table 3.3. Set of docking experiments with different conformers and blocked residues.

	Set #	Presence of ICL3	Blocked residues	# of complexes
Group A	1	Yes	NONE	2000
	2	Yes	ECL2[1] and ICL3[2] on one monomer	2000
	3	Yes	All loops[3] on one monomer	2000
	4	Yes	All loops on both monomers	2000
Group B	5	No	NONE	2000
	6	No	ECL2 on one monomer	2000
	7	No	All loops[3] on one monomer	2000
	8	No	All loops[3] on both monomers	2000

[1] Extracellular loop region incorporating residues from 172 to 196.
[2] Intracellular loop region incorporating residues from 231 to 262.
[3] All intra and extracellular regions from 61 to 66, 96 to 102, 137 to 147, 172 to 196, 231 to 266, 299 to 304, 329 to 342.

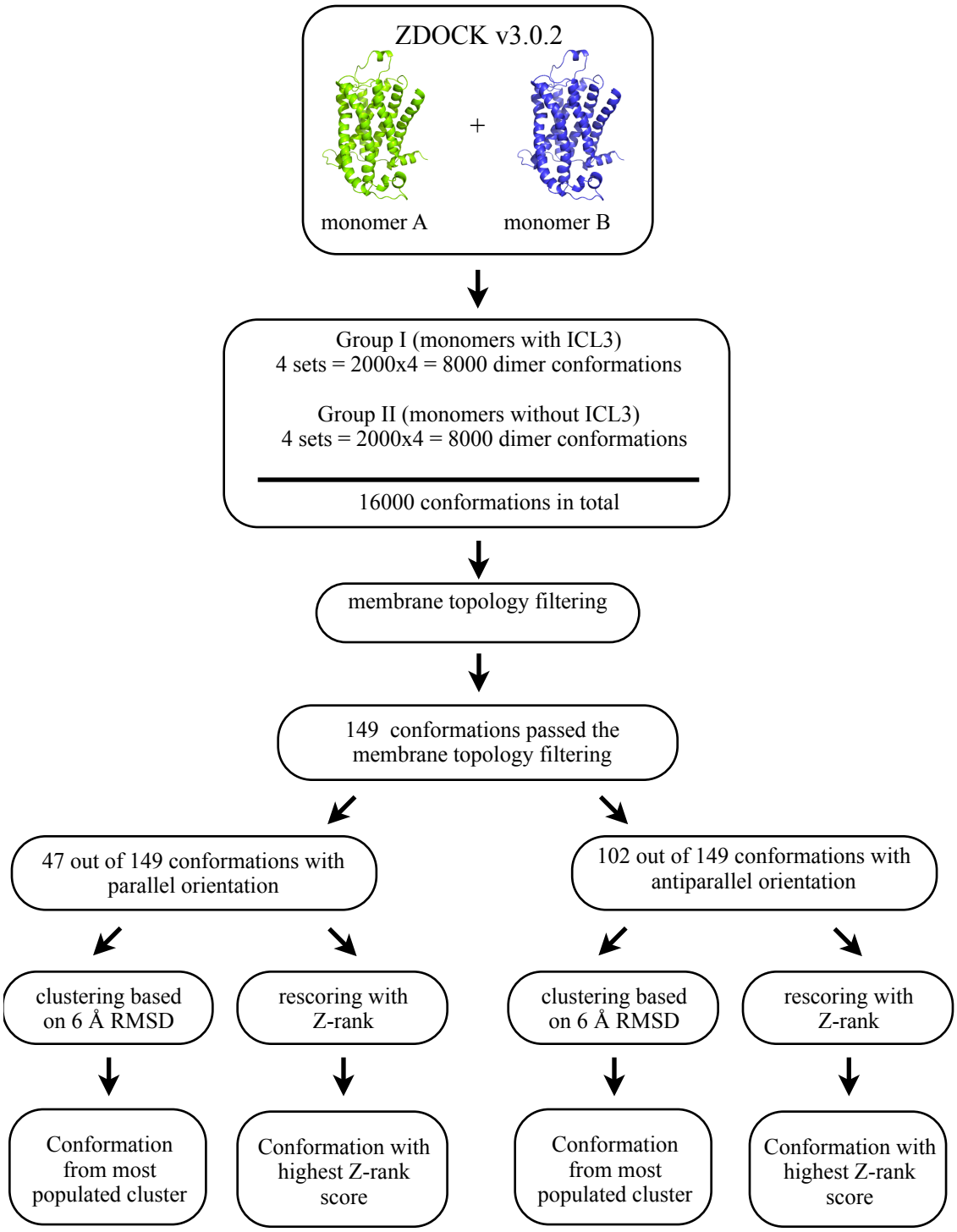


Figure 3.8. The flowchart that describes the procedure for determining the most plausible dimer conformation among the docked poses generated by ZDOCK. (115)

Thus, 149 complexes were separated into two groups based on their orientations, either parallel or antiparallel. Each group was further clustered based on an RMSD value of 6 Å using *k*-means clustering algorithm (105) which produced 20 and 56 clusters for parallel and antiparallel dimers, respectively. A representative structure from the highest-populated clusters were selected and inspected for the presence of TM6 at the interface. Furthermore, each group was reevaluated using a Z-rank scoring function which uses detailed electrostatics, van der Waals, and desolvation terms (119). The whole procedure described so far was summarized in a flowchart illustrated in Figure 3.8.

Table 3.4 shows the number of complexes conforming to the membrane topology requirements for each docking set listed in Table 3.3. While in sets #1 through #4, monomers have ICL3 region, in sets #5 through #8 monomers don't have it. On average, the percentage of the conformations with an acceptable tilt angle varies around 10% in every set as listed in the second column. In other words, neither the presence of ICL3 nor the blocked residues succeed to affect the tilt angle value.

On the other hand, however, the presence of ICL3 does not have a remarkable effect on the z-offset values, the number of conformations with an acceptable z-offset value varies with the choice of the blocked residues (see third column). In the case in which no residues were blocked as in set#1 (with ICL3) and #5 (without ICL3), the percentage is near 25% which is around 500 conformations out of 2000. When, only ECL2 and ICL3 were blocked on one monomer as in the case of set #2 (with ICL3) and #6 (without ICL3), the percentage is 20%. The smallest percentage of conformations satisfying the z-

Table 3.4. The percentage values of conformations that fit the membrane topology requirements. The last column shows the number of conformations that passed the filtering test. The actual values are given in parenthesis.

	Set #	tilt angle[1]	z-offset[2]	MemTop[3]
Group A	1	10.6 (212)	24.2 (484)	1.4 (28)
	2	10.75 (215)	20.15 (403)	1.0 (20)
	3	10.0 (200)	13.15 (263)	0.5 (10)
	4	12.75 (250)	25.5 (510)	0.85 (17)
Group B	5	9.65 (193)	22.9 (458)	1.25 (25)
	6	10.15 (203)	20.15 (403)	1.09 (22)
	7	10.40 (208)	14.45 (289)	0.65 (13)
	8	11.45 (229)	26.20 (524)	0.70 (14)

[1] Conformations that have a tilt angle value equals to or lower then 0.4 radians
[2] Conformations that have a z-offset value equals to or between -6 Å and 6 Å
[3] Conformations that fit both the tilt angle and z-offset

offset requirement was about 13% which belongs to set #3 (with ICL3) and #7 (without ICL3) for the case where all loops on one monomer were blocked.

When the area blocked from binding increased, the number of conformations that fulfilled the z-offset criterion diminished. Interestingly, when all loops on both monomers were blocked, the percentages increased to 25.5 and 26.2% for sets #4 (with ICL3) and #8 (without ICL3), which become the two highest percentage values for z-offset.

This might be explained by the fact that when all the loops on both monomers were blocked, the conformational search produced more realistic dimer conformations with a

correct membrane topology. This case was also observed for the tilt angle which was determined as the highest two values, 12.75 and 11.45% for sets #4 and #8, respectively.

Nevertheless the percentage of the conformations that have both acceptable tilt angle and *z*-offset drastically decreased in set #4 and #8 with respect to other percentages. The number of conformations that satisfied both criterion are listed on the last column of Table 3.4. The highest percentage values are 1.4 and 1.25 % corresponding to sets #1 and #5, respectively, where there were no blocked residues. The lowest percentage values belongs to set #3 and #7 and they are 0.5 and 0.65 %. These were the two cases where all the loops were blocked on one monomer only.

To understand how all 2000 conformations generated for each eight set were distributed based on their tilt angle, z-offset and Z-rank score value, a profile was created for each set as illustrated in Figure 3.9 and Figure 3.10. In all eight profiles, the dot represents each conformation's z-offset and tilt angle on the x- and y-axis, respectively, and the color of the dot illustrates the range in which the Z-rank score value was found (See the color bar in Figures 3.9 and 3.10).

As seen in the profiles, tilt angle values are highly scattered. Z-rank scores varies between -125 and 55 for all 16000 conformations. While the amount of the blocked residues increase, it is observed that the color of the profiles are changing from red to blue, that is Z-rank scores decrease. Set #1, #2, #5, #6 (Figure 3.9a, b, and Figure 3.10a, b) have relatively higher Z-rank scores. In cases of set #4 and #8 (Figure 3.9d and Figure

3.10d), they have the lowest Z-rank scores and the distribution of z-offset values is narrower.

Intersection of the black lines in the profiles designate the zone which contains the complexes conforming to the membrane topology requirements. It is also clearly seen that sets #1 and #5 (Figure 3.9a and Figure 3.10a) have the highest amounts of conformations in the MemTop zone.

