Chemogenomic Approaches to Drug Design: Docking-based virtual screening of Nematode GPCRs for potential anthelmintic agents

Raban Wilfred Masuka

Department of Chemistry
University of Cape Town
Rondebosch 7701
South Africa

February 2016

Student: Raban W. Masuka

Signed by candidate

Co-supervisors: Kelly Chibale

Chemogenomic Approaches to Drug Design: Docking-based virtual screening of Nematode GPCRs for potential anthelmintic agents

Raban Wilfred Masuka

Thesis Presented for the degree of
DOCTOR OF PHILOSOPHY
in the Department of Chemistry
UNIVERSITY OF CAPE TOWN
South Africa

Supervisor:	Professor Graham E. Jackson
Co-Supervisor:	Professor Kelly Chibale

Department of Chemistry
University of Cape Town
Rondebosch 7701
South Africa

February 2016

Declaration

I declare that this thesis, "*Chemogenomic approaches to Drug Design: Docking-based virtual screening of nematode GPCRs for potential anthelmintic agents*," is my own work that was performed under the co-supervision of Professor G. E. Jackson and Professor Kelly Chibale. This work is proprietary and has not been submitted before for a degree or examination at this or any other university. In addition, all resources I have used have been acknowledged or referenced.

Raban Wilfred Masuka February 2016

Signature__ _____

Signed by candidate

Dedication

This work is dedicated to my awesome parents, Julius Tongai Masuka and Bertha Abigail Masuka and the Omnipotent God they introduced me to at a tender age.

Acknowledgements

The journey of a thousand miles is attained by slow degrees. Had it not been for the Lord God who was on my side, I would have been consumed during the course of my studies. Psalms 116:12 "*What shall I render unto Jehovah for all His benefits toward me*" (ASV).

The enthusiasm, support and guidance of my supervisors, Professor Graham Ellis Jackson and Professor Kelly Chibale will be esteemed forever. Prof. Jackson was patient with me as I struggled with NMR and molecular dynamic simulations. He always availed himself whenever I required his assistance. Prof Chibale was ever on hand to help me understand medicinal chemistry and the application of chemistry in drug discovery. Their leadership and insight during my studies was a pillar as I ventured into the uncharted waters of computational chemistry. Their commitment and serenity through the studies was a beacon of light for me throughout my studies.

Dr. Grace Mugumbate was always on hand to answer even the worst of questions. Her support and encouragement through the studies and as the storm raged calmed my nerves. I am grateful and thank you for believing in me and showing me direction. May God enlarge your territory and bless you indeed. I am also grateful to Dr. Gadzikano Munyuki for his assistance with GROMACS during the course of this study.

The Medicinal Chemistry research group members provided unwavering support and a homely environment during my time with there ensuring a stress-free stay. The CADD team particularly Dr. Elumalai Pavadai and Stephen Fienberg, were timeous in their assistance throughout my studies. The long deliberations and scientific arguments with them ensured that I think outside the box. The stay in the research would not have been as rosy had it not been for the commitment of Elaine Rutherford-Jones, Deirdre Brooks and Gerald Heselink. They helped me a lot with administration, software maintenance and upgrades on my computers. Their dedication to duty is out of this world.

Professor Timothy Geary, Victoria Muise and Yun Hui were very helpful with the *in vitro* assays. I am also grateful for the invaluable commitment of Vernita Reid who assisted in the interpretation of the biological data and arrangement of Chapter 5.

I am forever grateful to God for providing me with a supportive family and all-weather friends. Special thanks go to Bella Mavhuro, Oswell Munyoro, Silence Chigariro, Arnold Mudovozi, Yolandah Hall, Elton Mukonda, Lorraine Phiri, Takura Mafaifi, Edmore Munyoro, Kudzai Kagoro, and Clyde Makamure. I am thankful for their love, support and dedication throughout my studies. The Department of Chemistry at the University of Cape Town has been very supportive to me and I am very thankful to everyone.

Professor John Overington, Anne Hersey and the ChEMBL team offered me incredible encouragement during my internship there.

I am indebted to the Medical Research Council (MRC) and National Research Foundation (NRF) for funding my studies through the Professor Kelly Chibale. Had it not been for the financial support I would not have been where I am today.

Outputs from Study

1. Medicinal Chemistry and Computer Aided Drug Design, Omics Conferences, Atlanta, Georgia, USA, 2-4Nov 2015. Awarded Best poster prize
2. Pre-doctoral Fellowship at European Molecular Biology Laboratory-European Bioinformatics Institute, Cambridge, United Kingdom under the Computational Chemical Biology Group led by Dr. John Overington (October 2014- February 28, 2015)
3. Advances in Drug Design Conference, Marriot Hotel, London, UK, Feb 2015
4. Poster Presentation at the Wellcome Trust Resources for Computational Drug Discovery, November 17-21, 2014 at European Molecular Biology Laboratory-European Bioinformatics Institute, Cambridge, United Kingdom
5. Poster presentation at the Young Chemists' Symposium October 23, 2014, University Of Cape Town, Cape Town, South Africa.
6. Oral presentation at the Young Chemists' Symposium October 23, 2014, University Of Cape Town, Cape Town, South Africa.
7. Poster presentation at the 41th SACI National Convention, Dec 1-6, 2013, River Park Conference Centre, East London, South Africa.

Abstract

Chemogenomic Approaches To Drug Design: Docking-Based Virtual Screening Of Nematode G-Protein Coupled Receptors For Potential Anthelmintic Agents.

Raban Wilfred Masuka

Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa.
February 2016

Among common problems affecting human health and veterinary medicines, helmintic infections are major. The pathogens affect 550-750 Million people worldwide, and affect childhood growth, pregnancies, and development of the intellect. Helminths affects the well-being of animals as well including livestock and reduce the animal populations. However, the current anthelmintics are no longer as effective and some strains have developed resistance thus increasing the need for new anthelmintics. Unfortunately, not too much information is available detailing the physiology of helminths.

The published genomic sequence of nematode *Caenorrhabdtis elegans* as well the primary sequence of the FLP18R1 G-Protein Coupled Receptor are available. GPCRs play a significant role as targets for therapeutics and are responsible for signal transduction in cells. Thus, nematode GPCRs offer an alternative target to design new anthelmintics. Unfortunately, very little information exists about these targets and there are no known x-ray or NMR structures.

In this work, the 3D structure of nematode GPCR receptor (FLP18R1) was determined using homology modeling using the beta-2-adrenergic receptor as a template. The homology model developed had 24.87 % sequence identity with the template. Explicit membrane molecular dynamic simulations were used to optimize and refine the helices of the model over 100 ns. The homology model was of acceptable quality.

In addition, solution structures of four known agonist neuropeptides that bind to FLP18R1; *af3, af4, af20* and *flp18-6* were determined using NMR-restrained molecular dynamics in DPC using GROMACS. Cyclic conformations of the

neuropeptide structures were observed. Blind docking of the neuropeptides into FLP18R1 model was performed using AutoDock4.2 and Glide using SP-peptide docking. The docked complexes with Glide scores of -11.2, -10.9, -14.1 and -9.9 kcal/mol respectively were further optimized in explicit POPC membrane molecular dynamics simulations for 100 ns. The optimized complexes were used in docking-based virtual screening of the NCI database to identify potential FLP18R1 inhibitors. *In vitro* evaluation of the top 10% compounds confirmed three FLP18R1 inhibitors with IC_{50} below 100μM. Further titrations indicated that only NCI 327396 inhibits yeast growth at 10 μM. These compounds are suggested to provide insight into research on novel anthelmintics.

Table of Contents

List of Tables

Preface

Nematode (Helmintic) infections are major problems in veterinary medicine and human health. Nematode infections are part of the Neglected Tropical Diseases (NTDs).[1] The world has responded to the growing threat posed by neglected tropical diseases, which have destroyed nearly a billion of the world population, posing negative impact on health, education, and socio-economic development. Several resolutions by the World Health Assembly (WHA), seek to find ways of control and eradication of NTDs and for the mitigatory measures to ensure control and reduction programs.[2–4] The scientific community and health authorities in the public and private sector are increasingly working towards initiatives to manage and subsequently eradicate neglected tropical diseases.[2,5]

Hookworms are a class of nematodes that are mostly host specific, hematophagous, causing gastro-intestinal bleeding, and anaemia in most chronic infections.[6–10] Although helmintic diseases, that are caused by parasitic nematodes affect about 550-750 million people worldwide,[4,9,11,12] more work and research initiatives seeking to find and develop new anthelmintics is required. Parasitic nematodes have devastating effects on the growth of children, nutrition, healthy pregnancies as well as developing the intellect and memory functions in young children.[7,9,11,13,14] It is understood that the intermediation of the FMRF amide-Like Peptides (FLP) signalling pathway will lead to a reduction in nematode infections in humans. In some work reported by Kubiak (2008), candidate (G-Coupled Protein Receptors (GPCR) receptors were expressed matching specific FLP ligands in heterologous cells or *xenopus* oocytes.[6,7]

This thesis discusses the GPCRs taken from the *Caenorrhabdtis Elegans* (*C. Elegans*) genome and how they were utilized in a drug discovery program to identify potential anhelmintic agents. The structure of *(C. Elegans)* is shown in Figure 1.1.

Figure 1.1: The pictorial representation of *Caenorrhabdtis Elegans*[15] nematode

Elucidation of the three dimensional structure of GPCRs is critical to demystify the molecular mechanism underlying diseases and syndromes caused by mutations in GPCRs.[15–22] The drugs that activate or inhibit GPCRs, are understood do so by either initiating or blocking the receptor, which results in a transition from an inactive, non-signalling receptor conformation to an active state, which induces G-Protein mediated signalling. This conformational change ensures that signal is transduced via the GPCR into the cell, as a response to the interaction of drugs or endogenous ligands at the cells' exterior.[23–27]

The broad aim of this thesis is therefore the elucidation of the three-dimensional structure of FLP18R1 from the *C. elegans* genome using molecular dynamics refined homology modeling and Beta-2-Adrenergic receptor (2RH1) and Rhodopsin (1F88) as templates with a view to identifying the putative binding site. To mimic the receptor's true environment, extensive molecular dynamics simulations were performed in an explicit membrane to determine stability and conformational changes during molecular dynamics.

1 Chapter 1: Introduction

1.1 Parasitism

Parasites are some of the causative agents of disease in humans. A parasite is defined as an organism that spends all or in other instances, part of its growth and development cycle in or on another host to obtain its nutrition for development and/or reproduction at the expense of the host.[3,28–31] Therefore, parasitic diseases are infectious diseases that have parasites as causative or transmission agents. However, parasitic infections and related diseases have an adverse effect on all living organisms, including plants and animals.[12,14,29]

1.1.1 Types of parasites

Parasitic infections/diseases are caused by mainly three types of organisms, viz: protozoa, helminths, and ectoparasites.[13,27,32,33] Protozoa are parasites, which are single-celled organisms and live and multiply inside of the host, in this case, humans. Protozoa are known to cause some infections, which include giardiasis.[12,13,34] Helminths are parasitic worms that are multi-celled organisms that can live alone or in humans. Helminths are believed to be the most infectious agents in human beings in most developing countries. The life cycle of a hookworm, an example of a helminth, is shown in Figure 1.2. The diagrammatic representation also shows that hookworms as parasites grow in the host. Two types of helminths are commonly known to exist. These are the nematodes (round worms), which are primarily intestinal worms and filial worms which are responsible for causing lymphatic filariasis.[13,35]

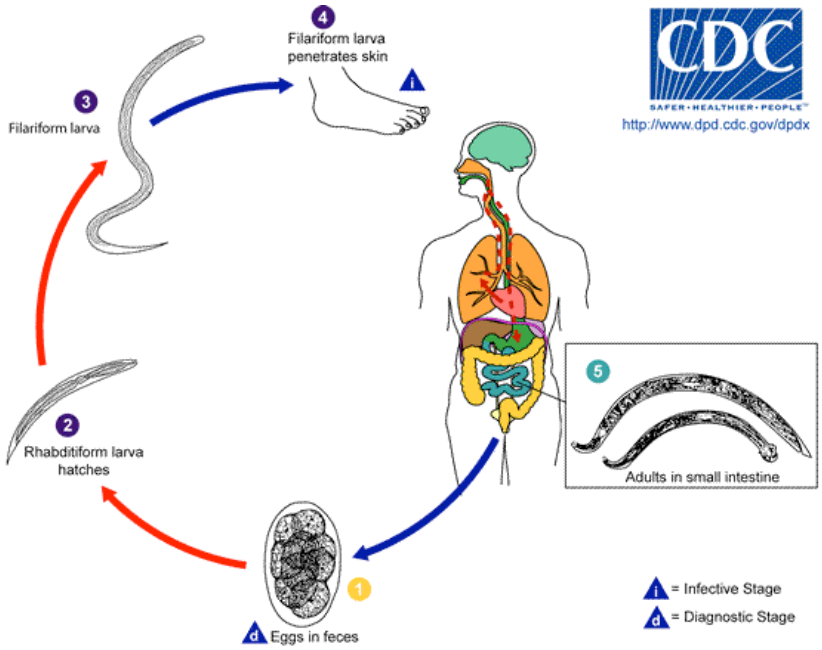


Figure 1.2: The life cycle of a hookworm. Much gratitude to the Centre for Disease Control for the pictorial representation.

Ectoparasites are multi-celled organisms, such as mosquitoes, bedbugs, ticks, which feed off the skin of humans and live in them.[37]

1.1.2 Global Geographical distribution of Nematode (Hookworm) infections

Global epidemiological estimates suggest that 3.5 billion people in the world are at risk of contracting neglected tropical diseases, and the reports in 2011 revealed that nearly 1 billion people were infected with one or several NTDs concomitantly, with helmintic diseases having the highest prevalence rates.[1,5,32,36,37]Various initiatives have been undertaken to compile the distribution of helmintic infections. The evaluation of the health of the population, relative to the importance of different diseases and epidemiology, and understanding the epidemiology across countries and trends over time are of paramount importance in order provide a framework for determining the cost-effectiveness of the relevant intervention programs.[5,30,34,36–38]
The risk of parasitic infections in humans is now generally limited to distinct geographical areas because of the specificity of host-parasite relationships, climatic factors and poverty.[13,39]

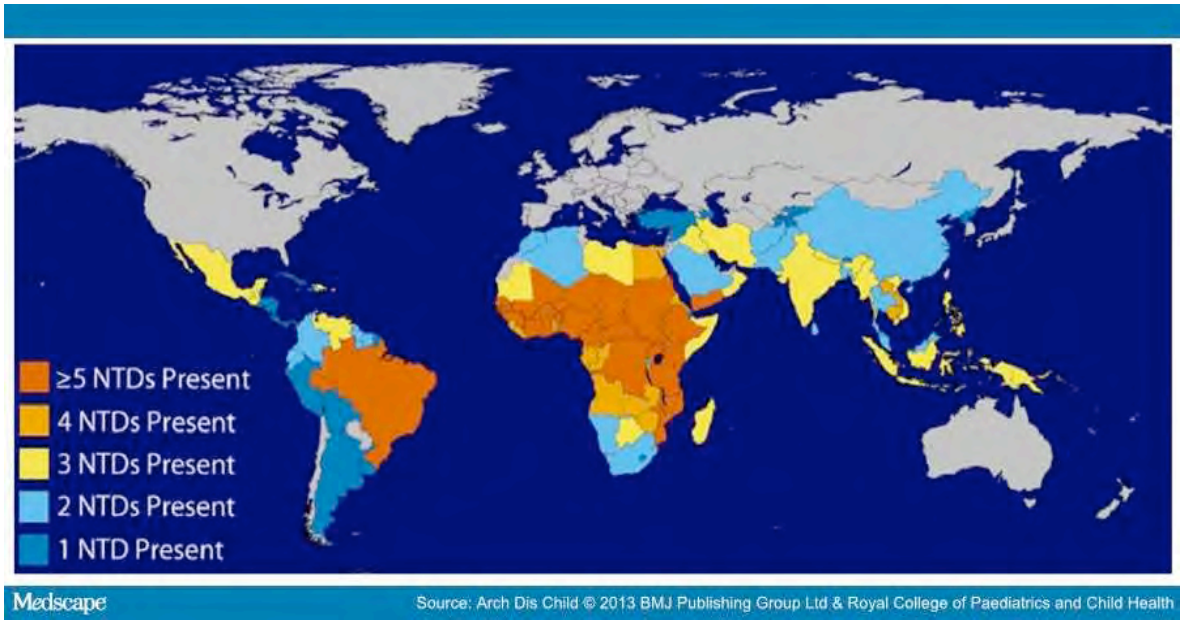


Figure 1.3: The global distribution of nematode infections and the map also shows the concurrence of nematode diseases across the world.[36]

Figure 1.3 shows that parasitic infections are more prevalent in rural or developing regions of Africa, Latin America and parts of Asia and less common in developed areas. Thus the most affected areas are the tropical and subtropical regions of the world. However, these infections are also present in the United States albeit to a limited degree.[3,32,40,41]

1.1.3 *Diseases caused by helminthes and treatment*

Helminths are causative agents of most human helmintic diseases/ infections, some of which involve the musculoskeletal system. These agents may be classified into nematodes or roundworms, trematodes or flatworms, and cestodes or tapeworms[5,34,42]. The diseases caused by helminths include among others, elephantiasis (lymphoedema), river blindness, African eye worm disease, dracunculiasis, filariasis, pinworm and chinese liver fluke[28,31,36–38,40,43]. The treatment of these diseases is mostly through drugs such as albendazole, mebendazole, pyrantel pamoate, praziquantel, azithromycin, flubendazole, levamisole, diethylcarbamazide among others[39,44]. Some of these compounds are shown in Figure 1.4.

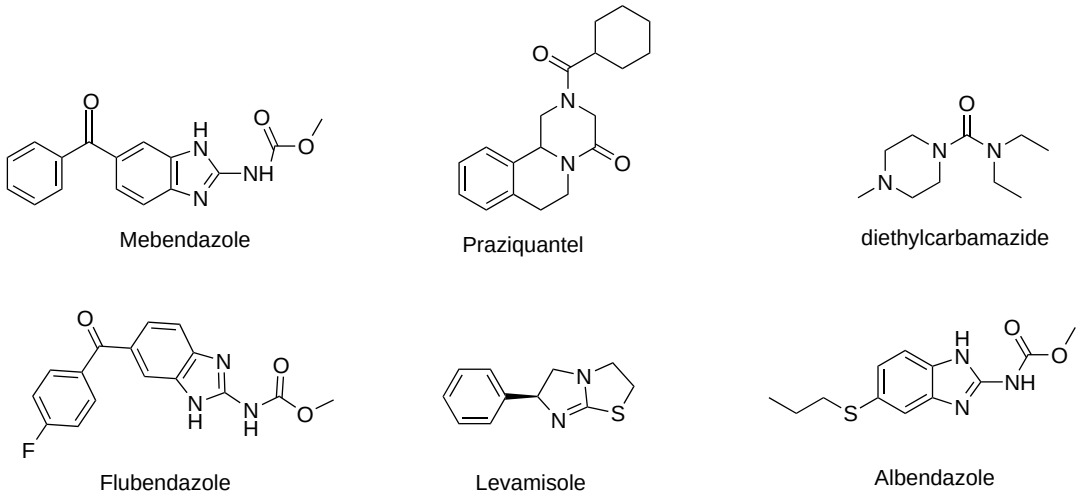


Figure 1.4: Structures of some of the current anthelmintic drug compounds

Current anthelmintics are known to selectively affect the neurotransmitter receptors in the nematodes. Anthelmintics act on diverse species across considerable phylogenetic space of the phylum.[31,38,45,46] However, there has also been relatively poor efficacy of these available anthelmintics against some human nematode pathogens. Thus, the incidence of these diseases continue to grow as new resistant strains are identified, which cannot be treated using current drugs. This means new initiatives are required to develop new drugs to counter the effect of these new strains. The resistance[4,12,34] of the current anthelmintics has been attributed to the following:

- changes in drug translocation
- receptor modification or post receptor modification
- changes in receptor numbers

Like most drugs on the market, the serious occupational exposure and environmental impacts of some anthelmintics cannot go unnoticed. Neuropeptides (agonists) are susceptible to degradation and hence they are protected by modifying them using amidation at either the N and or C terminal and in most instances this modification confers biological activity to the peptides including in *C. elegans.*[7,27] FaRPs share common residues at the C terminus which are found in C elegans which is Arg-Phe-NH_2[6,26]. In work reported by Kubiak (2008), candidate GPCR receptors were expressed matching specific FLP ligands in heterologous cells or xenopus oocytes.[7]

1.2 G-Protein Coupled Receptors (GPCRs)

GPCRs are drug targets of the integral proteins in the human body and are believed to be the major therapeutic target for over 50% of current marketed drugs.[47–49] GPCRs work as minute machines in nanoscale terms to help bind extracellular ligands initiating and mediating the signal transduction and information in the cell.[23,46,50–52] These ligands include neurotransmitters, proteins, photons, amines and hormones among others.[53–55] Ligand binding results in conformational changes in the GPCR as the signal is transduced into the intracellular trimeric guanine nucleotide binding regulatory proteins (G-proteins).[14,23–25,52,56]

Thus, GPCRs are actively involved in cellular communication and hence all the mediation of senses such as smell, taste, vision, secretion, pain, metabolism, and neurotransmission.[18] As such, malfunction of GPCRs is involved in many diseases. These include HIV, ulcers, cardiovascular diseases, nocturnal heartburn, migraine headaches and anxiety disorders[18,49,57,58].

Despite this importance and relevance to body function, very few 3-dimensional structures of GPCRs are known to date. There are various reasons for this, chief among which is the difficulty in crystallizing and handling GPCRs experimentally. It may also be because these proteins require a membrane mimetic[59–62] environment to maintain the protein folding and thus retain the shape. The last decade has seen massive growth and development of computational technology and its increasing use in drug discovery initiatives has also aided the elucidation of three dimensional[20,63,64] structures of GPCRs.

Literature has shown that there are various ligands that bind to GPCRs as agonists or antagonists.[15,65–68] There are also others that are allosteric modulators[23,52,69,70] for GPCRs. The compounds targeting GPCRs are not strictly specific. GPCR ligands target other proteins rather than being limited to their own family GPCR and this results in the description of the GPCRs as promiscuous targets.[20,22,71] This is an important attribute as these multi-targeted compounds through polypharmacology[71–73]

help in the effectiveness of treatment of diseases. The GPCRs being an important target also benefit significantly from the polypharmacology.

1.2.1 Structure of GPCRs

GPCRs are membrane proteins arranged as seven transmembrane proteins (7TM) and helical regions in an anticlockwise arrangement when viewed from the extracellular N-terminus.[20,51,74–76] The helical regions are linked one to another by three intracellular loops in the C-terminal domain and three extracellular loops in the N-terminal domain.[75] The loops vary in the number of residues from one GPCR to the next. The selectivity of binding of the ligands on the extracellular side is affected by conserved residues whereas the conserved residues in the intracellular (cytoplasmic side) has an effect on G-protein coupling selectivity.[20,66,77]

Research has revealed that there is small correlation between the size of the extracellular ligand and the length of the N-terminus of the GPCR. For instance, class A GPCRs predominantly have short N-terminus and possess highly conserved helical regions.

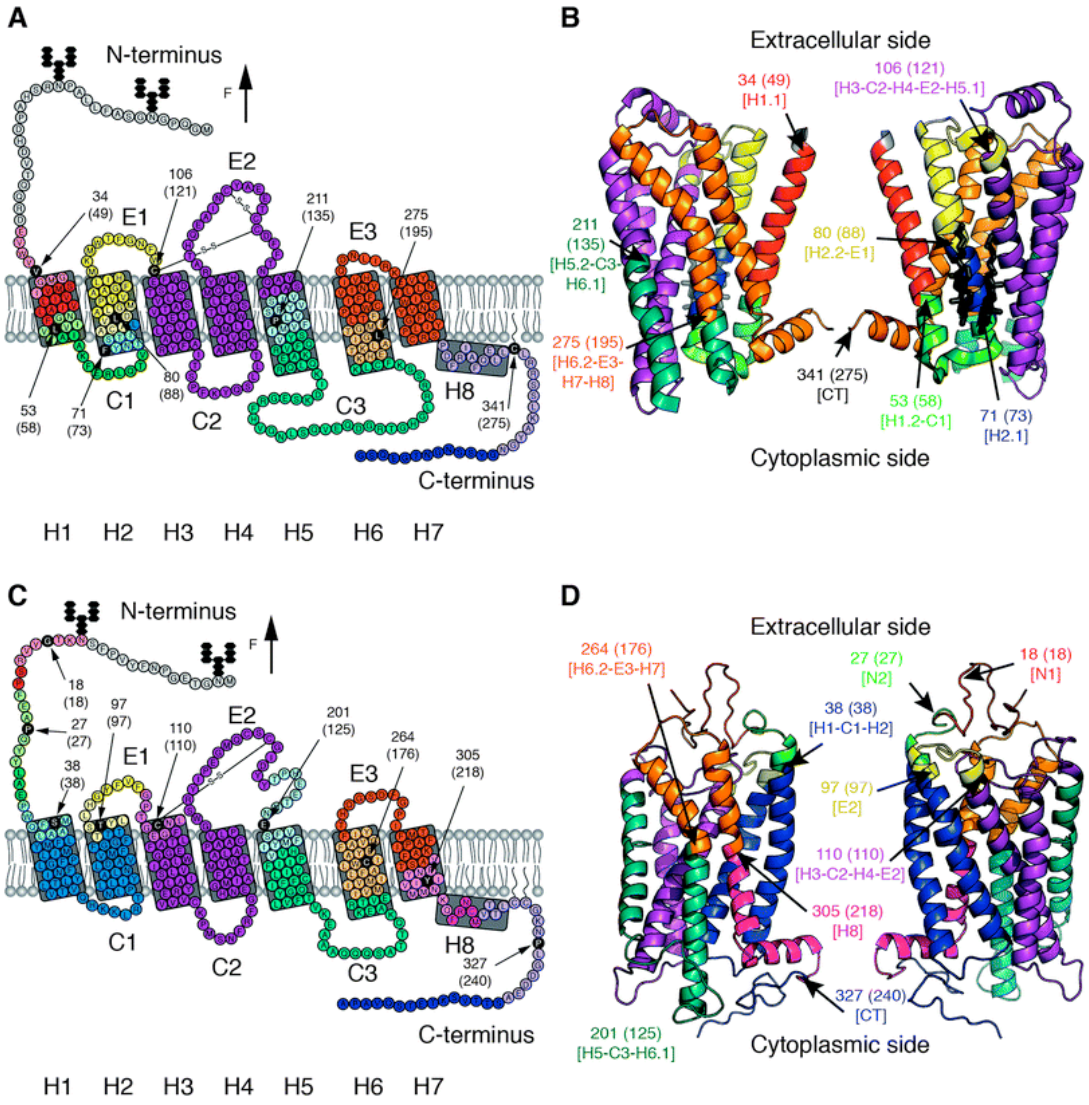


Figure 1.5: (A, C) Amino acid sequence of the human beta 2 adrenergic receptor (PDB code 3D4S) and bovine rhodopsin (PDB ID code 1U19). Secondary and (B, D) tertiary structure models of the human β_2AR and of the bovine rhodopsin, respectively. Each structural segment denoted by a distinct colour, is stabilized by intermolecular and intramolecular interactions. (A, C) Black amino acid residues (aa) signify the end of the stable structural segment and the beginning of the next stable structural segment in the sequence. The GPCR α-helical regions of both GPCRs are labeled H1–H7. The cytoplasmic (intracellular loops) and extracellular loops are coded as C1, C2, C3 and E1, E2, E3, respectively. Most class A GPCRs have a small helix usually indicated as H8 on the C-terminal.[78]

The N-terminus houses the putative binding site for most large polypeptides and glycoprotein hormones.[19,79,80] However, some family A GPCRs with a long N-terminal domain have an affinity for glycoprotein hormones.[56,59,68]

Protein or peptide ligands are believed to bind to the extracellular loops, particularly EC2 or generally the N-terminus of the GPCRs.[19,70,81–84] Currently, about 109 GPCRs have been drugged,[21] most of which have ligands binding in the putative binding site of the transmembrane region. Functional selectivity of extracellular ligands is of paramount importance in elucidating the signal transduction pathway[24] for both the therapeutic actions and the side effects of drugs.[21,73]

Docking and binding affinity studies have postulated that similar sequence motifs could recognize chemical features of similar ligands in the same biological space.[20,51,85,86] As such, the physicochemical properties of residues in the putative binding site can be compared. Thus the conservation of residues in the transmembrane regions explain this noticeable trend in class A GPCRs. Extracellular loop 2 (EC2) is important in the stability of GPCRs.[56,68,82,87] This is because in most class A GPCRs, there is a conserved cysteine residue,[81,88,89] which binds to the CWxP motif in TM6. This bond is an ionic lock,[55] which basically forms a “lid” over the active binding site. This ionic lock tends to keep helices 3 and 6 together such that they do not move apart and thus maintain the protein structure.[55,82]

The best example is the subfamily of opioid and chemokine receptors that share a similar β-hairpin conformation[90] in the second extracellular loop, which make it a common peptidic-binding motif in many GPCRs. In spite of the five classes of GPCRs sharing 7TM topology in humans, sequence identity across the five families is low (below 25%) and there is variation in the N terminal domain.[22,90] However, within subfamilies, the sequence structural similarity is better between sub types.[51,91] Thus templates can be used within sub types that permit predictions by comparative modeling, which is accurate enough for some applications.

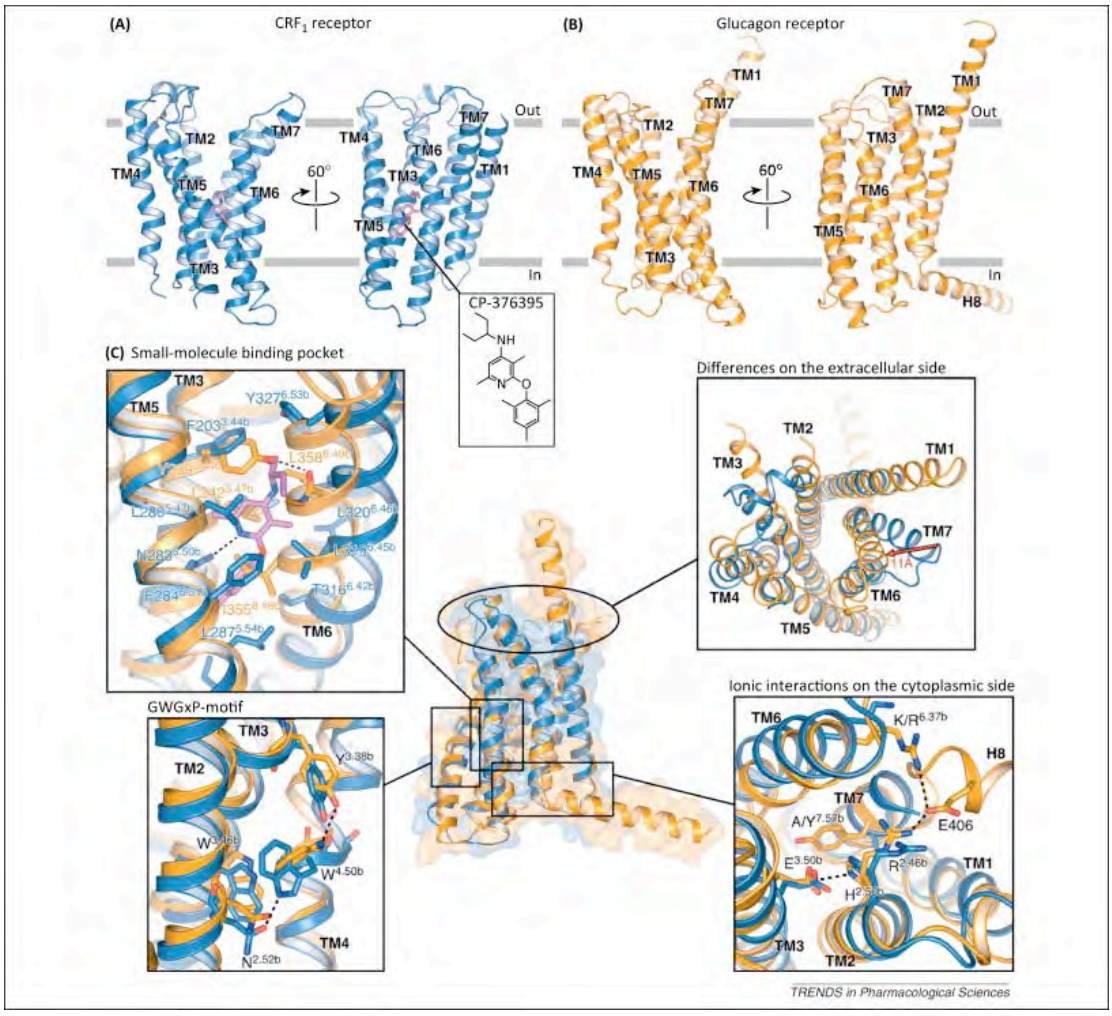


Figure 1.6: Example (A, B) Typical crystal structures of class B GPCR corticotropin-releasing factor receptor 1 (CRF1) (PDB ID code 4K5Y) and glucagon receptor (PDB ID code 4L6R) are displayed in blue and orange ribbons, respectively, in two conformations within the membrane. The disulfide bond between extracellular loop 2 (ECL2) and TM3 are indicated as purple sticks. (C) Alignment of the CRF1 and glucagon receptor, with insets that show the regions of interest. To highlight the structural differences in the extracellular halves of CRF_1 and glucagon receptor, the distance of almost 11Å between the Ca-atoms of residues 7.33b at the N-terminal end of TM7 is indicated with a red arrow. Superposition of the two receptors around CP-376395, highlight the binding site and show the antagonist binding configuration and the significant structural transformations observed for TM6 of the two receptors[92].

1.2.2 Classification of GPCRs

GPCRs are classified into 5 major families[15,92–94] based on sequence similarity and functionality. These are:

- Class A (Rhodopsin family),
- Class B (Secretin receptors)

- Class C (Metabotropic glutamate/Pheromone)
- Class D (Frizzled)
- Class E (Taste-like)

Of these, class A has the most number of known structures. The first class A GPCR to have its structure determined was rhodopsin in 2000.[18,95] The successive class A GPCRs were built based on the rhodopsin structure as the template. In 2007, Beta-2 Adrenergic receptor[55,93,96]was elucidated and subsequent GPCRs were determined. As of June 2015, the ChEMBL database has recorded 857 GPCRs, of which only 30 crystal structures are known. According to ChEMBL_20 version, there are 688 Class A; 44 Class B; 35 Class C; 4 Class D and only 1 class E GPCR.[97–101]

1.2.3 *GPCR Ligands*

GPCRs ligands are classified according to their functionality viz. stimulating (agonism), block (antagonism), or reduce (inverse agonism) the mediation of signal transduction in the cell.[15,52,65,67,68,102] Most are peptides and small molecules. Small molecules are subdivided into further classes such as monoamines and its derivatives, lipid-like, nucleotide-like, small organic, and photon sensing molecules.[53,98–101] Peptides are subdivided into short peptide, chemokine, opioid, dopamine, protease activated and glycohormone among others.[97,99,103–105] G-Coupled Protein Receptors possess a ligand-binding sub-pocket. This can be detected for proteins with absolutely different folds and catalytic activities. Compounds targeting GPCRs are not strictly specific.[71,73] They target other proteins and this results in the description of the GPCRs as promiscuous targets.[71,106]

When the ligand binds to the GPCR, the chemical conformation of this receptor protein is altered and this conformational state determines the functional state of the protein.[19,107] Ligand binding results in conformational changes in the GPCR as the signal is transduced into the intracellular trimeric guanine nucleotide binding regulatory proteins (G-proteins).[23–25] Site-directed mutagenesis studies,[51,108,109] molecular modeling and structural data suggest that most class A GPCRs share a putative binding pocket for small molecules at a similar spatial position comparable to retinol in rhodopsin and beta-2 Adrenergic bound to carazolol.[60,102]

Due to sequence homology among GPCRs, it is widely believed that similar sequence motifs recognize similar ligands or fragments at a common binding site. Recent advances in chemogenomics have provided novel insights into predictions of protein-ligand interactions as well as molecular recognition through analysis of bioactivity databases[97,110] Knowledge of common structures of ligands[89,111] can enhance understanding of protein-ligand binding and interactions. Even though the different GPCR families have low sequence homology[112], functional selectivity of their ligands is seen.

A study by Erguner in 2010[113] highlighted that about 88% of ligands possess common substructures and that these are shared among various GPCR families. These substructures are commonly called privileged motifs (Figure 1.7).[89,111] This follows the general understanding that molecules, which have pharmacological behaviour that is similar, tend to have binding affinity towards same family of GPCRs. These include among others the benzodiazopinones, arylpiperazines and certain biphenyl motifs.[89,111,113] These are basically known to target family A GPCRs and are commonly used as starting structures when one designs ligands towards this family of GPCRs.

However, having a privileged motif does not necessarily translate into a desired pharmacological[113] activity and the presence of privileged motifs does not unambiguously define a target receptor. The spiropiperidine moiety is an example of a target-family preferring motif even though it is not restricted to any particular class.[111,114] This is common in most inhibitors for GPCRs and also found in ligands of biogenic amine receptors and is also common within ligands of chemokine and peptide-binding GPCRs. However, other compounds also have cross family targets. This is an important attribute as these multi-targeted compounds through polypharmacology facilitate the effectiveness of treatment of diseases.