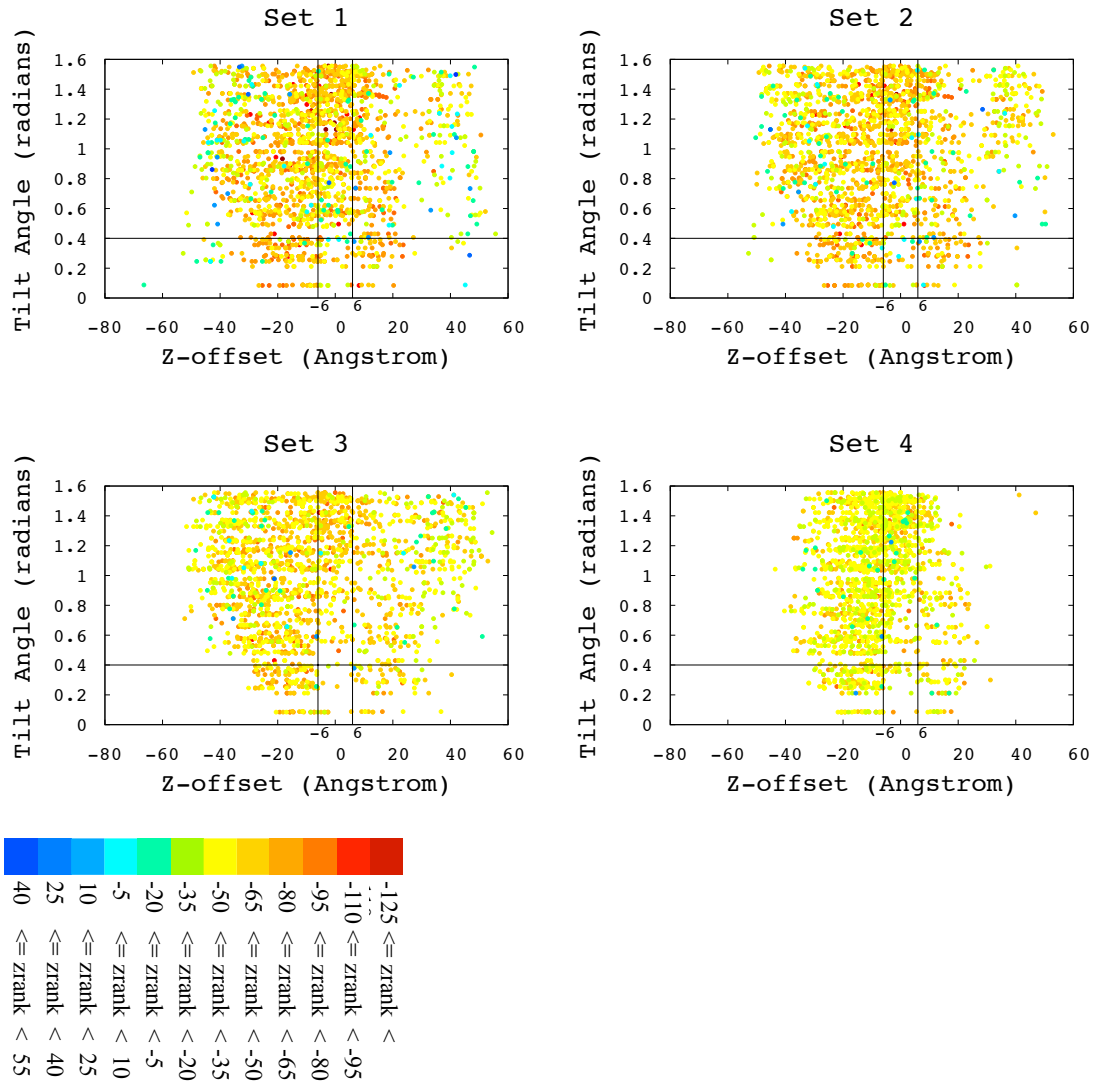


Figure 3.9. The z-offset, tilt angle and Z-rank score values of all 2000 conformations for Group A sets. The color bar indicates the Z-rank score values.

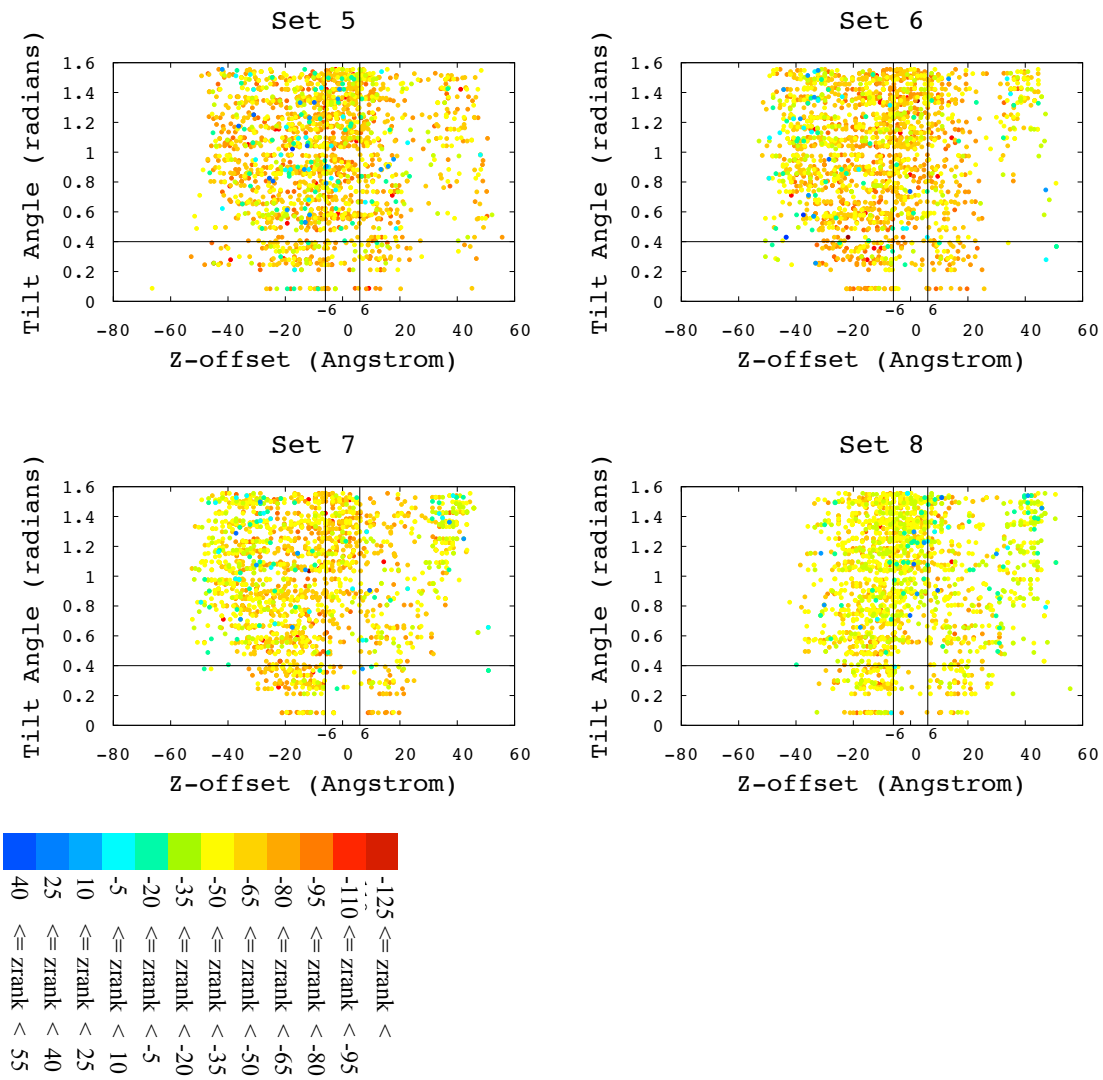


Figure 3.10. The z-offset, tilt angle and Z-rank score values of all 2000 conformations for Group B sets. The color bar indicates the Z-rank score values.

Cluster analysis was based on mutual similarity among the outcome conformations of the filtering. For the parallel and antiparallel dimer structures, the K-means algorithm (*k-clust*) implanted in the MMTSB Tool Set was used for an RMSD value of 6 Å. Cluster analysis produced 20 and 56 clusters for parallel and antiparallel dimers, respectively.

To determine the interface region between two monomers in a dimer, a quantitative approach was used. First, a representative complex from each cluster was selected as being the closest conformer to the cluster's average structure (or cluster's centroid). Next, the two monomers in the complex were set apart without being subject to any local conformational change, leaving them completely disconnected. Finally, the amount of change in solvent accessible surface area (SASA) of each residue *i* upon complex formation was determined using the following equation,

$$dSASA_i = \frac{SASA_{u,i} - SASA_{c,i}}{SASA_{u,i}} \qquad (3.1)$$

where $SASA_{u,i}$ and $SASA_{c,i}$ are the SASA values of residue *i* in the monomer before and after complex formation, respectively. The value of dSASA varies between 0 (apart from the interface) and 1 (completely buried in the interface). If a residue is not located at the interface, its dSASA is expected to be zero since no change in its SASA value is observed upon dimerization, whereas, if a residue *i* is completely buried at the interface, the change in SASA value is expected to be close to one. All SASA calculations were conducted in VMD using the measure SASA module with a probe radius of 1.4 Å larger than the van der Waals radius (120).

Figure 3.11a and Figure 3.11b illustrate two of these $dSASA_i$ profiles that represent the most-populated clusters obtained for parallel and antiparallel dimers, respectively. Accordingly, a TM region was considered to be at the interface, if any one of its residue's $dSASA_i$ was greater than 0. For parallel dimers, both monomers have their TM6 at the interface, whereas, only one of the monomer's TM6 region exists at the interface for antiparallel dimers, in their most populated cluster. Besides from all these, the sum of $dSASA_i$ values of amino acids in each individual TM regions were calculated for both monomers . These values thus quantitatively expose the TM regions that take part in the interface either in both or a single monomer. In parallel associations, TM5 and TM6 regions of both monomers are found in the interface. While TM5 region of Monomer A (the green colored monomer in Figure 3.11) has a total $dSASA_i$ value of 2.4, this value is 6.16 for the TM5 region of Monomer B (the blue colored monomer in Figure 3.11). The values for TM6 regions are 4.14 -1.97 in monomer A and monomer B, respectively (See in Figure 3.11b). TM7 region participates in the interface with a total $dSASA_i$ value of 3.38, solely via monomer A.

In antiparallel associations TM3 and TM4 regions of Monomer A are found in the interface with a a total $dSASA_i$ values of 2.21 and 5.34, respectively (See in Figure 3.11d). The total $dSASA_i$ values of TM5, TM6 and TM7 regions of Monomer B are 1.79, 320, and 4.20.

Profiles for the conformations representing the second populated clusters of parallel and antiparallel structures were generated by employing the same approach used in Figure 3.11 (Figure 3.12). A symmetrical dimerization was not observed in the parallel

conformation (Figure 3.12a). TM3, TM5, and TM6 of monomer A participate in the interface by the $dSASA_i$ values of 3.37, 3.34, and 1.69 (Figure 3.12b). TM6 contributes the interface region by only 3 residues (not shown in figure). Whereas, TM1 of monomer B dominantly takes part in the interface with a value of 4.05. In antiparallel conformation, TM5 and TM6 regions in both monomers are found in the interface. Especially, TM5 have high $dSASA_i$ values (2.84 - 4.47) on both monomers (Figure 3.12d).

Furthermore, based on the presence of TM6 at the interface evaluated by $dSASA_i$ values, each cluster was assigned to one of three states colored by red, yellow or blue, as illustrated in Figure 3.13a and Figure 3.13b where each vertical bar indicates the existence of TM6 at the interface as well as the number of elements in that cluster. According to the color code, the red clusters would indicate a dimer formation where TM6 regions of both monomers are found at the interface, whereas in the yellow clusters, only one of the monomer's TM6 would be present at the interface. Finally, the blue clusters would suggest that no TM6 is located at the interface.

In parallel associations, the most-populated cluster (6^{th} cluster) with 7 dimers corresponded to the "red state" where both TM6 were detected at the interface. Furthermore, in 11 out of 20 clusters, TM6 was not observed at the interface. On the other hand, the total number of conformers in the remaining 9 clusters in which at least one TM6 is present is 26, which is greater than the total number of conformers in 11 clusters, which is determined as 21.

In antiparallel associations, the most-populated cluster had 8 dimers and was colored in yellow suggesting the presence of one monomer's TM6 at the interface. There exist a total of 26 clusters with 48 dimers in which one TM6 is at the interface. About 18 dimers found in 9 clusters have both TM6 at the interface, whereas 21 clusters with 36 dimers did not have a TM6 region at the interface. Overall, the number of dimers with at least one TM6 at the interface is greater than the number of dimers with no TM6, in both parallel and antiparallel associations.

Finally, each 47 parallel and 102 antiparallel dimer conformations were reevaluated using their Z-rank score value, which is a positive number for plausible protein-protein associations. As illustrated in Figure 3.14a and Figure 3.14b each conformer's Z-rank score is designated with a vertical bar colored based on the presence of TM6 at the interface as in Figure 3.13a and Figure 3.13b. It is noteworthy that for parallel dimers, the conformers having two TM6 regions at the interface (red cluster) also had significantly high score values on the average. For parallel dimers, the average Z-rank score values for the red, yellow, and blue clusters are determined as 69.3 ± 20.1, 43.3 ± 18.9, 39.5 ± 9.7, respectively. On the other hand, the average Z-rank score values for antiparallel dimers are calculated as 51.8 ± 19.5, 54.8 ± 21.4, and 55.7 ± 14.9 for the red, yellow and blue clusters, respectively. Clearly, all three values are found to be close to each other and near 50. Thus, the antiparallel dimers are not as distinguishable as parallel dimers with respect to their score values.

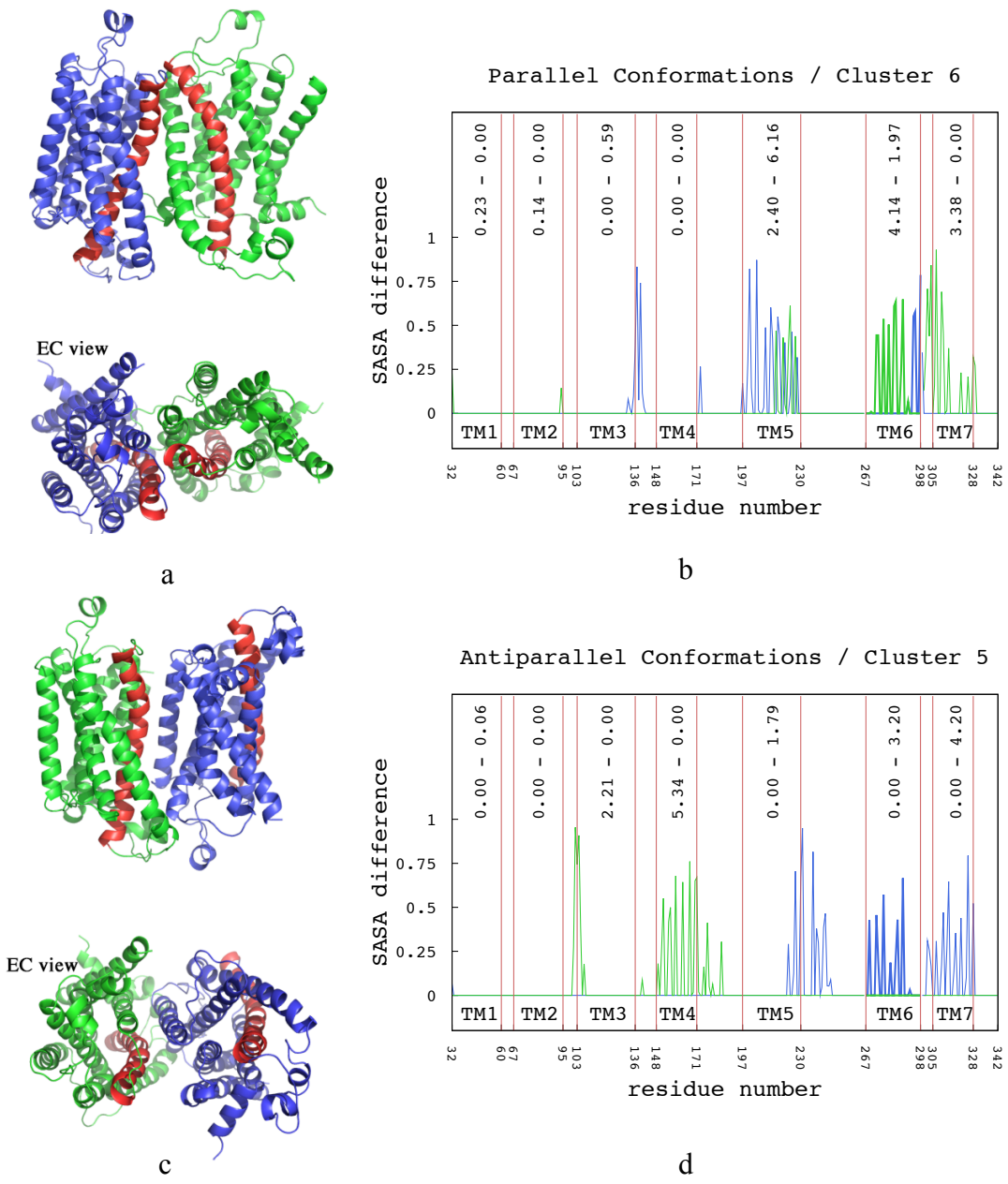


Figure 3.11 Representative snapshots of the most-populated cluster for (a) parallel and (c) antiparallel dimers with the views form the side and extracellular side of the receptor (EC view). Their corresponding $dSASA_i$ profiles illustrated in (b) and (d), respectively. The total $dSASA_i$ values were calculated for each TM regions in monomers (e.g., In parallel dimers, the total $dSASA_i$ value of TM6 in Monomer A is 4.14 and in Monomer B is 1.97). (115)

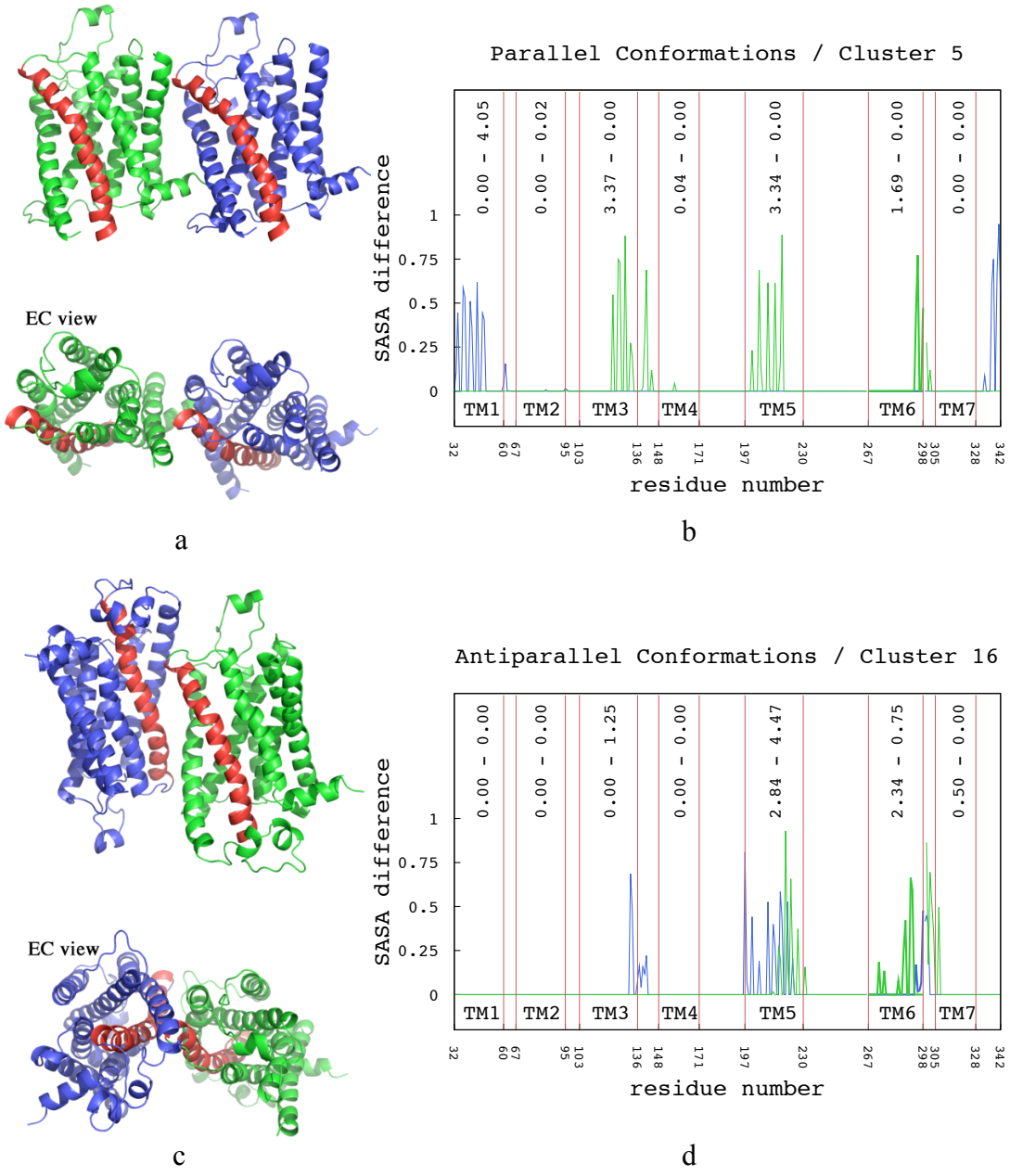
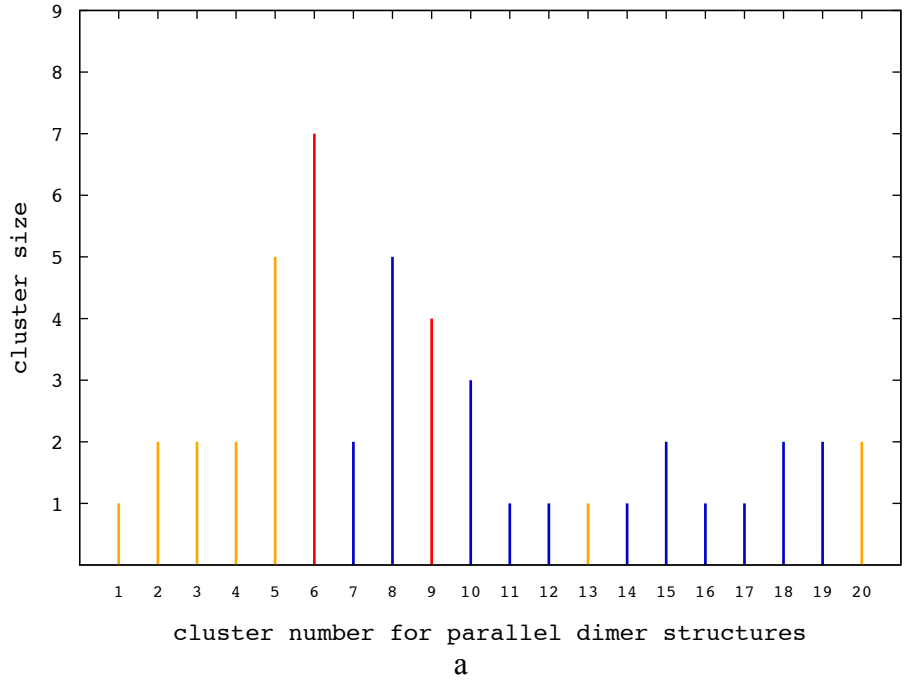
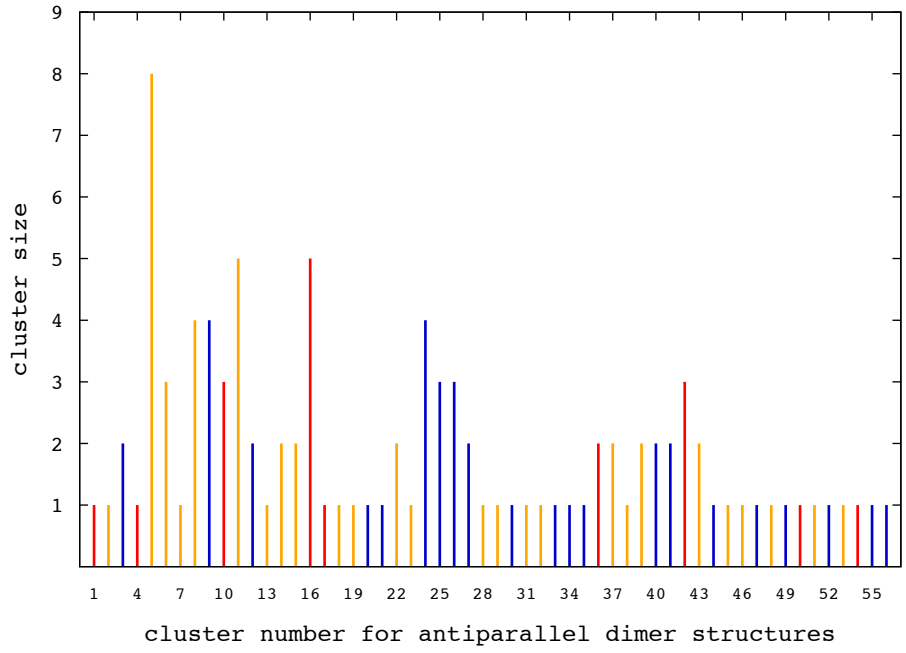