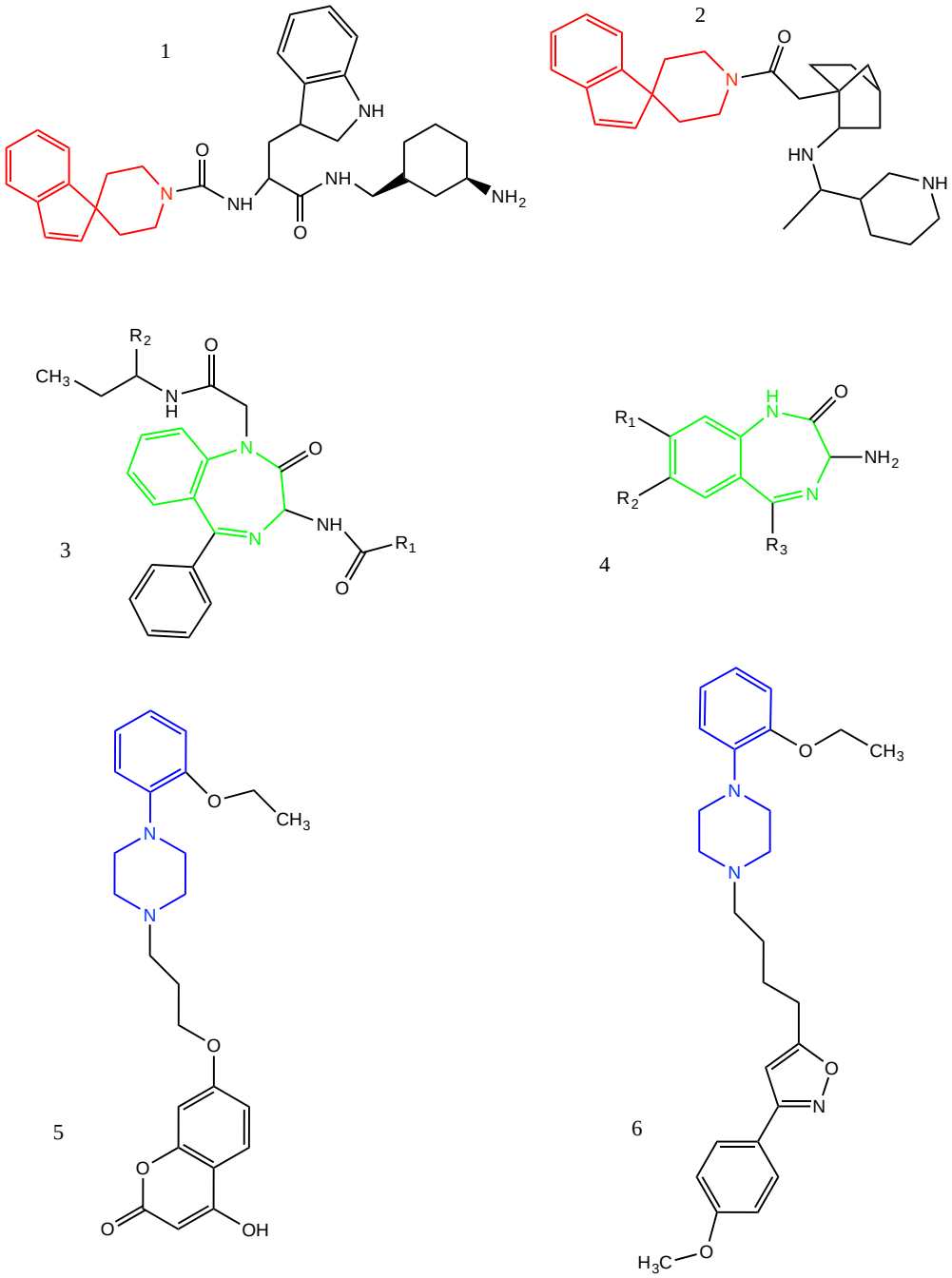


Figure 1.7: Privileged motifs in GPCR ligand shown in different colours. The red spiropiperidine motif in 1 and 2, green motif in 3 and 4 show benzodiazopine motifs, and blue Arylpiperazine motif in 5 and 6 are common in GPCRs.[113]

Figure 1.7 highlights some of the GPCR privileged substructures: biphenyl, spiropiperidines (1, 2), 1, 1-diphenylmethane, 4-aryl-piperidines, 4-arylpiperazines (3,4). Benzodiazepine substructures are commonly found in receptors of the central nervous system as well as in ion channels.[89,113]

1.2.4 GPCRs in Caenorhabditis elegans genome and state of the art of FLPs

The availability of sequence information following recent advances in genomic sequencing of invertebrates such as *Caenorrhabdtis elegans*[6,115–117] commonly known as *C. elegans*, has vastly improved drug discovery initiatives in the GPCR field of research. Despite its small genome, *C. elegans* is about 3% (about 97Mb in size) the size of the human genome and other mammalian genomes. It contains nearly 20,000 genes.[6,118–120] *C. elegans* is a heterotrophic eukaryotic organism (metazoan) and as such its genome was thought to be significantly more complex than those of single-celled eukaryotes such as yeast and protozoans.[116,118] However, its genome in reality about three times the size of the human malaria parasite *Plasmodium falciparum* (*Pf*) genome and only about eight times that of fission yeast.[118]

It is against this background that this metazoan genome of *C. elegans* has instructions required for the development and functionality of a fully differentiated animal.[7,27,118] Advances in genomic sequencing of *C. elegans* have resulted in research showing that the *C. elegans* genome encodes nearly 1100 GPCR genes.[118,121] Most of these genes (~1000) encode putative neural receptors that respond to changes in chemical stimuli, commonly called chemoreceptors.[118] To date, 113 neuropeptide genes encoding over 250 distinct neuropeptides have been identified.[6,7,14,38] More than 70 FLP genes are understood to have been predicted to target GPCRs.[6] Although the majority of FLPs and NLPs are likely to signal through GPCRs and the insulin-like ligands to signal through receptor tyrosine kinases, there is still more work to be done in identifying the signaling pathways.[6] Given limitations in this regard to date, neuropeptides are envisioned to be involved in all behaviors in *C. elegans.*[7,14,27]

Several other research groups have also reported that FLP18 receptor activity is regulated by matching FLP18 peptides.[14,116,117] Physiological experiments concluded that FMRFamide-like Peptides (FaRPs) have sufficient potency with an overall effect on neuromuscular systems of nematodes.[7,14,26,27,79,117] Nematodes have FMRFamide related peptides, which play a central role in nematode motor and sensory capabilities thus making the FLP signaling an appealing target for novel parasiticides.[6,7,14,79]

Studies reveal that classical neurotransmitter gated ion channels in the helminth nerve or muscle are targets for current anthelmintics[5,12,13,44] which lead to the paralysis or death of the nematodes. Avermectins are a family of compounds, which are well known anthelmintics that work as glutamate agonists by enhancing the opening of glutamate gated chloride channels, thus paralysing pharyngeal pumping in the host.[26,44,76] The classical example is ivermectin, which binds to the glutamate receptor present in the membrane chloride channel. Other anthelmintics work by affecting the nervous system of the parasites in order to counter the infections caused by the nematodes.[7,13,28,38,122] Still, others affect glucose uptake[26] of the parasites and in so doing affect the supply of energy and paralysing the parasite. The essential attributes of a potential drug target include:

- Importance of target in the functionality of the parasite[17–19,21]
- exclusive to the parasite and matchless in the host or variable to host homologue so that there is adequate pharmacological distinction.[7,38,71,83]

Most work done on the *C. elegans* genome using neuropeptide precursor algorithms has shown that 23 FLP genes have been identified.[6,7,14,27,117] However, RNA_i experiments[6,7,46,123] to date have been unsuccessful in the quest to identify phenotypes of most peptide GPCRs in *C. elegans.* Consideration has been made on the variation in the number of GPCRs in other species such as *Drosophila melanogaster*, humans, mouse and *Anopheles gambiae*.[46,124] It was found that the *C. elegans* genome encodes more GPCRs than other organisms. Studies have shown that organisms like the former encode 200, 750, 1000, and 276 GPCR genes respectively.[125] There are two large families of neuropeptides namely the insulin-like peptides and the FMRFamide (Phe-Met-Arg-Phe-NH_2)-related neuropeptides also commonly called FaRPs, which are referred to as FLPs in *C. elegans.*[6,7,14]

It is currently thought that most anthelmintics affect the functioning of nematode chemoreceptors selectively, and act on many species of considerable phylogenetic space within the phylum.[1,9,118] The main role played by FMRFamide-Like Peptides (FLPs) in motor and sensory capabilities of nematodes make FLP signaling a potential target for new parasiticides.[6,14] It is believed that intervention in the

signaling pathway would result in the reduction of the currently high nematode infections in humans.

Despite some available drugs for treating illnesses associated with nematodes, treatment is hampered by the increasing and widespread occurrence of nematode strains that have been selected for anthelmintic drug resistance.[42,44] Therefore, there is still need for identification of novel drugs that are not prone to resistance

1.3 Rationale of study: Drug Design

New disease indications have been identified over time. With the advent of new diseases, it is also essential to identify new modern and technological advances in drug research to counter these new diseases. Drug discovery processes involve the design and include target identification, lead identification and subsequent optimization as well as the introduction of new drugs to the public.[126–128] A number of approaches have been used in the process to discover new drug entities. Developments in the pharmaceutical sector have seen the implementation of new approaches to drug design. Two major approaches are Structure Based Drug Design (SBDD) and Ligand-Based Drug Design (LBDD)[56,127,129–133] as shown in Figure 1.8. Pharmacophore and structure based methods are believed to be the answer to current issues such as conformational flexibility of various molecules in Quantitative Structure Activity Relationship (QSAR) models.[130,134–138]

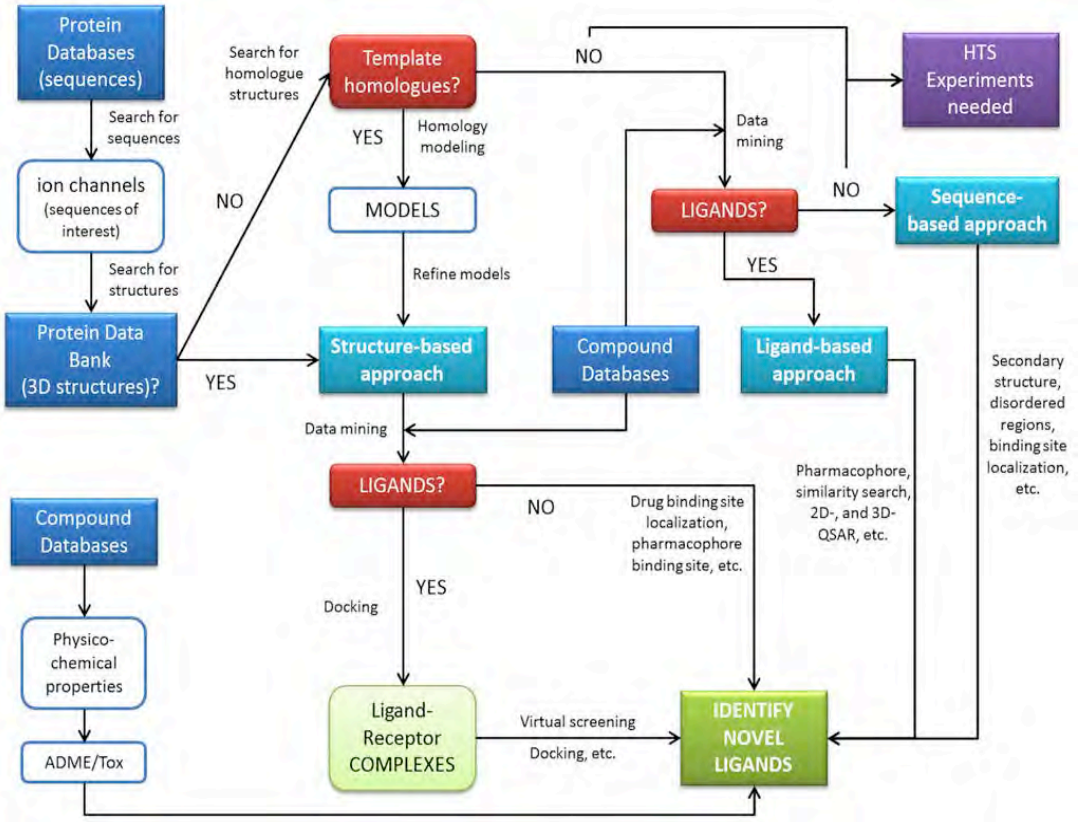


Figure 1.8: The flow chart in Computer Aided Drug Design and decision making on selection of approach to be followed[139]

These approaches are explained in detail in the next section and they are determined by the amount of information available in any new drug discovery initiative. Structure based and ligand based drug design approaches exploit scoring functions,[132,140–142] which are mathematical algorithms to evaluate binding affinities and poses of the protein-ligand complexes after docking. These algorithms are used to rank the results based on these binding scores. Scoring functions are usually challenging and it is very difficult to identify the right binding pose, as some parameters such as molecular interactions of some molecular species are hard to calculate and challenging to rank.[110,143] This ranking can be used to discriminate ligands by their docking scores[129,140–142] and evaluate energetically feasible poses.

1.4 Structure-Based Drug Design (SBDD)

The design and subsequent optimization of a protein target, as a clinical drug target is what is known as Structure-Based Drug Design.[82,129,132,134,144] Identification of drug

targets is an important step in any drug discovery initiative.[111,145] Different approaches are employed in elucidating the structures of targets. These can be crystal structures or homology models. The three-dimensional structure of targets can be deduced by various approaches. When the structures are known, X-ray crystallography and Nuclear Magnetic Resonance techniques are employed.[48,50,56,62,90,120,146] However, in some cases, the structures are not known and very limited information about the protein is available. Depending on the available information, homology modeling and threading approaches are used.

Sufficient information about the drug's three-dimensional structure and how its shape, conformation and charge cause it to interact and binding with its biological target that triggers a therapeutic effect.[20,64,147,148] Proteins are major drug targets and to some extent nucleic acids are also targets. Small molecules with potentially high biological activity could be accurately and reliably screened using computational algorithms and virtual libraries.[49,82,98,114,149–152] If the compounds are found to be experimentally active, they are either purchased or synthesized. Advances in drug design have seen the development of fast and reliable structure based virtual screening methods, such as docking, to obtain hits from large molecular libraries.[153–155] Searching algorithms are used to generate different docking conformations of ligands into protein targets.[80,83,105,107,140,156]

1.4.1 *Homology Modeling*

Using Computer-Aided Drug Design (CADD) approaches, primary sequences of amino acids can be used to build homology models. Primary sequences can be obtained from Genome databases such as GenBank.[14] Homology models are constructed based on careful selection of the template. The protein to be built in this regard is the target.[129] As such, secondary structure prediction tools are explored such as Psi-blast Secondary Structure Prediction (PSIPRED).[64]

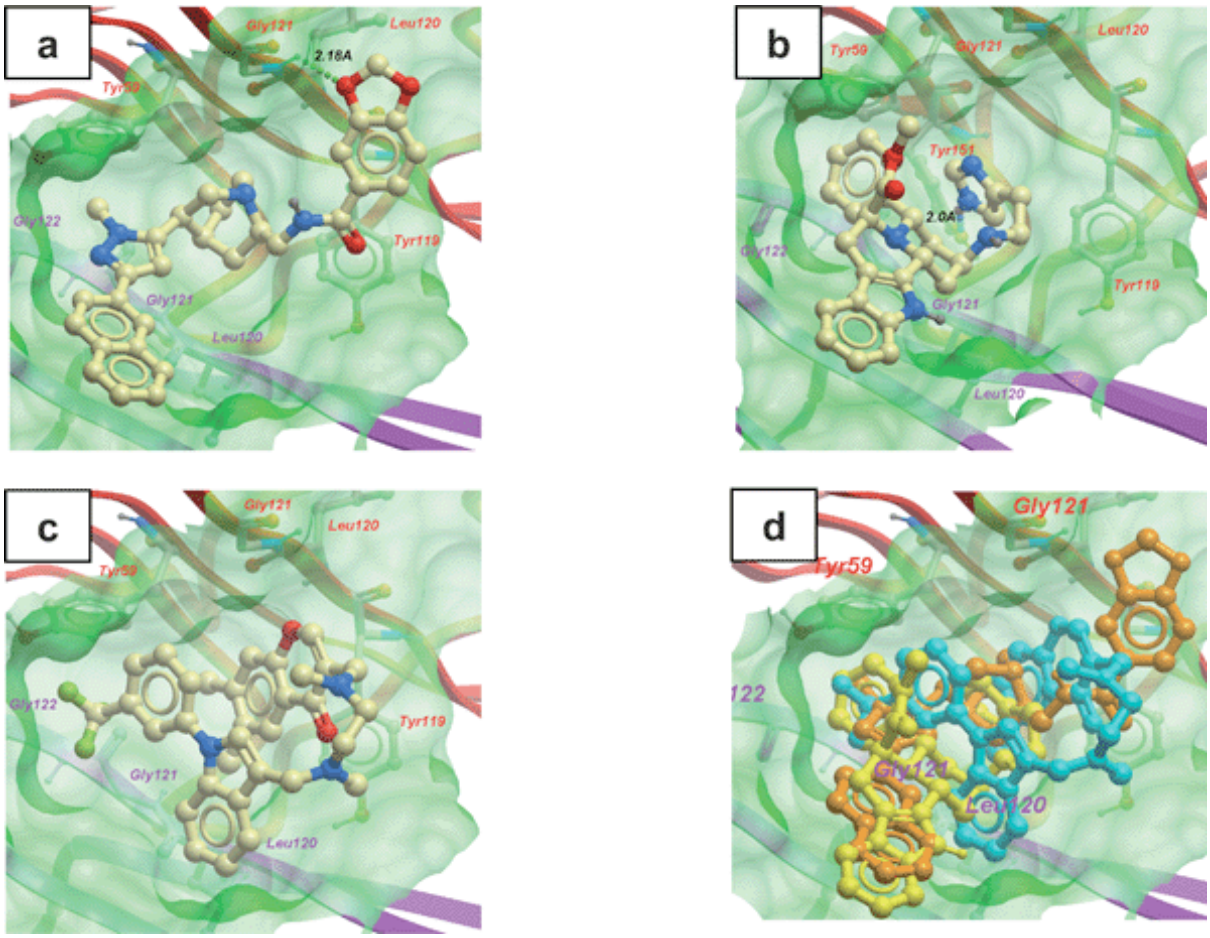


Figure 1.9: Low-energy binding poses of a) screening hit quinuclidine, b) screening hit indolo [2,3-*a*] quinolizidine, c) SPD304 bound to TNF-α dimer and d) superimposition of quinuclidine (orange), SPD304 (blue), and indolo [2,3-*a*] quinolizidine (yellow) and generated by virtual ligand docking[157]

The accuracy of the homology model is dependent on the similarity of the target protein to the chosen template. The first stage in homology modeling is aligning the primary sequence in a protein database such as the Protein Data Bank (PDB), where identical and similar proteins are identified.[20,64,82,158–161] These are ranked according to the identity score as a percentage and also give information such as sequence coverage and the number of amino acid residues in the sequences identified. The top identified proteins giving sufficient coverage and Expected value (E-value) are used to choose the protein sequences that are used in protein sequence profiling,[125,162] which is basically identifying the conserved residues in the helical regions.

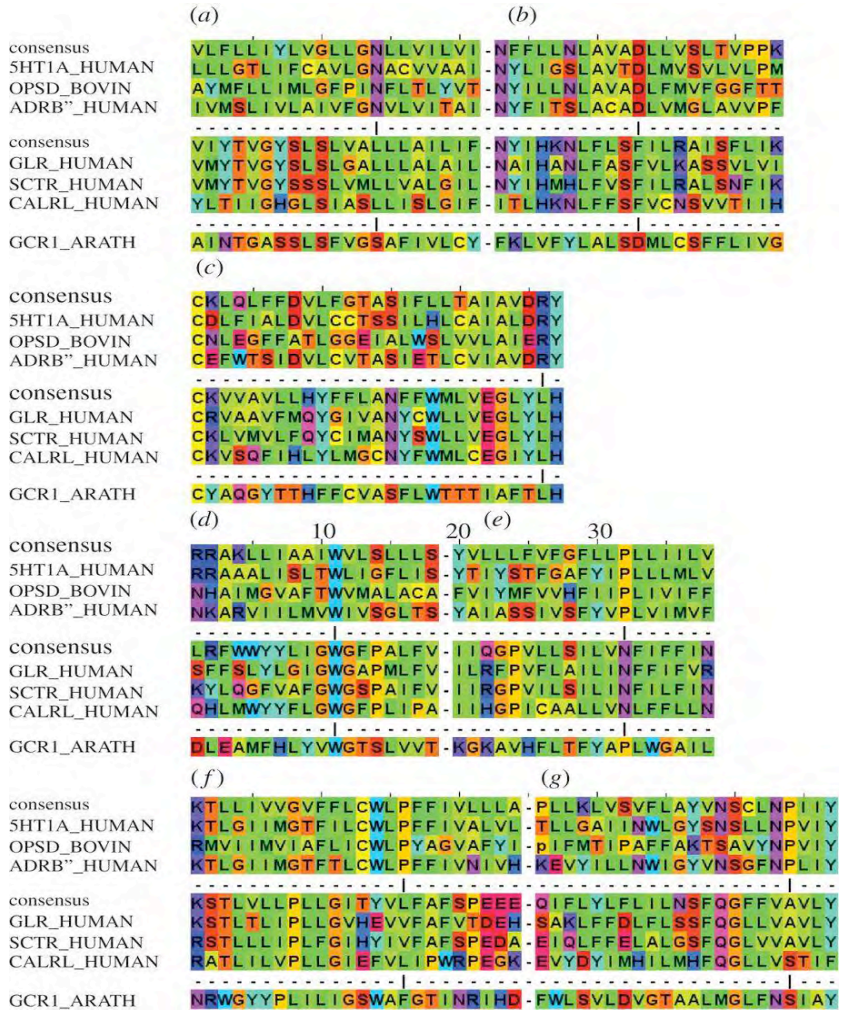


Figure 1.10: Multiple sequence alignment of Class A GPCRs used as templates in homology modeling showing the conserved residues in each helix indicates as (*a*-*h*).[51]

Expected values above one mean that the proteins cannot be used as templates. Ideal templates have an E-value close to zero. Hence, E-values decrease with increase in sequence similarity.[91]

Alignment tools such as ClustalW[64,163] are then used to align the primary sequence of the target to that of the template with the highest coverage and identity score.[20,64,82,158] Several considerations are made in choosing the best template, which include high sequence similarity, high sequence identity and similarity in binding site/pocket or the region in which binding takes place.[91,112,164]

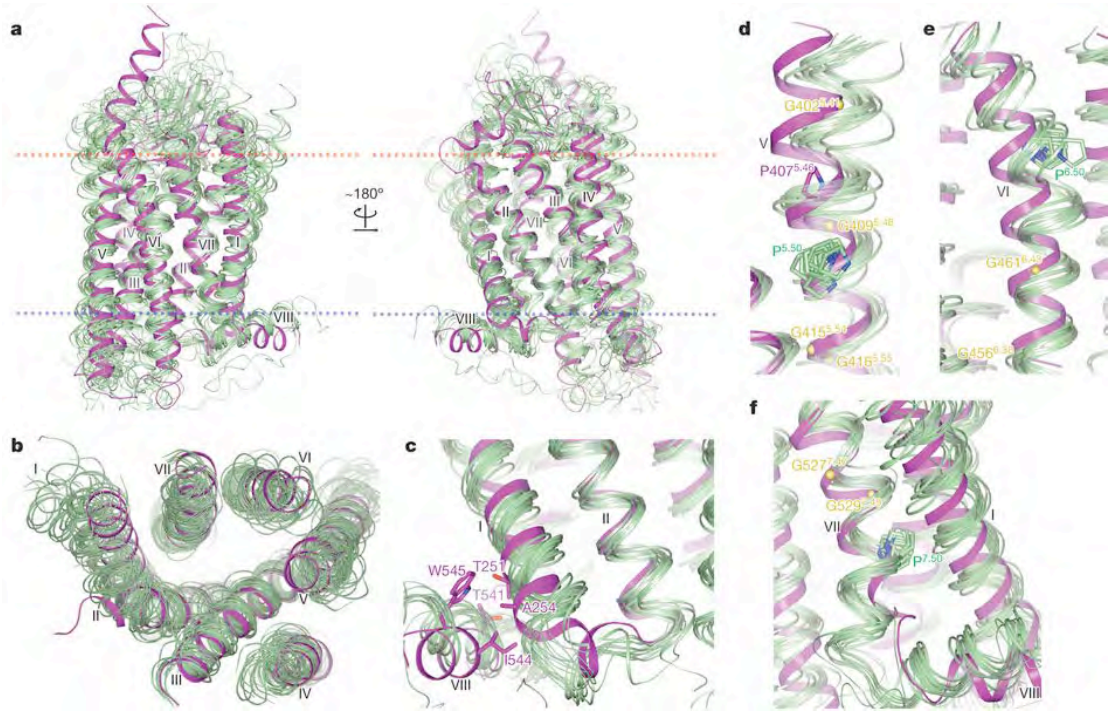


Figure 1.11: Homology model showing alignment to template (purple) and target models (green) shown in different views (*a-f*)[69]**.**

In some complex cases, multiple templates[52,112,158,160] can be aligned with the target in cases where the variability in the loop regions is apparent; ligand binding conformations and activity states are variable. Typical alignment is shown in Figure 1.10.

Studies have shown that docking scores on homology models (Figure 1.11) made using multiple templates have less than 2Å Root Mean Standard Deviation (RMSD) from the crystal structures used as templates compared to using single templates.[20,112,158] Thus multiple templates, coupled with sequence protein profiling, can be used to generate acceptable homology models, for example in GPCR homology modeling where sequence similarity is exceptionally low.[112,164] Careful selection of templates for homology modeling may lead to poor model quality and docking calculations with a poor model may lead to poor conclusions.[16,83,149] Orientations of side chains, which lead to structural differences in the proteins that may look highly similar, if different, result in completely different models.[85,108,165]

Prolines and glycines are responsible for most of the kinks and bulges in transmembrane regions and are suggested as essential mediators for hinge-bending.[82,83,160] For example, proline is understood to disrupt the helical hydrogen bonding and affect this normal bonding and may also effect repulsive steric interactions with the adjacent backbone atoms.[166,167] The shift of one residue in the alignment can result in a distortion of up to 3.8Å in the target backbone from the template structure.[20,82,83,158,159] This is seen in the beta-2 Adrenergic and beta-1 Adrenergic structures.[158,159,168]

Literature suggests the use of multiple sequences generate homology models with kinks and bulges predicted in the helical regions of the transmembrane proteins.[158,160,169] MODELER tools are then used to model the homologous target protein against template.[164,170] Model quality decreases with declining sequence identity.[56,83,129,171] Theoretically, a homology model has ~1-2Å RMSD match from Cα atoms at 70% sequence identity. However, at low sequence identities of ≤ 25%, the match with Cα is only 2-4Å.[112] MODELER can build biased models, which rely on the template chosen, even some with ambiguous spatial restraints.[16,61,83] In all this, it is important to constantly monitor that there are no gaps in the helical regions. GPCRs have significantly different loop regions from their templates. Loops are very flexible and difficult to predict.[75,82,172] Various approaches such as *de novo* design, molecular dynamics and geometrical sampling of dihedral angles starting from transmembrane regions by construction of the residues to select loop backbone, can be used to construct and optimize loop regions in proteins.[16,22,59,173,174]

The models generated by MODELER tools use various scoring functions depending on the algorithms used. Some of the commonly used scoring functions include the Discrete Optimized Protein Energy (DOPE) and GA341 score.[160] The GA341 score shows the reliability of a model that is derived from statistical potentials where a reliable model has a value higher than the prescribed cut-off of 0.7. Homology modeling results in structures with unfavourable contacts and bonds.[56,175] Running energy minimization can fix the unfavourable angles, torsions and bond lengths.[60,127,176,177] In some cases, homology modeling exploits threading to improve alignments (Figure 1.12) and incorporates *ab initio* folding for loop predictions with molecular dynamic simulations for optimizing the structure.[176,178]

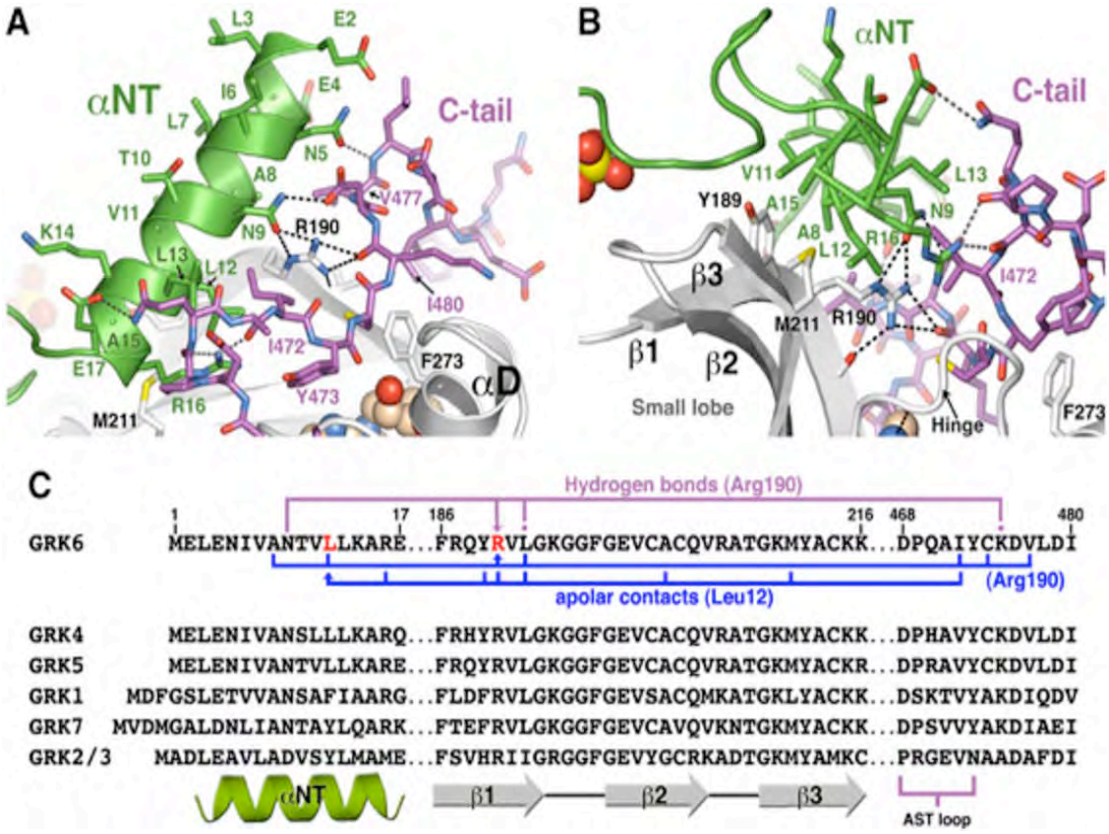


Figure 1.12: Effect of aligning conserved residues in the binding pocket in Beta1, Beta 2 and Beta3 adrenergic receptors. Arg190 and Leu13 are conserved in A and B. C shows the alignment of helices in the receptors.[168]

Alignment of conserved residues is critical as some conserved residues are involved in the ligand binding pocket.[158] Thus the side chain must be accurately positioned in the binding site to aid catalytic mechanisms. When a model has been generated, it has to be validated. Several tools can be used for model validation including mutagenesis studies and a Ramachandran Plot of psi (Ψ) and phi (Φ), which determines the stereo chemical quality of the target, considering the positioning of residues in energetically feasible regions in the structure.[173,179] Critically, a good model should have at least 90% of the residues in allowed regions.[16,173,179] Typically, errors in a model are significantly minimized if they are localized. Figure 1.13 illustrates a typical Ramachandran plot.

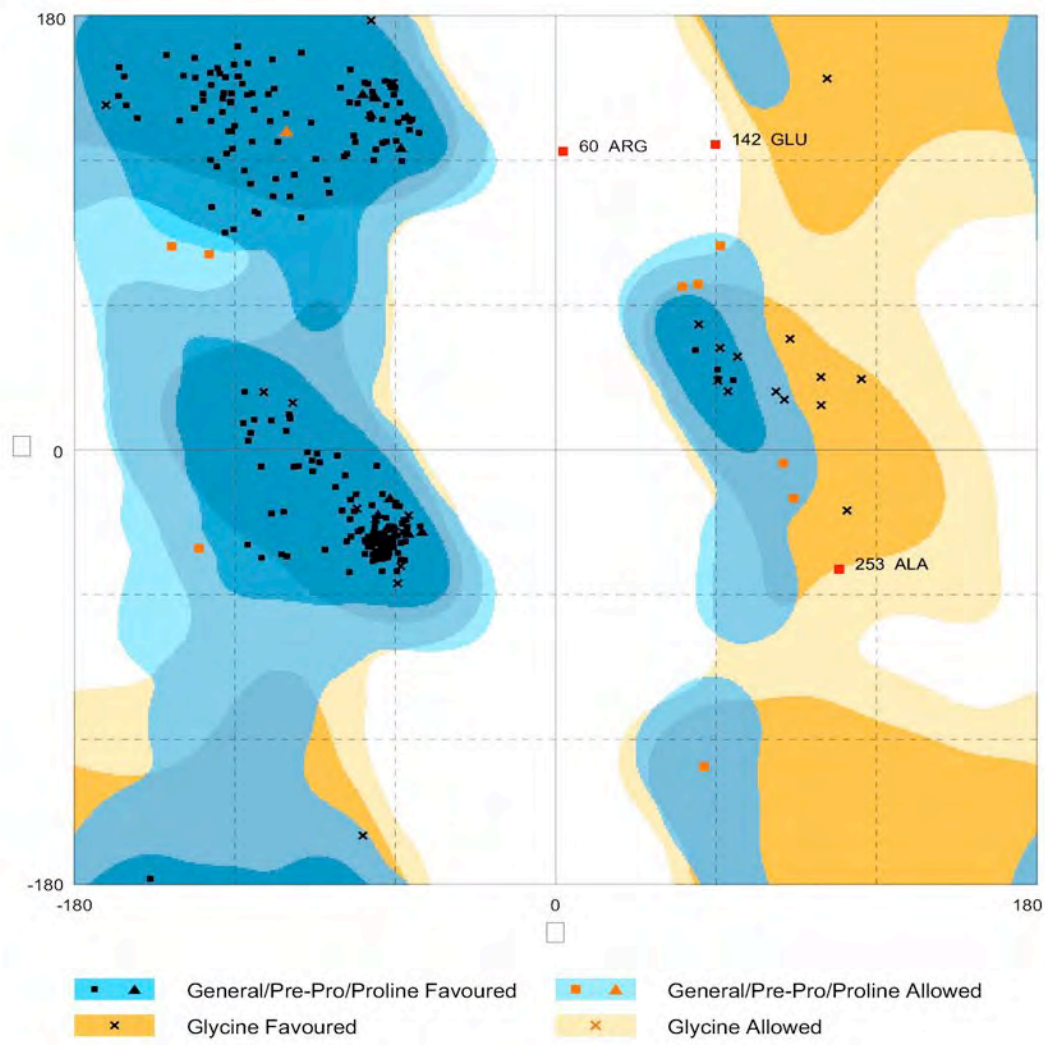


Figure 1.13: Ramachandran plot showing how protein structures are validated following orientation of amino acid residue backbone and side chains.[179,180]

However, in the assessment, it is not necessarily true to state that a clean Ramachandran plot validates a model.

1.4.2 Molecular Dynamics Simulations

Multiple conformations of protein structures are explored close to the lowest possible free energy. Different conformers also exist, which are separated by energy barriers greater than k_BT (where k_B is the Boltzmann constant and T is the temperature).[60,127,176,181] Evolution between substrate bound conformation ensue on moderately lengthy time scales. Consequently, it is essential to develop time-saving

techniques, which permit an easy search of conformational space. Several methodologies have recently been suggested to enrich the effectiveness of the conformational space searches.[74] The future of sampling methods depend on numerous approaches using hybrid, and fast sampling methods and molecular dynamic simulations to obtain a realistic ensemble.[127,140]

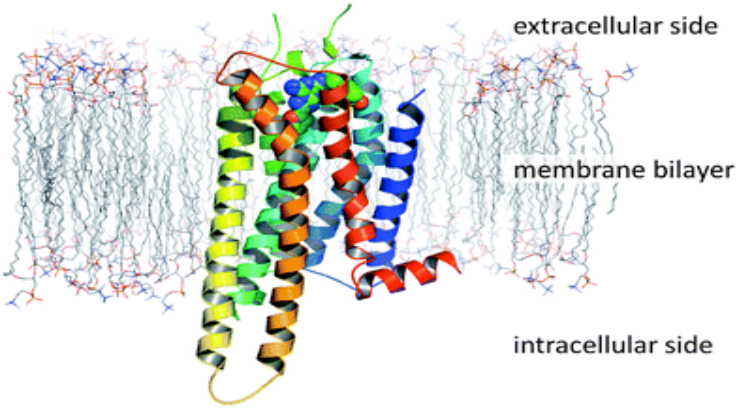


Figure 1.14: Mimetic membrane molecular dynamic simulations are carried out in a water box using GROMACS. This shows how the GPCR looks in the membrane. The loops are not embedded in the POPC membrane bilayer and the helical regions are covered by the membrane.[60]

Optimization of homology models to correct possible distortions in the bond angles, lengths, dihedral angles as well as other electronic properties of structures can be enhanced using molecular dynamics.[60,86,90,177] When running molecular dynamic simulations, time steps are considered in femto-seconds and the specificity of small time steps aids the trajectory to cover limited proportions of phase.[182] In contrast, large time steps introduce instability and artifacts in the system integration algorithms as a result of high-energy overlaps. However, long time scales (160-200 nanoseconds) in a mimetic membrane improves the quality of homology models.[60,140,176]

Molecular dynamic simulations also help to understand the flexibility of the homology models or protein targets as well as to remove any bumps and kinks that may be found in the structures.[82,160] Molecular dynamics can be performed in vacuum, water and DPC-micelle solution.[60,82,90,124,176] Molecular simulations can be performed for the protein alone, or for the peptide and / or for the protein-ligand complex in a mimetic lipid membrane as shown in Figure 1.14 for membrane receptors. Several steps are involved in molecular dynamic simulations. These are:

- Simulated annealing
- Equilibration
- Production

Molecular dynamic simulations exploit the time dependent behaviors of the atoms in the molecular system by manipulating and integrating Newtonian equations of motion and potential energy functions.[60,74,107] Careful parameterization of the force field is critical in understanding molecular dynamics. Force fields help understand, and elaborate the atomic interactions.[183] There are many force fields and care should be taken to evaluate which force field is most appropriate for the system under investigation.[60,124]

1.4.2.1 *Simulated annealing*

In molecular dynamics, the behaviour and position of each atom in a system is computed as a function of time using algorithms based on classical mechanics.[184,185] Simulated annealing optimizes the behaviour of molecules in a given system to overcome local minima during simulations.[60,171] In this case, the temperature is increased to explore parameter space in the system where the macroscopic behaviour of interactions is computed from microscopic behaviour.[171] This is followed by gradual reduction of temperature to approach global minima of potential energy.[124] At higher temperatures like 600K, molecular species move freely. The system is gradually cooled to a lower temperature i.e. 300K to allow atoms to settle into positions that may be close to global minima.[60,171,176]

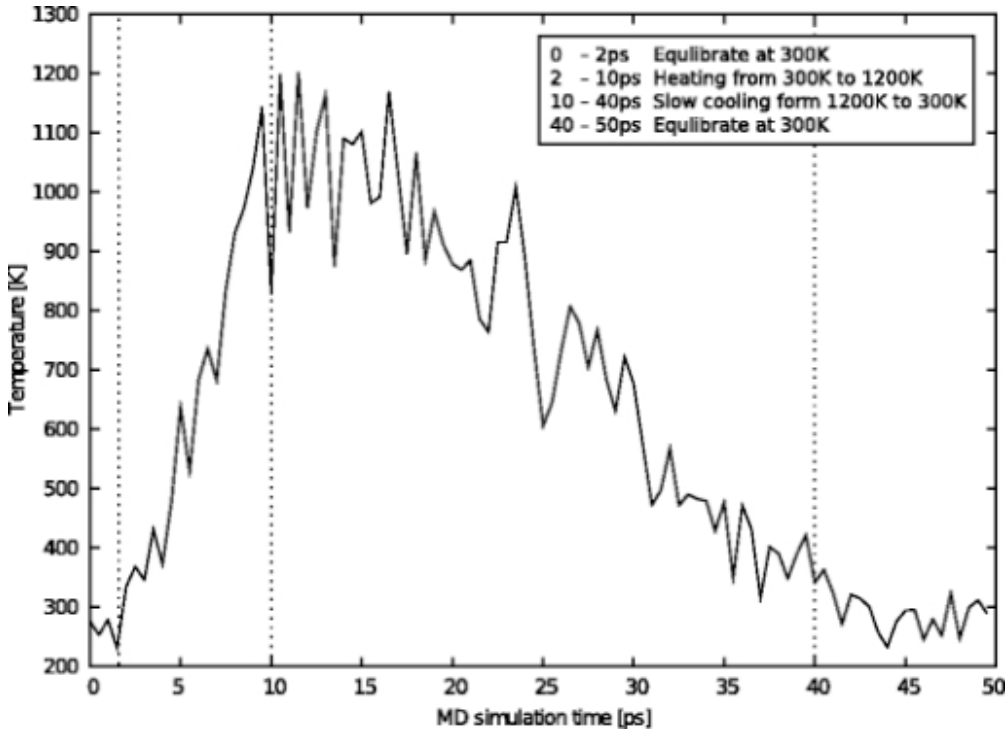


Figure 1.15: The graph showing simulated annealing from a maximum of 1200K to of 300K over 50ps[171]

The Monte Carlo algorithm is used in performing simulated annealing.[60,107,171] Simulated annealing (Figure 1.15) uses an annealing schedule, which specifies how the gradual cooling takes place against given time steps during the conformational search.

1.4.2.2 *Equilibration*

The system size and starting structure used in simulations affect the length of equilibration of a system.[186] This follows the general principle that the smaller the system, the faster the equilibration and vice versa. To run molecular dynamic simulations, the whole system must have achieved a uniform temperature of interest.[60,171] The general rule of thumb for any equilibration will be to employ position restraints on the protein, and add solvent sufficient enough to soak the protein and neutralize the ionic charge by adding anions or cations into the system.[187,188] Monitoring the simulation box dimensions, variations in energies and evaluating the equilibration, helps to determine if the protocol is meeting

expectations. However, the potential energy of the system is not a good indicator of progress in the equilibration.