Figure 3.12 Representative snapshots of the second most-populated cluster for (a) parallel and (c) antiparallel dimers with the views form the side and extracellular side of the receptor (EC view). Their corresponding $dSASA_i$ profiles illustrated in (b) and (d), respectively.



a



b

Figure 3.13 Cluster profiles for parallel (a) and antiparallel (b) dimer structures passed the membrane topology filter. K-means cluster was performed with a radius of 6 Å (115)

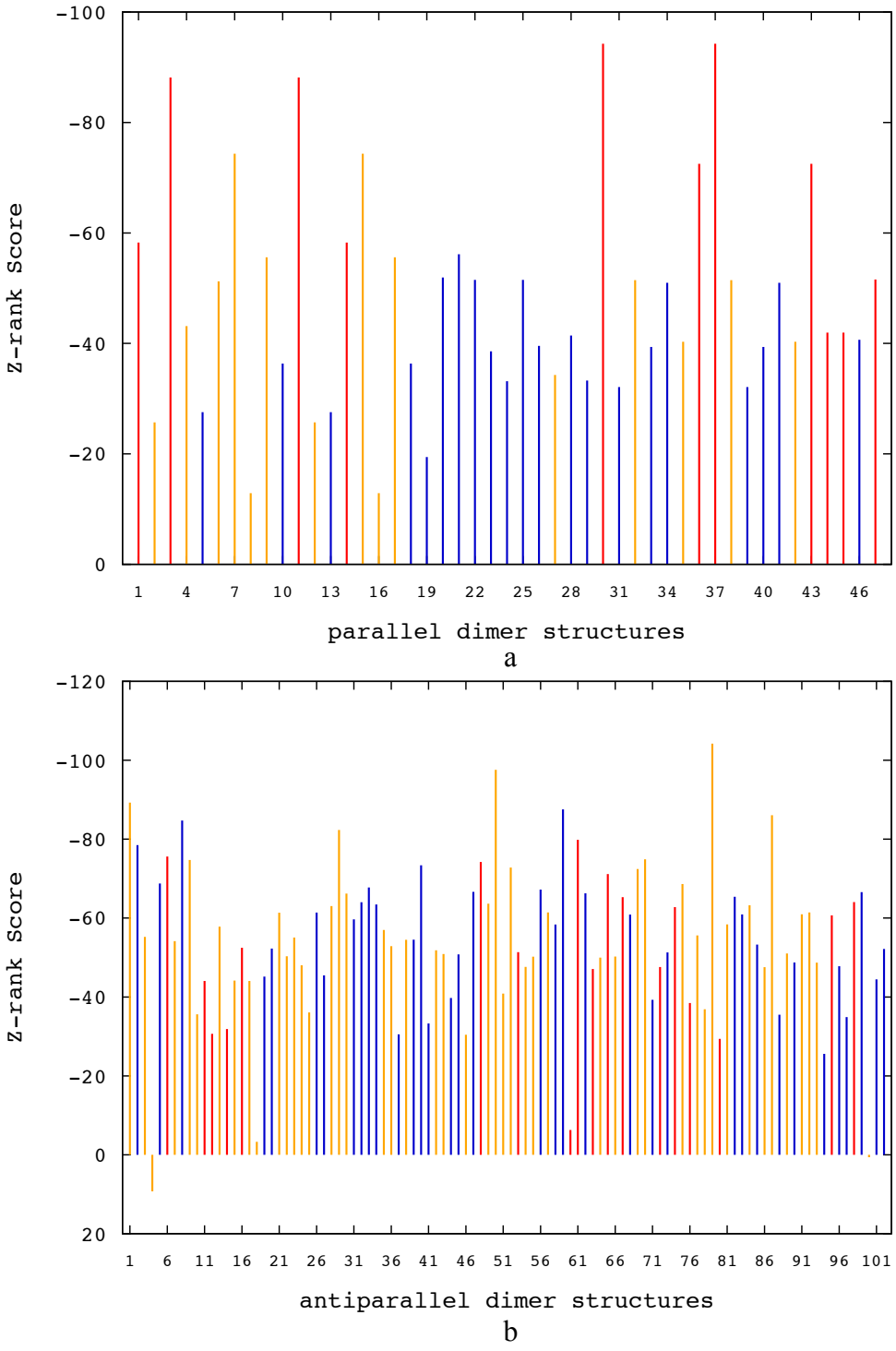


Figure 3.14 Z-rank scores for 47 parallel (a) and 102 antiparallel (b) dimer structures passed the membrane topology filter. (115)

3.2.2 HADDOCK (High Ambiguity Driven protein–protein DOCKing)

HADDOCK is driven by biochemical and/or biophysical interaction data, mutagenesis data or bioinformatics predictions. Ambiguous Interaction Restraints (AIRs) are derived from all this data prior to the docking process. An AIR is basically a collective input which includes the ambiguous intermolecular distances between the interacted residues in a complex structure (117). Before docking, a list of "active residues" (the known interacted residues within the dimer structure) and a list of "passive residues" (potential interacted residues) must be supplied for each molecule to generate AIRs. Since the inadequate experimental data on β_2AR dimer, a SASA calculation was applied on the monomeric state of β_2AR lacking of ICL3 region. Each residue of a transmembrane region that has a SASA value higher then 60 was introduced as the "active residues" (Table 3.5). "passive residues" were automatically generated by HADDOCK. A total of seven distinct runs each returning 200 complexes were carried out via HADDOCK2.2 web server interface for seven transmembrane regions. By this means, the evaluation of different TM interface regions with respect to each other was been possible. All SASA calculations were conducted in VMD using the measure SASA module with a probe radius of 1.4 Å larger than the van der Waals radius.

Table 3.5. Seven docking runs with the given active residues

TM #	Presence of ICL3	active residues	# of complexes
1	No	32, 33, 34, 36, 38, 39, 41, 42, 45, 49, 52 , 59, 60	200
2	No	84, 88, 94	200
3	No	103, 104, 105, 107, 108, 122, 134, 135	200
4	No	148, 149, 151, 152, 155, 156, 159, 163, 164, 166, 167, 170, 171	200
5	No	197, 198, 201, 202, 205, 206, 210, 213, 214, 216, 217, 220, 221, 223, 224, 225, 227, 228, 229, 230	200
6	No	270, 273, 277, 284, 287, 288, 291, 298	200
7	No	306, 310, 313, 321, 324, 328	200

The docking protocol takes the PDB files (in our case the two identical β_2AR monomers without ICL3) and AIRs to proceed the run in three steps. The first step applies random rotations to the monomers and then performs a rigid body energy minimization in which translations and rotations are allowed. By default 1000 complexes are generated in this stage. The best 200 complexes according to their energy score pass the second step where the structures are refined by a semi-flexible simulated annealing in torsion angle space. Both backbone and side-chain flexibility are allowed in the interface. The third step consisting of a refinement in Cartesian space with explicit solvent (e.g. water molecules). The resulting structures then are clustered with a cutoff 7.5 Å and these clusters are sorted according to the average HADDOCK score and average buried surface area (BSA) (121). HADDOCK score is a weighted sum of van der Waals, electrostatic, desolvation and restraint violation energies together with BSA. BSA calculation gives the size of the interface in a complex by subtracting the SASA of the

complex from the sum of SASA of the separate molecules. Each 200 complexes derived from seven runs filtered according to the same membrane topology criterion used in ZDOCK experiments. The runs consisting of active residues found in TM1, TM2, TM5 and TM7 had none of their complexes corresponded the membrane topology requirements. On the other hand, complexes of TM3 (4 complexes) and TM6 (2 complexes) runs passed the filtering test in low percentages whereas TM4 (14 complexes) run had a relatively high percentage of complexes with a membrane topology fitness (Table 3.6).

Table 3.7 summarizes the HADDOCK scores for the complexes that passed the membrane topology filtering. HADDOCK generates clusters whit a size higher then 4 therefore some complexes doesn't belong to any clusters. These complexes denoted as "none" in the table. The complex generated by TM6 run (Table 3.7 last row) presents in the second highest populated cluster with the highest average HADDOCK score (-100.937) among the other complexes in the table. Since HADDOCK score is positively correlated with BSA score, this complex has also the highest BSA score among others. In the context of orientation, TM4 run generated 3 parallel oriented complexes. Their HADDOCK score underlined at Table 3.7. All other complexes were found to be in an antiparallel orientation.

Table 3.6. The percentage values of complexes that fit the membrane topology requirements. The last column shows the number of conformations that passed the filtering test. The actual values are given in parenthesis.

Run	tilt angle[1]	z-offset[2]	MemTop[3]
TM1	0.0 (0)	19.5 (39)	0.0 (0)
TM2	4.5 (9)	43.5 (87)	0.0 (0)
TM3	15.5 (31)	18.5 (37)	2.0 (4)
TM4	20.5 (41)	49.0 (98)	7.0 (14)
TM5	0.0 (0)	68.5 (137)	0.0 (0)
TM6	7.0 (14)	60.0 (120)	1.0 (2)
TM7	4.0 (8)	69.0 (138)	0.0 (0)

[1] Conformations that have a tilt angle value equals to or lower then 0.4 radians
[2] Conformations that have a z-offset value equals to or between -6 Å and 6 Å
[3] Conformations that fit both the tilt angle and z-offset

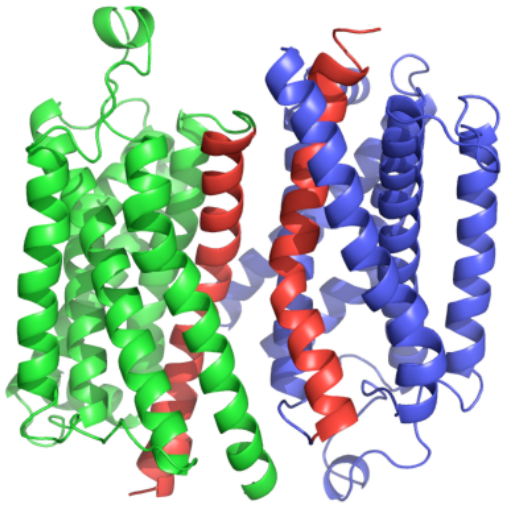


Figure 3.15 Representative snapshot for the complex that have the highest cluster score of -100.937. Red regions denoting TM6.