Equilibration of a system can be at NVT or NPT ensembles where N is particle number, V is Volume, T is temperature and P is pressure.[124] These parameters are controlled and the cell parameters are relaxed. Temperature coupling is essential during NVT ensemble and there is pressure coupling during NPT.[60,171]

1.4.2.3 *Production run*

Prior to a production run, energy minimizations are run to eliminate and relax potential steric clashes in the protein structures and mimetic membranes. These steric clashes may be caused by addition of hydrogen atoms or solvent molecules, which are added into the system.[124,167,189] Simulations require specifications such as coordinate file of all atoms in the system, interactions between the atoms such as dihedral angles, bond orders, bond angles, bond types and charge distribution.[167] The parameters to perform as well as control the simulations in a dynamic system are also required. Parameter sets can include use of periodic boundary conditions, box type, size, distance to nearest neighbour, electrostatics, pressure and temperature coupling where the molecular dynamic simulations are taking place.[60,171]

Thermodynamic laws are fulfilled during molecular dynamic simulations resulting in new configurations. Production run during molecular dynamics solve classical molecular mechanics such as Newton's equations of motion:[60,190]

$$\mathbf{F}_i = \mathbf{m}_i\mathbf{a}_i, \qquad \text{(Equation 1.1)}$$

where $\mathbf{F}_i$ is the force,

$\mathbf{m}_i$ is the mass,

and $\mathbf{a}_i$ is the acceleration of atom *i*.

The force on atom *i* in equation 1.1 can be computed from the derivative of the potential energy V with respect to the coordinates r_i,

$$\mathbf{F}_i = -\partial\mathbf{V}/\partial\mathbf{r}_i. \qquad (2)$$

and

$$\mathbf{m}_i\,\partial^2\mathbf{r}_i/\partial\mathbf{t}^2 = -\partial\mathbf{V}/\partial\mathbf{r}_i \text{……......(3)}$$

Integration of Newton's equation of motion gives a trajectory that describes the acceleration of each atom; its topological coordinates and velocity. Such a simulation can be animated to describe how the system varies with time.

1.5 Ligand Based Drug Design (LBDD)

In drug discovery efforts, the availability of information such as the ligand molecules that bind to a biological target of interest can be used to design or search for similar ligands.[70,191] In this case, LBDD approaches can be exploited. These molecules have known bioactivities to the target. Some of the compound libraries that can be screened to obtain similar compounds include Maybridge, ZINC, ChEMBL, NCI, DrugBank and other in-house compound libraries in the pharmaceutical industries as well as natural products libraries.[99,100,103–105,192]

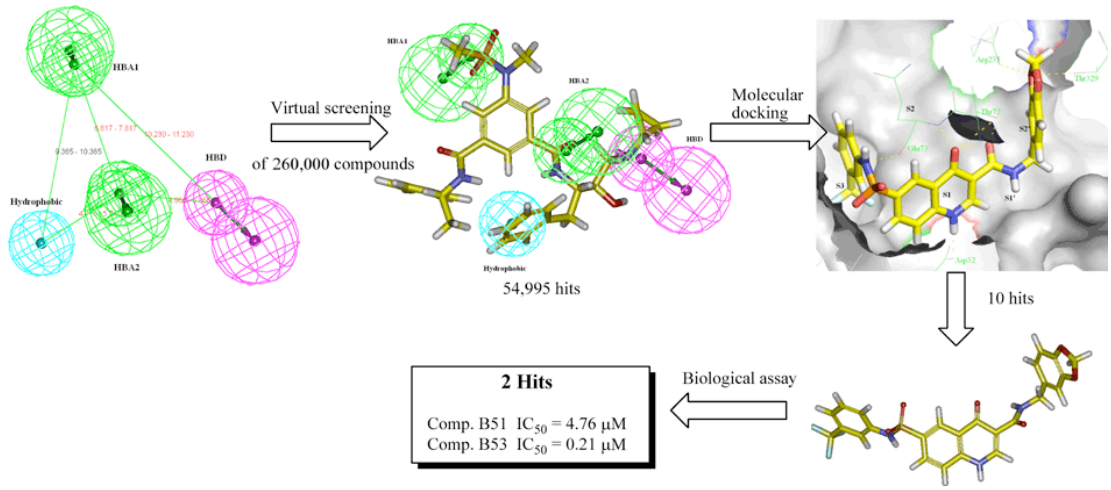


Figure 1.16: Beta-Secretase pharmacophore model showing a typical Hypo4 model. Chemical features are shown as green-hydrogen bond acceptor, and magenta-hydrogen bond donor, and cyan-hydrophobic feature.[193]

Drug-like compounds possess favourable properties for Absorption, Distribution, Metabolism, Excretion, and Toxicology (ADMET).[63,99,101,126,134,194] Thus favourable compounds exhibit membrane permeability, low toxicity, oral bioavailability and bioactivity as well as metabolic stability.[73,101] Some of the approaches used for ligand-based drug design include substructure and similarity searching, QSAR modeling, pharmacophore modeling and three dimensional shape matching.[130,135,140,156,195]

1.5.1 QSAR Modeling

QSAR models exploit the correlation between chemical properties of a compound's structure, which are the chemical descriptors, and biological activity of compounds.[130,136] Some of the information about the compounds with biological activity such as inhibitory concentrations (IC_{50}) and binding affinity (K_D) is essential. Similarity search of compounds from databases is performed using known biologically active compounds as references. Fingerprints such as Extended Connectivity Fingerprint (ECFP)[103,105] are used, along with variable diameters of nearest neighbours such as ECFP4 and ECFP6 to identify compounds that exhibit biological activity as their neighbours. However, the availability of substructural or structural similarity of known bioactive compounds to nearest neighbours limits similarity searches. QSAR models are validated by assessing the linearity of the correlation factor (R^2) as well as Root Mean Square Error (RMSE).

1.6 Conformational studies of peptides

Three mechanisms are suggested for peptide binding. These include conformation selection, induced fit and preferred conformation[90,107]. A classical model of ligand binding in protein receptors and subsequent activation of the receptor is induced when the requisite binding bioactive conformation of a peptide is achieved in the docking calculations. However, the binding site in the receptors such as GPCRs is limited to the extracellular side since the protein is bound in a lipid membrane, thus limiting the surface area and binding pocket of the ligands.[61,90,107] Therefore, a particular conformation of the peptide is essential to ensure that it binds to the receptor to give the required signal. It has been shown that pathways have been proposed where the peptides have been demonstrated to have an intrinsic ability to pre-orient and also pre-organize in solution to a conformation that facilitates binding to a receptor.[90]

Studies have also shown that GPCR peptides adopt a common pattern using shape recognition where a common "turn" motif is adopted in aqueous solution[70,90,107]. This conformation is associated with binding to, as well as mediation of bioactivity of, GPCRs. The peptide "turn"[124] may be one of many common ones in peptides such as

a gamma turn where 3 residues are involved, or a (beta) ß turn where 4 residues are involved as well as an α-turn where the peptide has 5 residues in the turn. Multiple ß turns are frequently described as 3_{10} helices if the ß turns are linked consecutively[9048,107]. Disulfide cross-links may lead to ß turns as does backbone cyclization of peptide amide bonds. This cyclization enables the peptide to easily penetrate membrane cell walls.

Signal transfer into intracellular trimeric G-proteins occurs when ligands such as small molecules and peptides bind to the GPCRs thereby causing conformational changes of the receptor.

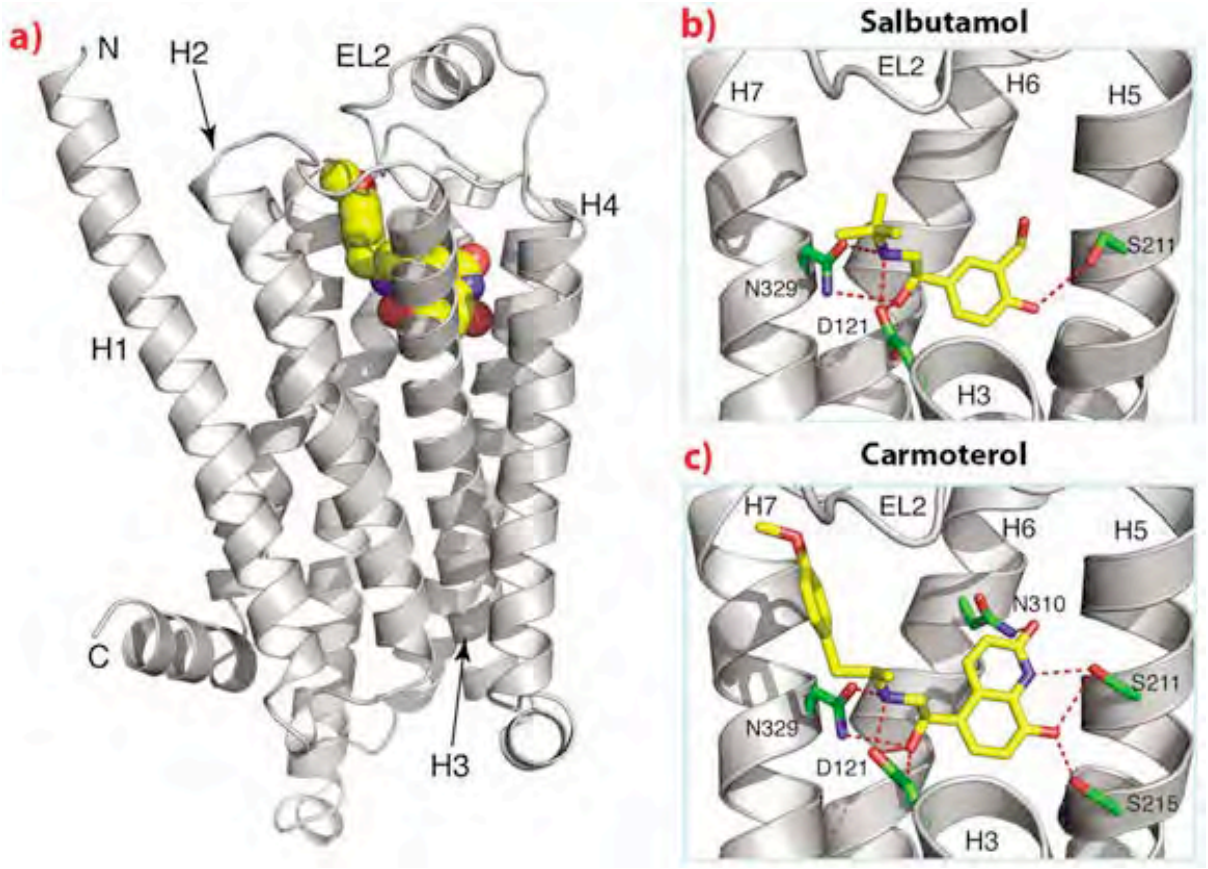


Figure 1.17: Agonist bound conformation of β_1AR used in the treatment of asthma. a) Overall structure of the receptor, b) Bound conformational structure of β_1AR with salbutamol (Ventolin) and c) bound structure with carmoterol. Some parts of α-helices H2, H3 and H4 have been removed in (b) and (c) for clarity[25]

Conformational studies have been investigated in aqueous solutions as well as bound conformations of the peptides. These environments in which conformational searches were done include vacuum, water, DPC/micelle, a mimetic membrane, as well as in

dimethyl sulphoxide (DMSO).[124,196] NMR restraints have been used in determining conformational structures of peptides in solution. This technique followed other techniques such as dichroism[197] and fluorescence spectroscopy,[107] which even though very useful, had some limitations on indicating exact regions of secondary structure. These techniques can be aided by incorporating NMR spectroscopy and molecular modeling in conformational studies.[198,199] Interproton distances are calculated from the integration of peak areas using NOE crosspeaks in 2D NMR and these distances are used as NMR restraints during molecular dynamic simulations [48,90]. The ultimate aim will be to determine the dominant peptide conformation space accessible using theoretical and experimental methods as complimentary techniques.

Some experiments carried out in *C.elegans* and *Ascaris suum* using Ca^{2+} mobilization assays in FLIPR format have matched peptides to FLP18 receptor proteins at 5μM. The strongest readout receptor activation signal in Ca^{2+} assays was identified for peptides with a C-terminal -VPGVLRF-NH_2 motif[6]. These were expressed functionally using FLP18R1a in Chinese Hamster Ovary (CHO) cells. The peptides that will be discussed in this thesis are *flp18-6, af3, af4* and *af20*. Studies conducted on protein ligand matching identified *af3, af4* and *af20* as potential agonists for FLP18 GPCR receptor, but mostly in *Ascaris suum*.

1.6.1 flp18-6

The sequence of this peptide is **DVPGVLRF-NH_2**. This implies that the peptide is amidated on the C-terminus to protect it against potential degradation.

1.6.2 af3

The sequence is **AVPGVLRF-NH_2.** Ca^{2+} mobilization assays on CHO cells performed using the *af3* peptide, AVPGVLRF-NH_2, was identified as the most highly potent agonist for FLP18R1a. The potency was found to be (EC_{50} 7.6– 13.1 nM)[6].

1.6.3 af4

Like the *af3* peptide, the *af4* peptide with this sequence **GDVPGVLRF-NH_2**, is an endogenous ligand for FLP18 receptors. The C-terminal possesses a similar motif with the *af3* and *af20* peptides[6]. This peptide was found to be an agonist for FLP18R1a with an EC_{50} of 13.5nM[6] in Ca^{2+} mobilization assays.

1.6.4 af20

When isoforms of FLP18R1 were matched with *flp18* peptides, the *af20* peptide was also found to induce a very high activation signal at 5μM in Ca^{2+} assays carried out[6]. The peptide sequence is **GMPGVLRF-NH_2**. However, this peptide is functionally expressed in *ascaris suum* (*A.suum*)[6].

1.7 Aim

To identify potential anthelmintic hit compounds through docking-based virtual screening on the target, FLP18R1.

1.8 Objectives

- To investigate the solution structure of *af3*, *af4*, *af20* and *flp18-6* hormones from nematodes using NMR-distance restrained molecular dynamics
- To build a homology model of FLP18R1a using its primary sequence from GenBank
- To optimize the loops of the homology model using molecular dynamic simulations
- To identify the essential active site on the receptor (FLP18R1)
- To perform docking calculations of the peptides to the receptor
- To perform molecular dynamic simulation of protein-peptide complexes in mimetic membrane environment
- To perform docking structure based virtual screening on the protein-ligand complexes
- To build a pharmacophore model of the docked protein-ligand complex[7]
- To perform pharmacophore-based virtual screening of the protein-ligand

complex in order to identify virtual hits

- To submit identified virtual hits for biological evaluation
- To use docking studies to rationalise biological screening results

1.9 References

(1) Brooker, S. Estimating the Global Distribution and Disease Burden of Intestinal Nematode Infections: Adding up the Numbers -a Review. *Int. J. Parasitol.* **2010**, *40* (10), 1137–1144.

(2) Kitamura, T.; Obara, H.; Takashima, Y.; Takahashi, K.; Inaoka, K.; Nagai, M.; Endo, H.; Jimba, M.; Sugiura, Y. World Health Assembly Agendas and Trends of International Health Issues for the Last 43 Years: Analysis of World Health Assembly Agendas between 1970 and 2012. *Health Policy (New. York).* **2013**, *110* (2-3), 198–206.

(3) Coles, G. C.; Bauer, C.; Borgsteede, F. H.; Geerts, S.; Klei, T. R.; Taylor, M. A.; Waller, P. J. World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) Methods for the Detection of Anthelmintic Resistance in Nematodes of Veterinary Importance. *Vet. Parasitol.* **1992**, *44* (1-2), 35–44.

(4) Kaplan, R. M. Drug Resistance in Nematodes of Veterinary Importance: A Status Report. *Trends in Parasitology*, 2004, *20*, 477–481.

(5) Hürlimann, E.; Schur, N.; Boutsika, K.; Stensgaard, A. S.; de Himpsl, M. L.; Ziegelbauer, K.; Laizer, N.; Camenzind, L.; Pasquale, A.; Ekpo, U. F.; Simoonga, C.; Mushinge, G.; Saarnak, C. F. L.; Utzinger, J.; Kristensen, T. K.; Vounatsou, P. Toward an Open-Access Global Database for Mapping, Control, and Surveillance of Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.* **2011**, *5* (12), 1–11.

(6) Kubiak, T. M.; Larsen, M. J.; Bowman, J. W.; Geary, T. G.; Lowery, D. F. FMRFamide-like Peptides Encoded on the Flp-18 Precursor Gene Activate Two Isoforms of the Orphan *Caenorhabditis Elegans* G-Protein-Coupled Receptor Y58G8A.4 Heterologously Expressed in Mammalian Cells. *Biopolym. - Pept. Sci. Sect.* **2008**, *90* (3), 339–348.

(7) Geary, T. G.; Kubiak, T. M. Neuropeptide G-Protein-Coupled Receptors, Their Cognate Ligands and Behavior in *Caenorhabditis Elegans*. *Trends Pharmacol. Sci.* **2005**, *26* (2), 56–58.

(8) World Health Organization; Organization, W. H. *World Health Statistics 2008*; **2008**.

(9) Chan, M. S. The Global Burden of Intestinal Nematode Infections - Fifty Years on. *Parasitology Today*, **1997**, *13*, 438–443.

(10) World Health Organization. *World Malaria Report*; **2010**; Vol. WHO/HTM/GM.

(11) Kitamura, T.; Obara, H.; Takashima, Y.; Takahashi, K.; Inaoka, K.; Nagai, M.; Endo, H.; Jimba, M.; Sugiura, Y. World Health Assembly Agendas and Trends of International Health Issues for the Last 43 Years: Analysis of World Health Assembly Agendas between 1970 and 2012. *Health Policy* **2013**, *110* (2-3), 198–206.

(12) Savioli, L.; Gabrielli, A. F. Helmintic Diseases: Intestinal Nematode Infection. In *International Encyclopedia of Public Health*; **2008**; pp 332–339.

(13) Lustigman, S.; Prichard, R. K.; Gazzinelli, A.; Grant, W. N.; Boatin, B. a.; McCarthy, J. S.; Basáñez, M. G. A Research Agenda for Helminth Diseases of Humans: The Problem of Helminthiases. *PLoS Negl. Trop. Dis.* **2012**, *6* (4).

(14) McVeigh, P.; Leech, S.; Mair, G. R.; Marks, N. J.; Geary, T. G.; Maule, A. G. Analysis of FMRFamide-like Peptide (FLP) Diversity in Phylum Nematoda. *Int. J. Parasitol.* **2005**, *35* (10), 1043–1060.

(15) Chitwood, D. J. Research on Plant-Parasitic Nematode Biology Conducted by the United States Department of Agriculture-Agricultural Research Service. *Pest Manag. Sci.* **2003**, *59* (6-7), 748–753.

(16) Du, P.; Salon, J. a; Tamm, J. a; Hou, C.; Cui, W.; Walker, M. W.; Adham, N.; Dhanoa, D. S.; Islam, I.; Vaysse, P. J.; Dowling, B.; Shifman, Y.; Boyle, N.; Rueger, H.; Schmidlin, T.; Yamaguchi, Y.; Branchek, T. a; Weinshank, R. L.; Gluchowski, C. Modeling the G-Protein-Coupled Neuropeptide Y Y1 Receptor Agonist and Antagonist Binding Sites. *Protein Eng.* **1997**, *10* (2), 109–117.

(17) Kryshtafovych, A.; Fidelis, K. Protein Structure Prediction and Model Quality Assessment. *Drug Discov. Today* **2009**, *14* (7-8), 386–393.

(18) Kuhn, M.; Campillos, M.; González, P.; Jensen, L. J.; Bork, P. Large-Scale

Prediction of Drug-Target Relationships. *FEBS Lett.* **2008**, *582* (8), 1283–1290.

(19) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W. a. Prediction of Structure and Function of G Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12622–12627.

(20) Skolnick, J.; Zhou, H.; Gao, M. Are Predicted Protein Structures of Any Value for Binding Site Prediction and Virtual Ligand Screening? *Curr. Opin. Struct. Biol.* **2013**, *23* (2), 191–197.

(21) Bissantz, C.; Logean, A.; Rognan, D. High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building, and Receptor Library Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1162–1176.

(22) Adams, R.; Worth, C. L.; Guenther, S.; Dunkel, M.; Lehmann, R.; Preissner, R. Binding Sites in Membrane Proteins - Diversity, Druggability and Prospects. *Eur. J. Cell Biol.* **2012**, *91* (4), 326–339.

(23) van der Horst, E.; Peironcely, J. E.; Ijzerman, A. P.; Beukers, M. W.; Lane, J. R.; van Vlijmen, H. W. T.; Emmerich, M. T. M.; Okuno, Y.; Bender, A. A Novel Chemogenomics Analysis of G Protein-Coupled Receptors (GPCRs) and Their Ligands: A Potential Strategy for Receptor de-Orphanization. *BMC Bioinformatics* **2010**, *11*, 316.

(24) Taylor, C. M.; Barda, Y.; Kisselev, O. G.; Marshall, G. R. Modulating G-Protein Coupled Receptor / G-Protein Signal Transduction by Small Molecules Suggested by Virtual Screening. **2008**, 5297–5303.

(25) Brogi, S.; Tafi, A.; Désaubry, L.; Nebigil, C. G. Discovery of GPCR Ligands for Probing Signal Transduction Pathways. **2014**, *5* (November), 1–14.

(26) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of. *Nature* **2013**, *494* (7436), 185–194.

(27) Novozhilova, E.; Kimber, M. J.; Qian, H.; McVeigh, P.; Robertson, A. P.; Zamanian, M.; Maule, A. G.; Day, T. A. FMRFamide-like Peptides (FLPs) Enhance Voltage-Gated Calcium Currents to Elicit Muscle Contraction in the

Human Parasite Schistosoma Mansoni. *PLoS Negl. Trop. Dis.* **2010**, *4* (8).

(28) Frooninckx, L.; Rompay, L.; Temmerman, L.; Sinay, E.; Beets, I.; Janssen, T.; Husson, S. J.; Schoofs, L. Neuropeptide GPCRs in *C. Elegans*. *Front. Endocrinol. (Lausanne).* **2012**, *3* (DEC), 1–18.

(29) Castillo, J. C.; Reynolds, S. E.; Eleftherianos, I. Insect Immune Responses to Nematode Parasites. *Trends in Parasitology*, 2011, *27*, 537–547.

(30) Kutz, S. J.; Hoberg, E. P.; Nagy, J.; Polley, L.; Elkin, B. "Emerging" Parasitic Infections in Arctic Ungulates. *Integr. Comp. Biol.* **2004**, *44* (2), 109–118.

(31) Colditz, I. G. Six Costs of Immunity to Gastrointestinal Nematode Infections. *Parasite Immunology*, **2008**, *30*, 63–70.

(32) Knopp, S.; Steinmann, P.; Hatz, C.; Keiser, J.; Utzinger, J. Nematode Infections: Filariases. *Infectious Disease Clinics of North America*, 2012, *26*, 359–381.

(33) Knopp, S.; Steinmann, P.; Keiser, J.; Utzinger, J. Nematode Infections. Soil-Transmitted Helminths and Trichinella. *Infectious Disease Clinics of North America*, **2012**, *26*, 341–358.

(34) Basáñez, M. G.; French, M. D.; Walker, M.; Churcher, T. S. Paradigm Lost: How Parasite Control May Alter Pattern and Process in Human Helminthiases. *Trends Parasitol.* **2012**, *28* (4), 161–171.

(35) Ezenwa, V. O.; Etienne, R. S.; Luikart, G.; Beja-Pereira, A.; Jolles, A. E. Hidden Consequences of Living in a Wormy World: Nematode-induced Immune Suppression Facilitates Tuberculosis Invasion in African Buffalo. *Am. Nat.* **2010**, *176* (5), 613–624.

(36) Barry, M. a; Simon, G. G.; Mistry, N.; Hotez, P. J. Global Trends in Neglected Tropical Disease Control and Elimination: Impact on Child Health. *Arch. Dis. Child.* **2013**, *98* (8), 635–641.

(37) Crompton, D. W. The Public Health Importance of Hookworm Disease. *Parasitology* **2000**, *121 Suppl*, S39–S50.

(38) Centers for Disease Control and Prevention. CDC - Malaria http://www.cdc.gov/malaria/about/biology/.

(39) Brooker, S.; Kabatereine, N. B.; Tukahebwa, E. M.; Kazibwe, F. Spatial Analysis of the Distribution of Intestinal Nematode Infections in Uganda. *Epidemiol. Infect.* **2004**, *132* (6), 1065–1071.

(40) Bishop, S. C.; Stear, M. J. Modeling of Host Genetics and Resistance to Infectious Diseases: Understanding and Controlling Nematode Infections. In *Veterinary Parasitology*; **2003**; Vol. 115, pp 147–166.

(41) Bethony, J.; Brooker, S.; Albonico, M.; Geiger, S. M.; Loukas, A.; Diemert, D.; Hotez, P. J. Soil-Transmitted Helminth Infections: Ascariasis, Trichuriasis, and Hookworm. *Lancet*, **2006**, *367*, 1521–1532.

(42) Tembely, S.; Lahlou-Kassi, A.; Rege, J. E. O.; Sovani, S.; Diedhiou, M. L.; Baker, R. L. The Epidemiology of Nematode Infections in Sheep in a Cool Tropical Environment. *Vet. Parasitol.* **1997**, *70* (1-3), 129–141.

(43) Viney, M. How Do Host Immune Responses Affect Nematode Infections? *Trends in Parasitology*, **2002**, *18*, 63–66.

(44) Martin, R. J. Modes of Action of Anthelmintic Drugs. *Vet. J.* **1997**, *154* (1), 11–34.

(45) Kumar, S.; Koutsovoulos, G.; Kaur, G.; Blaxter, M. Toward 959 Nematode Genomes. *Worm*, **2012**, *1*, 41–49.

(46) Meeusen, T.; Mertens, I.; Clynen, E.; Baggerman, G.; Nichols, R.; Nachman, R. J.; Huybrechts, R.; De Loof, A.; Schoofs, L. Identification in Drosophila Melanogaster of the Invertebrate G Protein-Coupled FMRFamide Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (24), 15363–15368.

(47) Shore, D. M.; Reggio, P. H. The Therapeutic Potential of Orphan GPCRs, GPR35 and GPR55. *Front. Pharmacol.* **2015**, *6* (April), 1–22.

(48) Tapaneeyakorn, S.; Goddard, A. D.; Oates, J.; Willis, C. L.; Watts, A. Solution- and Solid-State NMR Studies of GPCRs and Their Ligands. *Biochim. Biophys. Acta - Biomembr.* **2011**, *1808* (6), 1462–1475.

(49) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known

Drug Databases. *J. Comb. Chem.* **1999**, *1* (1), 55–68.

(50) Wang, C.; Wu, H.; Katritch, V.; Han, G. W.; Huang, X.-P.; Liu, W.; Siu, F. Y.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the Human Smoothened Receptor Bound to an Antitumour Agent. *Nature* **2013**, *497* (7449), 338–343.

(51) Vohra, S.; Taddese, B.; Conner, A. C.; Poyner, D. R.; Hay, D. L.; Barwell, J.; Reeves, P. J.; Upton, G. J.; Reynolds, C. a. Similarity between Class A and Class B G-Protein-Coupled Receptors Exemplified through Calcitonin Gene-Related Peptide Receptor Modelling and Mutagenesis Studies. *J R Soc Interface* **2013**, *10* (79), 20120846.

(52) Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric Sodium in Class A GPCR Signaling. *Trends Biochem. Sci.* **2014**, *39* (5), 233–244.

(53) Hiller, C.; Kühhorn, J.; Gmeiner, P. Class A G-Protein-Coupled Receptor (GPCR) Dimers and Bivalent Ligands. *J. Med. Chem.* **2013**, *56* (17), 6542–6559.

(54) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The Structure and Function of G-Protein-Coupled Receptors. *Nature* **2009**, *459* (7245), 356–363.

(55) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of Two Distinct Inactive Conformations of the beta2-Adrenergic Receptor Reconciles Structural and Biochemical Observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4689–4694.

(56) Yarnitzky, T.; Levit, A.; Niv, M. Y. Homology Modeling of G-Protein-Coupled Receptors with X-Ray Structures on the Rise. *Curr. Opin. Drug Discov. Devel.* **2010**, *13* (3), 317–325.

(57) Hollenstein, K.; De Graaf, C.; Bortolato, A.; Wang, M. W.; Marshall, F. H.; Stevens, R. C. Insights into the Structure of Class B GPCRs. *Trends Pharmacol. Sci.* **2014**, *35* (1), 12–22.

(58) Bd, L.; Wy, T. A Nematode Model for Mitochondrial Diseases. *West Coast Worm Meeting*, **2000**.

(59) Gardner, M. R.; Kattenhorn, L. M.; Kondur, H. R.; Von Schaewen, M.;

Dorfman, T.; Chiang, J. J.; Haworth, K. G.; Decker, J. M.; Alpert, M. D.; Bailey, C. C.; Neale Jr, E. S.; Fellinger, C. H.; Joshi, V. R.; Fuchs, S. P.; Martinez-Navio, J. M.; Quinlan, B. D.; Yao, A. Y.; Mouquet, H.; Gorman, J.; Zhang, B.; Poignard, P.; Nussenzweig, M. C.; Burton, D. R.; Kwong, P. D.; Piatak Jr, M.; Lifson, J. D.; Gao, G.; Desrosiers, R. C.; Evans, D. T.; Hahn, B. H.; Ploss, A.; Cannon, P. M.; Seaman, M. S.; Farzan, M. AAV-Expressed eCD4-Ig Provides Durable Protection from Multiple SHIV Challenges. *Nature* **2015**.

(60) Schlegel, B.; Sippl, W.; Höltje, H. D. Molecular Dynamics Simulations of Bovine Rhodopsin: Influence of Protonation States and Different Membrane-Mimicking Environments. *J. Mol. Model.* **2005**, *12* (1), 49–64.

(61) Mondal, S.; Johnston, J. M.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H. Membrane Driven Spatial Organization of GPCRs. *Sci. Rep.* **2013**, *3*, 2909.

(62) Lindert, S.; Maslennikov, I.; Chiu, E. J. C.; Pierce, L. C.; McCammon, J. A.; Choe, S. Drug Screening Strategy for Human Membrane Proteins: From NMR Protein Backbone Structure to in Silica- and NMR-Screened Hits. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 724–733.

(63) Schmidt, T.; Bergner, A.; Schwede, T. Modelling Three-Dimensional Protein Structures for Applications in Drug Design. *Drug Discov. Today* **2014**, *19* (7), 890–897.

(64) Reddy, C. S.; Vijayasarathy, K.; Srinivas, E.; Sastry, G. M.; Sastry, G. N. Homology Modeling of Membrane Proteins: A Critical Assessment. *Comput. Biol. Chem.* **2006**, *30* (2), 120–126.

(65) 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* **2011**, *332* (6027), 322–327.

(66) Malo, M. Selectivity of Dopamine D1 and D2 Receptor Agonists – A Combined Computational Approach; **2012**.

(67) Evers, A.; Klabunde, T. Structure-Based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the

alpha1A Adrenergic Receptor. *J. Med. Chem.* **2005**, *48* (4), 1088–1097.

(68) Jayanthi, S.; Kang, S. W.; Bingham, D.; Tessaro, B. a; Suresh Kumar, T. K.; Kuenzel, W. J. Identification of Antagonists to the Vasotocin Receptor Sub-Type 4 (VT4R) Involved in Stress by Molecular Modelling and Verification Using Anterior Pituitary Cells. *J. Biomol. Struct. Dyn.* **2014**, *32* (4), 648–660.

(69) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2013**, *53*, 531–556.

(70) Gruber, C. W.; Muttenthaler, M.; Freissmuth, M. Ligand-Based Peptide Design and Combinatorial Peptide Libraries to Target G Protein-Coupled Receptors. *Curr. Pharm. Des.* **2010**, *16* (28), 3071–3088.

(71) Haupt, V. J.; Daminelli, S.; Schroeder, M. Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key. *PLoS One* **2013**, *8* (6).

(72) Reker, D.; Rodrigues, T.; Schneider, P.; Schneider, G. Identifying the Macromolecular Targets of de Novo-Designed Chemical Entities through Self-Organizing Map Consensus. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (11), 4067–4072.

(73) Lavecchia, A.; Di Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.

(74) Sale, K.; Faulon, J. L.; Gray, G. A; Schoeniger, J. S.; Young, M. M. Optimal Bundling of Transmembrane Helices Using Sparse Distance Constraints. *Protein Sci.* **2004**, *13* (10), 2613–2627.

(75) Tastan, O.; Klein-Seetharaman, J.; Meirovitch, H. The Effect of Loops on the Structural Organization of α-Helical Membrane Proteins. *Biophys. J.* **2009**, *96* (6), 2299–2312.

(76) Kunishima, N.; Shimada, Y.; Tsuji, Y.; Sato, T.; Yamamoto, M.; Kumasaka, T.; Nakanishi, S.; Jingami, H.; Morikawa, K. Structural Basis of Glutamate Recognition by a Dimeric Metabotropic Glutamate Receptor. *Nature* **2000**, *407* (6807), 971–977.

(77) Gloriam, D. E.; Foord, S. M.; Blaney, F. E.; Garland, S. L. Definition of the G

Protein-Coupled Receptor Transmembrane Bundle Binding Pocket and Calculation of Receptor Similarities for Drug Design. *J. Med. Chem.* **2009**, *52* (14), 4429–4442.

(78) Zocher, M.; Bippes, C. a; Zhang, C.; Müller, D. J. Single-Molecule Force Spectroscopy of G-Protein-Coupled Receptors. *Chem. Soc. Rev.* **2013**, *42* (19), 7801–7815.

(79) Mühlfeld, S.; Schmitt-Wrede, H. P.; Harder, A.; Wunderlich, F. FMRFamide-like Neuropeptides as Putative Ligands of the Latrophilin-like HC110-R from Haemonchus Contortus. *Mol. Biochem. Parasitol.* **2009**, *164* (2), 162–164.

(80) Hetényi, C.; Van Der Spoel, D. Toward Prediction of Functional Protein Pockets Using Blind Docking and Pocket Search Algorithms. *Protein Sci.* **2011**, *20* (5), 880–893.

(81) Sanders, M. P. a; Verhoeven, S.; De Graaf, C.; Roumen, L.; Vroling, B.; Nabuurs, S. B.; De Vlieg, J.; Klomp, J. P. G. Snooker: A Structure-Based Pharmacophore Generation Tool Applied to Class A GPCRs. *J. Chem. Inf. Model.* **2011**, *51* (9), 2277–2292.

(82) de Graaf, C.; Rognan, D. Customizing G Protein-Coupled Receptor Models for Structure-Based Virtual Screening. *Curr. Pharm. Des.* **2009**, *15* (35), 4026–4048.

(83) Dalton, J. a R.; Jackson, R. M. Homology-Modelling Protein-Ligand Interactions: Allowing for Ligand-Induced Conformational Change. *J. Mol. Biol.* **2010**, *399* (4), 645–661.

(84) Tyndall, J. D. A; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2005**, *105* (3), 793–826.

(85) Worth, C. L.; Kleinau, G.; Krause, G. Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models. *PLoS One* **2009**, *4* (9).

(86) Yoshikawa, Y.; Oishi, S.; Kubo, T.; Tanahara, N.; Fujii, N.; Furuya, T. Optimized Method of G-Protein-Coupled Receptor Homology Modeling: Its Application to the Discovery of Novel CXCR7 Ligands. *J. Med. Chem.* **2013**,

56 (11), 4236–4251.

(87) Schlegel, B. A Molecular Modelling Study Of A G-Protein Coupled Receptor. **2005**.

(88) Baldwin, J. M. The Probable Arrangement of the Helices in G Protein-Coupled Receptors. *EMBO J.* **1993**, *12* (4), 1693–1703.

(89) Bondensgaard, K.; Ankersen, M.; Thøgersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of Privileged Structures by G-Protein Coupled Receptors. *J. Med. Chem.* **2004**, *47* (4), 888–899.

(90) Daura, X.; Gademann, K.; Schäfer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. The Beta-Peptide Hairpin in Solution: Conformational Study of a Beta-Hexapeptide in Methanol by NMR Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **2001**, *123* (10), 2393–2404.

(91) Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. *Curr. Protoc. Bioinforma.* **2013**, No. SUPL.43.

(92) Hollenstein, K.; De Graaf, C.; Bortolato, A.; Wang, M. W.; Marshall, F. H.; Stevens, R. C. Insights into the Structure of Class B GPCRs. *Trends in Pharmacological Sciences*, **2014**, *35*, 12–22.

(93) Millar, R. P.; Newton, C. L. The Year in G Protein-Coupled Receptor Research. *Mol. Endocrinol.* **2010**, *24* (1), 261–274.

(94) Filmore, D. It's a GPCR World. *Mod. drug Discov.* **2004**, *7* (11), 24–27.

(95) Orry, A. J.; Wallace, B. A. Modeling and Docking the Endothelin G-Protein-Coupled Receptor. *Biophys. J.* **2000**, *79* (6), 3083–3094.

(96) Heifetz, A.; Schertler, G. F. X.; Seifert, R.; Tate, C. G.; Sexton, P. M.; Gurevich, V. V.; Fourmy, D.; Cherezov, V.; Marshall, F. H.; Storer, R. I.; Moraes, I.; Tikhonova, I. G.; Tautermann, C. S.; Hunt, P.; Ceska, T.; Hodgson, S.; Bodkin, M. J.; Singh, S.; Law, R. J.; Biggin, P. C. GPCR Structure, Function, Drug Discovery and Crystallography: Report from Academia-Industry International Conference (UK Royal Society) Chicheley Hall, 1–2 September 2014. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **2015**.

(97) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.;

Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42* (D1).

(98) Mok, N. Y.; Brenk, R. Mining the ChEMBL Database: An Efficient Chemoinformatics Workflow for Assembling an Ion Channel-Focused Screening Library. *J. Chem. Inf. Model.* **2011**, *51* (10), 2449–2454.

(99) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1), 1–8.

(100) Wassermann, A. M.; Bajorath, J. BindingDB and ChEMBL: Online Compound Databases for Drug Discovery. *Expert Opinion on Drug Discovery*, **2011**, *6*, 683–687.

(101) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1).

(102) Vilar, S.; Ferino, G.; Phatak, S. S.; Berk, B.; Cavasotto, C. N.; Costanzi, S. Docking-Based Virtual Screening for Ligands of G Protein-Coupled Receptors: Not Only Crystal Structures but Also in Silico Models. *J. Mol. Graph. Model.* **2011**, *29* (5), 614–623.

(103) Willighagen, E. L.; Waagmeester, A.; Spjuth, O.; Ansell, P.; Williams, A. J.; Tkachenko, V.; Hastings, J.; Chen, B.; Wild, D. J. The ChEMBL Database as Linked Open Data. *J. Cheminform.* **2013**, *5* (5).

(104) Southan, C.; Sitzmann, M.; Muresan, S. Comparing the Chemical Structure and Protein Content of ChEMBL, DrugBank, Human Metabolome Database and the Therapeutic Target Database. *Mol. Inform.* **2013**, *32* (11-12), 881–897.

(105) Heikamp, K.; Bajorath, J. Large-Scale Similarity Search Profiling of ChEMBL Compound Data Sets. *J. Chem. Inf. Model.* **2011**, *51* (8), 1831–1839.

(106) Levit, A.; Beuming, T.; Krilov, G.; Sherman, W.; Niv, M. Y. Predicting GPCR Promiscuity Using Binding Site Features. *J. Chem. Inf. Model.* **2014**, *54* (1),

184–194.

(107) Bierzyński, A. Methods of Peptide Conformation Studies. *Acta Biochim. Pol.* **2001**, *48* (4), 1091–1099.

(108) Kooistra, A. J.; Kuhne, S.; De Esch, I. J. P.; Leurs, R.; De Graaf, C. A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *Br. J. Pharmacol.* **2013**, *170* (1), 101–126.

(109) Flower, D. R. Modelling G-Protein-Coupled Receptors for Drug Design. *Biochim. Biophys. Acta - Rev. Biomembr.* **1999**, *1422* (3), 207–234.

(110) Huang, C. C.; Tesmer, J. J. G. Recognition in the Face of Diversity: Interactions of Heterotrimeric G Proteins and G Protein-Coupled Receptor (GPCR) Kinases with Activated GPCRs. *Journal of Biological Chemistry*, **2011**, *286*, 7715–7721.

(111) Schnur, D. M.; Hermsmeier, M. A.; Tebben, A. J. Are Target-Family-Privileged Substructures Truly Privileged? *J. Med. Chem.* **2006**, *49* (6), 2000–2009.

(112) Tramontano, A. Homology Modeling with Low Sequence Identity. *Methods* **1998**, *14* (3), 293–300.

(113) Erguner, B.; Hattori, M.; Kanehisa, M. Characterizing Common Substructures of Ligands for GPCR Protein Subfamilies, **2010**

(114) GPCR-Targeted Library. http://www.chemdiv.com

(115) Janssen, T.; Husson, S. J.; Lindemans, M.; Mertens, I.; Rademakers, S.; Ver Donck, K.; Geysen, J.; Jansen, G.; Schoofs, L. Functional Characterization of Three G Protein-Coupled Receptors for Pigment Dispersing Factors in *Caenorhabditis Elegans*. *J. Biol. Chem.* **2008**, *283* (22), 15241–15249.

(116) Li, C.; Kim, K. Family of FLP Peptides in *Caenorhabditis Elegans* and Related Nematodes. *Front. Endocrinol. (Lausanne).* **2014**, *5* (October), 1–16.

(117) Mertens, I.; Clinckspoor, I.; Janssen, T.; Nachman, R.; Schoofs, L. FMRFamide Related Peptide Ligands Activate the *Caenorhabditis Elegans* Orphan GPCR Y59H11AL.1. *Peptides* **2006**, *27* (6), 1291–1296.

(118) Bargmann, C. I. Neurobiology of the *Caenorhabditis Elegans* Genome. *Science* **1998**, *282* (5396), 2028–2033.

(119) Blaxter, M. *Caenorhabditis Elegans* Is a Nematode. *Science* **1998**, *282* (5396), 2041–2046.

(120) Gerstein, M. B.; Lu, Z. J.; Van Nostrand, E. L.; Cheng, C.; Arshinoff, B. I.; Liu, T.; Yip, K. Y.; Robilotto, R.; Rechtsteiner, A.; Ikegami, K.; Alves, P.; Chateigner, A.; Perry, M.; Morris, M.; Auerbach, R. K.; Feng, X.; Leng, J.; Vielle, A.; Niu, W.; Rhrissorrakrai, K.; Agarwal, A.; Alexander, R. P.; Barber, G.; Brdlik, C. M.; Brennan, J.; Brouillet, J. J.; Carr, A.; Cheung, M.-S.; Clawson, H.; Contrino, S.; Dannenberg, L. O.; Dernburg, A. F.; Desai, A.; Dick, L.; Dosé, A. C.; Du, J.; Egelhofer, T.; Ercan, S.; Euskirchen, G.; Ewing, B.; Feingold, E. A.; Gassmann, R.; Good, P. J.; Green, P.; Gullier, F.; Gutwein, M.; Guyer, M. S.; Habegger, L.; Han, T.; Henikoff, J. G.; Henz, S. R.; Hinrichs, A.; Holster, H.; Hyman, T.; Iniguez, A. L.; Janette, J.; Jensen, M.; Kato, M.; Kent, W. J.; Kephart, E.; Khivansara, V.; Khurana, E.; Kim, J. K.; Kolasinska-Zwierz, P.; Lai, E. C.; Latorre, I.; Leahey, A.; Lewis, S.; Lloyd, P.; Lochovsky, L.; Lowdon, R. F.; Lubling, Y.; Lyne, R.; MacCoss, M.; Mackowiak, S. D.; Mangone, M.; McKay, S.; Mecenas, D.; Merrihew, G.; Miller, D. M.; Muroyama, A.; Murray, J. I.; Ooi, S.-L.; Pham, H.; Phippen, T.; Preston, E. A.; Rajewsky, N.; Rätsch, G.; Rosenbaum, H.; Rozowsky, J.; Rutherford, K.; Ruzanov, P.; Sarov, M.; Sasidharan, R.; Sboner, A.; Scheid, P.; Segal, E.; Shin, H.; Shou, C.; Slack, F. J.; Slightam, C.; Smith, R.; Spencer, W. C.; Stinson, E. O.; Taing, S.; Takasaki, T.; Vafeados, D.; Voronina, K.; Wang, G.; Washington, N. L.; Whittle, C. M.; Wu, B.; Yan, K.-K.; Zeller, G.; Zha, Z.; Zhong, M.; Zhou, X.; Ahringer, J.; Strome, S.; Gunsalus, K. C.; Micklem, G.; Liu, X. S.; Reinke, V.; Kim, S. K.; Hillier, L. W.; Henikoff, S.; Piano, F.; Snyder, M.; Stein, L.; Lieb, J. D.; Waterston, R. H. Integrative Analysis of the *Caenorhabditis Elegans* Genome by the modENCODE Project. *Science* **2010**, *330* (6012), 1775–1787.

(121) Nathoo, A. N.; Moeller, R. A.; Westlund, B. A.; Hart, A. C. Identification of Neuropeptide-like Protein Gene Families in *Caenorhabditis elegans* and Other Species. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (24), 14000–14005.

(122) Elphick, M. R.; Emson, R. H.; Thorndyke, M. C. FMRFamide-like Immunoreactivity in the Nervous System of the Starfish *Asterias Rubens*. *Biol. Bull.* **1989**, *177* (1), 141–145.

(123) Arenas, N. E.; Salazar, L. M.; Soto, C. Y.; Vizcaíno, C.; Patarroyo, M. E.; Patarroyo, M. a; Gómez, A. Molecular Modeling and *in Silico* Characterization of *Mycobacterium Tuberculosis* TlyA: Possible Misannotation of This *Tubercle Bacilli-Hemolysin*. *BMC Struct. Biol.* **2011**, *11* (1), 16.

(124) Mugumbate, G.; Jackson, G. E.; Van Der Spoel, D.; Kövér, K. E.; Szilágyi, L. Anopheles Gambiae, Anoga-HrTH Hormone, Free and Bound Structure-A Nuclear Magnetic Resonance Experiment. *Peptides* **2013**, *41*, 94–100.

(125) Vassilatis, D. K.; Hohmann, J. G.; Zeng, H.; Li, F.; Ranchalis, J. E.; Mortrud, M. T.; Brown, A.; Rodriguez, S. S.; Weller, J. R.; Wright, A. C.; Bergmann, J. E.; Gaitanaris, G. A. The G Protein-Coupled Receptor Repertoires of Human and Mouse. *Proc Natl Acad Sci U S A* **2003**, *100* (8), 4903–4908.

(126) Rognan, D. Chemogenomic Approaches to Rational Drug Design. *Br. J. Pharmacol.* **2007**, *152* (1), 38–52.

(127) Kalyaanamoorthy, S.; Chen, Y. P. P. Modelling and Enhanced Molecular Dynamics to Steer Structure-Based Drug Discovery. *Prog. Biophys. Mol. Biol.* **2014**, *114* (3), 123–136.

(128) Ekins, S.; Mestres, J.; Testa, B. *In Silico* Pharmacology for Drug Discovery: Methods for Virtual Ligand Screening and Profiling. *Br. J. Pharmacol.* **2007**, *152* (1), 9–20.

(129) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. II. Are Homology Models of G-Protein Coupled Receptors Suitable Targets? *Proteins Struct. Funct. Genet.* **2003**, *50* (1), 5–25.

(130) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48* (17), 5448–5465.

(131) Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J.

Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications. *J. Chem. Inf. Model.* **2009**, *49* (10), 2333–2343.

(132) Kalyaanamoorthy, S.; Chen, Y. P. P. Structure-Based Drug Design to Augment Hit Discovery. *Drug Discov. Today* **2011**, *16* (17-18), 831–839.

(133) Zhou, H.; Skolnick, J. FINDSITEX: A Structure-Based, Small Molecule Virtual Screening Approach with Application to All Identified Human GPCRs. *Mol. Pharm.* **2012**, *9* (6), 1775–1784.

(134) Moroy, G.; Martiny, V. Y.; Vayer, P.; Villoutreix, B. O.; Miteva, M. A. Toward *in Silico* Structure-Based ADMET Prediction in Drug Discovery. *Drug Discov. Today* **2012**, *17* (1-2), 44–55.

(135) Schafferhans, A.; Klebe, G. Docking Ligands onto Binding Site Representations Derived from Proteins Built by Homology Modelling. *J. Mol. Biol.* **2001**, *307* (1), 407–427.

(136) Huang, H. J.; Yu, H. W.; Chen, C. Y.; Hsu, C. H.; Chen, H. Y.; Lee, K. J.; Tsai, F. J.; Chen, C. Y. C. Current Developments of Computer-Aided Drug Design. *J. Taiwan Inst. Chem. Eng.* **2010**, *41* (6), 623–635.

(137) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. a F.; Klomp, J. P. G.; De Vlieg, J.; Del Rio, A. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, *52* (6), 1607–1620.

(138) Fidom, K.; Isberg, V.; Hauser, A. S.; Mordalski, S.; Lehto, T.; Bojarski, A. J.; Gloriam, D. E. A New Crystal Structure Fragment-Based Pharmacophore Method for G Protein-Coupled Receptors. *Methods* **2015**, *71*, 104–112.

(139) Fernández-Ballester, G.; Fernández-Carvajal, A.; González-Ros, J. M.; Ferrer-Montiel, A. Ionic Channels as Targets for Drug Design: A Review on Computational Methods. *Pharmaceutics*, 2011, *3*, 932–953.

(140) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins Struct. Funct. Genet.* **2002**, *47* (4), 409–443.

(141) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev.*

Drug Discov. **2004**, *3* (11), 935–949.

(142) Carlsson, J.; Ranganathan, A.; Irwin, J. J. Exercise 3 : Molecular Docking to (All) Crystal Structures of G Protein-Coupled Receptors - Does Molecular Docking Work ? **2012**, 2–4.

(143) Bharatham, N.; Bharatham, K.; Shelat, A. A.; Bashford, D. Ligand Binding Mode Prediction by Docking: Mdm2/Mdmx Inhibitors as a Case Study. *J. Chem. Inf. Model.* **2014**, *54* (2), 648–659.

(144) Barillari, C.; Marcou, G.; Rognan, D. Hot-Spots-Guided Receptor-Based Pharmacophores (HS-Pharm): A Knowledge-Based Approach to Identify Ligand-Anchoring Atoms in Protein Cavities and Prioritize Structure-Based Pharmacophores. *J. Chem. Inf. Model.* **2008**, *48* (7), 1396–1410.

(145) Pausch, M. H. G-Protein-Coupled Receptors in Saccharomyces Cerevisiae: High-Throughput Screening Assays for Drug Discovery. *Trends Biotechnol.* **1997**, *15* (12), 487–494.

(146) Make, N.; Sequence, M.; Hsqc, S.; Mode, P. Assigning Proteins Using Sparky.

(147) Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discov. Today Technol.* **2010**, *7* (4).

(148) Wolber, G. 3D Pharmacophore Elucidation and Virtual Screening. *Drug Discov. Today Technol.* **2010**, *7* (4), e203–e204.

(149) Fan, H.; Irwin, J. J.; Sali, A. Virtual Ligand Screening Against Comparative Protein Structure Models. 1–23.**2009**

(150) Lindstrom, W.; Morris, G. M.; Weber, C.; Huey, R. Using AutoDock 4 for Virtual Screening. *Screening* **2008**, No. January, 1–37.

(151) Mugumbate, G.; Newton, A. S.; Rosenthal, P. J.; Gut, J.; Moreira, R.; Chibale, K.; Guedes, R. C. Novel Anti-Plasmodial Hits Identified by Virtual Screening of the ZINC Database. *J. Comput. Aided. Mol. Des.* **2013**, *27* (10), 859–871.

(152) Vass, M.; Schmidt, É.; Horti, F.; Keseru, G. M. Virtual Fragment Screening on GPCRs: A Case Study on Dopamine D3 and Histamine H4 Receptors. *Eur. J. Med. Chem.* **2014**, *77*, 38–46.

(153) Chen, Z.; Li, H.; Zhang, Q.; Bao, X.; Yu, K.; Luo, X.; Zhu, W.; Jiang, H. Pharmacophore-Based Virtual Screening versus Docking-Based Virtual Screening: A Benchmark Comparison against Eight Targets. *Acta Pharmacol. Sin.* **2009**, *30* (12), 1694–1708.

(154) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; David, S.; Olson, A. J.; Farmaceutica, C.; Napoli, S. De. Virtual Screening with AutoDock : Theory and Practice. **2011**, *5* (6), 597–607.

(155) Studi, D.; Milano, D. I.; Facolt, B.; Matematiche, S. Development of Methodologies for Molecular Docking and Their Applications. **2009**.

(156) Dobi, K.; Hajdú, I.; Flachner, B.; Fabó, G.; Szaszkó, M.; Bognár, M.; Magyar, C.; Simon, I.; Szisz, D.; Lõrincz, Z.; Cseh, S.; Dormán, G. Combination of 2D/3D Ligand-Based Similarity Search in Rapid Virtual Screening from Multimillion Compound Repositories. Selection and Biological Evaluation of Potential PDE4 and PDE5 Inhibitors. *Molecules* **2014**, *19* (6), 7008–7039.

(157) Ma, D. L.; Chan, D. S. H.; Leung, C. H. Molecular Docking for Virtual Screening of Natural Product Databases. *Chemical Science*, 2011, *2*, 1656.

(158) Pierri, C. L.; Parisi, G.; Porcelli, V. Computational Approaches for Protein Function Prediction: A Combined Strategy from Multiple Sequence Alignment to Molecular Docking-Based Virtual Screening. *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804* (9), 1695–1712.

(159) Isberg, V.; de Graaf, C.; Bortolato, A.; Cherezov, V.; Katritch, V.; Marshall, F. H.; Mordalski, S.; Pin, J. P.; Stevens, R. C.; Vriend, G.; Gloriam, D. E. Generic GPCR Residue Numbers – Aligning Topology Maps While Minding the Gaps. *Trends Pharmacol. Sci.* **2015**, *36* (1), 22–31.

(160) Latek, D.; Pasznik, P.; Carlomagno, T.; Filipek, S. Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison. *PLoS One* **2013**, *8* (2), 1–10.

(161) Roumen, L.; Sanders, M. P. A.; Vroling, B.; de Esch, I. J. P.; de Vlieg, J.; Leurs, R.; Klomp, J. P. G.; Nabuurs, S. B.; de Graaf, C. *In Silico* Veritas: The Pitfalls and Challenges of Predicting GPCR-Ligand Interactions. *Pharmaceuticals* **2011**, *4* (9), 1196–1215.

(162) Heilker, R.; Wolff, M.; Tautermann, C. S.; Bieler, M. G-Protein-Coupled Receptor-Focused Drug Discovery Using a Target Class Platform Approach. *Drug Discov. Today* **2009**, *14* (5-6), 231–240.

(163) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; Mcgettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.

(164) Jung, J.; Lee, B. Use of Residue Pairs in Protein Sequence-Sequence and Sequence-Structure Alignments. *Protein Sci.* **2000**, *9* (8), 1576–1588.

(165) Werner, T.; Morris, M. B.; Dastmalchi, S.; Church, W. B. Structural Modelling and Dynamics of Proteins for Insights into Drug Interactions. *Adv. Drug Deliv. Rev.* **2012**, *64* (4), 323–343.

(166) Eilers, M.; Hornak, V.; Smith, S. O.; Konopka, J. B. Comparison of Class A and D G Protein-Coupled Receptors: Common Features in Structure and Activation. *Biochemistry* **2005**, *44* (25), 8959–8975.

(167) Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G. W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; Zhang, W.; Xie, X.; Yang, H.; Jiang, H.; Cherezov, V.; Liu, H.; Stevens, R. C.; Zhao, Q.; Wu, B. Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* **2013**, *341* (6152), 1387–1390.

(168) Hu, L. A.; Chen, W.; Martin, N. P.; Whalen, E. J.; Premont, R. T.; Lefkowitz, R. J. GIPC Interacts with the beta1-Adrenergic Receptor and Regulates beta1-Adrenergic Receptor-Mediated ERK Activation. *J. Biol. Chem.* **2003**, *278* (28), 26295–26301.

(169) Fuchs, P. F. J.; Alix, A. J. P. High Accuracy Prediction of β-Turns and Their Types Using Propensities and Multiple Alignments. *Proteins Struct. Funct. Genet.* **2005**, *59* (4), 828–839.

(170) Schneider, G.; Baringhaus, K. H. *Molecular Design. Concepts and Applications*; **2008**; Vol. 3.

(171) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. Modeling Activated States of GPCRs: The Rhodopsin Template. *J. Comput. Aided. Mol. Des.* **2006**,

20 (7-8), 437–448.

(172) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure of a Human A_2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (5905), 1211–1217.

(173) Hooft, R. W.; Sander, C.; Vriend, G. Objectively Judging the Quality of a Protein Structure from a Ramachandran Plot. *Comput. Appl. Biosci.* **1997**, *13* (4), 425–430.

(174) Friesner, R. A.; Abel, R.; Goldfeld, D. a.; Miller, E. B.; Murrett, C. S. Computational Methods for High Resolution Prediction and Refinement of Protein Structures. *Curr. Opin. Struct. Biol.* **2013**, *23* (2), 177–184.

(175) De, A.; Veulens, N.; Rodríguez, R. G Protein-Coupled Receptors as Targets for Drug Design. *Biotecnol. Apl.* **2009**, *26* (1), 24–33.

(176) Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M. High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5* (10).

(177) Systems, I. An Introduction into “ Docking ” and “ Molecular Dynamics Simulations ” Amino Acids. **2012**.

(178) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9* (1), 71.

(179) Lakshmi, B.; Ramakrishnan, C.; Archunan, G.; Sowdhamini, R.; Srinivasan, N. Investigations of Ramachandran Disallowed Conformations in Protein Domain Families. *Int. J. Biol. Macromol.* **2014**, *63*, 119–125.

(180) Arenas, N. E.; Salazar, L. M.; Soto, C. Y.; Vizcaíno, C.; Patarroyo, M. E.; Patarroyo, M. A.; Gómez, A. Molecular Modeling and in Silico Characterization of Mycobacterium Tuberculosis TlyA: Possible Misannotation of This Tubercle Bacilli-Hemolysin. *BMC Struct. Biol.* **2011**, *11*, 16.

(181) Schneider, M.; Wolf, S.; Schlitter, J.; Gerwert, K. The Structure of Active

Opsin as a Basis for Identification of GPCR Agonists by Dynamic Homology Modelling and Virtual Screening Assays. *FEBS Lett.* **2011**, *585* (22), 3587–3592.

(182) Wang, Y. T.; Chan, C. H.; Su, Z. Y.; Chen, C. L. Homology Modeling, Docking, and Molecular Dynamics Reveal HR1039 as a Potent Inhibitor of 2009 A (H1N1) Influenza Neuraminidase. *Biophys. Chem.* **2010**, *147* (1-2), 74–80.

(183) Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. *J. Phys. Chem. B* **2013**, *117* (28), 8408–8421.

(184) White, J. A.; Román, F. L.; González, A.; Velasco, S. Periodic Boundary Conditions and the Correct Molecular-Dynamics Ensemble. *Phys. A Stat. Mech. its Appl.* **2008**, *387* (27), 6705–6711.

(185) Taylor, J. S.; Burnett, R. M. DARWIN: A Program for Docking Flexible Molecules. *Proteins Struct. Funct. Genet.* **2000**, *41* (2), 173–191.

(186) Schmidt, T. H.; Kandt, C. LAMBADA and InflateGRO2: Efficient Membrane Alignment and Insertion of Membrane Proteins for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2012**, *52* (10), 2657–2669.

(187) Janosi, L.; Gorfe, A. A. Simulating POPC and POPC/POPG Bilayers: Conserved Packing and Altered Surface Reactivity. *J. Chem. Theory Comput.* **2010**, *6* (10), 3267–3273.

(188) Schneider, E. H.; Schnell, D.; Strasser, A.; Dove, S.; Seifert, R. Impact of the DRY Motif and the Missing “Ionic Lock” on Constitutive Activity and G-Protein Coupling of the Human Histamine H4 Receptor. *J. Pharmacol. Exp. Ther.* **2010**, *333* (2), 382–392.

(189) Kalyaanamoorthy, S.; Chen, Y. P. P. Modelling and Enhanced Molecular Dynamics to Steer Structure-Based Drug Discovery. *Prog. Biophys. Mol. Biol.* **2014**, *114* (3), 123–136.

(190) Fleishman, S. J.; Ben-Tal, N. Progress in Structure Prediction of α-Helical Membrane Proteins. *Curr. Opin. Struct. Biol.* **2006**, *16* (4), 496–504.

(191) Yang, S. Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discov. Today* **2010**, *15* (11-12), 444–450.

(192) Hu, Y.; Bajorath, J. Growth of Ligand-Target Interaction Data in ChEMBL Is Associated with Increasing and Activity Measurement-Dependent Compound Promiscuity. *J. Chem. Inf. Model.* **2012**, *52* (10), 2550–2558.

(193) Niu, Y.; Ma, C.; Jin, H.; Xu, F.; Gao, H.; Liu, P.; Li, Y.; Wang, C.; Yang, G.; Xu, P. The Discovery of Novel β-Secretase Inhibitors: Pharmacophore Modeling, Virtual Screening, and Docking Studies. *Chem. Biol. Drug Des.* **2012**, *79* (6), 972–980.

(194) Niu, M.; Dong, F.; Tang, S.; Fida, G.; Qin, J.; Qiu, J.; Liu, K.; Gao, W.; Gu, Y. Pharmacophore Modeling and Virtual Screening for the Discovery of New Type 4 cAMP Phosphodiesterase (PDE4) Inhibitors. *PLoS One* **2013**, *8* (12), 1–15.

(195) Xiang, X.; Taufer, M. Molecular Docking Based on Shape Complementary Methods. *PowerPoint Presentation*, **2009**

(196) Smondyrev, A. M.; Berkowitz, M. L. Molecular Dynamics Simulation of DPPC Bilayer in DMSO. *Biophys. J.* **1999**, *76* (5), 2472–2478.

(197) Langelaan, D. N.; Bebbington, E. M.; Reddy, T.; Rainey, J. K. Structural Insight into G-Protein Coupled Receptor Binding by Apelin. *Biochemistry* **2009**, *48* (3), 537–548.

(198) Castellani, F.; Rossum, B. Van; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H.; Reynolds, W. F.; Enriquezi, R. G. NMR Spectroscopy. *Nature* **2002**, *420*, 98–102.

(199) Cross, T. A.; Opella, S. J. Solid-State NMR Structural Studies of Peptides and Proteins in Membranes. *Current Opinion in Structural Biology*, **1994**, *4*, 574–581.

2 Chapter 2: NMR and Molecular Modeling of Peptides

2.0 Summary

Work done by Kubiak and co-workers matched the FLP18R1 GPCR receptor to putative peptide ligands.[1] Experiments[1,2] carried out on *C.elegans* and *Ascaris suum (A. suum),* using Ca^{2+} mobilization assays in Fluorescence Imaging Plate Reader (FLIPR) format, have matched peptides to FLP18 receptor proteins at 5μM.[1,3] The strongest receptor activation signal in the Ca^{2+} assays was identified for peptides with a C-terminal -PGVLRF-NH_2 motif.[1] These were expressed functionally using FLP18R1a in Chinese Hamster Ovary (CHO) cells.[1] The primary sequence of *flp18-6, af3, af4* and *af20* peptides has been published and are shown in Table 2.1.[1]

Table 2.1 Published primary sequences of peptides[1]

Peptide	**Primary sequence**
af3	Ala-Val-Pro-Gly-Val-Leu-Arg-Phe-NH_2
af4	Gly-Asp-Val-Pro-Gly-Val-Leu-Arg-Phe-NH_2
af20	Gly-Met-Pro-Gly-Val-Leu-Arg-Phe-NH_2
flp18-6	Asp-Val-Pro-Gly-Val-Leu-Arg-Phe-NH_2

The FLPs are characterized by decreasing amino acid conservation moving from the C to N terminals with the motif Pro-Gly-Val-Leu-Arg-Phe-NH_2 common in all peptides.[2] Since there are no experimental 3D structures of the FMR-like peptides (FLP), Nuclear Magnetic Resonance (NMR) distance-restrained molecular dynamics was used to determine their solution conformational structure.[4,5] Structure determination using NMR techniques was performed using the Nuclear Overhauser Effect (nOe) technique,[6,7] which is a population-weighted average that depends on the inverse sixth power of the internuclear distance of the nuclei of the interacting atoms.[7] The conformational differences in these FLPs correlate with the bioactivity variations.[1,8] Molecular dynamics simulations were performed in vacuum, water and DPC-micelle solution[5] using GROMACS[9] version 4.6.3. The peptide structures selected from cluster analysis in this chapter will be used in docking calculations in Chapter 4 and virtual screening in Chapter 5.

The chemical shifts of each of the peptides were compared to random coil values[10,11] to ascertain regions that comprise substantial populations of secondary structure .

2.1 Nuclear Magnetic Resonance (NMR)

NMR spectroscopy has evolved to be one of the most potent analytical techniques in solution structures determination of biological molecules.[12–14] This is done by mapping carbon-hydrogen frameworks within molecules, among which are proteins and peptides, even in membrane bilayers.[12,13,15] The procedure for assignment of peptide and protein NMR spectra requires the use of experiments such as homonuclear Correlation Spectroscopy (COSY), Total Correlation Spectroscopy (TOCSY), and Nuclear Overhauser Effect Spectroscopy (NOESY).[11,16,17] One and two-dimensional spectra have been expended to determine the structures and dynamics of the molecules in solution and membrane. The phenomenon used includes chemical shifts, amide proton chemical exchange[18], coupling constants and amide proton chemical shift temperature dependence as well as nOe intensities.[18–20]

The diversity of the physical environments or conditions of the experiments allow the technique to be used with versatility in the study of biological molecules. These physical conditions may be concentration, solvent, solvent composition, pH and temperature at which the analysis is executed.[18,19] Thus to interpret chemical shifts precisely, calibration with reference to temperature and pH is required.[12] NMR experiments of neuropeptides that analyse amide protons are often investigated below pH ~ 6 in order to reduce the amide proton exchange rate.[8] Protons in distinctive conditions absorb at marginally different frequencies, hence are distinguishable by NMR.

TOCSY and the published primary sequence of the peptide are used to assign proton chemical shifts.[1,21] The correlation experiments, COSY and TOCSY, denote the information on the scalar connectivity used for identifying the individual amino acid spin system.[16,21] Proton resonances are usually determined by variances in chemical shifts. Chemical shifts are affected by the local environment of a nucleus and consequently vary according to the primary, secondary, and tertiary structure of the protein.[12] NMR studies have established correlation between the chemical shifts of

Hα protons and the secondary structure of a peptide.[4] Thus from the resonance peak assignments, the secondary structure of the peptide can be predicted.

The critical NMR phenomenon in protein structure determination is NOESY.[17] The nOe's obtained from NOESY experiments provide essential 'through space' or sequential walking connectivity information[12] via spin-lattice relaxation. NOESY correlates protons in close spatial proximity to each other by incoherent magnetization transfer through dipole-dipole cross relaxation.[6,17] Correlations between protons relies on the distance between the protons but generally, a signal is only detected if the proton distance is smaller than 5Å.[17,22] Superimposition of the NOESY spectrum onto TOCSY spectrum as it has been assigned, results in the sequential pattern of $d_{NN}(i, i+1)$ and $d_{\alpha N}(i, i+1)$ cross peaks being predicted.[6,16,17,23]The correlation of neighbouring peaks can be ascertained in NOESY experiments.

Equation 2.1 highlights that interproton distances from the 2D nOe cross peak intensities are estimated by the Isolated Spin Pair Approximation (ISPA)[7]:

$$r_{ij} = r_{ref} (a_{ref}/a_{ij})^{1/6} \quad \text{(Equation 2.1)}$$

Where r_{ij} is the interproton distance to be estimated and a_{ij} is the corresponding 2D nOe cross peak intensity. NMR restraints are calculated as interproton distance. The restraints are effectively used in conformational search of peptides during NMR distant-restrained molecular dynamics. Floating the chirality in preliminary distance geometry can be calculated to assign the crosspeaks.[22,24]

2.2 Experimental Methods

2.2.1 NMR Sample Preparation of the peptides

The published primary sequence of the neuropeptides[1] *flp18-6, af3, af4* and *af20* (Table 2.1)**,** were used and the peptides were synthesized by GL Biochem Ltd (Shanghai) Ltd, China. The purity of the white powder was checked by HPLC and NMR and found to be >98%. NMR samples were run either in aqueous or DPC micelle solution. Typically ~ 1 mg of peptide was dissolved in either 0.5 ml of 90:10 (v/v) H_2O: D_2O or 150 mM deuterated dodecylphosphocholine (DPC-d_{38}) solution, buffered at pH 4.5 by 20 mM sodium phosphate. In both cases 1% of 4,4-dimethyl-4-silapentane-1-sulphonic acid **(**DSS) was used as the chemical shift reference.

2.2.2 NMR Experiments

^{1}H NMR experiments of the peptides were performed on a Bruker Avance Systems 700 MHz (Department of Chemistry, University of Washington, Seattle, USA) at 280K or a Bruker Advance 600 MHz NMR spectrometer equipped with a PRODIGY cryoprobe (Department of Chemistry, University of Cape Town, South Africa) and Bruker Avance II 500 MHz (Department of Chemistry, University of Debrecen, Hungary). The two dimensional NMR spectra of the peptides in water and DPC were collected in the phase sensitive mode from TOCSY (mixing time, 80 ms), NOESY (mixing time, 150 ms) and ROESY (mixing time, 150 ms)[17,25,26] using the dipsi2esgpph for TOCSY and noesyesgpph for NOESY spectra. These pulse sequences use excitation sculpting to suppress the water resonance. The data were processed with NMRpipe and spectra assigned using Sparky.[27] Spectral assignments were based on the method of Wüthrich[25] and ISPA[7] was used to calculate the interproton distances.

2.2.3 Restrained MD simulations

PyMOL[28] was used to build the starting structures for the conformational search and energy minimization was performed for 3000 steps using the l-BFGS integrator in GROMACS 4.6.3. Distance restrained molecular dynamics simulations[29] were performed for 10 ns in vacuum and water using GROMACS 4.6.3 in a cubic box with dimensions 70 x 70 x 70 Å. The OPLS-AA/L all-atom force field was used in molecular simulations using a time step of 2 fs, at constant temperature, volume and number of particles (NVT). A cut-off of 12 Å was used for van der Waals interactions and electrostatic interactions calculated by the PME summation for real space calculations using the simple leapfrog integrator.

2.2.3.1 In vacuum

The energy minimised starting structure was inserted into a simulation box and energy minimised for 3000 steps using a l-BFGS minimiser.[5] The calculated NMR interproton distances were applied as distance restraints between specific protons. The listed distances were categorized as weak, medium and strong increasing the values by ~ 10% accordingly. The peptide was subjected to gradual heating from 0 to 600 K over 10 ns to overcome energy barriers and reach a global minimum.[5] Periodic simulated annealing from 600 to 300 K was performed for 4 ns. Neighbour lists were

updated every fifth integration time step under periodic boundary conditions.[30] Angular centre of mass removal was performed.[30] Thus, simulated annealing[5,31] was performed over 7000000 steps, collecting 100 structures. The resultant conformations were energy minimised and cluster analysis was performed on all trajectories using a cut-off distance of <0.25 nm based on the RMSD of backbone atoms. The best structure with the lowest energy conformation was identified and collected.

2.2.3.2 *In water*

The lowest energy conformation from the vacuum molecular dynamics was used as the starting structure for 10 ns MD simulations in water. The peptide was position-restrained and Simple Point Charge (SPC) water added to soak the peptide fully. All simulations were performed with periodic boundary conditions.[30] Treatment of Lennard-Jones[32] and Coulombic non-bonded interaction terms was performed with a cut-off distance of 12 Å. Neighbour lists were updated every fifth integration time step under periodic boundary conditions. Angular centre of mass removal was performed. V-rescale temperature coupling method[33] of the peptide and solvent separately at 300K was achieved at a coupling constant, tau (τ), of 1 x 10^{-4} ns[5,9,29]. The constraint used for covalent bonds involving the hydrogen atom was the LINCS algorithm. NMR-distance restrained MD simulation was performed for 10 ns and 100 structures were collected. Cluster analysis of the trajectories of the MD simulation structures was done using a cut-off value of <0.25 nm by superimposing the backbone atoms. A minimum energy conformation was selected from the largest cluster to give gave structure of the peptide.

2.2.3.3 *In DPC*

2.2.3.3.1 *Energy minimization and Equilibration*

All MD simulations were carried out using GROMACS version 4.6.3. Cluster analysis of MD simulations in water yielded a lowest energy conformation that was taken as a starting structure in DPC micelle simulation for 10 ns. The initial configuration was energy minimized using steepest gradient for 1000 steps and embedded in a cubic DPC water box with dimensions measuring 70 x 70 x 70 Å. The system was energy minimized for 20000 steps. The first 10000 steps used the steepest

descent method and the last 10000 used the conjugate gradient method with no restraints. The box with 46 DPC molecules was solvated with 10282 water molecules. The solvated system was heated to 300 K in 200 ps in the NVT ensemble[30] with DPC position restraints using the MD leapfrog integrator with a set time step of 2 fs. Temperature control was done using velocity rescaling and Periodic Boundary Conditions (PBC) were used. The system was then subjected to NPT dynamics at constant temperature, T = 300 K for 1 ns and Nose-Hoover temperature coupling. Pressure coupling was performed with a Parrinello-Rahman barostat with linear centre of mass removal.

2.2.3.3.2 Production Run

NMR distance restraints with a force constant of 1 000 kJ mol^{-1} K^{-1} were applied for the production run for 10 ns with a time step of 2 fs using the MD leapfrog integrator. A cut-off of 12 Å was applied for short-range van der Waals interactions. Particle Mesh Ewald (PME) summation method was used to calculate long-range electrostatics with grid spacing of 12 Å and short-range Coulombic electrostatics were performed with a cutoff of 12 Å. LINCS constraints was applied to the bonds. The peptide was coupled to the DPC micelle while the solvent was not coupled. V-rescale thermostat was implemented for temperature coupling at 300 K and a coupling constant, τT, of 1 x 10^{-4} ns. The pressure coupling was not applied. Neighbour lists were updated every fifth integration time step under periodic boundary conditions. Angular centre of mass removal was performed. The MD simulations were performed on the Dell C6145 series600 UCT High Performance Cluster with 64 nodes and 4 CPUs per core and 128GB RAM.

2.3 Results and Discussion

2.3.1 flp18-6 Peptide

2.3.1.1 NMR Spectral assignment and interproton distances

Sparky[34] was successfully used for the ^{1}H chemical shifts assignments of *flp18-6* in water and the results are given in Table 2.2. Chemical shift tables aided peak assignment and primary sequence of the peptides (Table 2.1) was referred to when

identifying individual amino acid spin system from TOCSY spectra.

Table 2.2 ^{1}H chemical shift (ppm) assignments for *flp18-6*, DVPGVLRF-NH2 in H2O-D2O 9:1 pH ca 4.5 (Phosphate buffer), T=290K

#	Res	HN	Hα	Hββ'	Hγγ'	Hδδ'	Hε	Hζ	Hη
1	Asp		4.290	2.658, 2.658					
2	Val	8.723	4.482	2.113	1.002, 0.941				
3	Pro		4.403	2.332	2.043, 1.922	3.893, 3.706			
4	Gly	8.566	3.939						
5	Val	7.991	4.077	2.058	0.914				
6	Leu	8.413	4.329	1.601, 1.480		0.861, 0.929			
7	Arg	8.331	4.245	1.671	1.486, 1.435	3.123	7.204		7.163, 7.619
8	Phe	8.310	4.613	3.715, 2.993					

The Isolated Spin Pair Approximation (ISPA)[5,7] method was used to calculate interproton distances from the observed nOe crosspeaks. Corrections for the pseudo atom were done according to Wüthrich (1986)[25] (Appendices 2). ISPA is considered to be only an approximate model of nOe cross peak volumes and does not encompass all mechanisms that have a bearing on relaxation rates.[7,35] Furthermore, ISPA underestimates interproton distances. The mixing times of 150 ms for both the

NOESY and ROESY[6,35] are short and thus the ISPA is considered as the suitable method with only a small effect from spin diffusion.[5] The interproton distances were classified as strong, medium and weak.[36] It is common practice in MD simulations to restrain the distance to a range of intervals rather than to a unique value. The practice minimizes systematic bias because of theoretically inaccurate calculated distance restraints. The geminal[6,35] proton-proton distance Pro3 Hβ1-Hβ2 internuclear distance, 0.178 nm (1.78 Å), was selected as the reference for the *flp18-6*.

2.3.2 NMR Experiments of flp18-6 peptide

2.3.2.1 Peptide flexibility and structural prediction

The Random Coil Index (RCI) server[37] tool available on http://wishart.biology.ualberta.ca, was used to determine the NH vector order parameters, S^2, of *flp18-6*. The RCI calculations group the chemical shift data acquired from six different nuclei (^{1}Hα, ^{1}HN, ^{15}N, ^{13}Cα, ^{13}Cß and ^{13}CO or any permutations thereof) into a single parameter that relates the amplitudes of backbone protein motions. This includes the order parameter (S^2) and root mean square fluctuations (RMSFs) of protein structural ensembles.

Figure 2.1 highlights the random coil deviations and the flexibility of the peptide as calculated using the RCI tool[37]. Differences in chemical shifts of the random coil values in literature as deduced by Wishart and Berjanskii.[11,37,38]

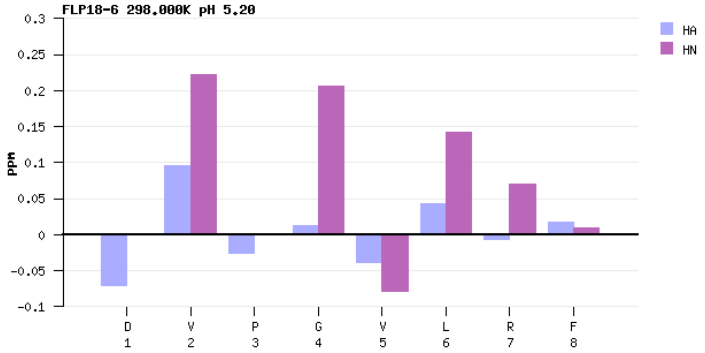


Figure 2.1: Experimental Random Coil deviation of the *flp18-6* peptide at pH 5.20

The RCI values were calculated from the corresponding chemical shifts to predict the flexibility of the peptide. The slightly negative random coil deviations for *flp18-6* and the marginal positive deviations are indicative of the decreased mobility of the peptide. The order parameters shown in Figure 2.2 indicate that the variations of the

peptide are smallest at the N-terminus and are larger at the C-terminus end of the peptide than at the centre. The order parameters of *flp18-6* are between 0.4 and 0.8, which reflects that the peptide is less well-structured and a freely fluctuating peptide chain,[39] especially Asp1 to Val5. The experimental order parameters from leucine Leu6 to Phe8 are close to 0.8. It therefore explains the rigid conformation in this part of the peptide and thus is more structured due to the presence of a salt bridge between Arg7 and Asp1. The other parameter is low for Asp1 as the H_α and H_N protons for Asp1 lie beyond the salt bridge at the free N-terminus (this will be clear in the structure presented later – Figure 2.5)

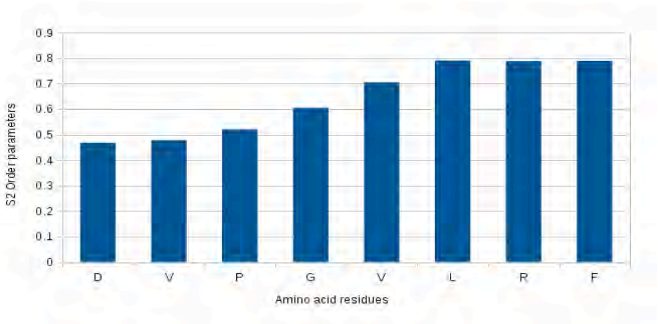


Figure 2.2: ^{1}H-$^{13}C\alpha$ Model-free Squared Generalized Order parameters, S^2, of the *flp18-6* calculated from the RCI server

The *flp18-6* peptide does not have a rigid conformation and it is flexible mostly in the N-terminus. The Hα chemical shifts (Table 2.1) from both NMR (NMR_RMSD) and molecular dynamics simulations (MD_RMSD) are shown in Figure 2.3, highlighting that the theoretical trend is consistent with the experimental observations.