Table 3.7.HADDOCK results for the complexes fit membrane topology requirements

Run	amount of clusters[1]	cluster #	cluster size	cluster score	HADDOCK score	BSA[2]
TM3	20	6	8	-79.783	-121.4095	2480.75
		11	7	-76.605	-99.2595	2469.48
		13	6	-63.071	-94.0175	2270.71
		5	10	-63.933	-81.5857	2319.29
TM4	13	none	none	none	-116.53109	2358.64
		6	7	-71.125	-91.97653	2207.73
		none	none	none	-75.9008	2060.13
		none	none	none	**-73.25767**	1696.29
		none	none	none	-62.631408	1760.44
		3	19	-43.363	-48.0061	1644.11
		2	31	-42.874	**-47.67074**	1645.22
		2	31	-42.874	-40.61999	1369.25
		2	31	-42.874	-39.15158	1448.65
		8	7	-37.277	**-31.23355**	1610.76
		none	none	none	-29.91022	1589.88
		8	7	-37.277	-28.541716	1204.71
		5	10	-53.365	-12.11014	1334.71
		4	14	-34.402	-9.90614	1310.84
TM6	11	11	4	-75.983	-91.58812	2220.9
		2	29	-100.937	-85.184445	2510.48

[1] Total number of clusters generated in the run
[2] BSA: Buried Surface Area

Chapter 4

Conclusions

Certainly, determination of the dimer interface and the structural organization of the dimers could be a powerful tool in the development of more specific drugs. In this context, this study aimed to suggest a plausible interface candidate for β_2AR. Various computational and experimental studies (17-22) have directed a focus on TM6 as the interface region. Inspired by the peptide experiments conducted by Hebert and his coworkers (8), docking calculations were carried out using peptides derived from different transmembrane regions in the receptor. First, a peptide consisting of 23 residues derived from TM6 region just as the one used in Hebert et al.'s peptide experiments was blindly docked to β_2AR monomer. This blind docking approach enabled the peptide to bind arbitrarily to the TM bundle. The results of 200 distinct docking runs showed that the peptide preferred to bind to TM5 and TM6 regions with considerably high AD, DSX/CSD and DSX/PDB score values in both the first and the most-populated clusters.

Afterwards, the same peptide was also docked to various transmembrane regions which are confined in 4 different grid boxes. Especially, when the grid box was covering the TM6 region, almost 99% of all docked poses preferred to bind to TM5 and TM6 regions with relatively high score values. Interestingly, when the grid box did not incorporate

TM6, then all the score values and the number of conformations in the first cluster reduced significantly. Overall, the peptide strongly preferred TM6 region almost all the time when TM6 was inside the grid box, and interacted with other TM sites with less affinity and less frequently when TM6 was left outside the grid box.

Along with the short peptide docking experiments, each TM helix was derived into individual peptides (seven peptides in total), and was again blindly docked to β_2AR monomer. All seven TM-derived peptides preferred TM6 regions persistently with relatively high score values as observed in the previous docking results. Especially, it is noteworthy that the best pose of TM6- derived peptide interacted specifically with TM6 and TM7, and with a DSX/PDB score value of −121.5, which was the highest negative one observed in all peptide docking experiments.

The second part of our study was composed of protein– protein dockings where two identical β_2AR monomers were rigidly docked to each other. The analysis of 16,000 dimer complexes generated by these docking runs revealed that TM6 presents in the interface region of the dimer conformations with the highest Z-rank score values and in the highest populated clusters. TM6 was seen at the interfaces of both parallel and antiparallel orientations . The average Z-rank score for parallel dimer conformations was calculated as 69.3 ± 20.1, which was the highest average score among the docking results and TM6 region was found to be at the interface in both monomers.

A semi-flexible information driven docking was also employed for examining the docking behavior of β_2AR monomers. In this step, residues found in the surface of seven

TM regions were supplied as the interacted residues on both monomers. Right after the membrane topology filtering, the complex with a TM6 interface was found to be having the highest HADDOCK score.

The crystal structures of dimeric GPCRs (64, 88-94) have been obtained so far have various interface regions and this regions have symmetrical orientations. Moreover the crystal structures showed that the same GPCR is enable to inherent multiple interface regions in a row of dimers. Our results definitively point a TM5/TM6 interface which is also observed in the crystal structure of CXCR4 (PDB:3ODU) and μ-opioid (PDB: 4DKL), members of Class A.

Dimerization of a GPCR is a process which may occur in both the membrane (e.g. ligand-determined, constitutive dimerization) and the interior of the cell (internalization, ontogeny). Despite the notion that a docking study which entirely excludes a membrane environment is capable to generate an interface for only the dimerization occur in the cytoplasm, the found interface region by our docking experiments is also plausible for a membrane environment.

Evidently, TM6 has a key role in the receptor activation with the contribution of an outward movement which might also be involved in the dimerization since its effect on the membrane distortion eventually cause an energy cost. Therefore two monomers may form associations in order to alleviate this energy disruption in the membrane. This situation suggests a spontaneous nature for the dimerization.

On the other hand, It is still a subject of much debate whether one dimer activate one or two G proteins. The crystal structure of β_2AR with G_s G-protein and agonist (PDB : 3SN6) shows a contact point between Gα and helices TM5 and TM6. This cytoplasmic contact point is known to be too small for a simultaneous binding process. Therefore the proposition that two receptor molecules might be necessary to satisfy the binding requirements for G protein (29) was suggested. If this arrangement of one GPCR dimer interacting with a single heterotrimeric G protein is the functional complex in cells, an interface consisting of TM5/TM6 as in this study would favor a potent model for an active dimeric β_2AR structure.

References

(1) Filmore, D., "It's a GPCR world", *Modern Drug Discovery (American Chemical Society)*, 24-28 (November 2004).

(2) Overington, J. P., Al-Lazikoni, B., Hopkins, A. L., "How many drug targets are there?", *Nature Reviews Drug Discovery* **5**, 993-996 (2006).

(3) Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., Dixon, R. A., "Structure and function of G-protein coupled receptors", *Annual Review of Biochemistry* **63**, 101-132 (1994).

(4) Fung, B. K., Hurley, J. B., Stryer, L., "Flow of information in the light-triggered cyclic nucleotide cascade of vision", *Proceedings of the National Academy of Sciences* **78**, 152-156 (1981).

(5) Whorton, M. R., Bokoch, M. P., Rasmussen, S. G., Haung, B., Zare, R. N., Kobilka, B. K., Sunahara, R.K., "A monomeric G-protein coupled receptor isolated in a high density lipoprotein particle efficiently activates its G-protein.", *Proceedings of the National Academy of Sciences* **104**, 7682-7687 (2007).

(6) Angers, S., Salahpour, A., Joly, E., Hilairet, S., Chelsky, D., Dennis, M., Bouvier, M., "Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET)", *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3684-3689 (2000).

(7) Bouvier, M., "Oligomerization of G-protein-coupled transmitter receptors", *Nature Reviews Neuroscience* **2**, 274–286 (2001).

(8) Hebert, T. E., Moffett, S., Morello, J. P., Loisel, T. P., Bichet, D. G., Barret, C., Bouvier, M., "A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation.", *Journal of Biological Chemistry* **271**, 16384–16392 (1996).

(9) Milligan, G., "G protein-coupled receptor dimerization: Molecular basis and relevance to function", *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1768**, 825–835 (2007).

(10) Ianoul, A., Grant, D. D., Rouleau, Y., Bani-Yaghoub, M., Johnston, L. .J., and Pezacki, J. P., "Imagining nanometer domains of ß-adrenergic receptor complexes on the surface of cardiac myocytes", *Nature Chemical Biology* **1**, 196-202 (2005).

(11) Lavoie, C., Mercier, J. F., Salahpour, A., Umapathy, D., Breit, A., Villeneuve, L. R., ... Hebert, T. E., "β1/ β2-Adrenergic receptor heterodimerization Regulates β2-Adrenergic receptor internalization and ERK Signaling Efficacy.", *Journal of Biological Chemistry* **277**, 35402–35410 (2002).

(12) Mercier, J. F., Salahpour, A., Angers, S., Breit, A., Bouvier, M., "Quantitative assessment of beta 1- and beta 2-adrenergic receptor homo- and heterodimerization by bioluminescence resonance energy transfer", *Journal of Biological Chemistry* **277**, 44925–44931 (2002).

(13) Salahpour, A., Angers, S., Mercier, J. F., Lagace, M., Marullo, S., and Bouvier, M., "Homomerization of the beta 2-adrenergic receptor as a prerequisite for cell surface targeting", *Journal of Biological Chemistry* **279**, 33390–33397 (2004).

(14) Dorsch, S., Klotz, K. N., Engelhardt, S., Lohse, M. J., Bünemann, M., "Analysis of receptor oligomerization by FRAP microscopy", *Nature Methods* **6**, 225–230 (2009).

(15) Fung, J. J., Deupi, X., Pardo, L., Yao, X. J., Velez-Ruiz, G. A., DeVree, B. T., ... Kobilka, B. K., "Ligand-regulated oligomerization of β2-adrenoceptors in a model lipid bilayer", *EMBO Journal* **28**, 3315–3328 (2009).

(16) Gouldson, P. R., Snell, C. R., Bywater, R. P., Higgs, C., Reynolds, C. A., "Domain swapping in G protein- coupled receptor dimers", *Protein Engineering Design and Selection* **11**, 1181–1193 (1998).

(17) Gouldson, P. R., Bywater, R. P., and Reynolds, C. A., "Correlated mutations amongst the external residues of G protein-coupled receptors", *Biochemical Society Transactions* **25**, 529S (1997).

(18) Gouldson, P. R., and Reynolds, C. A., "Simulations on dimeric peptides: Evidence for domain swapping in G protein-coupled receptors?", *Biochemical Society Transactions* **25**, 1066–1071 (1997).

(19) Gouldson, P. R., Dean, M. K., Snell, C. R., Bywater, R. P., Gkoutos, G., and Reynolds, C. A., "Lipid-facing correlated mutations and dimerization in G-protein coupled receptors", Protein Engineering Design and Selection **14**, 759–767 (2001).

(20) Nilsson, A. M., Wijaywardene, M., Gkoutos, G., Wilson, K. M., Fernandez, N., Reynolds, C. A., "Correlated mutations in the HLA class II molecule.", *International Journal of Quantum Chemistry* **73**, 85–96 (1999).

(21) Oliveira, L., Paiva, A. C. M., and Vriend, G., "A common motif in G protein-coupled seven transmembrane helix receptors", *Journal of Computer-Aided Molecular Design* **7**, 649–658 (1993).

(22) Pazos, F., Helmer-Citterich, M., Ausiello, G., and Valencia, A., "Protein docking with sequences. Correlated mutations contain information about protein-protein interaction.", *Journal of Molecular Biology* **271**, 511–523 (1997).

(23) Nemoto, W., and Toh, H., "Prediction of interfaces for oligomerizations of G-protein coupled receptors", *Proteins* **58**, 644–660 (2005).

(24) Livingstone, C. D., and Barton, G. J. "Protein-sequence alignments: A strategy for the hierarchical analysis of residue conservation", *Computer Applications in the Biosciences* **9**, 745–756 (1993).

(25) Dean, M. K., Higgs, C., Smith, R. E., Bywater, R. P., Snell, C. R., Scott, P. D., ... Reynolds, C. A., "Dimerization of G-protein-coupled receptors", *Journal of Medicinal Chemistry* **44**, 4595–4614 (2001).

(26) Gkoutos, G. V., Higgs, C., Bywater, R. P., Gouldson, P. R., and Reynolds, C. A., "Evidence for dimerization in the β2-adrenergic receptor from the evolutionary trace method", International Journal of Quantum Chemistry **74**, 371–379 (1999).

(27) Ciruela, F., Casado, V., Mallol, J., Canela, E. I., Lluis, C., and Franco, R., "Immunological identification of A1 adenosine receptors in brain cortex", *Journal of Neuroscience Research* **42**, 818–828 (1995).

(28) Filizola, M., and Weinstein, H., "The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics", *FEBS Journal* **272**, 2926–2938 (2005).

(29) Liang, Y., Fotiadis, D., Filipek, S., Saperstein, D. A., Palczewski, K., and Engel, A., "Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes", *Journal of Biological Chemistry* **278**, 21655–21662 (2003).

(30) Fredriksson, R., Schioth, H. B., "The repertoire of G- protein-coupled receptors in fully sequenced genomes", *Mol Pharmacol* **67**(5), 1414–1425 (2005).

(31) Bjarnadóttir, T. K., Gloriam, D. E., Hellstrand, S. H., Kristiansson, H., Fredriksson, R., Schiöth, H. B., "Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse", *Genomics* **88** (3), 263–73 (September 2006).

(32) Fridmanis, D., et al "Formation of new genes explains lower intron density in mammalian rhodopsin G protein-coupled receptors", *Mol Phylogenet Evol* **43**, 864–880 (2007).

(33) Pierce, K., Premont, R., Lefkowitz, R., "Seven- transmembrane receptors", *Nat Rev Mol Cell Biol* **3**, 639–650 (2002).

(34) Kenakin, T., Miller, L. J., "Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery", *Pharmacol Rev* **62**(2), 265–304 (2010).

(35) Morris, A. J., Malbon, C. C., "Physiological Regulation of G Protein-Linked Signaling", *Physiological Reviews* **79(**4), 1373-1430 (1999).

(36) Mirzadegan, T., Benko, G., Filipek, S., Palczewski, K., "Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin", *Biochemistry* **42**(10), 2759–2767 (2003).

(37) Katritch, V. I., Cherezov, V., Stevens, R. C., "Diversity and modularity of G protein-coupled receptor structures", *Trends Pharmacol Sci.* **33**(1), 17-27 (2012).

(38) Mobarec, J., Sanchez, R., Filizola, M., "Modern homology modeling of G-protein coupled receptors: which structural template to use?", *J Med Chem* **52**, 5207–5216 (2009).

(39) Olivella, M., Gonzalez, A., Pardo, L., Deupi, X., "Relation between sequence and structure in membrane proteins", *Bioinformatics* **29**(13), 1589–1592 (2013).

(40) Ballesteros, J. A., Weinstein, H., "Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors", *Methods Neurosci* **25**, 366–428 (1995).

(41) Vaidehi, N., Bhattacharya, S., Larsen, A. B., "Structure and Dynamics of G-Protein Coupled Receptors", in Filizola, M. (ed.), *G Protein-Coupled Receptors - Modeling and Simulation,* 37-54, Springer Science Business Media Dordrecht, 2014.

(42) Ballesteros, J., Kitanovic, S., Guarnieri, F., Davies, P., Fromme, B. J., Konvicka, K., Chi, L., Millar, R. P., Davidson, J. S., Weinstein, H., "Functional microdomains in G-protein- coupled receptors", *J Biol Chem* **273**(17), 10445–10453 (1998).

(43) Visiers, I., Ebersole, B. J., Dracheva, S., Ballesteros, J., Sealfon, S.,C., Weinstein, H., "Structural motifs as functional microdomains in G-protein coupled receptors: energetic considerations in the mechanism of activation of the

serotonin 5HT2A receptor by disruption of the ionic lock of the arginine cage", *Int J Quantum Chem* **88**(1), 65–75 (2002).

(44) Prioleau, C., Visiers, I., Ebersole, B. J., Weinstein, H., Sealfon, S.,C., "Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT2C receptor", *J Biol Chem* **277**(39), 36577–36584 (2002).

(45) Barak, L., S., Menard, L., Ferguson, S. S .G., Colapietro, A. M., Caron, M. G., "The conserved seven- transmembrane sequence NP (X) 2, 3Y of the G- protein-coupled receptor superfamily regulates multiple properties of the beta 2-adrenergic receptor", *Biochemistry* **34**(47), 15407–15414 (1995).

(46) Fritze, O., Filipek, S., Kuksa, V., Palczewski, K., Hofmann, K. P., Ernst O. P., "Role of the conserved NPxxY (x) 5, 6F motif in the rhodopsin ground state and during activation", *Proc Natl Acad Sci U S A* **100**(5), 2290– 2295 (2003).

(47) Lagerstrom, M.C., Schioth, H. B., "Structural diversity of G protein-coupled receptors and significance for drug discovery", *Nat Rev Drug Discov* **7**(4), 339–357 (2008).

(48) Rosenbaum, D. M., Rasmussen, S. G., Kobilka, B. K., "The structure and function of G-protein-coupled receptors", *Nature* **459**(7245), 356–363 (2009).

(49) Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Trong, I. L., Teller, D. C., Okada, T., Stenkamp, R. E., "Crystal structure of rhodopsin: a G protein-coupled receptor", *Science* **289**(5480), 739–745 (2000).

(50) Warne, T., Serrano-Vega, M. J., Baker, J.G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G. W., Tate, C. G., Schertler, G. F. X., "Structure of a beta1-adrenergic G-protein-coupled receptor", *Nature* **454** (7203), 486–491 (2008).

(51) Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W.I., Kobilka, B. K., "High-resolution crystal structure of an engineered human beta2-adrenergic G protein coupled receptor", *Science* **318**(5854),1258–1265 (2007).

(52) Gonzalez, A., Cordomí, A., Matsoukas, M., Zachmann, J., Pardo, L., "Modeling of G Protein-Coupled Receptors Using Crystal Structures: From Monomers to Signaling Complexes", in Filizola, M. (ed.), *G Protein-Coupled Receptors - Modeling and Simulation,* 15-33, Springer Science Business Media Dordrecht, 2014.

(53) Klco, J. M., Nikiforovich, G. V., Baranski, T. J. “Genetic analysis of the first and third extracellular loops of the C5a receptor reveals an essential WXFG motif in the first loop.”, *J Biol Chem* **281**(17), 12010–12019 (2006).

(54) Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G.F., Babu, M. M., “Molecular signatures of G-protein-coupled receptors”, *Nature* **494**(7436), 185–194 (2013).

(55) Hurst, D. P., Grossfield, A., Lynch, D. L., Feller, S., Romo, T. D., Gawrisch, K., Pitman, M.C., Reggio, P. H., “A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor”, *J Biol Chem* **285**,17954–17964 (2010).

(56) Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., Xu, H., Shaw, D. E., “Pathway and mechanism of drug binding to G-protein-coupled receptors”, *Proc Natl Acad Sci U S A* **108**(32), 13118–13123 (2011).

(57) Gonzalez, A., Perez-Acle, T., Pardo, L., Deupi, X., “Molecular basis of ligand dissociation in beta- adrenergic receptors”, *PLoS One* **6**, e23815 (2011).

(58) Johnston, J. M., Aburi, M., Provasi, D., Bortolato, A., Urizar, E., Lambert, N. A., Javitch, J. A., Filizola, M., “Making structural sense of dimerization interfaces of delta opioid receptor homodimers”, *Biochemistry* **50**(10), 1682–1690 (2011).

(59) Sansuk, K., Deupi, X., Torrecillas, I. R., Jongejan, A., Nijmeijer, S., Bakker, R. A., Pardo, L., Leurs, R., “A structural insight into the reorientation of transmembrane domains 3 and 5 during family a G protein-coupled receptor activation”, *Mol Pharmacol* **79**(2), 262–269 (2011)

(60) Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., Devree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., et al “Structure of a nanobody- stabilized active state of the beta(2) adrenoceptor.”, *Nature* **469**, 175–180 (2011a).

(61) Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., et al “Crystal structure of the beta2 adrenergic receptor-Gs protein complex”, *Nature* **477**, 549–555 (2011b).

(62) Peeters, M. C., van Westen, G. J., Li, Q., Ijzerman, A. P. “Importance of the extracellular loops in G protein- coupled receptors for ligand recognition and receptor activation”, *Trends Pharmacol Sci* **32**(1), 35–42 (2011).

(63) Niesen, M., Bhattacharya, S., Vaidehi, N., “Conformational selection upon ligand binding in G-protein coupled receptors”, *J Am Chem Soc* **133**(33), 13197–13204 (2011).

(64) Vaidehi, N., Bhattacharya, S., "Multiscale computational methods for mapping conformational ensembles of G-protein coupled receptors", *Adv Protein Chem Struct Biol* **85**, 253–280 (2011).

(65) Bokoch, M. P., Zou, Y., Rasmussen, S. G., Liu, C. W., Nygaard, R., Rosenbaum, D. M., Fung, J. J., Choi, H. J., Thian, F. S., Kobilka, T. S., Puglisi, J. D., Weis, W. I., Pardo, L., Prosser, R. S., Mueller, L., Kobilka, B. K., "Ligand-specific regulation of the extracellular surface of a G-protein- coupled receptor", *Nature* **463**(7277), 108–112 (2010).

(66) Choe, H. W., Kim, Y. J., Park, J. H., Morizumi, T., Pai, E. F., Krauss, N., Hofmann, K. P., Scheerer, P., Ernst, O. P., "Crystal structure of metarhodopsin II", *Nature* **471**, 651–655 (2011).

(67) Standfuss, J., Edwards, P. C., D'Antona, A., Fransen, M., Xie, G., Oprian, D. D., Schertler, G. F., "The structural basis of agonist-induced activation in constitutively active rhodopsin", *Nature* **471**, 656–660 (2011).

(68) Xu, F., Wu, H., Katritch, V., Han, G. W., Jacobson, K. A., Gao, Z. G., Cherezov, V., Stevens, R. C., "Structure of an agonist bound human A2A adenosine receptor", *Science* **332**(6027), 322–327 (2011).

(69) Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Yao, X. J., Weis, W. I., Stevens, R. C., Kobilka, B. K., "GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function", *Science* **318**(5854), 1266–1273 (2007).

(70) Maurice, P., Guillaume, J. L., Benleulmi-Chaachoua, A., Daulat, A. M., Kamal, M., Jockers, R. "GPCR- interacting proteins, major players of GPCR function", *Adv Pharmacol* **62**, 349–380 (2011).

(71) Bockaert, J., Perroy, J., Becamel, C., Marin, P., Fagni, L., "GPCR interacting proteins (GIPs) in the nervous system: Roles in physiology and pathologies", *Annu Rev Pharmacol Toxicol* **50**, 89–109 (2010).

(72) Magalhaes, A. C., Dunn, H., Ferguson, S. S., "Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins", *Br J Pharmacol* **165**, 1717–1736 (2012).

(73) Oldham, W. M., Hamm H. E., "Heterotrimeric G protein activation by G-protein-coupled receptors", *Nat Rev Mol Cell Biol* **9**(1), 60–71 (2008).

(74) Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y., Velasquez, J., Kuhn, P., Stevens, R. C., "A specific cholesterol binding site is

established by the 2.8 A structure of the human beta2-adrenergic receptor", *Structure* **16**(6), 897-905 (June 2008).

(75) Chung, K. Y., Rasmussen, S. G., Liu, T., Li, S., DeVree, B. T., Chae, P. S., Calinski, D., Kobilka, B. K., Woods, V. L. Jr., Sunahara, R. K., "Conformational changes in the G protein Gs induced by the beta2 adrenergic receptor", *Nature* **477**(7366), 611–615 (2011).