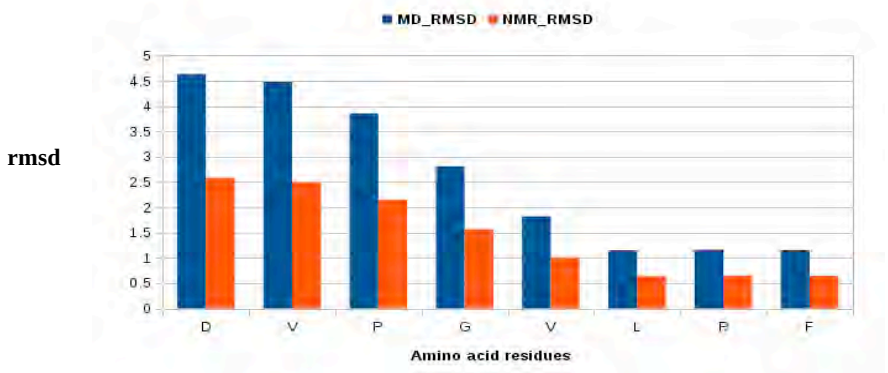


Figure 2.3: Correlation between MD_RMSD and NMR_RMSD in *flp18-6*, calculated using chemical shifts on the RCI server[37]

The large MD_RMSD from Asp1 to Val5 of (1.8 - 4.6) corresponds to the

unstructured conformation that was explained in relation to the S^2 order parameters in Figure 2.2. Interestingly, the large deviation and fluctuation help to validate the claim that the N-terminus of the peptide is highly flexible and has an unstructured conformation. The C-terminus has a significantly lower MD_RMSD and fluctuations confirming the rigidity of this part of the peptide.

2.3.2.2 Temperature dependence on the Amide Chemical Shifts

The amide proton (H^N) chemical shifts were plotted against the temperature (Figure 2.4) to determine their temperature coefficients. The calculated temperature coefficient indicates probability of the formation of hydrogen bond by amide protons.

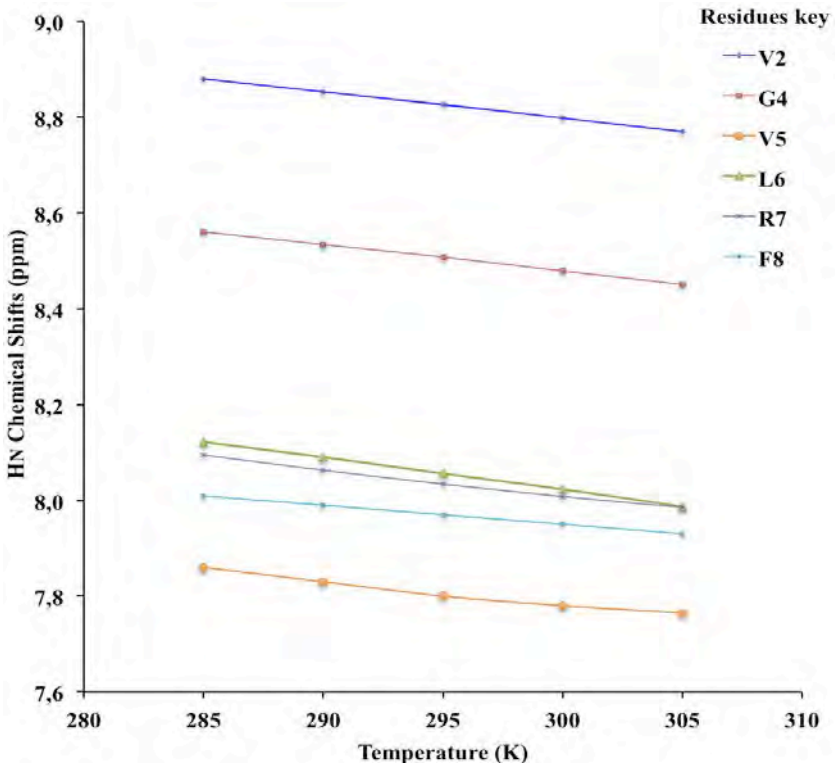


Figure 2.4: The linear plot of amide proton chemical shifts as a function of the temperature of the *flp18-6* peptide in DPC at pH 4.5.

The plot of H^N chemical shifts against temperature was linear as shown in Figure 2.4. The average correlation coefficient from the plot is 0.99. The dependence of chemical shifts on the temperature may affect the structure as hydrogen bonds will weaken with temperature in a protein.[10,19] The peptide is flexible and the amplitude of protein motion increases with temperature.[11,37] This means that as temperature increases, the H^N signal shifts upfield. This is normally referred to as negative temperature coefficient.[19] This trend affects the hydrogen-bonded amide group, which has a

peptide bond resulting in the amide proton in the peptide to be shifted downfield.[40] The H^N temperature coefficient (Table 2.3) can be used to predict hydrogen bond donors.[10,19,40]

The measured H^N temperature coefficients of *flp18-6* in DPC, shown in Table 2.3, are between -6.75 to -3.9 ppb/K. Most of the residues in the peptide are hydrophobic. According to Baxter (1997),[19] temperature coefficients of H^N in water from -10 ppb/K to -6 ppb/K reveal the presence of transient hydrogen bonds.[19] However, the *flp18-6* peptide under investigation in this study was found to be insoluble in water but reveals intramolecular hydrogen bonding in DPC. Therefore, the $^1H^N$ temperature coefficients in DPC reveal the presence of transient hydrogen bonds in the peptide. The hydrogen bond strength weakens with an increase in temperature and the amide protons slightly shift downfield (i.e. a relative upfield shift)[19,40] due to exposure to solvent. If the amide is H-bonded it exchanges at a slow rate with solvent and so will shift less. Also, if it is screened from solvent, by say a hydrophobic group, it will also shift less.

Table 2.3 Amide proton (H^N) chemical shift (δ) dependencies on Temperature in *flp18-6* peptide at pH 4.5

Residue ($^1H^N$)	**Δδ H^N (ppm)**	**Δδ $H^N/\Delta T$ *(ppb/K)***
Val2	- 0.110	-5.50
Pro3	-	-
Gly4	- 0.110	-5.50
Val5	- 0.095	-4.75
Leu6	- 0.135	-6.75
Arg7	- 0.109	-5.45
Phe8	- 0.079	-3.95

The observed temperature coefficients are slightly higher than those commonly deduced for peptides.[10,19,41] This can be attributed to the solvent inaccessibility of some of the residues, e.g. Leucine (Leu6), with a temperature coefficient ~ -6.75 ppb/K. The H^N temperature coefficient of the aromatic phenylalanine (Phe8) was

lowered to -3.95 ppb/K. The conformational preference of the Phe8 aromatic side chain may be affected by the orientation at the C-terminal. The aromatic ring tends to deshield amide protons downfield. Another possibility may be that the Phe8 aromatic ring, which is bulky, and the glycine (Gly4) amide protons are close to each other, resulting in an electrostatic interaction (H···π interaction) between the two residues. This interaction could be affected because of lack of steric hindrance in the glycine. In a study by Levitt and Perutz (1988), the weak interactions of residues with the π-electron cloud in the aromatic ring affect the orientation of the bulky aromatic side chains.[18]

The observations from the NMR distance restrained molecular dynamics simulations confirm that the *flp18-6* peptide possesses a cyclic conformation[5] (Figure 2.5). The Arg7 forms a salt bridge with the aspartic acid (Asp1) and thus there is a minimal solvent accessible area to the residues at the centre of the peptide with the temperature coefficients ranging from -4.75 to 6.75 ppb/K. The peptide structure conformation has an effect on the higher temperature coefficients. Therefore, the NMR analysis in aqueous solution supports the hypothesis that small peptides are structured in aqueous solutions.[7]

2.3.3 MD simulations of flp18-6

2.3.3.1 Conformational Search

A conformational search for the *flp18-6* peptide was performed using interproton distances and restraints calculated from nOe cross peak volumes in the NMR experiments. Distance restrained MD was performed for 14 ns in vacuum from 0 K to 600 K and simulated annealing to 300 K to search for low energy conformations of the peptide. The system was heated to 600 K over 10 ns so that system will attain a global energy optimum. The system was gradually cooled over 4 ns from 600 K to 300 K. 100 Conformations of the peptide structures were collected from the MD simulations. Cluster analysis of the structures was carried out at a cut-off distance of 0.25 nm, giving one cluster, and the structure with lowest energy conformation was extracted. This lowest energy conformation was used, as the starting structure during distance restrained MD in water at 300 K. Another 100 structures were collected and each of the structures was energy minimised using the low memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) integrator over 3000 steps. Cluster analysis of

the results provided only one cluster, which indicates that the conformation of the peptide is conserved in water (Figure 2.5a).

The lowest energy structure has a cyclic conformation (Figure 2.5b). The basic nitrogen in Arg7 and the oxygen of the acidic Asp1 form a salt bridge[42]. This salt bridge (Figure 2.5b) is responsible for the cyclic conformation of the peptide in vacuum, water and DPC micelle solution[5]. Salt bridges in proteins and peptides have been investigated and it has been concluded that they provide for the specificity of the conformations.[5,42–44] Also, Asp1 has an oxygen atom of its carboxyl moiety forming an attractive electrostatic interaction with the guanidinium ion of the basic Arg7

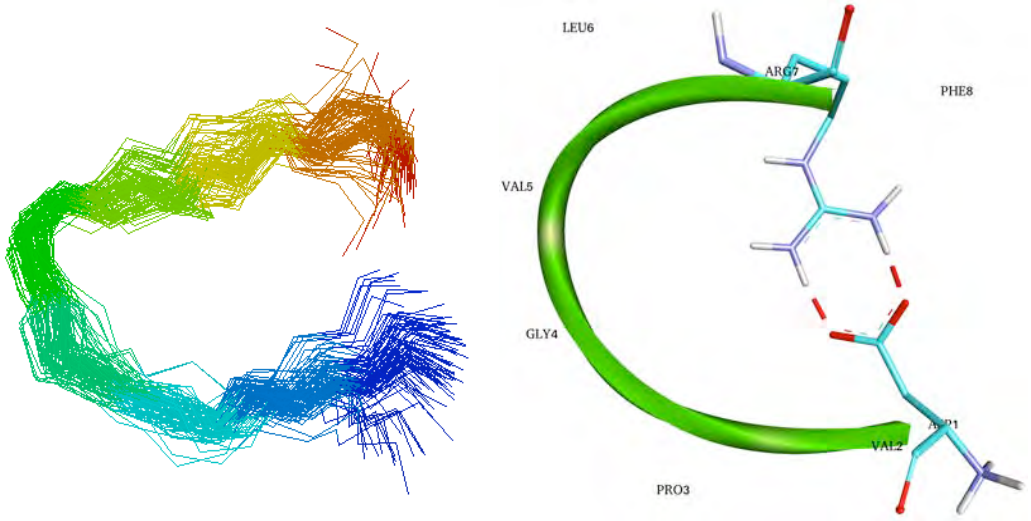


Figure 2.5: a) View of the backbone superimposition of 100 structures from molecular dynamic simulations of *flp18-6* in DPC at 300K. The top right represents the N-terminus of the structure. b) Interactions in the *flp18-6* indicating the salt bridge between the guanidinium ion of Arg7 and Asp1 carboxylate ion

The hydrogen atom on the Cδ of the proline (Pro3) pyrrolidine ring interacts with the carbonyl oxygen atoms through a 2.9Å hydrogen bond. The aromatic ring of the Phe8 behaves as a hydrogen acceptor as the approaching amide proton will polarize the π electrons. Studies undertaken by Worth and Wade, using Molecular Mechanics (MM) calculations, concluded that amide protons are predominantly parallel to aromatic rings in either protein environment or in aqueous solution.[45] In our study, this deduction supports the realization that there are interactions between the amide protons and the solvent as well as with the protein molecules. Phe8 forms an aromatic

π-alkyl interaction with Leu6, thus stabilizing the structure further.

In the *flp18-6* peptide, a type II beta (ß) turn exists in the non-helical central residues of the motif, Val-Pro-Gly-Val. The ß-turn is a secondary structure of these peptides, which effects change in the direction of the polypeptide chain. The ß-turn was predicted using the COUDES server[46] available on http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::COUDES. These observations are in agreement with the S^2 order parameters in Figure 2.2 and random coil values in Figure 2.3. Therefore, the conformation of the *flp18-6* from the molecular dynamics simulations should be close to the real structure. Figure 2.6 shows the variation of Cα RMSD over the 10 ns trajectory in water and DPC solution.

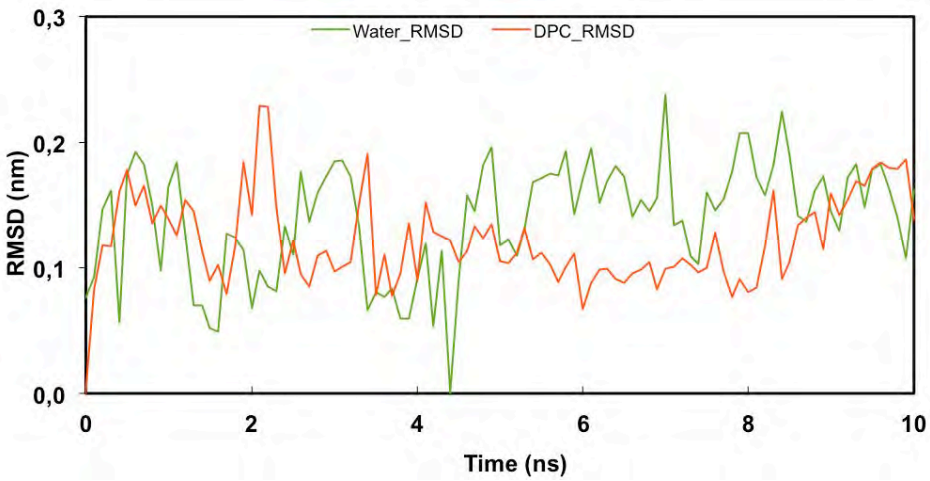


Figure 2.6: The variation of the Cα RMSD of *flp18-6* in water and DPC over 10 ns

The initial 0.3 ns show an exponential increase in RMSD from 0 to 0.2 nm in water and DPC. The average RMSD in DPC is 0.11 nm compared to the average in water of 0.15 nm. The simulation reaches a steady state after 4.4 ns in DPC. This stability is maintained up to 8.3 ns, where the RMSD increases again to 0.18 nm and converging with the water RMSD after 9.3 ns. Thus, the stable conformation of the peptide can be deduced as semi rigid, as given from the S^2 order parameters (Figure 2.2).

2.4 *af3* Peptide

2.4.1 NMR Spectral assignment and interproton distances

The protocol that was used for assigning NMR spectra and calculating interproton distances was highlighted in section 2.4.3. Table 2.4 outlines ^{1}H chemical shifts

assignments for *af3* in water. It has long been established that the local chemical environment of an atom has a highly sensitive effect on the chemical shift value of an atom.[20]

Table 2.4 ^{1}H chemical shift (ppm) assignments for *af3*, AVPGVLRF-NH_2 in H_2O-D_2O 9:1 pH ca 4.5 (Phosphate buffer), T=290K

#	Res	HN	Hα	Hββ'	Hγγ'	Hδδ'	Hε	Hζ	Hη
1	Ala		4.107						
2	Val	8.789	4.330	2.063	0.952, 0.994				
3	Pro		4.390	2.360, 2.019	1.859, 2.012	3.932, 3.613			
4	Gly	8.655	3.905						
5	Val	8.163	4.023	2.135	0.945				
6	Leu	8.169	4.248	1.624, 1.654	1.529	0.854, 1.070			
7	Arg	7.926	4.142	1.853, 1.657	1.311		3.037		
8	Phe	8.019	4.513	2.939, 3.205		7.578	7.181	7.301	

2.4.2 NMR Experiments

2.4.2.1 Temperature dependence on the HN amide proton chemical shifts

The H^N temperature coefficients were estimated from the plot of chemical shifts of the amide protons per residue of *af3* as a function of temperature (Figure 2.7). From

the work done by Cierpicki (2001), temperature coefficients can be measured from the plot and reveal the correlation between the temperature coefficients with hydrogen bonds.[18]

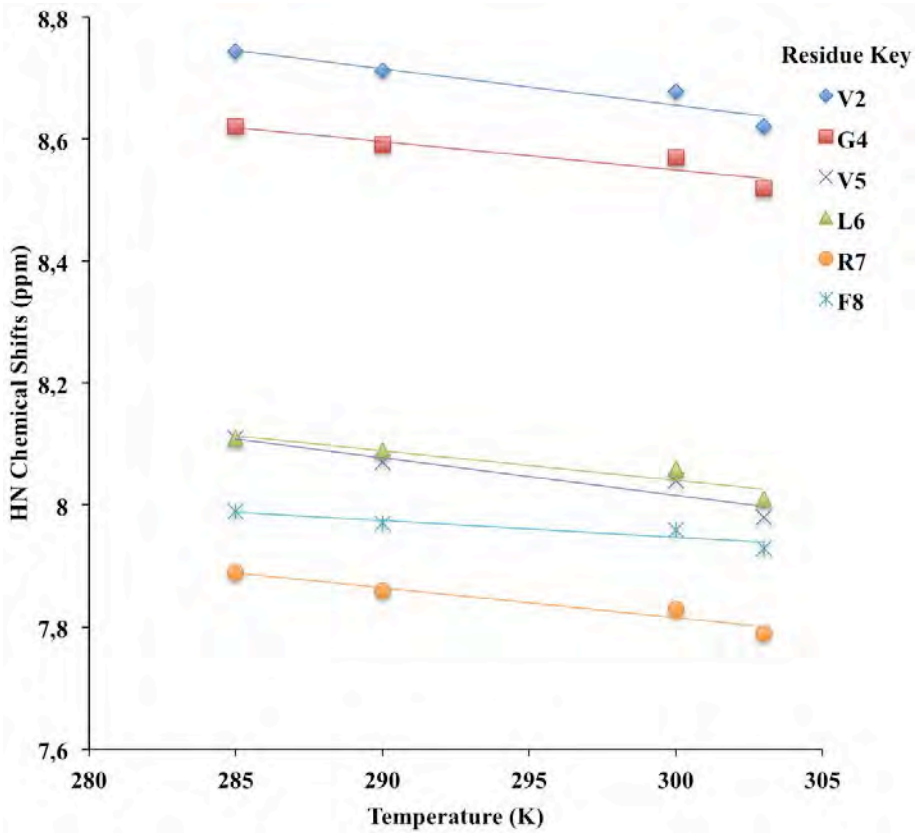


Figure 2.7: The plot of amide proton chemical shifts as a function of temperature of the *af3* peptide in DPC at pH 4.5.

As temperature increases, the amide proton signal shifts upfield (increment of $\Delta\delta$ H^N of an average of 0.1 ppm). For example, in Figure 2.7, the Arg7 amide proton shift at 285 K was observed as 7.91 ppm and at 303 K, the chemical shift was 7.79 ppm. Guanidinium protons of Arg7 indicate a semi linear dependence on temperature. The plot in Figure 2.7 show that the linearity is lost as temperature increases.[18,19] This is believed to be the effect of unfolding of the peptide at high temperatures thus weakening the strength of hydrogen bonds. The average correlation coefficient[41] of the plot gives a R = 0.89. The $^1H^N$ temperature coefficients of the *af3* in DPC solution at pH 4.5 help to predict intramolecular hydrogen bonding formation.[5,19] Table 2.5 gives the amide proton temperature coefficients[18] (-ppb/K) and the residue numbers in *af3*.

The data presented in Table 2.5 indicate that the H^N temperature coefficients have

longer magnitude than those anticipated for hydrogen bond formation. However, as was noted in the work by Langelaan (2009) and Mugumbate (2013), the GPCR ligand studies using exhaustive NMR revealed a beta turn structure and thus are stabilized by hydrogen bonding.[5,47] Thus as can be deduced from NMR studies on the *af3* peptide, which also display a beta turn, some residues in the peptide are masked from the solvent and show high H^N temperature coefficients.[19] The deshielding of the Val2 and Val5 amide protons may have caused the high magnitude temperature coefficient of -6.89 and -7.22 ppb/K. The aromatic ring in Phe8, with a temperature coefficient of -3.33 ppb/K, is thought to cause the downfield shift of the amide protons in Gly4 as earlier explained in Section 2.4.

Table 2.5 Amide proton (H^N) chemical shift (δ) dependencies on temperature in *af3* peptide at pH 4.5 in DPC

Residue ($^1H^N$)	**Δδ H^N (ppm)**	**Δδ $H^N/\Delta T$ *(ppb/K)***
Val2	- 0.12	-6.89
Pro3	-	-
Gly4	- 0.10	-5.56
Val5	- 0.13	-7.22
Leu6	- 0.10	-5.56
Arg7	- 0.10	-5.56
Phe8	- 0.06	-3.33

2.4.2.2 Peptide flexibility and structural prediction

The Hα and H^N chemical shifts for *af3* in water were compared to their random coil values, as shown in Figure 2.8. The RCI calculations group the chemical shift data acquired from six different nuclei (^{1}Hα, $^1H^N$, ^{15}N, ^{13}Cα, ^{13}Cß and ^{13}CO or any permutations thereof) into a single parameter that relates the amplitudes of backbone protein motions.[11,37] This includes the model-free S^2 order parameters and root mean square fluctuations (RMSFs) of protein structural ensembles.[11,37]

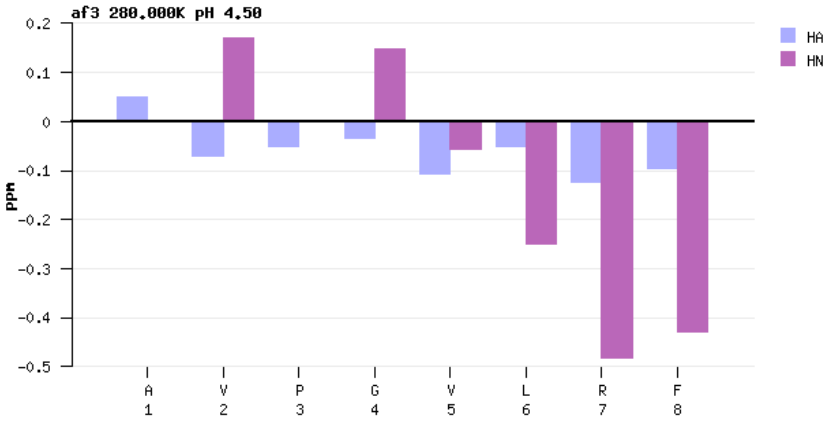


Figure 2.8: Backbone H^N and Hα chemical shift deviations calculated from random coil values for *af3* in water. Random coil chemical shifts were obtained from the RCI Server[37]

A secondary chemical shift is defined as the difference between the observed chemical shift of a given amino acid atom in a protein and the corresponding random coil value for the same atom.[48] The secondary chemical shifts have been employed to approximate the level of disorder in the protein structure and depict the probability of having a helical or β-sheet conformation.[18] This is shown by the negative RCI values in the secondary structure elements in Figure 2.8. It follows therefore that if the RCI values are positive, it means the conformation probably has β-sheet occurrence while negative values suggest a helical conformation.[48] The RCI values show that the secondary structure predicted has mostly a helical structure.

As indicated in Figure 2.8, the *af3* peptide has negative RCI values for the Hα of most amino acids except the alanine (Ala1). The assignment of Ala1 is not really completed as it is a terminal residue and thus not enough data was obtained in order to have chemical shifts. Thus, there is no information related the amide proton from Ala1. The H^N for Val2 and Gly4 are positive. The helical Hα protons and amide protons for the other residues from Val5 up to Phe8 have chemical shifts that are observed as less than their respective random coil values by nearly 0.3 ppm. β-sheets (positive RCI) indicate chemical shift values that are greater than their random coil values by about 0.3 ppm. The amide protons are thought to be involved in hydrogen bonding and this affects their chemical shifts, thus negating their use as a potential qualitative indicator for the presence of helix or β-structures.

Sequence and ring aromatic current effects[49] have been reported to affect the ^1H

chemical shifts.[50] The RCI values for Arg7 and Phe8 of the amide protons are greater than that of all the others. This trend may be attributed to the aromatic rings and sequence effects as well as the hydrogen bonding. The aromatic residue (Arg7) is involved in strong hydrogen bonding with Pro3 and this gives the cyclic conformation[5,51] of the *af3* shown in Figure 2.12 (Section 2.5.3). The order parameters, shown in Figure 2.9, indicate that the fluctuations of the peptide increase towards the centre and decrease towards the terminals of the peptide. The analysis of the RMSD of the peptide in water and DPC is shown in Figure 2.10.

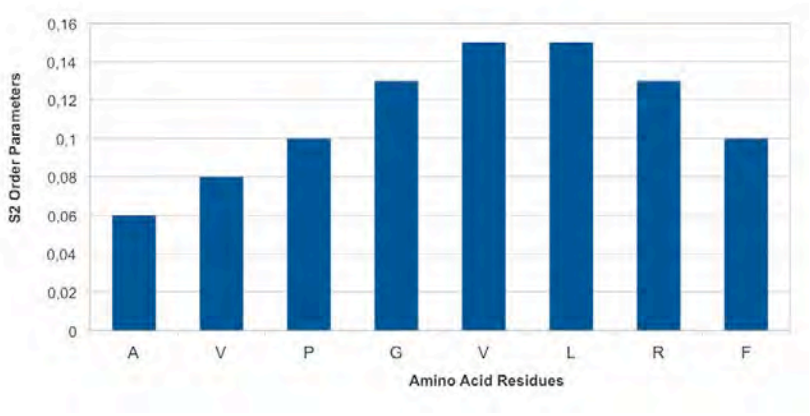


Figure 2.9: ^{1}H-$^{13}C\alpha$ Model-free Squared Generalized Order parameters, S^2, of the *af3* calculated from the RCI server

The order parameters in Figure 2.9 are below 0.2. This shows that the peptide is less structured and these results are confirmed by the RCI values in Figure 2.10. The RCI values ≤ 0.7 indicates that the peptide structure is highly disordered. Therefore the *af3* peptide is highly flexible as deduced from the RCI values.

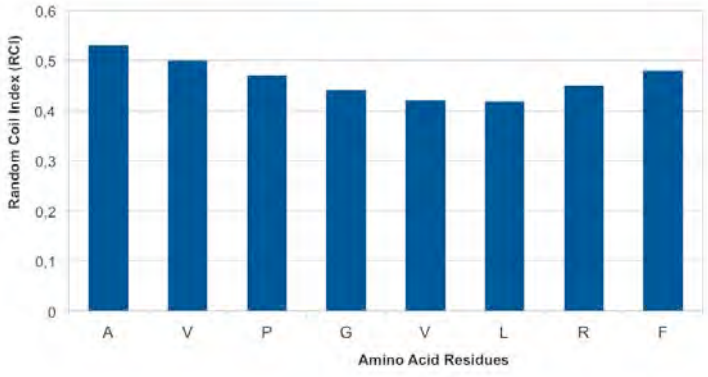


Figure 2.10: Random Coil values showing an estimate backbone root mean square fluctuation (RMSF) of the structural ensemble of *af3* calculated from chemical shifts.

The RCI values were calculated from the chemical shifts in Table 2.4 using the RCI server[37] as described in Section 2.4.3. The calculated average correlation coefficient between the RCI and the MD-RMSF (Figure 2.11) of the *af3* is 0.99. This depicts that the experimentally determined structure from the NMR and distance restrained molecular dynamics simulations are in total agreement. This validates the model relationship of the flexibility of the protein and the predicted RCI values from the chemical shifts. The closeness of the chemical shifts values to the RCI values calculated highlight a very mobile (flexible) protein.[37]

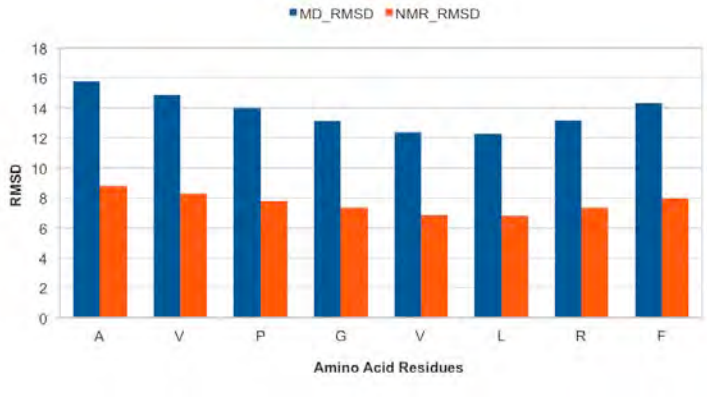


Figure 2.11: Correlation between MD_RMSD and NMR_RMSD in *af3* calculated using chemical shifts on the RCI server[37]

The large MD_RMSD and NMR_RMSF in the *af3* peptide (above 7) correspond to the unstructured conformation that was explained in relation to the S^2 order parameters in Figure 2.9. Interestingly, the large deviation and fluctuation help to validate the finding that the peptide is highly flexible and has an unstructured conformation.

2.4.3 *MD Simulations and Conformational Search*

Molecular dynamics simulations using NMR data were used to perform a conformational search of the *af3* polypeptide using the same protocol as explained in Section 2.4.3. The *af3* peptide has a similar C-terminal motif to the other three peptides, *flp18-6, af4* and *af20*. This motif that is conserved is –PVGVLRF. As discussed in Section 2.4.2.1, a type II beta (ß) turn exists in *af3* in the non-helical central residues of the motif, VPGV. The ß-turn is a type of secondary structure of peptide, which effects change in the direction of the non-regular polypeptide chain. Pro3 forms a type II ß-turn and it neighbours glycine, which is known to form kinks

in proteins.

Weak hydrophobic interactions were observed between the Cγ of the Pro3 pyrrolidine ring and backbone of Arg7. Figure 2.12 highlights the cyclic conformation and the interactions in the *af3* peptide. There are weak hydrogen bonds that are well defined from Hα of Pro3 and the amide proton of Gly4 to the carbonyl oxygen of the Leu6 with hydrogen bond lengths of 2.73 and 2.97 Å, respectively. These strong bonds facilitate the cyclic conformation of the *af3* peptide. Furthermore, the carbonyl oxygen of Val2 interacts with the amide proton in Gly4, forming a hydrogen bond of 2.8 Å, and helps to maintain the stability of the structure.

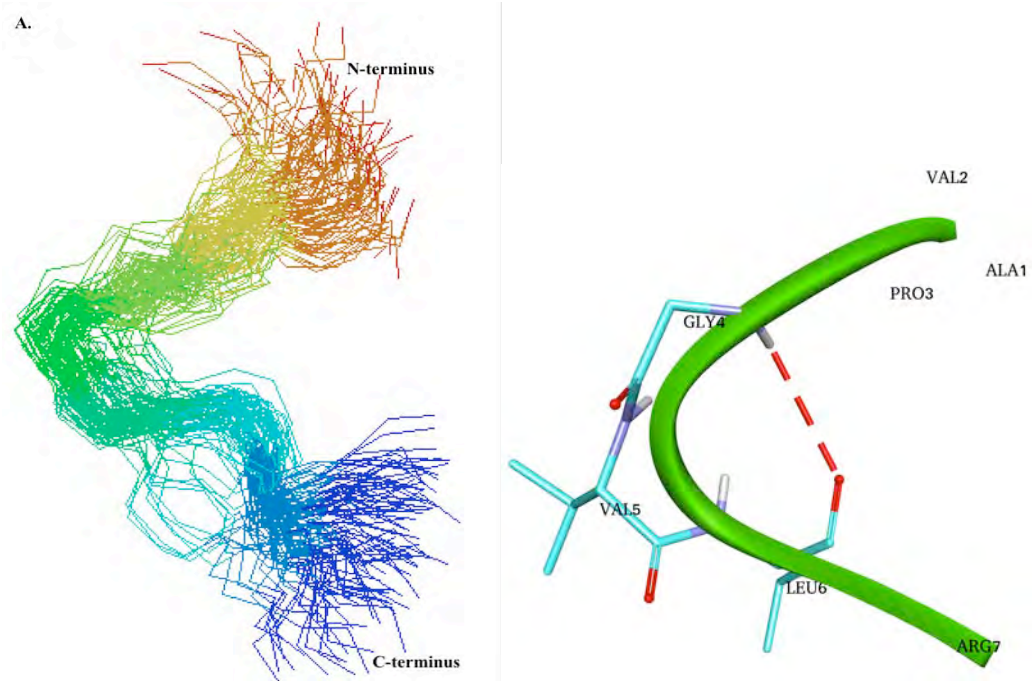


Figure 2.12: a) Pictorial representation of the overlay of 100 structures from molecular dynamic simulations of *af3* in water at 300K. b) The interactions in the conformation of *af3* that stabilize the structure showing hydrogen bond (red) between the NH of Gly4 interacting with the Leu6 carbonyl oxygen as the hydrogen bond acceptor.

Figure 2.13 shows that the peptide does not stabilise over the 10 ns of MD simulation. There is, however, consistent RMSD in DPC throughout the simulation, up to 8.5 ns. Furthermore, there is some convergence of the RMSD after 7.4 ns in both water and DPC.

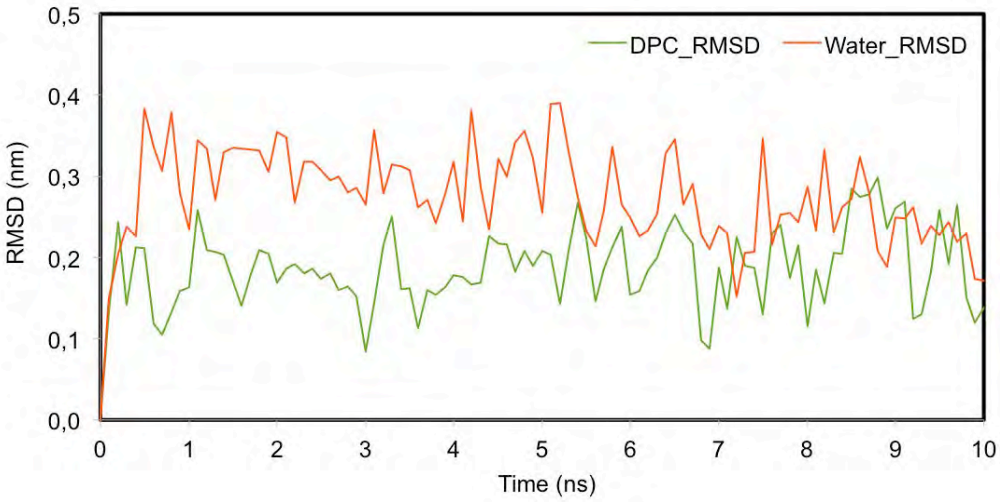


Figure 2.13: The Cα RMSD variation in the *af3* peptide in water and DPC

The average RMSD in water is 0.26 nm and in DPC it is 0.16 nm. There is a rapid increase in the RMSD from the starting structure in water and DPC in the first 0.2 ns, reaching an RMSD of 0.24 nm. To validate and explain this, and to investigate the compactness of the *af3* peptide over the 10 ns simulation, the radius of gyration (Rg) was calculated and is shown in 2.14. Rg measures the compactness of a molecule during MD simulations. It is calculated as:

$$Rg = (\sum_i \|r_i\|^2 m_i / \sum_i m_i)^{1/2} \qquad \text{(Equation 2.2)}$$

Where r_i and *i* indicate the position of the atom relative to the centre of mass of the molecule and m_i is the mass of the atom. Based on this calculation implemented in GROMACS version 4.6.3 and the plot by XMGRACE in Figure 2.14

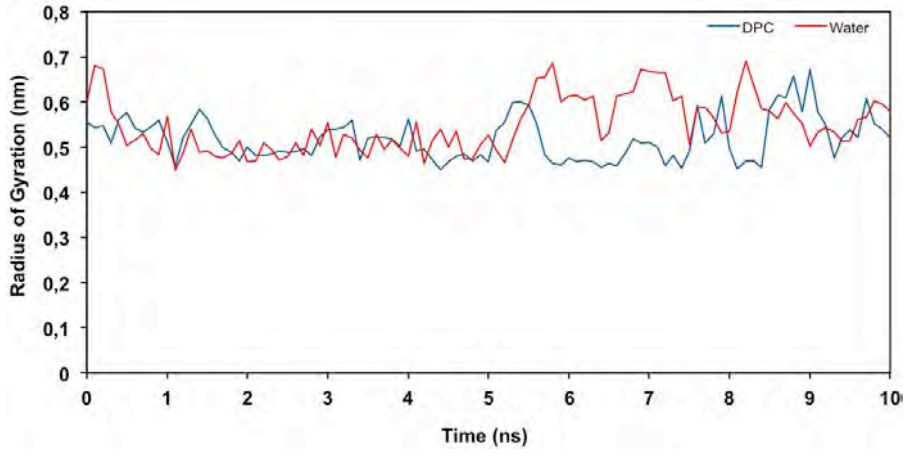


Figure 2.14: Backbone radius of gyration of *af3* peptide over 10 ns

According to the radius of gyration, the peptide is compact over the initial 5 ns and invariably changes after 5.4 ns. This helps to explain the highly disordered structure and the flexibility discussed using the observed and estimated random coil values. Without necessarily taking a compact shape, the Rg in water is more stable in the first 5 ns, albeit small. The Rg plot explains further that the *af3* is highly flexible.

2.5 *af4* Peptide

2.5.1 NMR Spectral assignment and interproton distances

The protocol that was used for assigning NMR spectra and calculating interproton distances was highlighted in section 2.4.1. Sparky was successfully used for the ^{1}H chemical shifts assignments for the *af4* in water and the results are given in Table 2.6

Table 2.6 ^{1}H chemical shift (ppm) assignments for *af4*, GDVPGVLRF-NH_2 in H_2O-D_2O 9:1 pH ca 4.5 (Phosphate buffer), T=290K

#	Res	HN	Hα	Hββ'	Hγγ'	Hδδ'	Hε
1	Gly						
2	Asp	8.895	4.674	2.607, 2.720			
3	Val	8.443	4.304	2.050	0.944		
4	Pro		4.379	1.861, 2.335	1.995	3.620, 3.924	
5	Gly	8.641	3.916, 3.616				
6	Val	8.188	4.029	2.128	0.938, 0.946		

7	Leu	8.193	4.251	1.635	1.514	0.923	
8	Arg	7.966	4.160	1.663	1.335		3.048
9	Phe	8.063	4.517	2.951, 3.194			

2.5.2 NMR Structural elucidation

Comparison of the ^{1}H and H^{N} chemical shifts with the random coil values generated from the RCI server[37] helps to understand the relationship that exists between the random chemical shifts and the flexibility of a protein. As has been explained in section 2.4.2, the chemical shifts obtained from Sparky assignments of the NMR spectra of the peptides are used to generate random coil values. Figure 2.15 highlights the backbone amide $^{1}H^{N}$ and ^{1}H chemical shift deviations. The $^{1}H^{N}$ shifts are lower than the random coil values.

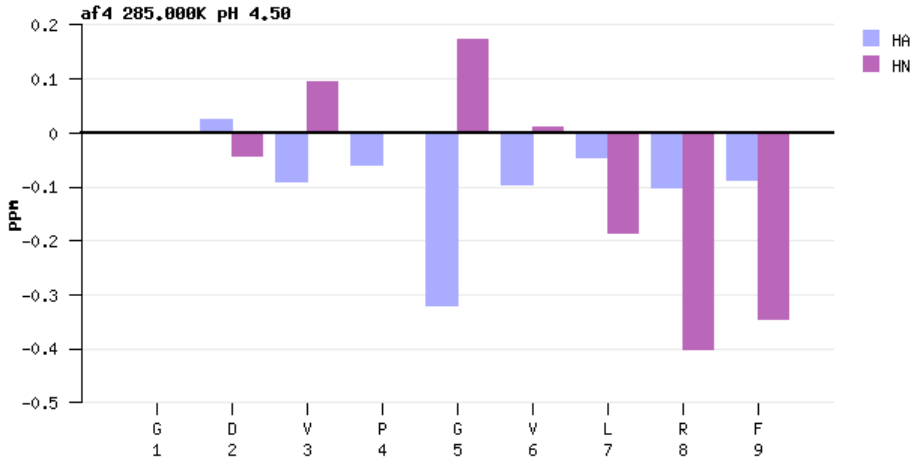


Figure 2.15: Backbone H^{N} and Hα chemical shift deviations from random coil values for *af4* in water. Random coil chemical shifts were obtained from the RCI Server[37]

The RCI values of the backbone amide and (Hα) alpha protons are mostly negative. The negative trend increases as we move towards the C-terminal of the peptide with mostly the amide protons in leucine, arginine and phenylalanine. These indicate random chemical shifts of -0.19 ppm, -0.4 ppm and -0.35 ppm, respectively. Since the random chemical shifts are predominantly negative, it implies that the protein is very

flexible. The chemical shifts of the amide protons are lower than RCI values. Figure 2.16 indicates that the RCI values of *af4* calculated from the chemical shifts that the peptide is very flexible and thus increased mobility. This is deduced because of the Random coil values being ≤ 0.7 ppm. The RCI values vary from between 0.45 and 0.53 ppm throughout the peptide. The random coil values are close to the chemical shift values.