(76) http://www.ebi.ac.uk/pdbe/quips?story=B2AR

(77) Gurevich, V. V., Gurevich, E. V., "The structural basis of arrestin-mediated regulation of G-protein-coupled receptors", *Pharmacol Ther* **110**(3), 465–502 (2006).

(78) Botelho, A. V., Gibson, N. J., Thurmond, R. L., Wang, Y., Brown, M. F., "Conformational energetics of rhodopsin modulated by nonlamellar-forming lipids", *Biochemistry* **41**(20), 6354–6368 (2002).

(79) Botelho, A. V., Huber, T., Sakmar, T. P., Brown, M. F., "Curvature and hydrophobic forces drive oligomerization and modulate activity of rhodopsin in membranes", *Biophys J* **91**(12), 4464–4477 (2006).

(80) Brown, M. F., "Modulation of rhodopsin function by properties of the membrane bilayer", *Chem Phys Lipids* **73**(1–2), 159–180 (1994).

(81) Gibson, N. J., Brown, M. F., "Role of phosphatidylserine in the MI-MIII equilibrium of rhodopsin*", *Biochem Biophys Res Commun* **176**(2), 915–921 (1991).

(82) Gibson, N. J., Brown, M. F., "Lipid headgroup and acyl chain composition modulate the MI-MII equilibrium of rhodopsin in recombinant membranes", *Biochemistry* **32**(9), 2438–2454 (1993).

(83) Mondal, S., Khelashvili, G., Shan, J., Andersen, O. S., Weinstein, H., "Quantitative modeling of membrane deformations by multihelical membrane proteins: application to G-protein coupled receptors", *Biophys J* **101**(9), 2092–2101 (2011).

(84) Mondal, S., Khelashvili, G., Wang, H., Provasi, D., Andersen, O. S., Filizola, M., Weinstein, H. "Interaction with the membrane uncovers essential differences between highly homologous GPCRs", *Biophys* J **102**(3), 514a (2012).

(85) Shan, J., Khelashvili, G., Mondal, S., Mehler, E. L., Weinstein, H., "Ligand-dependent conformations and dynamics of the serotonin 5-HT2A receptor

determine its activation and membrane-driven oligomerization properties", *PLoS Comput Biol* **8**(4), e1002473 (2012).

(86) Terrillon, S., Bouvier, M., "Roles of G-protein-coupled receptor dimerization", *EMBO Rep.* **5**(1), 30-4 (2004).

(87) Han, Y., Moreira, I. S., Urizar, E., Weinstein, H., Javitch, J. A., "Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation", *Nat Chem Biol* **5**(9), 688–695 (2009).

(88) Kusumi, A., Hyde, J. S., "Spin-label saturation-transfer electron-spin resonance detection of transient association of rhodopsin in reconstituted membranes", *Biochemistry* **21**, 5978–5983 (1982).

(89) Ryba, N. J. P., Marsh, D., "Protein rotational diffusion and lipid protein interactions in recombinants of bovine rhodopsin with saturated diacylphosphatidylcholines of different chain lengths studied by conventional and saturation-transfer electron-spin-resonance", *Biochemistry* **31**, 7511–7518 (1992).

(90) Periole, X., Huber, T., Marrink, S. J., Sakmar, T. P., "G protein-coupled receptors self-assemble in dynamics simulations of model bilayers", *J Am Chem Soc* **129**(33), 10126–10132 (2007).

(91) Periole, X., Knepp, A. M., Sakmar, T. P., Marrink, S. J., Huber, T., "Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers", *J Am Chem Soc* **134**(26), 10959–10965 (2012).

(92) Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., Granier, S., "Crystal structure of the mu-opioid receptor bound to a morphinan antagonist", *Nature* **485**, 321–326 (2012).

(93) Huang, J., Chen, S., Zhang, J. J., Huang, X. Y., "Crystal structure of oligomeric beta1-adrenergic G protein- coupled receptors in ligand-free basal state", *Nat Struct Mol Biol* **20**(4), 419–425 (2013).

(94) Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., et al "Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists", *Science* **330**, 1066–1071 (2010).

(95) Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. A., Huang, X. P., Carroll, F. I., et al "Structure of the human kappa-opioid receptor in complex with JDTic", *Nature* **485**, 327–332 (2012).

(96) Scheerer, P., Park, J. H., Hildebrand, P. W., Kim, Y. J., Krauss, N., Choe H. W., Hofmann, K. P., Ernst, O. P., "Crystal structure of opsin in its G-protein-interacting conformation", *Nature* **455**, 497–502 (2008).

(97) Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., Iwata, S., "Structure of the human histamine H1 receptor complex with doxepin", *Nature* **475** (7354), 65–70 (2011).

(98) Murakami, M., Kouyama, T., "Crystal structure of squid rhodopsin", *Nature* **453** (7193), 363–367 (2008).

(99) Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A., Palczewski, K., "Atomic-force microscopy: rhodopsin dimers in native disc membranes", *Nature* **421**, 127–128 (2003).

(100) Guo, W., Shi, L., Filizola, M., Weinstein, H., Javitch, J. A., "Crosstalk in G protein-coupled receptors: changes at the transmembrane homodimer interface determine activation", *Proc Natl Acad Sci U S A* **102**(48), 17495– 17500 (2005).

(101) Guo, W., Urizar, E., Kralikova, M., Mobarec, J. C., Shi, L., Filizola, M., Javitch, J. A., "Dopamine D2 receptors form higher order oligomers at physiological expression levels", *EMBO J* **27**(17), 2293–2304 (2008).

(102) Prasanna, X., Chattopadhyay, A., Sengupta, D., "Cholesterol Modulates the Dimer Interface of the b2-Adrenergic Receptor via Cholesterol Occupancy Sites", *Biophysical Journal Volume* **106**, 1920-1300 (2014).

(103) Ozcan, O., *Exploring the intrinsic dynamics of human beta-2 adrenergic receptor and its potential use in computational drug design studies*, M.S. Thesis, Bogazici University, 2011

(104) Narayanan, E., John, B., Mirkovic, N., Fiser, A., Ilyin, V. A., Pieper, U.,... Sali, A. , "Tools for comparative protein structure modeling and analysis", *Nucleic Acids Research* **31**, 3375–3380 (2003).

(105) Feig, M., Karanicolas, J., & Brooks, C. L., MMTSB Tool Set San Diego, CA: MMTSB NIH Research Resource, The Scripps Research Institute (2001).

(106) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J., "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility", *Journal of Computational Chemistry* **30**, 2785–2791 (2009).

(107) Neudert, G., & Klebe, G., "DSX: A knowledge-based scoring function for the assessment of protein-ligand complexes", *Journal of Chemical Information and Modeling* **51**, 2731–2745 (2011).

(108) Liu, R., Perez-Aguilar, J. M., Liang, D., & Saven, J. G., "Binding site and affinity prediction of general anesthetics to protein targets using docking", *Anesthesia and Analgesia* **114**, 947–955 (2012).

(109) Jois, S. D. S., & Siahaan, T. J., "A peptide derived from LFA-1 protein that modulates T-cell adhesion binds to soluble ICAM-1 protein", *Journal of Biomolecular Structure and Dynamics* **20**, 635–644 (2003).

(110) Song, S. X., Liu, D., Peng, J. L., Deng, H. W., Guo, Y., Xu, L. X., ... Xu, Y. H., "Novel peptide ligand directs liposomes toward EGF-R high-expressing cancer cells in vitro and in vivo", *The FASEB Journal* **23**, 1396–1404 (2009).

(111) Hetenyi, C., & Spoel, D. V. D., "Efficient docking of peptides to proteins without prior knowledge of the binding site", *Protein Science* **11**, 1737–1737 (2002).

(112) Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J., "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function", *Journal of Computational Chemistry* **19**, 1639–1662 (1998).

(113) Li, J., Edwards, P. C., Burghammer, M., Villa, C., & Schertler, G. F., "Structure of bovine rhodopsin in a trigonal crystal form", *Journal of Molecular Biology* **343**, 1409–1438 (2004).

(114) Warne, T., Moukhametzianov, R., Baker, J. G., Nehmé, R., Edwards, P. C., Leslie, A. G. W., ... Tate, C. G., "The structural basis for agonist and partial agonist action on a β1-adrenergic receptor", *Nature* **469**, 241–244 (2011).

(115) Koroglu, A., Akten, E. D., "Transmembrane helix 6 observed at the interface of β2AR homodimers in blind docking studies", *Journal of Biomolecular Structure and Dynamics* 33(7), 1503-1515 (2015).

(116) Chen, R., Li, L., & Weng, Z., "ZDOCK: An initial-stage protein-docking algorithm", *Proteins: Structure, Function, and Genetics* **52**, 80–87 (2003).

(117) Dominguez, C., Boelens, R., Bonvin, A. M. J. J., "HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information", *J. Am. Chem. Soc.* **125**, 1731-1737 (2003).

(118) de Vries, S. J., van Dijk, A.D.J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T., Bonvin, A. M. J. J., "HADDOCK versus HADDOCK:

New features and performance of HADDOCK2.0 on the CAPRI targets" *Proteins: Struc. Funct. & Bioinformatic* **69**, 726-733 (2007).

(119) Pierce, B., & Weng, Z., "ZRANK: Reranking protein docking predictions with an optimized energy function", *Proteins: Structure, Function, and Bioinformatics* **67**, 1078–1086 (2007).

(120) Humphrey, W., Dalke, A., & Schulten, K., "VMD: Visual molecular dynamics", *Journal of Molecular Graphics* **14**, 33–38 (1996).

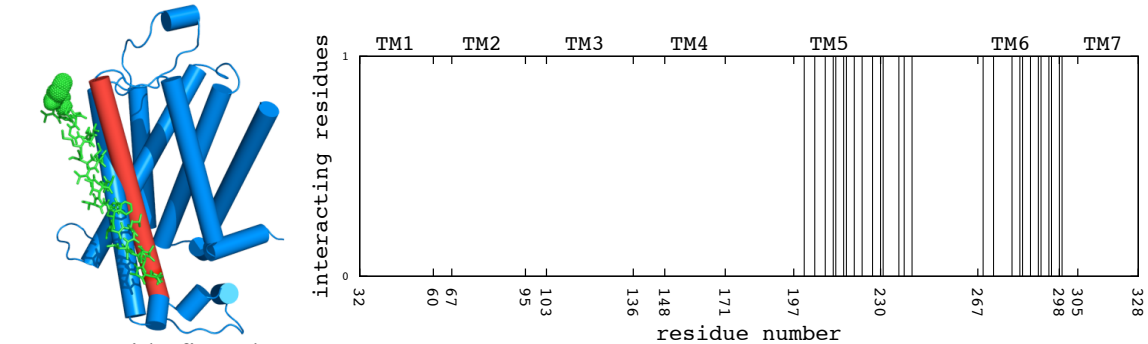
(121) Gardiner, E. J., Willett, P., and Artymiuk, P. J., "Protein docking using a genetic algorithm", *Proteins* **44***,* 44-56 (2001).