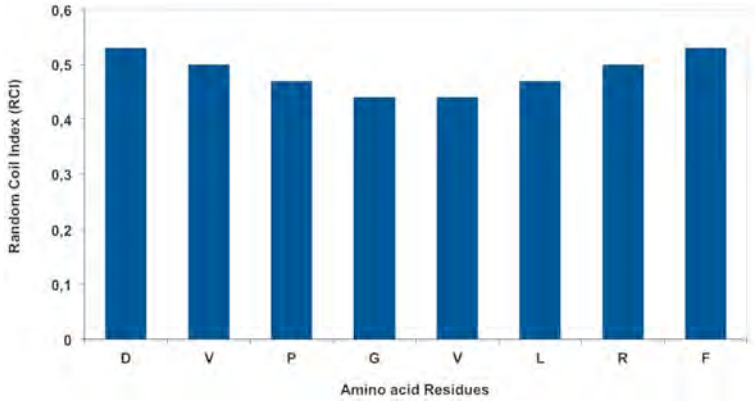


Figure 2.16: Random Coil values showing an estimated backbone root mean square fluctuation (RMSF) of the structural ensemble of *af4* calculated from chemical shifts.

The flexibility of the protein is further explained by the low order parameters, S^2, as shown in Figure 2.17. The order parameters are low, ranging from ~0.07 to 0.13 implying that the peptide is unstructured. Similar to *af3*, the C and N terminals of the *af4* are unstructured.

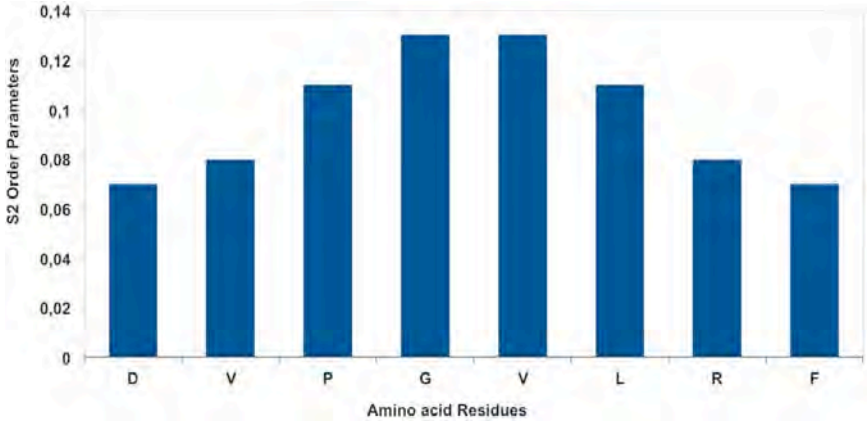


Figure 2.17: ^{1}H-^{13}Cα Model-free Squared Generalized Order parameters, S^2, of the *af4* calculated from the RCI server

The Hα chemical shifts (Table 2.6) are used to understand the NMR_RMSD and the MD_RMSD in Figure 2.18 highlights the observed trend, which is consistent between experimental and theoretical calculations. The variation of NMR_RMSD and

MD_RMSD validates the agreement between the molecular dynamics simulations and the NMR chemicals shifts. The NMR_RMSD and the MD_RMSD outlines the local variations per residue in the peptide. It is therefore concluded that they are consistent with the RCI values, which are higher at the N and C terminals than at the centre of the peptide. This is similar to the *af3* (Section 2.5).

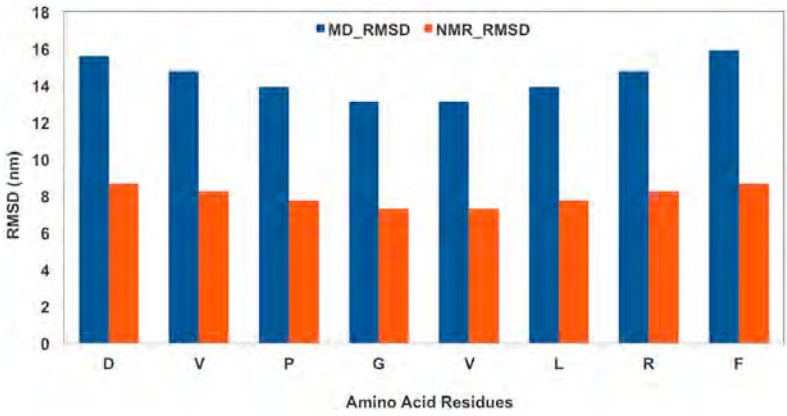


Figure 2.18: Correlation between MD_RMSD and NMR_RMSD in *af4* calculated using chemical shifts on the RCI server[37]

2.5.3 MD simulations and conformational search of af4

The protocol outlined in Section 2.4.3.1 was used for performing a conformational search in vacuum and water. The starting structure for water molecular dynamics simulations was extracted from the molecular dynamics simulations in vacuum after simulated annealing.[5,31] Figure 2.19 shows the cyclic conformation of the final solution structure obtained in DPC solution following the protocol in Section 2.4.3.1

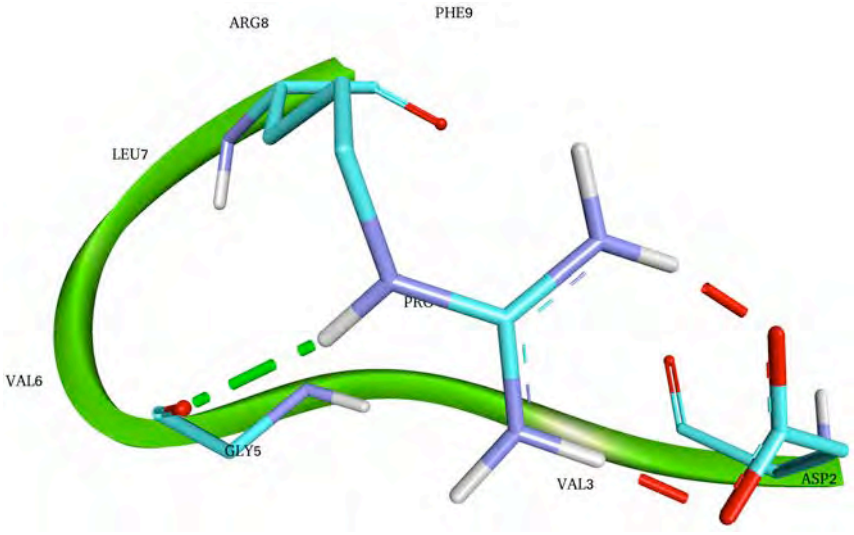


Figure 2.19: Interactions in the *af4* including salt bridge between Arg8 and Asp2 in red and, hydrogen bond (green)

A salt bridge (1.64 Å) between the carboxylate group of Asp2 and the guanidinium protons of Arg8 is very strong. This is formed as an electrostatic interaction between the arginine and the aspartic acid, as well as the hydrogen bond. The salt bridge stabilizes the cyclic conformation of *af4*. The carboxylate oxygen of Asp2 is also accepting a proton from the amide group of Val3, forming a strong hydrogen bond of 2.82 Å. The Arg8 Hε interacts with the carbonyl oxygen of Gly5 through a hydrogen bond that is 2.1 Å in length. The aromatic ring of Phe9 is involved in a π- alkyl hydrophobic interaction with the Cγ of the pyrrolidine ring of Pro4.

The stability of the peptide is shown by the moderate RMSD fluctuations in the MD per residue between 0.7 and 0.9 nm. During the 10 ns MD simulations, shown in Figure 2.20, in water and DPC, the RMSD throughout the simulation are 0.25 and 0.05 nm, respectively. The RMSD increases rapidly from 0 ns to about 1 ns in water. During MD simulations in water, the peptide was relatively stable after 7 ns. The RMSD variations in the MD simulations are slightly larger with an average of 0.09 nm in DPC, between 3.5 ns and 6 ns and the simulations converges from there.

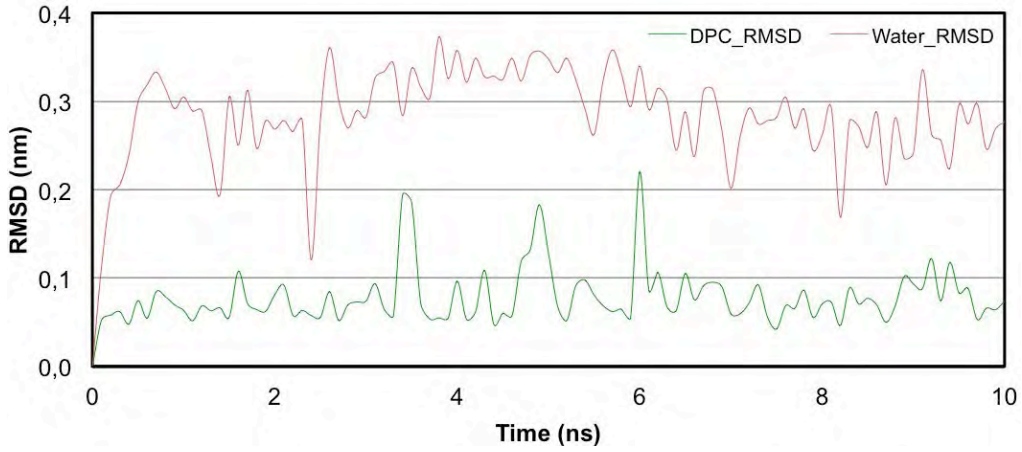


Figure 2.20: The Cα RMSD variation in the *af4* peptide in water and DPC over 10 ns

2.6 *af20* Peptide

The *af20* peptide has a conserved C-terminus as *af3*, *af4* and *flp18-6*. However, the *af20* was used in matching experiments in *A. suum*[1,2]. This section will present the results of this peptide and most discussions have been covered in section 2.5 and 2.6.

2.6.1 NMR Spectral assignment and interproton distances

The protocol that was used for assigning NMR spectra and calculating interproton distances was highlighted in section 2.4.1. Sparky was successfully used for the ^{1}H chemical shifts assignments for the *af20* in water and the results are given in Table 2.7.

Table 2.7 ^{1}H chemical shift (ppm) assignments for *af20*, GMPGVLRF-NH_2 in H_2O-D_2O 9:1 pH ca 4.5 (Phosphate buffer), T=290K

#	Res	HN	Hα	Hββ'	Hγγ'	Hδδ'	Hε	Hζ	Hη
1	Gly		3.807						
2	Met	9.047	4.704	1.986	2.625, 2.599		2.069		
3	Pro		4.451	2.357	2.013, 1.885	3.686, 3.900			
4	Gly	8.744	3.900, 3.684						
5	Val	8.146	4.037	2.167	0.950, 1.021				
6	Leu	8.089	4.224	1.669, 1.545	1.533	0.917, 0.859			
7	Arg	7.870	4.148	1.661	1.297	3.030	7.321		6.687, 7.007
8	Phe	7.970	4.498	2.945, 3.218		7.544	7.292	7.181	

The determination and prediction of the RCI values from chemical shifts is shown in Figure 2.21.

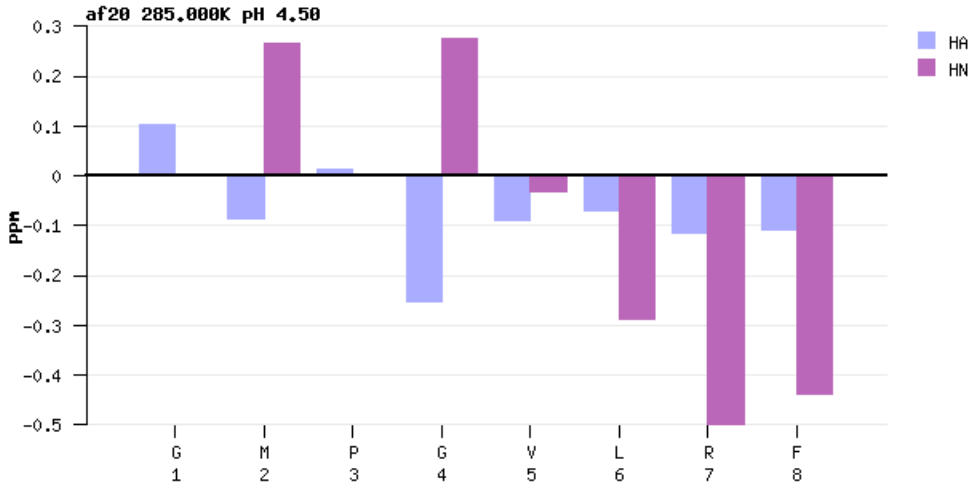


Figure 2.21: Backbone H^N and Hα chemical shift deviations from random coil values for *af20* in water. Random coil chemical shifts were obtained from the RCI Server[37]

The amide protons and Hα protons are negative from Figure 2.21 and this has been explained in Section 2.5 and 2.6. The results show a similar trend. A look at the RCI values (Figure 2.22) shows that the peptide is flexible with RCI values ranging between 0.42 and 0.55 ppm.

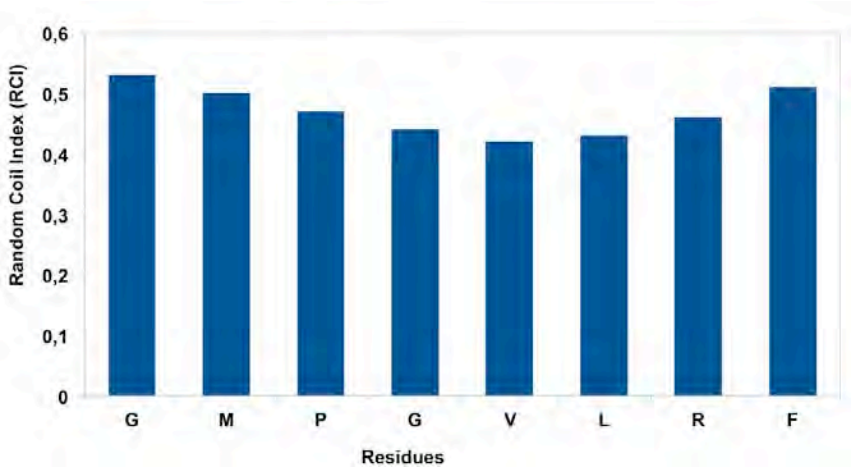


Figure 2.22: Random Coil values showing an estimate backbone root mean square fluctuation (RMSF) of the structural ensemble of *af20* calculated from chemical shifts.

There is a moderate difference between RCI values and the chemical shifts meaning that the peptide has moderate mobility. The flexibility of this protein is enunciated by the order parameters of the *af20* calculated from the chemical shifts on the RCI server and shown in Figure 2.23.

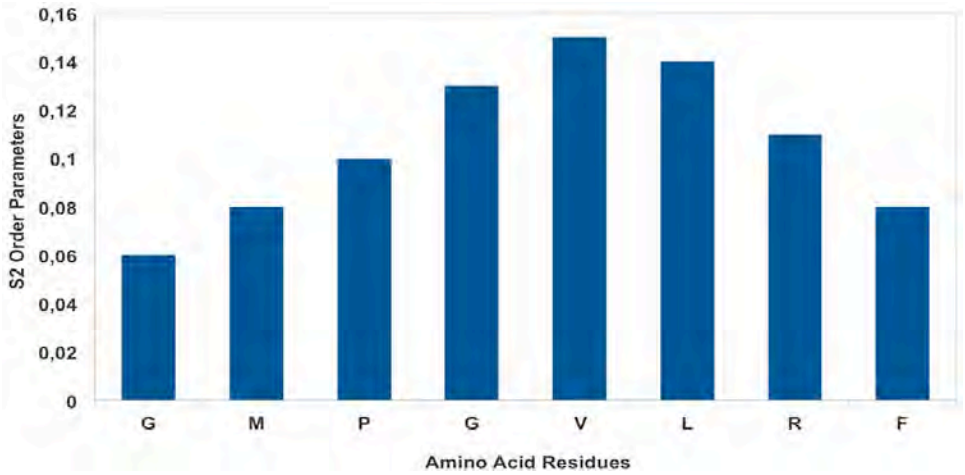


Figure 2.23: ^{1}H-$^{13}C\alpha$ Model-free Squared Generalized Order parameters, S^2, of the *af20* calculated from the RCI server.

The order parameters increase from the N-terminus to the centre of the peptide from Pro3 to Val5 and decreases towards the C-terminus. Order parameters reflect the local order of the structure. The order parameters in *af20* are very low and close to zero, reflecting an unstructured peptide. It can be deduced that the intermolecular hydrogen bonds do exist that maintain a cyclic conformation of the peptide. This is seen mostly around the areas where the order parameters are high, between Pro3 and Arg7 in this peptide.

To further explain the correlation between the random coil deviations and predicted RCI values in Figures 2.21 and 2.22 respectively, the MD_RMSD and NMR_RMSD were calculated. Figure 2.24 shows the relationship between the MD_RMSD and NMR_RMSD to expound on the fluctuations of the backbone in the experiments.

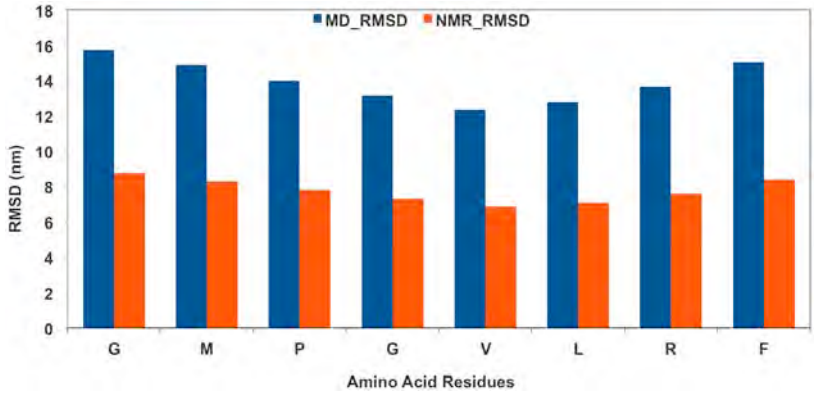


Figure 2.24: Correlation between MD_RMSD and NMR_RMSD in *af20* calculated using chemical shifts on the RCI server[37]

As has been shown in *af3* and *af4* in the previous sections 2.5 and 2.6, the correlation in this peptide is similar. The NMR_RMSD is consistent with the variation of the MD_RMSD that varies between 7 nm and 8.3 nm and 13 nm to 15.7 nm respectively. The relatively large deviation between the chemical shifts and NMR_RMSD is thought to justify the flexibility and unstructured conformation that was shown by the order parameters (Figure 2.23).

2.6.2 MD Simulations and Conformational Search of af20

Molecular dynamic simulations were performed to investigate the conformational changes[5] in the peptides over 10 ns. The protocol has been extensively explained in section 2.4.3. Cluster analysis of the final MD simulations in DPC was performed using a cut-off of 0.25 nm and one cluster was found.

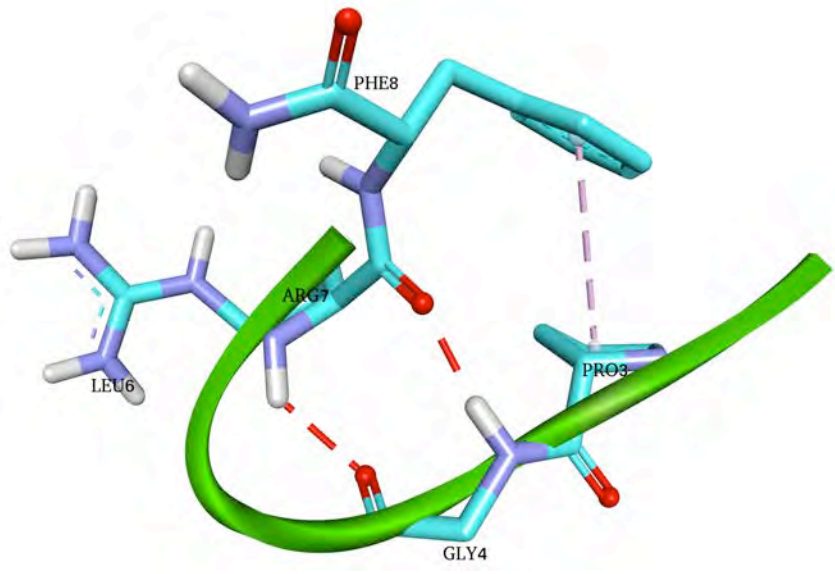


Figure 2.25: Intramolecular interactions in the *af20* peptide highlighting hydrophobic interaction between pyrrolidine ring of Pro3 and aromatic ring of Phe8 in magenta; hydrogen bonds in red between amide proton of Arg7 and carbonyl oxygen of Gly4 (2.11 Å) and NH of Gly4 with carbonyl oxygen of Arg7.

The Cγ on the Pro3 pyrrolidine ring and the π electron clouds on the aromatic ring of Phe8 are involved in a π- alkyl hydrophobic interaction. The Pro3 Hα interacts with the carbonyl oxygen through a 2.74 Å long hydrogen bond. The amide cap on the C-terminus has a proton that interacts with the carbonyl oxygen of Leu6. A strong

hydrogen bond of 1.98 Å forms between the amide hydrogen of Gly4 and the carbonyl oxygen of Arg7. Another hydrogen bond from the Arg7 Hα to the carbonyl oxygen of Gly4, which is 2.11 Å, is also very strong.

The global variation of the amplitude of protein motions was investigated over 10 ns and shown in Figure 2.26. Interestingly, the Cα RMSD in the peptide stabilizes after about 1 ns in both water and DPC. The average RMSD over the remaining 9 ns is 0.27 nm. As has also been seen in the *af3* and *af4*, the RMSD converges after 7 ns.

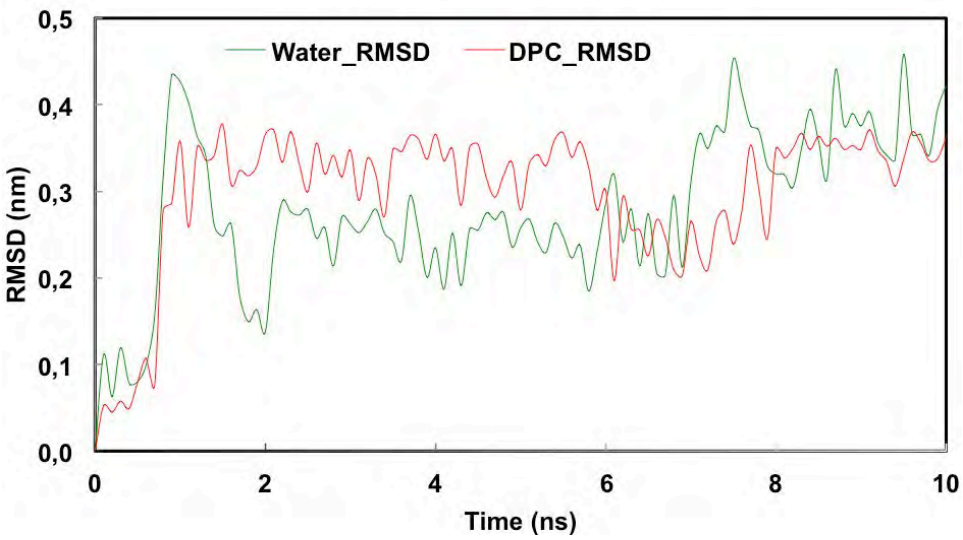


Figure 2.26: The Cα RMSD variation in the *af20* peptide in water and DPC over 10 ns

The MD simulations therefore validated the theoretical results. This means the theoretical and experimental results are in good agreement.

2.7 Relationship between the solution structures of the *af3, af4, af20* and *flp18-6* peptides

The peptides were taken from different nematode species with the *flp18-6* belonging to *C. Elegans* and *af3, af4* and *af20* taken from *Ascaris Suum* (*A. suum*).[2] The peptides share an invariant C-terminal moiety, which is PGVLRF-NH_2. The ^{1}H chemical shifts of PGVLRF-NH_2 conserved region in all the peptides are close to the random coil values and are comparable among all the 4 peptides. This intimates unstructured peptides in solution and the consequential flexibility is essential in ligand binding as the salt bridge breaks as established in chapter 4. The peptides differ in the N-terminal

residues and the sequences were given in Table 2.1. Sequence alignments of the *flp18-6* with *af3*, *af4* and *af20* indicate sequence identities of 77.78, 88.89 and 90 %.

NMR experiments were used to determine the solution structures of the peptides in aqueous solution. In all the peptides the amide proton on the N-terminal cannot be observed, due to the rapid exchange with the solvent. The amide proton chemical shifts were plotted against variable temperature from 280 K to 305 K. The observation from the plot is that the H^N chemical shift signal shifted upfield with temperature. This observation is consistent with many other previous studies on peptides.[5,19,20,40] The RCI values of the peptides indicate typically unstructured proteins that are flexible as validated by the order parameters of all the peptides.

The average correlation coefficient of the MD_RMSD and NMR_RMSD using the hydrogen and amide proton chemical shifts of the peptides was calculated and range from 0.86 to 0.99. The temperature coefficients of all peptides highlight that the peptides have transient hydrogen bonds and some intrinsically weak electrostatic interactions. Ring current effects significantly affect the chemical shifts of the glycine residue amide protons in the N-terminals.[18] Type II beta turns were found in all four peptides, which is also indicated by the experimental chemical shifts data. These were formed by a motif, VPGV in *af3, af4, flp18-6* and MPGV in *af20*. Even though our calculations for temperature coefficients are different from literature, the observed chemical shift data indicate that the peptides have some structure.

The peptides were simulated *in vacuo*, water, and DPC solution using NMR distance restraints. The peptides were gradually heated slowly up to 600 K followed by simulated annealing down to 300 K and peptide molecules were allowed to reach a global minimum. The MD simulations performed on the peptides in water and DPC over 10 ns display trajectories that are similar and consistent in this class of peptides. From these MD simulations, the results are consistent with the reports of mammalian peptides that over 100 GPCR peptide ligands have a turn structure.[5,47,52]

The structural and dynamic properties of these peptides are similar. The peptides have cyclic conformations that are a result of salt bridges in *af4 and flp18-6*, and strong hydrogen bonds in *af3* and *af20*. The bridges, coupled with hydrophobic π- alkyl

interactions, are believed to stabilize the conformations; these interactions are critical in explaining the NMR experimental results where the temperature coefficients are not sufficient to explain the hydrogen bonds that are formed. These peptides are similar to other GPCR peptide ligands, which typically have similar cyclic conformations.[5,52,53]

The *af4* and *af20* peptides have Glycine in the N-terminus and lack the Asp needed to form a salt bridge. However, this does not affect the cyclic conformation of *af4* but is believed to have an effect on the interactions in *af20*. This is caused by the close contact between the glycine and the aromatic ring of the Phenylalanine at the C-terminus. Suggestions by van der Spoel are that the aromatic ring prefers a particular conformational orientation towards at the C-terminus.[45] Since glycine has no side chain, there is no steric hindrance at the N-terminus. Therefore a strong interaction is suggested. The glycine amide proton will polarize the aromatic ring and the π electrons will act as hydrogen bond acceptor and result in the formation of a hydrogen bond.

2.8 Conclusions

The solution structures of four peptides: *af3, af4, af20 and flp18-6* were determined using NMR spectroscopy and molecular modeling. The peptide structures were well defined in 67 to 88 % of the residues. The amide proton chemical shifts decrease with temperature such that the amide proton signals shift upfield with temperature.

The calculated order parameters, S^2, using the proton chemical shifts were found to be between 0.4 and 0.8, which suggested that the peptide *(flp18-6)* is flexible and mobile. The RCI values however, are below 0.2 ppm, which highlights an unstructured protein.[11,39]. The peptide also has a type II beta turn. MD simulation results suggest that the RMSD fluctuations indicate a variably stable structure and that is not too compact, which is consistent with the NMR experiments.[11,37]

The residues in the peptides are mostly hydrophobic, resulting in hydrophobic interactions with the DPC micelle solution, and thus considerably influencing the

structure of the peptide.[54,55] The π- alkyl interactions and salt bridges help introduce stability in the otherwise unstructured peptide in aqueous solution. Salt bridges[43] exist between N-terminal aspartates and the C-terminal penultimate arginine resulting in a C-shape in *af4* and *flp18-6* peptides. The salt bridge is, however, very strong in *flp18-6* compared to *af4*, with bond lengths of 1.55 and 1.64 Å respectively. A highly persistent π- alkyl interaction that formed between the aromatic ring of phenylalanine and the pyrrolidine ring of Proline in *af3* and *af20* was detected. This gave these two peptides a cyclic conformation that may as well be referred to as a boat-shaped conformation.

In this work, molecular dynamic simulations were performed to reproduce the chemical shift data that was realized experimentally using NMR spectroscopy.[56] The NMR experiments were conducted in order to understand the structural and dynamic properties of small peptides in aqueous solution. The solution structures were determined using NMR restraints during molecular dynamic simulations. The investigations into peptide structures predicted from NMR experiments concluded that the peptides have turn/ cyclic conformations.[5] NMR experiments also confirmed that amide proton chemical shifts are temperature dependent and that the ring current effects affect the chemical shifts of amide protons. MD simulations were successful in revealing the structural similarities in the peptides, despite the characteristically high conformational space. The solution structures will be used in Chapter 4 and 5 for docking calculations and docking-based virtual screening respectively.

2.9 References

(1) Kubiak, T. M.; Larsen, M. J.; Bowman, J. W.; Geary, T. G.; Lowery, D. F. FMRFamide-like Peptides Encoded on the Flp-18 Precursor Gene Activate Two Isoforms of the Orphan *Caenorhabditis Elegans* G-Protein-Coupled Receptor Y58G8A.4 Heterologously Expressed in Mammalian Cells. *Biopolym. - Pept. Sci. Sect.* **2008**, *90* (3), 339–348.

(2) McVeigh, P.; Leech, S.; Mair, G. R.; Marks, N. J.; Geary, T. G.; Maule, A. G. Analysis of FMRFamide-like Peptide (FLP) Diversity in Phylum Nematoda. *Int. J. Parasitol.* **2005**, *35* (10), 1043–1060.

(3) Larsen, M. J.; Lancheros, E. R.; Williams, T.; Lowery, D. E.; Geary, T. G.; Kubiak, T. M. Functional Expression and Characterization of the *C. Elegans* G-Protein-Coupled FLP-2 Receptor (T19F4.1) in Mammalian Cells and Yeast. *Int. J. Parasitol. Drugs drug Resist.* **2013**, *3*, 1–7.

(4) O'Connell, M. R.; Gamsjaeger, R.; Mackay, J. P. The Structural Analysis of Protein-Protein Interactions by NMR Spectroscopy. *Proteomics*, **2009**, *9*, 5224–5232.

(5) Mugumbate, G.; Jackson, G. E.; Van Der Spoel, D.; Kövér, K. E.; Szilágyi, L. Anopheles Gambiae, Anoga-HrTH Hormone, Free and Bound Structure-A Nuclear Magnetic Resonance Experiment. *Peptides* **2013**, *41*, 94–100.

(6) Liu, H.; Spielmann, H. P.; Ulyanov, N. B.; Wemmer, D. E.; James, T. L. Interproton Distance Bounds from 2D NOE Intensities: Effect of Experimental Noise and Peak Integration Errors. *J. Biomol. NMR* **1995**, *6* (4), 390–402.

(7) Thomas, P. D.; Basus, V. J.; James, T. L. Protein Solution Structure Determination Using Distances from Two-Dimensional Nuclear Overhauser Effect Experiments: Effect of Approximations on the Accuracy of Derived Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (4), 1237–1241.

(8) Dossey, A. T.; Reale, V.; Chatwin, H.; Zachariah, C.; DeBono, M.; Evans, P. D.; Edison, A. S. NMR Analysis of *Caenorhabditis Elegans* FLP-18 Neuropeptides: Implications for NPR-1 Activation. *Biochemistry* **2006**, *45* (24), 7586–7597.

(9) Schlegel, B.; Sippl, W.; Höltje, H. D. Molecular Dynamics Simulations of Bovine Rhodopsin: Influence of Protonation States and Different Membrane-Mimicking Environments. *J. Mol. Model.* **2005**, *12* (1), 49–64.

(10) Merutka, G.; Jane Dyson, H.; Wright, P. E. "Random Coil" 1H Chemical Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series GGXGG. *J. Biomol. NMR* **1995**, *5* (1), 14–24.

(11) Berjanskii, M.; Wishart, D. S. NMR: Prediction of Protein Flexibility. *Nat. Protoc.* **2006**, *1* (2), 683–688.

(12) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013**, *18* (9), 10802–10828.

(13) Mattinen, M. Structural and Functional Studies of Fungal Cellulose-Binding Domains by NMR Spectroscopy. *Vtt Publ.* **1998**.

(14) Cross, T. A.; Opella, S. J. Solid-State NMR Structural Studies of Peptides and Proteins in Membranes. *Current Opinion in Structural Biology*, **1994**, *4*, 574–581.

(15) Castellani, F.; Rossum, B. Van; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H.; Reynolds, W. F.; Enriquezi, R. G. NMR Spectroscopy. *Nature* **2002**, *420*, 98–102.

(16) Hu, W.; Kakalis, L. T.; Jiang, L.; Jiang, F.; Ye, X.; Majumdar, A. 3D HCCH-COSY-TOCSY Experiment for the Assignment of Ribose and Amino Acid Side Chains in 13C Labeled RNA and Protein. *J. Biomol. NMR* **1998**, *12* (4), 559–564.

(17) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR Structure Determination with Automated NOE-Identification in the NOESY Spectra Using the New Software ATNOS. *Journal of Biomolecular NMR*, **2002**, *24*, 171–189.

(18) Cierpicki, T.; Otlewski, J. Amide Proton Temperature Coefficients as Hydrogen Bond Indicators in Proteins. *J. Biomol. NMR* **2001**, *21* (3), 249–261.

(19) Baxter, N. J.; Williamson, M. P. Temperature Dependence of ^{1}H Chemical Shifts in Proteins. *J. Biomol. NMR* **1997**, *9* (4), 359–369.

(20) Vranken, W. F.; Rieping, W. Relationship between Chemical Shift Value and Accessible Surface Area for All Amino Acid Atoms. *BMC Struct. Biol.* **2009**, *9* (1), 20.

(21) Sato, H.; Fukae, K.; Kajihara, Y. 2D Selective-TOCSY-DQFCOSY and HSQC-TOCSY NMR Experiments for Assignment of a Homogeneous Asparagine-Linked Triantennary Complex Type Undecasaccharide. *Carbohydr. Res.* **2008**, *343* (8), 1333–1345.

(22) Yang, J. H.; Ho, Y.; Tzou, D. L. M. A 13C Solid-State NMR Analysis of Steroid Compounds. *Magn. Reson. Chem.* **2008**, *46* (8), 718–725.

(23) Schanda, P.; Van Melckebeke, H.; Brutscher, B. Speeding up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **2006**, *128* (28), 9042–9043.

(24) Ho, Y.; Tzou, D. L. M.; Chu, F. I. Solid-State NMR Studies of the Molecular Structure of Taxol. *Magn. Reson. Chem.* **2006**, *44* (6), 581–585.

(25) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; **1986**; Vol. 32.

(26) Hinds, M. G.; Norton, R. S. NMR Spectroscopy of Peptides and Proteins. Practical Considerations. *Mol. Biotechnol.* **1997**, *7* (3), 315–331.

(27) Lee, W.; Westler, W. M.; Bahrami, A.; Eghbalnia, H. R.; Markley, J. I. PINE-SPARKY: Graphical Interface for Evaluating Automated Probabilistic Peak Assignments in Protein NMR Spectroscopy. *Bioinformatics* **2009**, *25* (16), 2085–2087.

(28) DeLano, W. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**.

(29) Makarewicz, T.; Kaźmierkiewicz, R. Molecular Dynamics Simulation by GROMACS Using GUI Plugin for PyMOL. *J. Chem. Inf. Model.* **2013**, *53* (5), 1229–1234.

(30) White, J. A.; Román, F. L.; González, A.; Velasco, S. Periodic Boundary Conditions and the Correct Molecular-Dynamics Ensemble. *Phys. A Stat. Mech. its Appl.* **2008**, *387* (27), 6705–6711.

(31) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. Modeling Activated

States of GPCRs: The Rhodopsin Template. *J. Comput. Aided. Mol. Des.* **2006**, *20* (7-8), 437–448.

(32) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72* (5), 2002–2013.

(33) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFFYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, *8* (6).

(34) Assigning Proteins Using Sparkyhttp://www.cgl.ucsf.edu/home/sparky/manual/

(35) Borgias, B.; Gochin, M.; Kerwood, D.; James, T. Relaxation Matrix Analysis of 2D NMR Data. *Prog. NMR Spectrosc.* **1990**, *22*, 83–100.

(36) Downing, A. K.; Watts, A.; Straus, S. K.; Grage, S. L.; Kamihira, M. TM Techniques Edited by. *Techniques 278*.

(37) Berjanskii, M. V.; Wishart, D. S. The RCI Server: Rapid and Accurate Calculation of Protein Flexibility Using Chemical Shifts. *Nucleic Acids Res.* **2007**, *35* (SUPPL.2), 531–537.

(38) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: A Web Server for Quantitative Evaluation of Protein Structure Quality. *Nucleic Acids Res.* **2003**, *31* (13), 3316–3319.

(39) Scheidt, H. A.; Morgado, I.; Rothemund, S.; Huster, D. Dynamics of Amyloid ß Fibrils Revealed by Solid-State NMR. *J. Biol. Chem.* **2012**, *287* (3), 2017–2021.

(40) Daley, M. E.; Graether, S. P.; Sykes, B. D. Hydrogen Bonding on the Ice-Binding Face of a Beta-Helical Antifreeze Protein Indicated by Amide Proton NMR Chemical Shifts. *Biochemistry* **2004**, *43* (41), 13012–13017.

(41) Fuchs, P. F. J.; Alix, A. J. P. High Accuracy Prediction of ß-Turns and Their Types Using Propensities and Multiple Alignments. *Proteins Struct. Funct. Genet.* **2005**, *59* (4), 828–839.

(42) Schneider, E. H.; Schnell, D.; Strasser, A.; Dove, S.; Seifert, R. Impact of the DRY Motif and the Missing “Ionic Lock” on Constitutive Activity and G-

Protein Coupling of the Human Histamine H4 Receptor. *J. Pharmacol. Exp. Ther.* **2010**, *333* (2), 382–392.

(43) Dunkin, C. M.; Pokorny, A.; Almeida, P. F.; Lee, H. S. Molecular Dynamics Studies of Transportan 10 (Tp10) Interacting with a POPC Lipid Bilayer. *J. Phys. Chem. B* **2011**, *115* (5), 1188–1198.

(44) Gil-Mast, S.; Kortagere, S.; Kota, K.; Kuzhikandathil, E. V. An Amino Acid Residue in the Second Extracellular Loop Determines the Agonist-Dependent Tolerance Property of the Human D3 Dopamine Receptor. *ACS Chem. Neurosci.* **2013**, *4* (6), 940–951.

(45) van der Spoel, D.; van Buuren, A. R.; Tieleman, D. P.; Berendsen, H. J. C. Molecular Dynamics Simulations of Peptides from BPTI: A Closer Look at Amide—aromatic Interactions. *J. Biomol. NMR* **1996**, *8* (3), 229–238.

(46) Fuchs, P. F. J.; Alix, A. J. P. High Accuracy Prediction of ß-Turns and Their Types Using Propensities and Multiple Alignments. *Proteins Struct. Funct. Genet.* **2005**, *59* (4), 828–839.

(47) Langelaan, D. N.; Bebbington, E. M.; Reddy, T.; Rainey, J. K. Structural Insight into G-Protein Coupled Receptor Binding by Apelin. *Biochemistry* **2009**, *48* (3), 537–548.

(48) Mauldin, K.; Lee, B. L.; Oleszczuk, M.; Sykes, B. D.; Ryan, R. O. The Carboxyl-Terminal Segment of Apolipoprotein A-V Undergoes a Lipid-Induced Conformational Change. *Biochemistry* **2010**, *49* (23), 4821–4826.

(49) Olubiyi, O. O. Investigation of the Interaction between Alzheimer' S Abeta Peptide and Aggregation Inhibitors Using Molecular Simulations. **2013**.

(50) Munyuki, G. (University of Cape Town), *Dissertation*. Molecular Mechanism of Action of Tyrocidine Antimicrobial Peptides Using NMR Spectroscopy and Computational Techniques, **2012**.

(51) Schiedel, A. C.; Hinz, S.; Thimm, D.; Sherbiny, F.; Borrmann, T.; Maass, A.; Müller, C. E. The Four Cysteine Residues in the Second Extracellular Loop of the Human Adenosine A2B Receptor: Role in Ligand Binding and Receptor Function. *Biochem. Pharmacol.* **2011**, *82* (4), 389–399.

(52) Ruiz-Gómez, G.; Tyndall, J. D. A; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Update 1 of: Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2010**, *110* (4), PR1–PR41.

(53) Tyndall, J. D. a; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2005**, *105* (3), 793–826.

(54) Khandelia, H.; Kaznessis, Y. N. Molecular Dynamics Simulations of Helical Antimicrobial Peptides in SDS Micelles: What Do Point Mutations Achieve? *Peptides* **2005**, *26* (11), 2037–2049.

(55) Leontiadou, H.; Mark, A. E.; Marrink, S. J. Antimicrobial Peptides in Action. *J. Am. Chem. Soc.* **2006**, *128* (37), 12156–12161.

(56) Yu, H.; Daura, X.; van Gunsteren, W. F. Molecular Dynamics Simulations of Peptides Containing an Unnatural Amino Acid: Dimerization, Folding, and Protein Binding. *Proteins* **2004**, *54* (1), 116–127.

3 Chapter 3: Protein Structure Determination

3.0 Summary

The three-dimensional (3D) structures of G-protein coupled receptors (GPCRs), which are targets for over 50% of current clinically approved drugs on the market, facilitate structure-based drug discovery.[1–3] As of August 2015, the Protein Data Bank (PDB) available at http://www.rcsb.org/pdb had 144 GPCR crystal structures deposited since the year 2000.[4] Unfortunately, none of these GPCRs are for a nematode. In the absence of any nematode GPCR X-ray crystal structures,[5] homology modeling can be used to construct their 3D structures.[6,7] The 3D structure of a GPCR consists of 4 main domains, the amino (N) terminus, 3 extracellular loops (ECL) outside the cell, 3 intracellular loops (ICL) and Carboxyl (C) terminus domain.[8,9]

In this chapter, homology modeling of FLP18R1, based on the crystal structure of Beta-2-adrenergic receptor (2RH1), was performed using Modeller version 9.11. To improve the quality of the homology models, the models were inserted in a mimetic membrane and optimized using molecular dynamic simulations. The homology model had a sequence identity of 24.87% to the beta-2-adrenergic receptor. This is of acceptable quality for a GPCR homology model. The conserved regions[10,11] in class A GPCRs, such as Asparagine (N) in helix 1 (H1), aspartic acid (D) in helix 2 (H2) and Aspartic acid-Lysine-Phenylalanine (DKF) in helix 3 (H3), a motif that is very important for receptor activation, were correctly predicted.[12] Despite the low sequence identity, the receptor sequence similarity that is fundamental in homology modelling is shown to be very essential in this investigation, particularly the disulfide bonds in extracellular loop 2 (EC2).[13]

The low sequence identity in homology modeling of GPCRs affects model quality.[14] The quality of the homology models, particularly in the loop regions is improved by optimization and refinement using molecular dynamic simulations in a mimetic membrane.[15] Therefore, this study confirms that homology modelling can be relied upon to determine high quality structures. The models can be employed to understand experimental results and develop structure-based hypotheses to understand the interactions and activities of the GPCRs.[16,17] The final model was used in structure-

based drug design towards identifying potential anthelmintic agents as discussed in chapters 4 and 5.

3.1 Molecular Modeling of GPCRs

The identification of nematode GPCRs relies on the heterologous expression of candidate orphan receptors and successively exploiting high throughput screening techniques of combinatorial libraries to search for the most active hits.[18] This may be challenging where the structures of the GPCRs are unknown and neither NMR nor X-ray structures are available or limited receptor information is available. Under such circumstances, homology modelling and molecular dynamic simulations can be used to determine the three dimensional structures that can be used for docking calculations and virtual screening of compound databases to identify possible ligands. Genomic sequencing and expressed sequence tag (EST) database screens were used to identify FLP genes.[18] It is from these that FLP18, a precursor gene, which encodes 8 FLPs was found[18–20].

Homology modeling is a structure-based method that has developed over the past decade. This approach together with docking calculations enhances the ability to investigate the molecular and atomic interactions between target proteins and their bound ligands[6,16,17,21–23]. This replaces the traditional methods such as high throughput screening of large chemical compound libraries. The accuracy of the homology model, its quality and refinement coupled with other structure-based searching strategies have proven that new chemical entities, which have good binding affinities for the targets can be identified.[5,10,13,22,24–27]

The important step in homology modeling is identifying suitable templates for the target. Comparing the target protein's biological information using algorithms such as Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST)[28] enables this. Biological information such as amino acid sequences can be used to screen large databases of different proteins. In this comparison, similarity of amino acids is also considered as a relationship to determine alignment. Secondary structure

prediction tools help to understand possible loop and helical regions in protein structures.[28–30]

3.2 Experimental Methods

3.2.1 Secondary Structure prediction

The primary sequence of the GPCR, FLP18R1, was obtained from GenBank with accession number Y58G8A.4a.[18,19] The protein is an isoform of NPR5 (WormBase focus) of the *C. Elegans* genome.[31] PSI-PRED[32] protein structure prediction server was employed for the prediction of the secondary structure of the GPCR receptor. Figure 3.1 shows the schematic diagram of the secondary structure prediction.

3.2.2 Identification and selection of template

Sequence similarity search was performed from the Protein Data Bank (PDB) using the PSIBLAST[28] online tools from European Bioinformatics Institute (EBI) available from http://www.ebi.ac.uk/Tools/sss/psiblast/. The proteins with the highest sequence identity and query coverage were selected with X-ray crystallography resolution below 2.8 Å.

3.2.3 Sequence alignment and homology modelling

Pairwise sequence alignment[6,33,34] (see Figure 3.2) of the target protein (FLP18R1) to the template selected from the PDB was carried out using ClustalW1.82 available from http://www.ebi.ac.uk/Tools/msa/clustalw2/.[34] The opening gap penalty used was 10 for the alignment in ClustalW and a gap extension of 0.05. The homology modelling was executed using Modeller v9.11,[35] which implements comparative modelling of the protein structure by satisfying the spatial restraints.[7,36] The **BLO**ck **SU**bstitution **M**atrix[37–39] (BLOSUM-62) in Chimera was used for homology modelling and the Needleman-Wunsch algorithm was implemented for local sequence alignment.[36]

3.2.4 Energy Minimization

The best homology model was taken as the starting structure for molecular dynamic simulations using GROMACS version 4.6.3. The starting structure was energy minimized to 300K at a time step of 2 fs, using the GROMOS96 ff43a1 force field.[40] This force field was applied throughout the simulation as it conserves the chemical and structural characters[41] of both the protein and the lipid bilayer throughout long simulation times of even up to 1μs.

The energy minimization was for 3000 steps using steepest gradient and conjugate gradient algorithms as well as constraints to check internal inconsistencies such as bad contacts, incorrect hydrogen bond interactions.[36] The cut-off selected for van der Waals interactions and electrostatic interactions was 0.1nm in order to perform real space calculations. All simulations were executed with periodic boundary conditions (pbc).[42] GROMACS version 4.6.3 was applied to apportion the protonation states to the titratable groups of FLP18R1 at the pH value of 7.

3.2.5 Embedding protein in POPC bilayer

The energy-minimized structure was embedded into a united-atom model-created pre-equilibrated Teilman's POPC[43,44]membrane with 128 lipid molecules and 2460 water molecules from http://moose.bio.ucalgary.ca/[43]using GROMOS87 force field. The distance to the centre of mass of neighboring lipids is 1 nm, so there are small areas in the lipid bilayer with overlaps. InflateGro script[45] was used to inflate the membrane and insert the FLP18R1 molecules. InflateGro slowly packs the lipid molecules around the protein (manually deleting lipids within 0.5Å of protein atoms) to bring the membrane to a physiological size of 65.8nm^2, which is the experimental size of POPC. The system is constantly minimized during the packing, while position restraining[45] the protein using a force constant of 100000 kJ $mol^{-1}nm^2$. The new protein box coordinates were assigned to the lipid box so that they are identical box sizes. Molecular coordinate space was checked using PyMOL.[46–48]

Solvation of the system was effected to increase the amount of water to 9045 molecules to allow water to soak the protein fully in the *z*-direction using SPC water. The lipid/water system was changed incrementally four fold (x4) in the *xy* planes. The system consisted of 37127 atoms. The system was energy minimized for 500 steps using the low-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) minimiser using a force constant of 1000 kJ $mol^{-1}nm^2$. The system was neutralized by adding 13 chloride (Cl^-) molecules, replacing equivalent water molecules.

3.2.6 *Equilibration*

MD simulations were used to investigate conformational changes in the protein and identify any structural features that are not obvious in inactive models. GROMOS87 force field was used in generating a pre-equilibrated POPC[43,44]membrane by Dr. Teilman and lipid parameters of the POPC were based on Berger.[41] GROMOS96 ff43a1 force field was applied throughout the simulations. It conserves the chemical and structural characters of both the protein and the lipid bilayer throughout simulation time of 100 ns. The system was equilibrated to facilitate its progression from the initial configuration to reach equilibrium.

Equilibrations were carried out in two stages starting with NVT ensemble at constant number of particles (N), volume (V), and energy (E) for 1 ns and a time step of 0.002 ps.[42] The simulation was performed at 300K, which is above the phase transition temperature of POPC.[49] Molecular dynamics (md) algorithm was the integrator used using LINCS constraints while position restraining the protein in the POPC membrane. Temperature coupling using velocity rescaling was executed for the different groups in the system, calculating electrostatic interactions using PME with Fourier spacing of 0.16 nm, generating velocities from Maxwell distribution at 300K with no pressure coupling.[43]

The second stage of equilibration was the NPT ensemble at constant N, pressure (P) and temperature (T) for 1 ns with a time step of 0.002 ps. Nose-Hoover temperature coupling was applied and Parrinello-Rahman for pressure coupling of the different groups in the system, under semi-isotropic scaling in the *xy* directions and centre of mass (COM) removal linearly at 300K. Molecular dynamics (md) algorithm was the

integrator used using LINCS constraints while position restraining the protein in the POPC membrane at constant NPT. Density and pressure variations were evaluated using XMGRACE.

3.2.7 Molecular dynamic simulations - Production run

Position restraints on the protein were removed for energy minimization for 20000 steps using the steepest descent algorithm. The molecular dynamic simulations were performed under NPT using periodic boundary conditions. Angular centre of mass removal was implemented for 100 ns with a 2 fs time step. A cut-off of 1.4 nm was used for short-range van der Waals interactions. Particle Mesh Ewald (PME) summation method was used to calculate long-range electrostatics with grid spacing of 0.9 nm. The protein was coupled to the POPC membrane[44,50] and the solvent was coupled to the chloride (Cl^-) ions. Velocity rescaling method was executed for temperature coupling at 300K and a coupling constant, τ_T, of 1×10^{-4} ns.

The pressure coupling was performed using Parrinello-Rahman with isotropic pressure coupling type at constant τ_p of 5.0 ps and an isothermal compressibility of 4.5×10^{-5} (bar^{-1}). Neighbour lists were updated every fifth integration time step. No constraints were applied throughout the 100 ns molecular dynamic simulations production run. The MD simulations were performed on the Dell C6145 series600 UCT High Performance Cluster with 64 nodes and 4 CPUs per core and 128GB RAM.

3.3 Results and Discussion

3.3.1 Secondary Structure prediction

The prediction of the secondary structure was carried out using the PSI-PRED server using the primary sequence of FLP18R1 receptor protein. PSIPRED integrates two artificial neural networks that executes analysis on output from PSI-BLAST.[32] The machine learning algorithm in PSIPRED recognizes the folding in the protein structure.[32] Fold recognition of protein secondary structure from amino acid sequences provide insight into the construction of useful tertiary models[28] in the

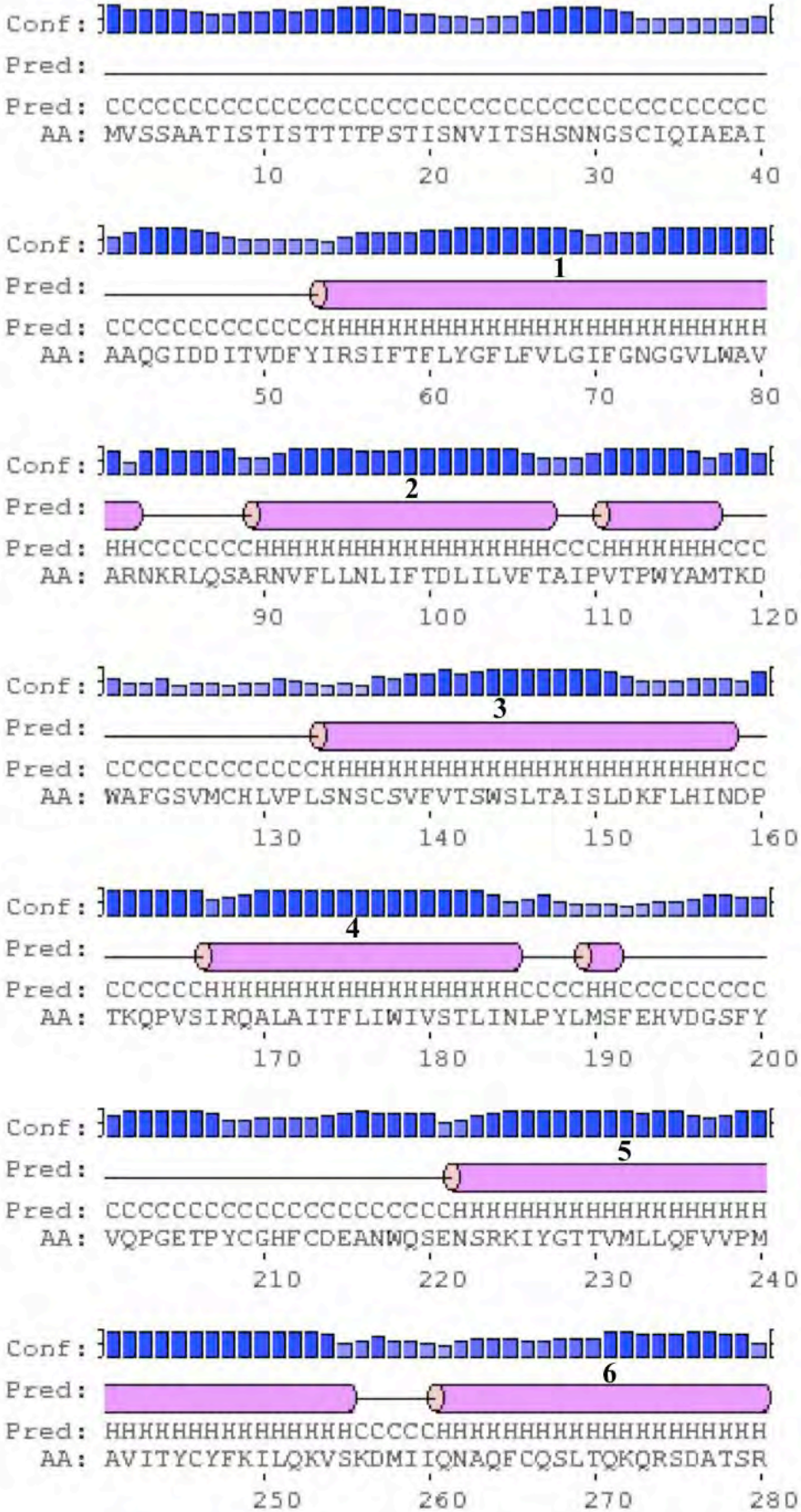
absence of homologous templates. The predicted structure is given in Figure 3.1 and shows the amino acids (AA) forming coils (C), helical regions (H) and strands. The protein has 8 α-helices connected to each other by loops. The 8th α-helix made up of 10 amino acids is significantly shorter than the other seven.

The proteins' 7 α-helices[29,51,52] correspond to the commonly referred seven transmembrane domain receptors, 7TM receptors that form a helical bundle. This finding shows that the protein was correctly predicted as a GPCR.[27,53–55] Typical GPCRs include chemokine receptors, beta 1 adrenergic, kappa opioid, dopamine D2 receptor, and serotonin receptors among many in class A of GPCRs.[56–59]

3.3.2 Sequence Similarity Search

The analysis of the primary sequence on Pfam (http://pfam.xfam.org/search/sequence) reveals that this protein belongs to the rhodopsin GPCR family and it has 7 transmembrane helices.[60,61] PSI-BLAST was implemented to characterize the primary sequence of FLP18R1, a class A GPCR and identify homologous protein sequences from the Protein Data Bank (PDB). The PSI-BLAST algorithm calculates the sequence similarity using a matrix that determines the common ancestry of the protein being characterized.[28,62,63] Therefore, the homology of the protein with the proteins in the same class is evaluated and given as percentage sequence identity and sequence similarity.[6,13,26]

The homologous proteins extracted from the PDB database are given in Table 3.1 indicating high confidence prediction scores in the helical areas. The beta-2-adrenergic receptor,[64] which also belongs to class A of the GPCRs, had the highest percent sequence identity (24.87%) to FLP18R1. Class A is the largest group of GPCRs (>90% of GPCRs)[1,65] and ligands bind in a typical cavity between the helices in this 7 transmembrane structure.[5]


Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
AA: MVSSAATISTISTTTTPSTISNVITSHSNNGSCIQIAEAI
10
20
30
40
Conf:
1
Pred:
Pred: CCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHH
AA: AAQGIDDITVDFYIRSIFTFLYGFLFVLGIFGNGGVLWAV
50
60
70
80
Conf:
2
Pred:
Pred: HHCCCCCCCHHHHHHHHHHHHHHHHHHCCCHHHHHHHCCC
AA: ARNKRLQSARNVFLLNLIFTDLILVFTAIPVTPWYAMTKD
90
100
110
120
Conf:
3
Pred:
Pred: CCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHCC
AA: WAFGSVMCHLVPLSNSCSVFVTSWSLTAISLDKFLHINDP
130
140
150
160
Conf:
4
Pred:
Pred: CCCCCCHHHHHHHHHHHHHHHHHHHCCCCHHCCCCCCCCC
AA: TKQPVSIRQALAITFLIWIVSTLINLPYLMSFEHVDGSFY
170
180
190
200
Conf:
5
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHH
AA: VQPGETPYCGHFCDEANWQSENSRKIYGTTVMLLQFVVPM
210
220
230
240
Conf:
6
Pred:
Pred: HHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHH
AA: AVITYCYFKILQKVSKDMIIQNAQFCQSLTQKQRSDATSR
250
260
270
280

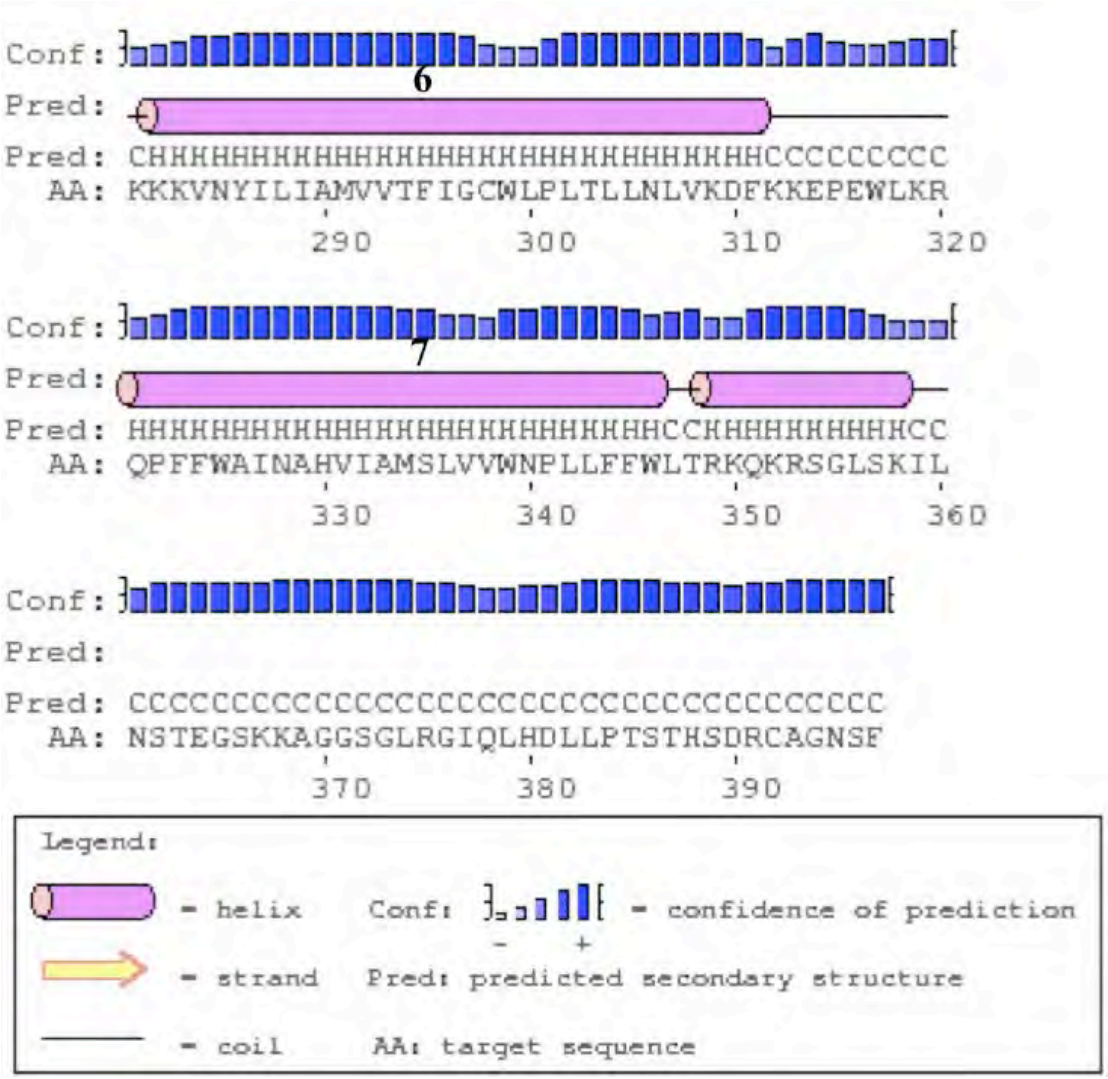


Figure 3.1: PSIPRED graphical output of the predicted secondary structure of FLP18R1 as produced by PSIPREDView.[32] *****The helices are indicated by numbers 1-7

3.3.3 Template selection and Sequence Alignment

Table 3.1 outlines the common ancestry of FLP18R1 obtained from PSIBLAST, which compares a primary sequence as the query to the protein database. ClustalW[6] was used for sequence alignment and computes pairwise alignment by searching the primary sequence against a library of domains. The PDB code and name of the protein are given specifically in parenthesis. The common ancestry of the protein is beta-2-adrenergic receptor, which is predicted as class A GPCR.[62,66] Therefore, the target protein is a 7 transmembrane (7TM) protein[29,67] and is shown in Figure 3.3.

Table 3.1 shows that the resolution of 4GBR_A, 3SN6_A, 4EA3_A,[68] 4DKL and 2R4R_A is high above that of 2RH1_A[64] and 4AMJ_A,[69] which is at 2.4Å.[27,56,58,70] Turkey beta-1 adrenergic receptor (4AMJ_A)[69] has a shorter chain (358 amino acids) than FLP18R1. Its sequence identity is lower than 2RH1. The chain lengths were also considered, and our target has 397 amino acid residues. Another reason why other templates were not selected is because of the absence of conserved residues in the helix 7. Beta-2-adrenergic receptor (**PDB ID code 2RH1** with a resolution of 2.4Å in the crystal structure)[64] with 422 amino acid residues was selected as the best template and the target (FLP18R1) with 397 amino acid residues. The chain length ensures that the target is fully covered. Table 3.1 gives the sequence identity as 24.87% after manual adjustment. The alignment using ClustalW was used for sequence alignment and is shown in Figure 3.2.

The aligned residues exclude the N terminal and C-terminal loops.[71] According to Yarnitzky (2010),[5] loopless[5] models postulate realistic complementary structures for docking and virtual screening applications as structure-based drug design approaches[5]. Therefore, the N-terminal and C-terminal loops were cut off and not used in the alignment (Figure 3.2).

Table 3.1 Statistical results from BLAST search showing common ancestry of FLP18R1. *The templates in blue were selected.

Template PDB ID_chain	**Max Score (%)**	**Resolution (Å)**	**Max identity (%)**	**Manual adjustment identity (%)**	**Query cover (%)**	**Chain length (AA)**
4GBR_A (beta-2-adrenergic)	81.6	3.99	27.18	22.98	77	309
3SN6_A (beta-2-adrenergic)	81.6	3.2	25.0	17.38	61	349
4EA3_A (Nociceptin opioid)	78.6	3.01	21.91	16.88	64	434

4DKL (mu-opioid)	75.9	2.80	23.68	21.41	47	464
3OAX_A (Bovine rhodopsin)	80.9	2.60	25.50	21.49	76	349
2R4R_A (beta-2-adrenergic)	77.8	3.4	24.66	22.47	77	365
*2RH1_A (beta-2-adrenergic)	70.1	2.4	25.18	24.87	59	500
4AMJ_A (turkey beta-1-adrenergic)	70	2.4	23.81	21.9	47	358
2Z73 (squid rhodopsin)	60.8	2.5	21.91	18.39	59	448

Knowledge-based information such as conserved regions or motifs (Table 3.2) was used for manual alignment of the automatically created pairwise alignment for the best transfer of match elements.

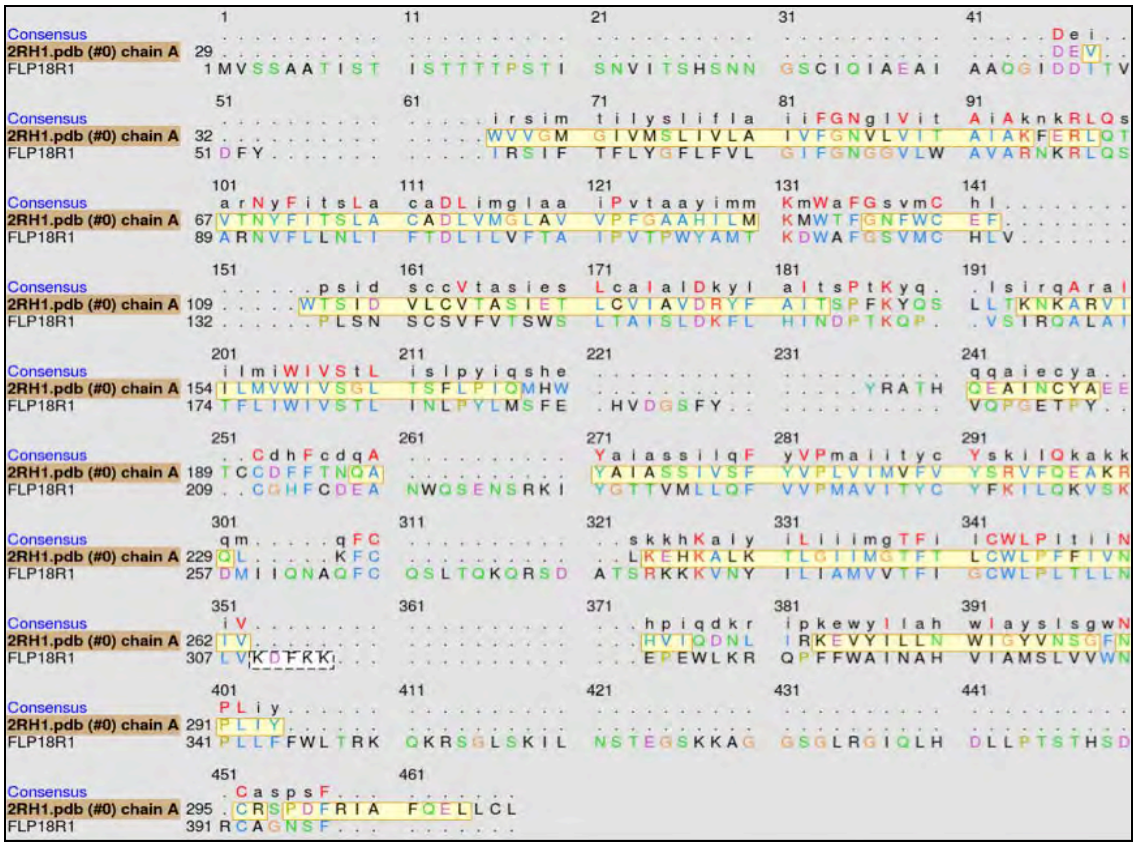


Figure 3.2: ClustalW alignment of FLP18R1 receptor (target) with beta-2-adrenergic (2RH1) receptor (template) with manual alignment

The sequence identity and sequence similarity of the GPCR per helix is shown in Table 3.3. Percent sequence identities and similarities between FLP18R1 and 2RH1 were calculated. The total number of the identical or similar residues was divided by number of residues in the shorter sequence to give % identity. The low sequence identity and similarity maybe due to the chain length of the target to that of the template.[6,13] The chain length of the template is longer than the target and this might have contributed to the low percent sequence identity.

The average sequence identity per helix is 25.31% and average sequence similarity is 53.39%. Similarity[39] depicts resemblance between the template, 2RH1, to the target, FLP18R1. The low sequence identity is possibly due to non-homology in the loop regions of the target to that of the template. The manual alignment focused more on the conserved residues in the helical regions.

Table 3.2 Highly conserved residues (bold) and patterns in Class A GPCRs and those in FLP18R1 receptor where *x* indicates an amino acid residue in the class A GPCRs[5]

Helix	Class A	FLP18R1

1	G*x* (3) **N**	**N**
2	L*x* (3) **D**	**D**
3	**E/ DRY**	**DKF**
4	**W** /W*x* (7, 8) P	**W**IVSTLINLP
5	**P**/ F*x*P*x* (7) Y	**P**
6	P/ F*x***CW***x*P	**CWL**
7	L*x* (5) **NP*x* (2) Y**	**NPLLF**

Each helix was aligned against each of the 7 transmembrane helices of beta-2 adrenergic receptor (Figure 3.2 and 3.3) The conserved residues Asparagine (N) in H1, Aspartic acid (D) in H2, Aspartic Acid (D), Lysine (K) and phenylalanine (F) in H3, tryptophan (W) in H4, Proline (P) in H5, Cysteine (C), Tryptophan (W), Leucine (L) in H6, and Asparagine (N), Proline (P), Leucine (L) and phenylalanine (F) in H7. These are important residues and some like the DKF motif are involved in receptor activation.[10,12]

Table 3.3 Sequence Similarity and sequence identity of FLP18R1 with 2RH1

Transmembrane (TM)	**Sequence Identity (%)**	**Sequence Similarity (%)**
1	20.7	52.0
2	24.7	59.7
3	17.9	46.4
4	37.2	51.9
5	19.4	52.9
6	34.5	65.3
7	22.8	45.5

Sequence alignment is also used to find characteristic motifs and conserved residues in protein families and to improve the secondary structure prediction elements as shown in Table 3.1. The observed high degree of conservation[5] among these key residues suggests that they have an essential role for either the structural or functional integrity of FLP18R1 receptor. The conserved residues in this study are shown in Table 3.2. The results of multiple sequence alignment with up to 3 templates and homology models are given in Table 3.1. Even though the sequence identity was

24.87%, which was considerably low, the model produced was of acceptable quality for GPCRs.

3.3.4 Homology model

The homology model of FLP18R1, in Figure 3.3, was built based on the atomic coordinates of beta-2-adrenergic receptor (2RH1). The model indicates that the receptor has an N-terminus loop that is 46 amino acids long, and a C-terminal intracellular loop with 37 amino acids, extending in the cytoplasm. The C-terminal has been shown to be responsible for binding to G-proteins in the cytoplasm.[57,71,72]

The absence of a short helix in the homology model on the extracellular loop 2 (EC2) was not anticipated. The construction of the ECL2 loop is extensive and has a concealed ß-sheet structure in the EC2 of rhodopsin. Structure optimization and refinement using 100 ns molecular dynamic simulations of the homology model result in this short helix appearing in EC2. This short helix has been found in EC2 of class A GPCRs in the transmembrane helices between H4 and H5 as depicted in Figure 3.4.[66,73] Some class A GPCRs with this short helix includes turkey beta-1-adrenergic[12] receptor (2VT4) and beta-2-adrenergic receptor (2RH1).[64] However, 2RH1 has different architecture in the EC2 as it is more exposed to solvent than in rhodopsin.[74]

The disulfide bond[59,65,75] observed in Class A GPCRs is also found in the FLP18R1 homology model between Cys128 on helix 3 to the Cys209 on extracellular loop 2 (ECL2). In rhodopsin and adrenoceptor GPCRs, the disulfide bonds that form between ECL2 and H3 are common. The disulfide bonds have been proven to exhibit important functional characteristics to the protein stability. The disulfide bond between ECL2 and H3 is believed to be responsible for maintaining the inactive GPCR state. This is a very important bond that breaks during docking of ligands into the binding pocket. For example, in rhodopsin[5,76,77] receptors, the ECL2 presents a cap to the binding pocket and thus make it inaccessible for docking.[75]

However, in beta-2-adrenergic receptors[59,74], the ECL2 loop does not directly cap the binding pocket and this maintains the disulfide bond in the inactive state, which

breaks when a ligand binds to the binding pocket.[59,77] However, class A GPCRs have variable ECL2 lengths, which tend to affect the position of the disulfide bonds usually found between a cysteine residue on ECL2 and H3. This is understood to be a very important bond, which stabilizes the accurate folding and structural pairing between the H3 and ECL2 of the protein between these conserved cysteine residues in these positions[5,57,76,78].

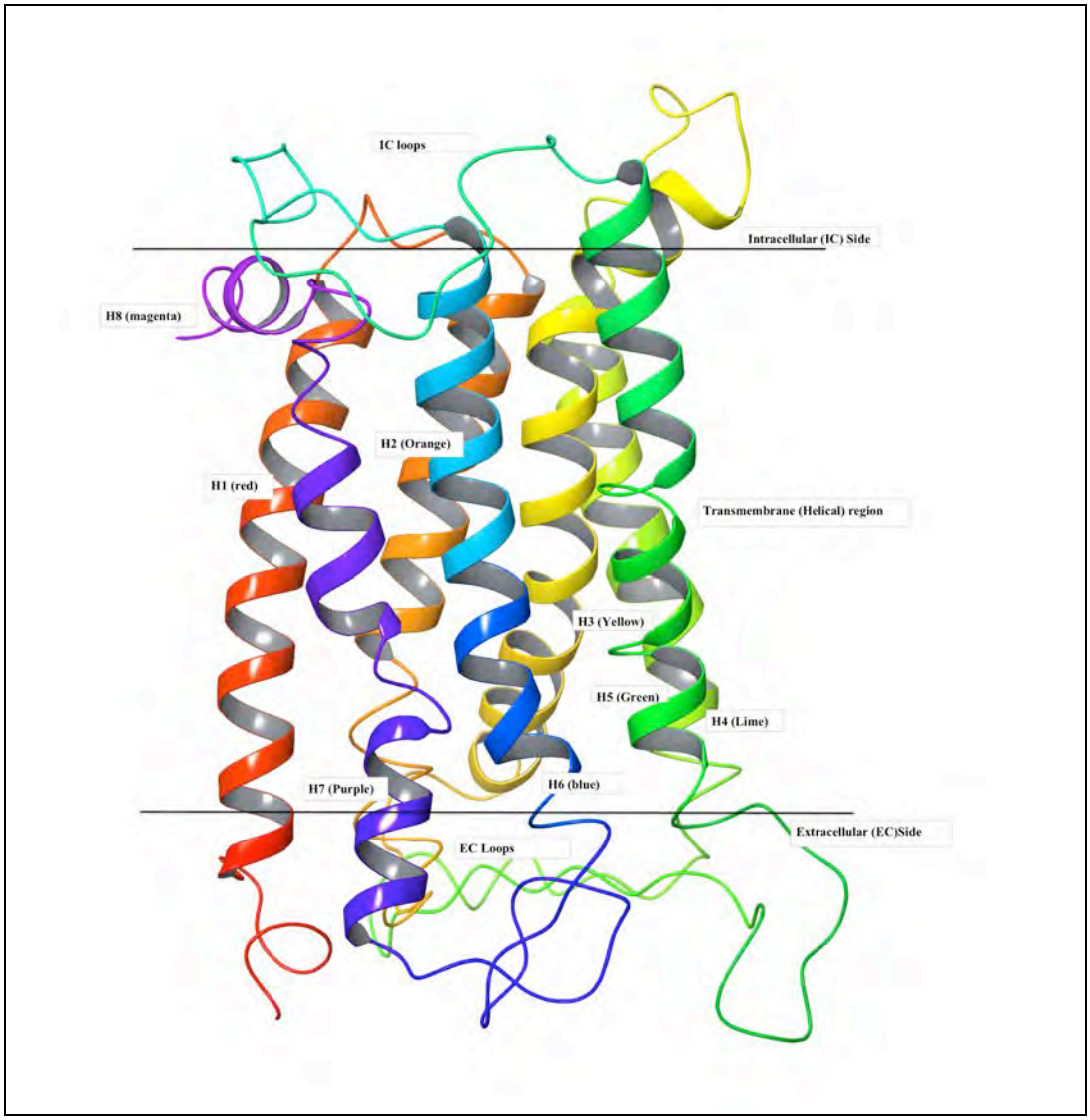


Figure 3.3: Homology model of FLP18R1 receptor showing extracellular loops (ECL) and intracellular loops (ICL) and the 7 transmembrane helices (H1-H7)

The second disulfide bond is between Cys202 and Cys213. However, this bond is not as important as the one involving Cys128 to Cys209. *In silico* mutation of Cys128 with an alanine shows that the stability of the receptor is affected as the folding is no

longer as it was using this important bond[79]. The disulfide bonds are preserved throughout the molecular dynamic simulations. The superimposition of the homology model (FLP18R1) with the template (2RH1) in Figure 3.4 helps explain the similarity and homology of the two proteins.

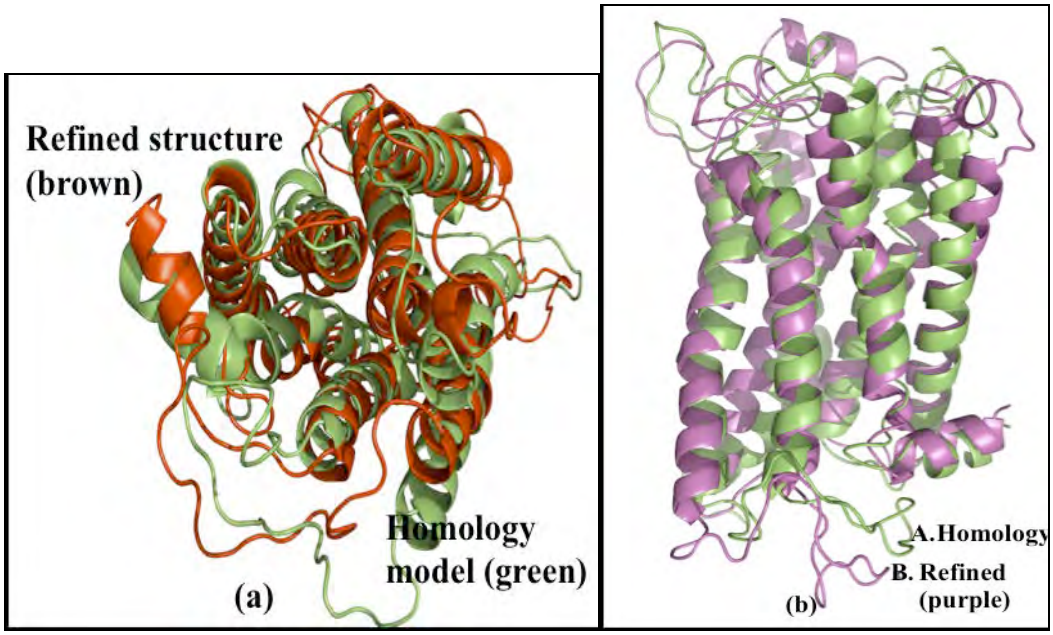


Figure 3.4 Superimposition of Homology model (green/A) with the template 2RH1 (brown/magenta /B) shown from intracellular side and side elevation. The N-terminal and C-terminal residues have been removed.[5]

Homology in the helical regions is shown in Figure 3.4. It can be deduced from this diagrammatic representation that extracellular loop 2 (EC2) in the FLP18R1 (target-green) in Figure 3.4 is structurally different from 2RH1. The 2RH1 has a short helix in ECL2, which is absent in the homology model of FLP18R1. The short beta helix is thought to have resulted from the ancestry in class A GPCRs of beta-2-adrenergic receptors. However, the 8th α-helix is also present in both template and target.