Appendix A : Sample Parameter File For AutoDock Runs

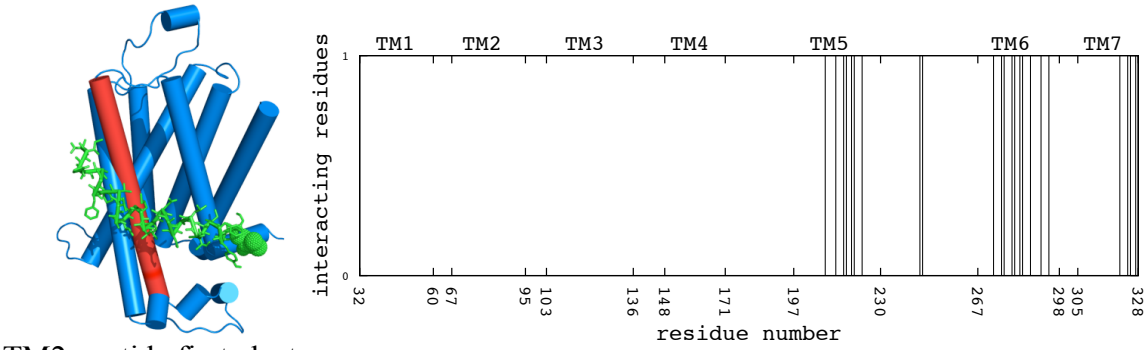
```
autodock_parameter_version 4.2     # used by autodock to validate parameter set
outlev 1                           # diagnostic output level
intelec                            # calculate internal electrostatics
seed pid time                      # seeds for random generator
ligand_types A C NA OA N SA HD     # atoms types in ligand
fld frame_03467.maps.fld           # grid_data_file
map frame_03467.A.map              # atom-specific affinity map
map frame_03467.C.map              # atom-specific affinity map
map frame_03467.NA.map             # atom-specific affinity map
map frame_03467.OA.map             # atom-specific affinity map
map frame_03467.N.map              # atom-specific affinity map
map frame_03467.SA.map             # atom-specific affinity map
map frame_03467.HD.map             # atom-specific affinity map
elecmap frame_03467.e.map          # electrostatics map
desolvmap frame_03467.d.map        # desolvation map
move ulti_peptide.pdbqt            # small molecule
about 10.9689 -6.4394 0.2885       # small molecule center
tran0 random                       # initial coordinates/A or random
axisangle0 random                  # initial orientation
dihe0 random                       # initial dihedrals (relative) or random
tstep 2.0                          # translation step/A
qstep 50.0                         # quaternion step/deg
dstep 50.0                         # torsion step/deg
torsdof 0                          # torsional degrees of freedom
rmstol 2.0                         # cluster_tolerance/A
extnrg 1000.0                      # external grid energy
e0max 0.0 10000                    # max initial energy; max number of retries
ga_pop_size 150                    # number of individuals in population
ga_num_evals 2500000               # maximum number of energy evaluations
ga_num_generations 27000           # maximum number of generations
ga_elitism 1                         # number of top individuals to survive to
next generation
ga_mutation_rate 0.02              # rate of gene mutation
ga_crossover_rate 0.8              # rate of crossover
ga_window_size 10                  #
ga_cauchy_alpha 0.0                # Alpha parameter of Cauchy distribution
ga_cauchy_beta 1.0                 # Beta parameter Cauchy distribution
set_ga                             # set the above parameters for GA or LGA
sw_max_its 300                     # iterations of Solis & Wets local search
sw_max_succ 4                      # consecutive successes before changing rho
```

```
sw_max_fail 4                    # consecutive failures before changing rho
sw_rho 1.0                       # size of local search space to sample
sw_lb_rho 0.01                   # lower bound on rho
ls_search_freq 0.06               # probability of performing local search on
individual
set_psw1                               # set the above pseudo-Solis & Wets
parameters
unbound_model bound              # state of unbound ligand
ga_run 200                       # do this many hybrid GA-LS runs
analysis                         # perform a ranked cluster analysis
```

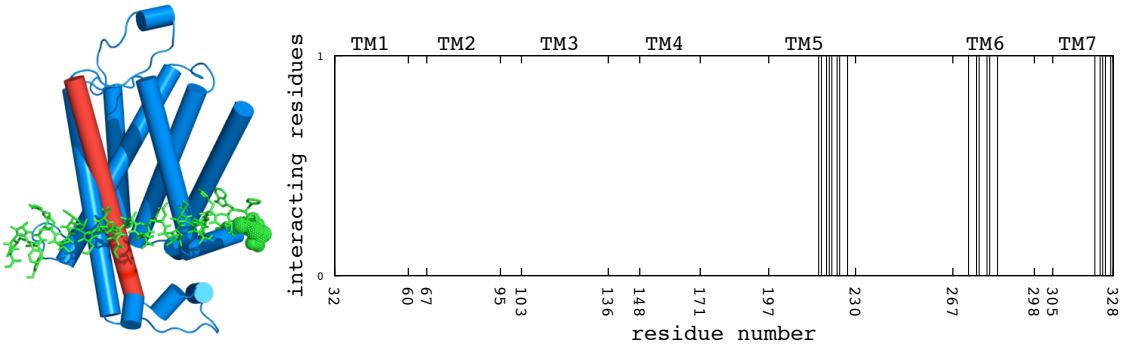
Appendix B: Long Peptide Docking Results



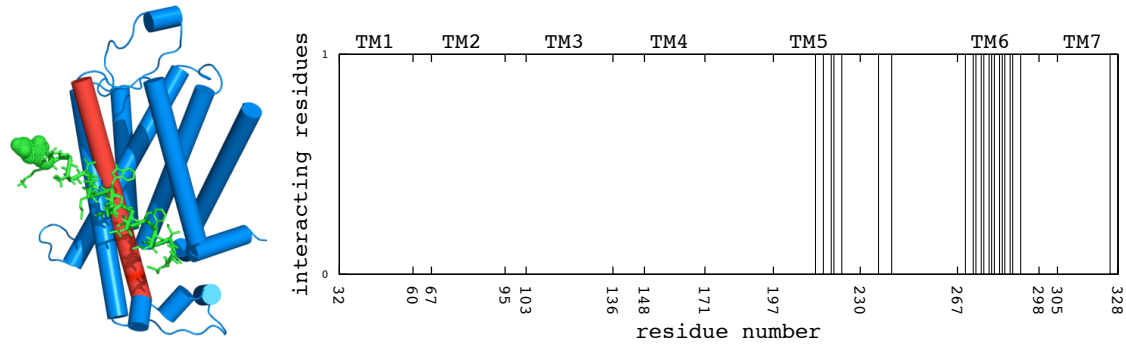
TM1 peptide first cluster



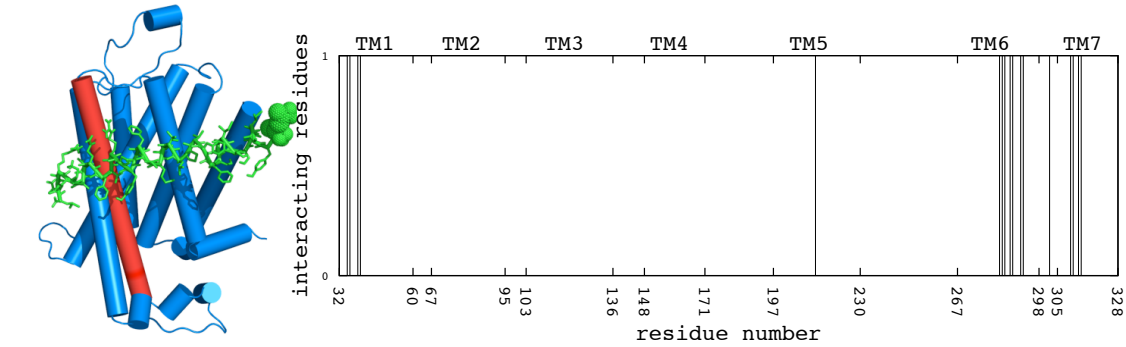
TM2 peptide first cluster



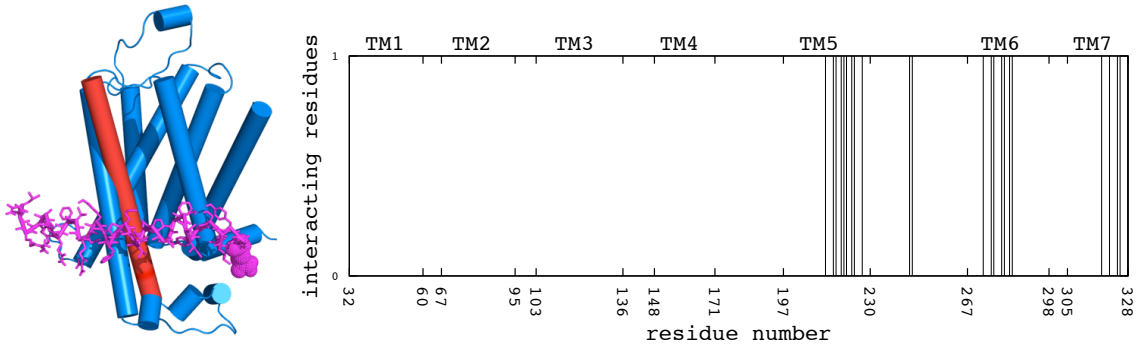
TM3 peptide first cluster



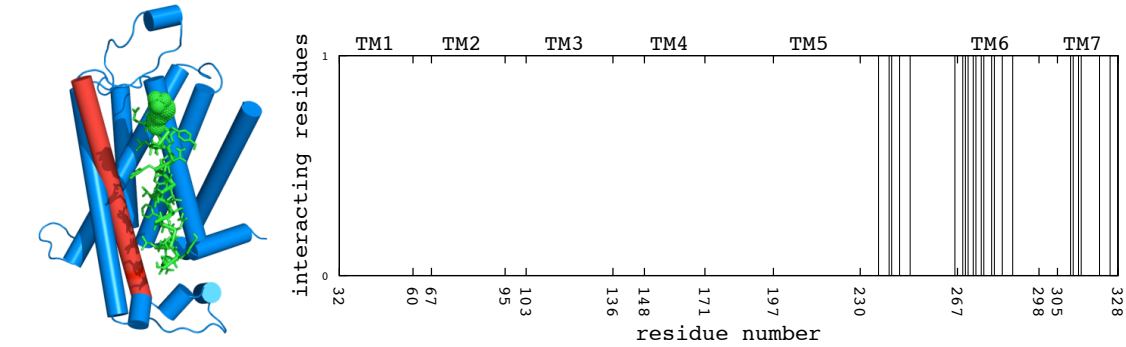
TM4 peptide first cluster



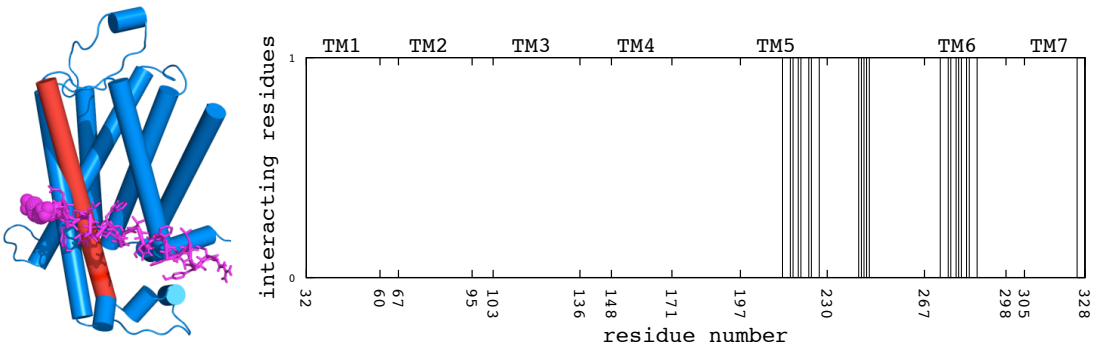
TM5 peptide first cluster



TM5 peptide highest populated cluster



TM7 peptide first cluster



TM7 highest populated cluster