Another important attribute of GPCR structures is the ionic lock that exists between H3 and H6 that is shown in Figure 3.5. This is exhibited in most inactive class A GPCRs. However, this ionic lock between DRY[10,12] in H3 and the H6 cytoplasmic end residues is not found in beta-2-adrenergic and in the crystal structure of μ-opioid receptor.[58] In 2RH1, arginine (Arg165) in H3 has been found to form a salt bridge instead with the aspartic acid of the DRY motif (Asp164). During molecular dynamic simulations, there is an inward movement of H6 towards H3 in the intracellular side.

This movement of the helices on the intracellular side has an average RMSD of 3.5Å and is shown in some of the class A GPCRs.[80]

The variations in the length of EC2 class A GPCRs including bovine rhodopsin, human H_4 histamine receptor and beta-2 adrenoceptor (ß2AR)[12] tend to have an effect on the position on the disulfide bonds, usually found between EC2 and H3. However, some receptors in class A such as A_2A adenosine receptor, beta-1 adrenergic and human beta-2 adrenergic receptor do not have an ionic lock between H3 and H6, which is mostly a salt bridge formed by non-covalent interactions[10,69,81,82]. Arg73 is conserved in H1 and forms hydrogen bonds to carbonyl backbone of Ala333 in H7. The conserved NPLLF in H7 plays a significant role. Arg73 also form a 2.8Å hydrogen bond with the OD1 of Asp101 on H2. This interaction is also observed in most class A GPCRs[10] and this residue is conserved class A GPCRs.

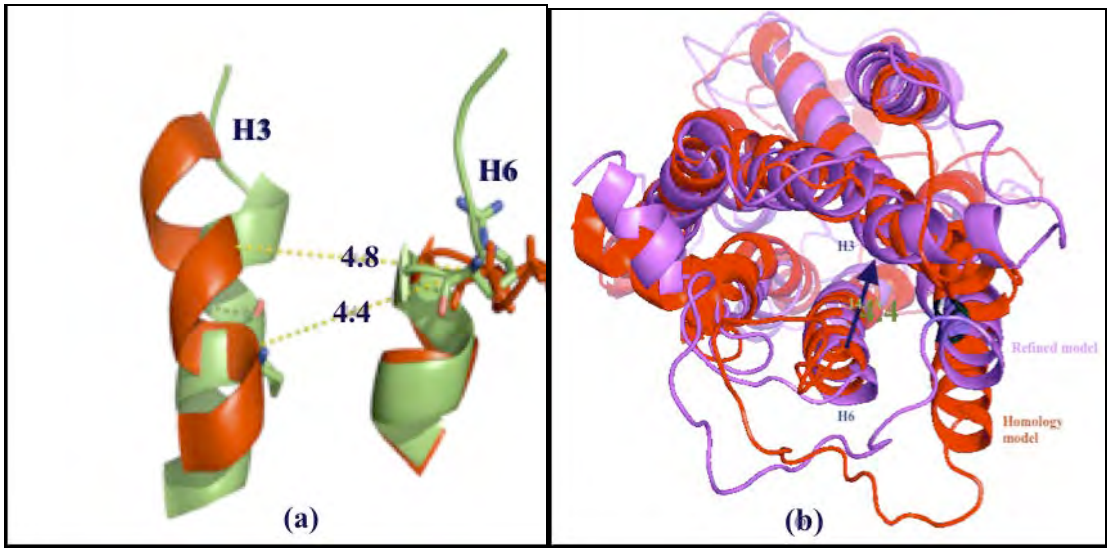


Figure 3.5: (a) The ionic lock in FLP18R1. The brown indicates the homology model and green is the MD refined structure. In (b), The inward movement of the H6 is shown towards H3 using a black arrow in the purple structure. The RMSD between the 2 structures is on average 3.5Å

The side chains of Gly63, Gly69, Gly74 and Gly75 of H1 are small and so are those of Ala329, Ala333 as well as Ala326 in H7. This allows the close packing of H1 and H7 that is commonly observed in class A GPCRs. This is thought to facilitate hydrogen bonding. The alanine and glycine residues are conserved in the helices.[66,73] The asparagine (Asn96) exhibits cation-pi (π) stacking of its guanidinium amide proton to the π-electrons of the aromatic ring tryptophan (Trp178) that is shown in

Figure 3.8. Tryptophan in H4 is conserved in 97% of the GPCRs in class A.[73] The template used in homology modeling, 2RH1, is partially responsible for the beta sheet inherited in ECL2 of the refined model, which is in good agreement with GPCRs in this class such as rhodopsin, μ-opioid, beta-2-adrenergic receptors.[68,77]

The aspartic acid (Asp152) of the canonical DKF motif in H3, which is conserved in most class A GPCRs in this class forms a bond with the arginine (Arg280) in helix 6 (H6) as in beta-2-adrenergic receptor (2RH1). The ionic lock is shown in Figure 3.5a. The Asp152 is conserved in FLP18R1 as part of the DKF motif. The salt bridge, which is an interaction between an electrostatic interaction and a hydrogen bond, in this instance, is between the anionic carboxylate ($RCOO^{-}$) of Asp152 in H3 and the positively charged side chain nitrogen (guanidinium ($RNHC(NH_2)_2^{+}$)[12,83] atom of Arg280 in H6.

The distance between the Asp152 and Arg280 in the homology model is 4.8Å and is 4.4Å in the refined structure. This reduction of the distance by 0.4Å explains the inward anticlockwise-clockwise rotation and inward movement of H3 towards H6 in the cytoplasmic side. Figure 3.6 show the orientation of the helices in the intracellular and extracellular side before and after molecular dynamic simulations. Studies[10,12] that have been published on GPCRs support this movement as shown in Figure 3.6.

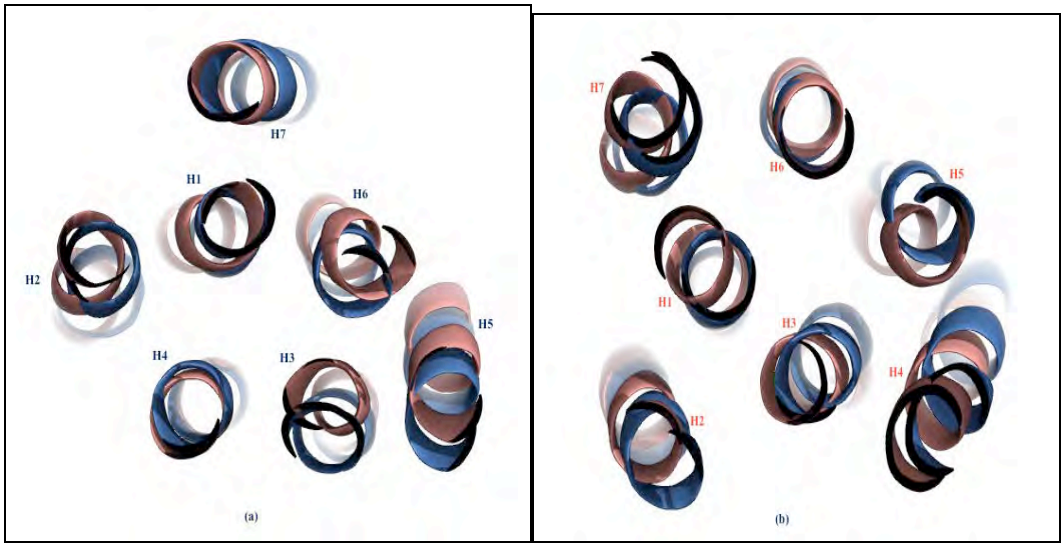


Figure 3.6: (a) Aerial view of extracellular side has shown residues 12Å from the extracellular side. (b) The intracellular side highlights the anticlockwise-clockwise and

inward movement of H6 towards H3. The blue indicates the refined structure after molecular dynamics and the brown highlights the homology model

GPCRs such as human H_4 histamine[12,84] have an ionic lock between the arginine residue of the DRY motif in H3 and either a Asp/ Glu motif in H6 of this receptor.[12] In chemokine receptor 5 (CCR5), it was found that the constitutive activity of the receptor is enabled by the integrity of the DRY motif in H3. This motif is needed for ligand binding in the CCR receptors.[85–87] The aspartic acid residue (Asp125) in H3 forms an ionic lock[83,85,86] with an arginine (Arg225) in H6, which is responsible for the binding of the TAK779 (N,N-dimethyl-N- [4-[[[2-(4-methylphenyl)-6,7-dihydro-5H-benzocyclohepten-8yl]carbonyl]amino]benzyl]tetrahydro-2H-pyran-4-aminium chloride), Figure 3.7 which has been shown to be a non-peptidic CCR5 ligand that inhibits HIV infection.

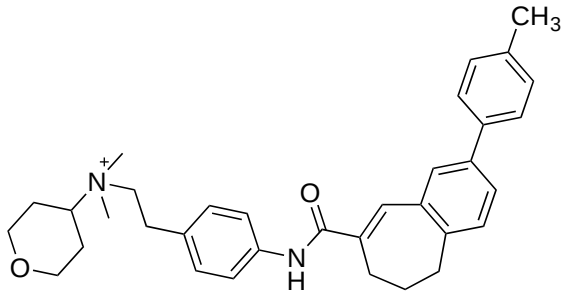


Figure 3.7: Non-peptidic CCR5 Ligand TAK779 (N, N-dimethyl-N- [4-[[[2-(4-methylphenyl)-6,7-dihydro-5H-benzocyclohepten-8yl] carbonyl] amino] benzyl] tetrahydro-2H-pyran-4-aminium chloride) inhibits HIV infection.

Figure 3.8 shows the FLP18R1 in the POPC membrane. It also shows water head groups in the solvated membrane. Figure 3.8b shows the different colours of the helices. Mimetic membranes used for performing explicit membrane protein molecular dynamic simulations include pre-equilibrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)[43,44,49,50], 1,2-dimyristoylsn-glycero-3-phosphocholine (DMPC) and dipalmitoyl phosphatidyl choline (DPPC)[49,88,89] molecules. In this work POPC was used for performing explicit membrane protein simulations in GROMACS.[43,44,50]

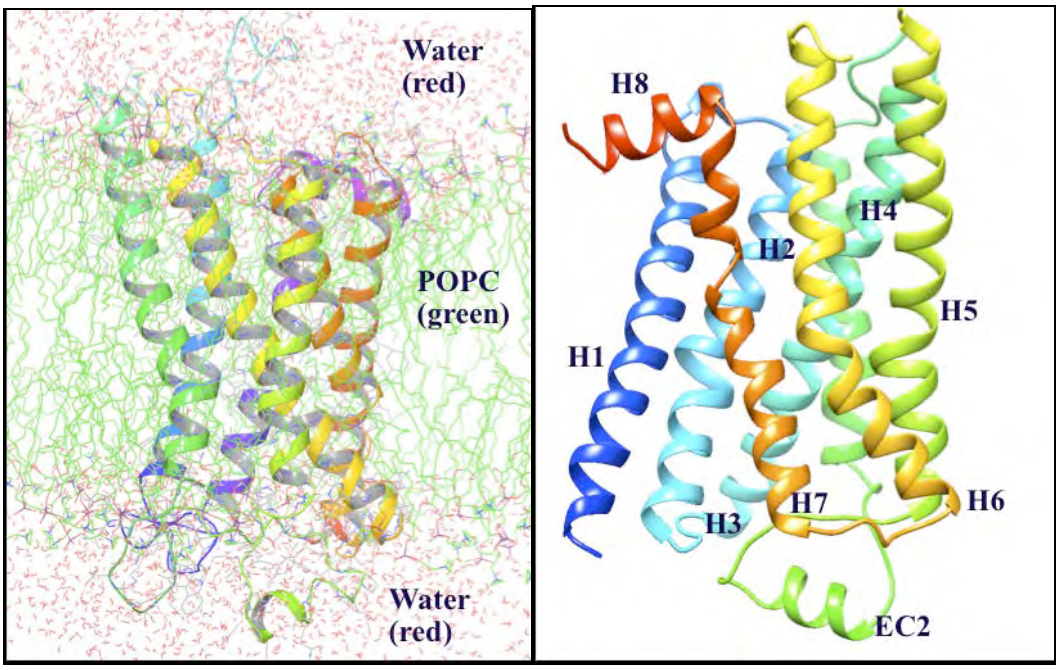


Figure 3.8: (a) FLP18R1 receptor (orange shows helices and green loops) in the POPC membrane (lime) and water molecules (red) and (b) show the FLP18R1 after molecular dynamics simulations. Note the short helix that forms on EC2 in between H4 and H5 [green in (b)], which is typical in GPCRs after MD

3.3.5 Loop Refinement using molecular dynamic simulations

Homology modeling was made by accurately modeling the helical regions without consideration of the loops. The exclusion of loops is mostly due to variations in loop regions of templates and target proteins such as length of loop and low sequence identity.[6,13] In careful modeling of helical regions, deleting gaps in helical regions is performed mostly using manual alignment. In addition, the alignment of conserved residues is done with the inherent hypothesis that ligand interaction occurs with transmembrane residues and thus homology models do not essentially encompass the loop regions. This hypothesis has since been dismissed following experimental results on the protein-ligand interactions with the kappa-opioid receptor,[58] where effectively the second extracellular loop takes part in ligand binding and is essentially part of the binding site[79]. Techniques that have been used to model loops include *ab initio*[90,91] computational approaches, database searching and molecular dynamic simulations.

During loop refinement, the helical packing improved. Table 3.4 (Section 3.6) shows

that the calculated volume of packing decreased to 44588.2 $Å^3$ from 44825.8 $Å^3$. The (Volume Area Dihedral Angle Reporter) VADAR version 1.8 server available on http://vadar.wishartlab.com/ that facilitates calculating volume, area and dihedral angles[92] was used.

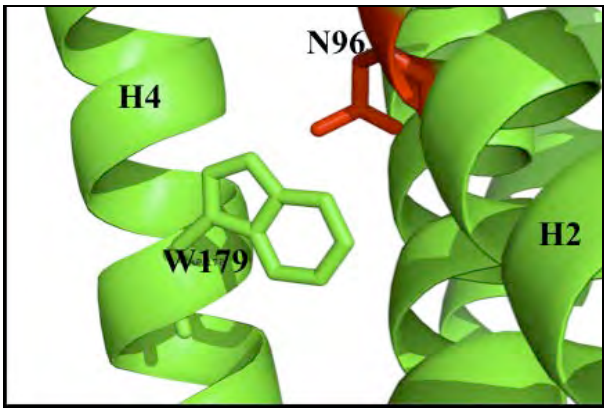


Figure 3.9: Cation-pi stacking between the amide proton of asparagine (Asn96) [red] to the negative pi (π) electrons of the aromatic tryptophan (Trp178) [green]

3.3.6 Model evaluation

Assessment of the stability of the homology model after it was refined by molecular dynamics was undertaken by calculating the Ramachandran plot,[93,94] time dependence of the root mean square fluctuation (RMSF) from the selected homology model-starting structure, and finding the radius of gyration (Rg) and the root mean square deviation; (see next paragraphs). Model evaluation was performed to approximate the stability of protein molecules of the receptor in water and in mimetic membrane.

RMSF evaluates the deviation of the position of the atoms in the receptor relative to the reference structure and is dependent on the time spent in that position.[95] The average rmsf of FLP18R1 is 0.1nm (Figure 3.10). The number of atoms in this receptor is 3443. The root mean standard deviation (RMSD) and RMSF are shown in Figure 3.10 and 3.11. This explains fully the deviations in the two structures during molecular dynamics.[96] High spikes in the loop regions were observed as shown in Figure 3.10. These are believed to result from structural instability because of inefficient restraints on this region; the main restraint with associating loop residues in the POPC membrane but allowed to simulate to optimize to its ideal structure.

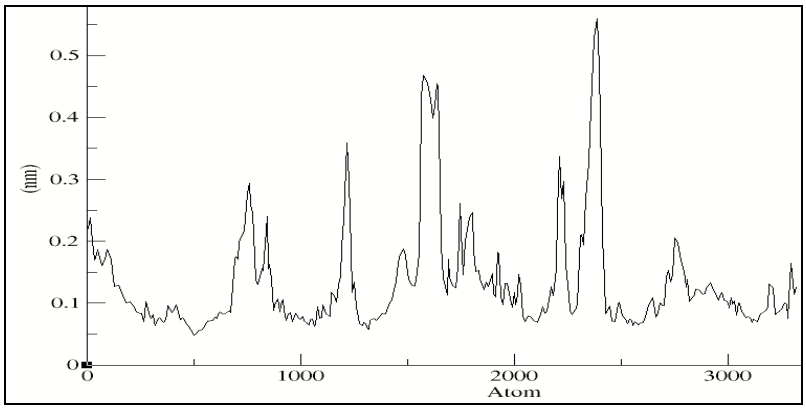


Figure 3.10: Root mean square fluctuation of the starting structure, FLP18R1, which is the reference structure. The spikes show the loop regions

Figure 3.11 shows that there was an exponential[22,41,46] increase in the deviation in the first 10 ns of the simulation with a RMSD of 0.3 ns. The RMSD stabilized after about 40 ns as shown in Figure 3.11. This may possibly be due to stabilization of the model after running molecular dynamic simulations of the model in a POPC-water box.

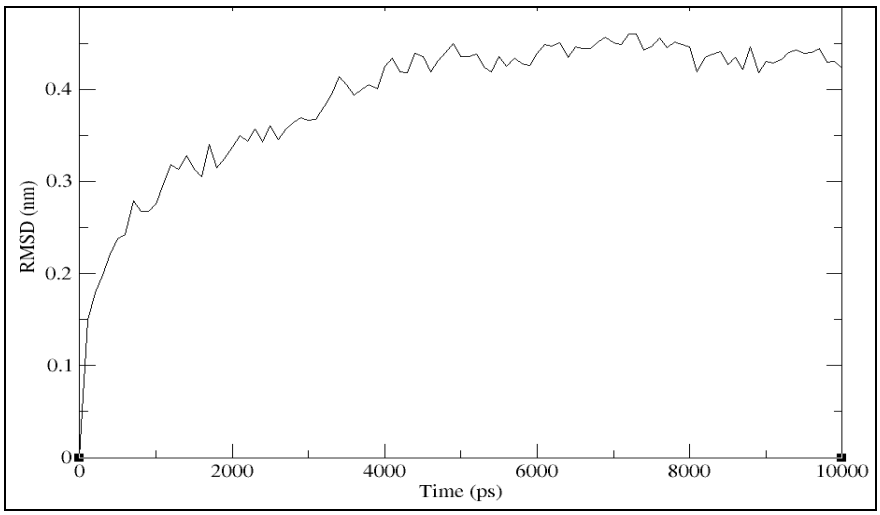


Figure 3.11: RMSD of the FLP18R1 from the starting structure after 100 ns.

The dynamic system reached a stable point after 40 ns and tailed off with an average RMSD of 0.4 nm. This observation correlates with the variation in the RMSD observed in Figure 3.10. The helices are opened up in the extracellular region as it simulates in the first 1.5 ns

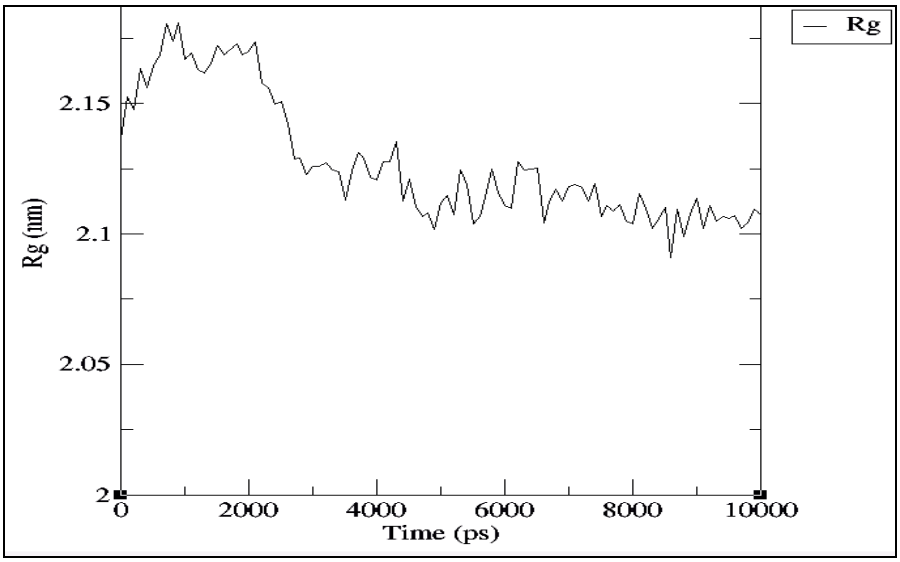


Figure 3.12: The radius of gyration (Rg) of the model after molecular dynamic refinement

The Rg denotes the firmness or compactness of the protein structures. The average Rg of the receptor was found to be 21.7Å (Figure 3.12). The Rg decreased during the 100 ns MD simulations and reached stability after about 35 ns as shown in Figure 3.11. Throughout the simulation, the compactness of the structures was noted.

The data in Table 3.4 show that MD refinement improved the packing of the residues in the structure, particularly in the loop regions. This information confirms the effectiveness of the molecular dynamic simulations.

Table 3.4 The packing calculations were based on the atomic radii using Shrake and Rupley (1973). (aa: amino acids)

Statistics	**Observed**	**MD refinement**
Helix	183 aa (58%)	216 aa (68%)
Beta (ß)	27 aa (8%)	10 aa (3%)
Coil	104 aa (33%)	88 aa (28%)
Turn	52 aa (10%)	36 aa (11%)

Table 3.5 elaborates the changes in the hydrogen bonds in the two structures, the starting (A) and the refined structure (B). Molecular dynamics helped in packing the structure as shown in the improvement of hydrogen bond distance that resembles the expected hydrogen bond length in X-ray and NMR structures (Table 3.5). In a study done by Morris (1992), the overall volume packing improved by 0.237 Å. It has been

mentioned that the protein was optimized in explicit molecular dynamic simulations in a water box.

Table 3.5 Hydrogen bond occupancy in the structures[97]

Statistics	**A**	**B**	**Expected (NMR/X-ray**
Mean h-bond distance	2.4 sd = 0.3	2.2	2.2 sd = 0.4
Mean h-bond energy	-1.4 sd = 0.8	-1.8 sd = -1.1	-2.0 sd = 0.8
Residues with h-bonds	240 (76%)	262 (83%)	235 (75%)
% Side Accessible Surface Area	51.26%	50.98%	

The interactions of the water-protein charged and polar head groups were calculated and are shown in Table 3.5. It follows from the table that hydrogen bond occupancy[97] increased to 83% from 76%, after performing molecular dynamics. That is an increase of 9.2% from the homology model. The 83% hydrogen bond occupancy shows an improvement of nearly 10% from the anticipated 75% with a mean hydrogen bond distance of 2.2 Å. Therefore, molecular dynamic simulations show that the loop refinement improved the packing in the protein homology model. The weakness of hydrogen bond occupancy is that it does not provide atomic details of the interactions. This information is, however, adequately enunciated using accessible surface area calculations.

The surface area of the protein that is exposed to the solvent was calculated using the VADAR[92] tools available on http://vadar.wishartlab.com/. The Accessible Surface Area (ASA) shown in Figure 3.13 is defined as the surface area of the protein (or residue) that an exposed water molecule could access or touch. The value calculated as the ASA column corresponds to the fractional accessible surface area, which helps in the assessment of protein structure quality. The fractional ASA ranges from 0.00 up to 1.00 $Å^2$. It follows from Figure 3.13 that the helical regions that are buried in the POPC membrane have ASA near zero, and the loop areas (Ala81-Ser88, Lys119-Ser125, Asp159-Gln169, Val195-Thr206, Phe212-Tyr227, Met258-Arg274, Phe311-Glu314, and Phe345-Lys350) that are more exposed in the cytoplasm and

extracellular domains are more exposed to the solvents and the ASA is typically between 0.5 to 1.00Å^2.

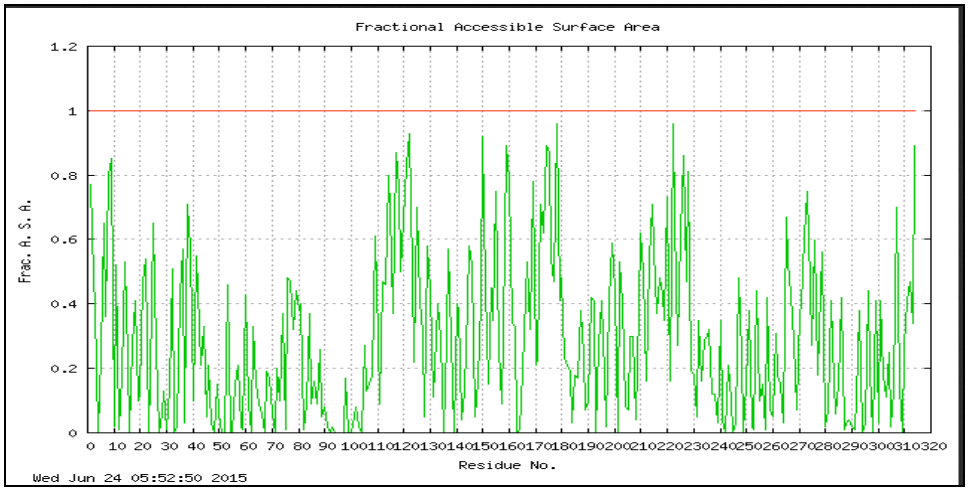


Figure 3.13: Fractional accessible surface area (ASA) of the receptor

The fractional ASA is used as a measure of the hydrophobicity[98] of the residues in a protein. There is a correlation between the fractional ASA and the hydrophobicity as revealed by the folding of the protein as shown in Figure 3.4 discussed earlier. From the fractional ASA of between 0.6-0.84 Å^2 (shown in Figure 3.13, the small beta strand in EC2 (Lys309-Glu314), this highlights the exposure of the region to the solvent and is not buried as the other helical areas. There is therefore proportionality between the hydrophobicity and the fractional ASA in this protein. This corroborates the known typical transmembrane GPCR structures that reveal that helices are hidden in membrane domains.

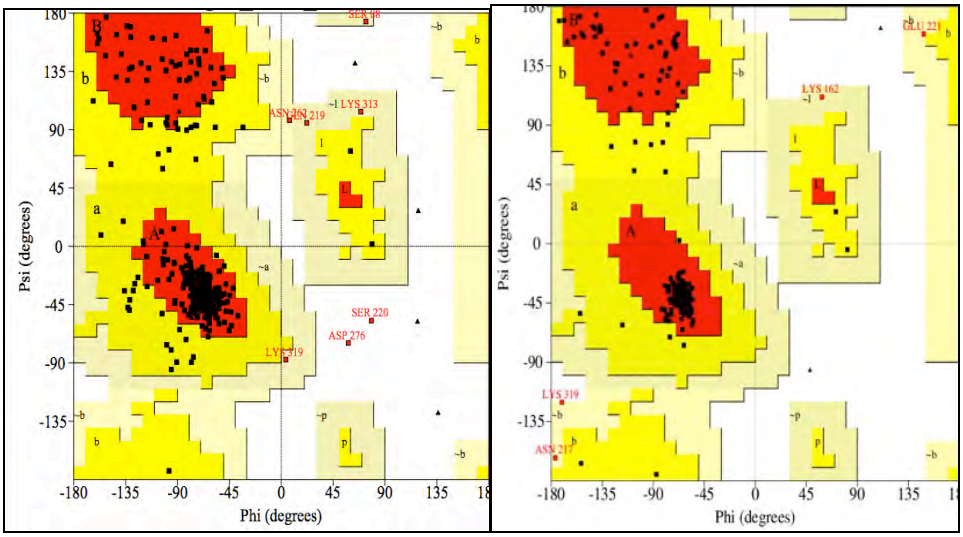


Figure 3.14: Ramachandran plots of the homology model (left) and MD refined structure (right).

The model validation was performed using a Ramachandran plot, Figure 3.14. Ramachandran plot help with the visualization of the backbone dihedral ψ against φ of the protein amino residues in the protein structures.[93,94] The Ramachandran plot of the homology model shows that 85.4% of the residues are in the core allowed regions, and that 9% are outliers. These include Ser88, Lys313, Gln219 and Ser220 among others. Three residues (Lys162, Gln221, Lys319) show poor chirality. When the structure was refined using molecular dynamic simulations over 100 ns, the structure improved significantly.

The allowed core region residues were observed to be 289 amino acid residues, which gives an average of 94.9% allowed residues in the Ramachandran plot (Figure 3.14) and only 6 residues were noted to be in the disallowed regions against 11 residues in the homology model. However, molecular dynamic simulations helped in refining the homology model as can be seen by the variations in the Ramachandran plot before and after molecular dynamic simulations. Therefore, after simulations, Lys162, Arg217, Gln221 and Lys319 are the only residues in disallowed regions. Therefore, the model is of acceptable quality and was used in further work reported in the following chapters.

3.4 Conclusion

The homology model of the FLP18R1 was successfully built and was of acceptable quality. The beta-2 adrenergic receptor was used as a template for homology modelling. The sequence identity was low at 24.87%. However, this is common for GPCR homology modelling. ClustalW[6,34,99] was used for pairwise alignment of the target and template and Modeller version 9.11 was used for homology modeling. Loop refinement was performed through molecular dynamic simulations over 100 ns. The structure stabilized after 35 ns as shown in the RMSD plot. The radius of gyration also indicates that the structure improved its compactness[100] with time, and it became constant after 35-40 ns and was at 0.2 nm on average. This may be attributed to the improvement in the hydrogen bonding during molecular dynamics and close packing.

The disulfide bonds between Cys128 and Cys209 stabilize the structure.[65,75,101] The DKF motif also plays a constitutive role by forming a salt bridge with Arg280 in H6 from Asp152. This ionic lock has been suggested to be formed between the conserved DRY[10,12] motif of H3 and Arg (R) in H6. The variation of this ionic lock was 0.4 Å.[59] This is a common anticlockwise-clockwise rotation of the H6 towards H3 in the intracellular side of the receptor. Other salt bridges[73,77] that stabilize the structure include the cation- π stacking of the amide proton in Arg96 and π electrons of the aromatic ring of Trp178 (W178).

Loop refinement suggested an inherited beta sheet on EC2, which is an important region in this GPCR as it has been found to be part of the binding pocket. Therefore, the protein structure determined in this study using homology modeling and refined using molecular dynamic simulations was found to be of acceptable quality. The final structure will be used for docking calculations to be discussed in Chapter 4.

3.5 References

(1) Jacob, L.; Hoffmann, B.; Stoven, V.; Vert, J. P. Virtual Screening of GPCRs: An *in Silico* Chemogenomics Approach. *BMC Bioinformatics* **2008**, *9*, 363.

(2) Roumen, L.; Sanders, M. P. A; Vroling, B.; de Esch, I. J. P.; de Vlieg, J.; Leurs, R.; Klomp, J. P. G.; Nabuurs, S. B.; de Graaf, C. *In Silico* Veritas: The Pitfalls and Challenges of Predicting GPCR-Ligand Interactions. *Pharmaceuticals* **2011**, *4* (9), 1196–1215.

(3) De, A.; Veulens, N.; Rodríguez, R. G Protein-Coupled Receptors as Targets for Drug Design. *Biotecnol. Apl.* **2009**, *26* (1), 24–33.

(4) Haupt, V. J.; Daminelli, S.; Schroeder, M. Drug Promiscuity in PDB: Protein Binding Site Similarity Is Key. *PLoS One* **2013**, *8* (6).

(5) Yarnitzky, T.; Levit, A.; Niv, M. Y. Homology Modeling of G-Protein-Coupled Receptors with X-Ray Structures on the Rise. *Curr. Opin. Drug Discov. Devel.* **2010**, *13* (3), 317–325.

(6) Reddy, C. S.; Vijayasarathy, K.; Srinivas, E.; Sastry, G. M.; Sastry, G. N. Homology Modeling of Membrane Proteins: A Critical Assessment. *Comput. Biol. Chem.* **2006**, *30* (2), 120–126.

(7) Sali, A. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Mol. Med. Today* **1995**, *1* (6), 270–277.

(8) Lundstrom, K. An Overview on GPCRs and Drug Discovery: Structure-Based Drug Design and Structural Biology on GPCRs. *Methods in molecular biology (Clifton, N.J.)*, **2009**, *552*, 51–66.

(9) Tapaneeyakorn, S.; Goddard, A. D.; Oates, J.; Willis, C. L.; Watts, A. Solution- and Solid-State NMR Studies of GPCRs and Their Ligands. *Biochim. Biophys. Acta - Biomembr.* **2011**, *1808* (6), 1462–1475.

(10) Vohra, S.; Taddese, B.; Conner, A. C.; Poyner, D. R.; Hay, D. L.; Barwell, J.; Reeves, P. J.; Upton, G. J.; Reynolds, C. A. Similarity between Class A and Class B G-Protein-Coupled Receptors Exemplified through Calcitonin Gene-Related Peptide Receptor Modelling and Mutagenesis Studies. *J R Soc*

Interface **2013**, *10* (79), 20120846.

(11) Bondensgaard, K.; Ankersen, M.; Thøgersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of Privileged Structures by G-Protein Coupled Receptors. *J. Med. Chem.* **2004**, *47* (4), 888–899.

(12) Schneider, E. H.; Schnell, D.; Strasser, A.; Dove, S.; Seifert, R. Impact of the DRY Motif and the Missing "Ionic Lock" on Constitutive Activity and G-Protein Coupling of the Human Histamine H4 Receptor. *J. Pharmacol. Exp. Ther.* **2010**, *333* (2), 382–392.

(13) Tramontano, A. Homology Modeling with Low Sequence Identity. *Methods* **1998**, *14* (3), 293–300.

(14) Kryshtafovych, A.; Fidelis, K. Protein Structure Prediction and Model Quality Assessment. *Drug Discov. Today* **2009**, *14* (7-8), 386–393.

(15) Kalyaanamoorthy, S.; Chen, Y. P. P. Modelling and Enhanced Molecular Dynamics to Steer Structure-Based Drug Discovery. *Prog. Biophys. Mol. Biol.* **2014**, *114* (3), 123–136.

(16) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discov. Today* **2009**, *14* (13-14), 676–683.

(17) Evers, A.; Klabunde, T. Structure-Based Drug Discovery Using GPCR Homology Modeling: Successful Virtual Screening for Antagonists of the alpha1A Adrenergic Receptor. *J. Med. Chem.* **2005**, *48* (4), 1088–1097.

(18) McVeigh, P.; Leech, S.; Mair, G. R.; Marks, N. J.; Geary, T. G.; Maule, A. G. Analysis of FMRFamide-like Peptide (FLP) Diversity in Phylum Nematoda. *Int. J. Parasitol.* **2005**, *35* (10), 1043–1060.

(19) Kubiak, T. M.; Larsen, M. J.; Bowman, J. W.; Geary, T. G.; Lowery, D. F. FMRFamide-like Peptides Encoded on the Flp-18 Precursor Gene Activate Two Isoforms of the Orphan *Caenorhabditis Elegans* G-Protein-Coupled Receptor Y58G8A.4 Heterologously Expressed in Mammalian Cells. *Biopolym. - Pept. Sci. Sect.* **2008**, *90* (3), 339–348.

(20) Geary, T. G.; Kubiak, T. M. Neuropeptide G-Protein-Coupled Receptors, Their

Cognate Ligands and Behavior in *Caenorhabditis Elegans*. *Trends Pharmacol. Sci.* **2005**, *26* (2), 56–58.

(21) Wang, Y. T.; Chan, C. H.; Su, Z. Y.; Chen, C. L. Homology Modeling, Docking, and Molecular Dynamics Reveal HR1039 as a Potent Inhibitor of 2009 A (H1N1) Influenza Neuraminidase. *Biophys. Chem.* **2010**, *147* (1-2), 74–80.

(22) Dalton, J. A. R.; Jackson, R. M. Homology-Modelling Protein-Ligand Interactions: Allowing for Ligand-Induced Conformational Change. *J. Mol. Biol.* **2010**, *399* (4), 645–661.

(23) Schmidt, T.; Bergner, A.; Schwede, T. Modelling Three-Dimensional Protein Structures for Applications in Drug Design. *Drug Discov. Today* **2014**, *19* (7), 890–897.

(24) Sanders, M. P. a; Verhoeven, S.; De Graaf, C.; Roumen, L.; Vroling, B.; Nabuurs, S. B.; De Vlieg, J.; Klomp, J. P. G. Snooker: A Structure-Based Pharmacophore Generation Tool Applied to Class A GPCRs. *J. Chem. Inf. Model.* **2011**, *51* (9), 2277–2292.

(25) Vilar, S.; Ferino, G.; Phatak, S. S.; Berk, B.; Cavasotto, C. N.; Costanzi, S. Docking-Based Virtual Screening for Ligands of G Protein-Coupled Receptors: Not Only Crystal Structures but Also *in Silico* Models. *J. Mol. Graph. Model.* **2011**, *29* (5), 614–623.

(26) Yoshikawa, Y.; Oishi, S.; Kubo, T.; Tanahara, N.; Fujii, N.; Furuya, T. Optimized Method of G-Protein-Coupled Receptor Homology Modeling: Its Application to the Discovery of Novel CXCR7 Ligands. *J. Med. Chem.* **2013**, *56* (11), 4236–4251.

(27) Schneider, M.; Wolf, S.; Schlitter, J.; Gerwert, K. The Structure of Active Opsin as a Basis for Identification of GPCR Agonists by Dynamic Homology Modelling and Virtual Screening Assays. *FEBS Lett.* **2011**, *585* (22), 3587–3592.

(28) Jones, D. T. Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *J. Mol. Biol.* **1999**, *292* (2), 195–202.

(29) Jones, D. T. Improving the Accuracy of Transmembrane Protein Topology

Prediction Using Evolutionary Information. *Bioinformatics* **2007**, *23* (5), 538–544.

(30) Nugent, T.; Jones, D. T. Transmembrane Protein Topology Prediction Using Support Vector Machines. *BMC Bioinformatics* **2009**, *10*, 159.

(31) *C. elegans* sequencing consortium. Genome Sequence of the Nematode C. Elegans: A Platform for Investigating Biology. The *C. Elegans* Sequencing Consortium [published Erratum Appears in Science 1999 Jan 1; 283(5398): 35]. *Science (80-).* **1998**, *282*, 2012–2018.

(32) McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000**, *16* (4), 404–405.

(33) Bissantz, C.; Logean, A.; Rognan, D. High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building, and Receptor Library Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1162–1176.

(34) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; Mcgettigan, P. a.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23* (21), 2947–2948.

(35) Sali, A.; Overington, J. P. Derivation of Rules for Comparative Protein Modeling from a Database of Protein Structure Alignments. *Protein Sci.* **1994**, *3* (9), 1582–1596.

(36) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815.

(37) Eddy, S. R. Where Did the BLOSUM62 Alignment Score Matrix Come From? *Nat. Biotechnol.* **2004**, *22* (8), 1035–1036.

(38) Styczynski, M. P.; Jensen, K. L.; Rigoutsos, I.; Stephanopoulos, G. BLOSUM62 Miscalculations Improve Search Performance. *Nature biotechnology*, 2008, *26*, 274–275.

(39) Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. *Curr. Protoc. Bioinforma.* **2013**, No. SUPL.43.

(40) Chiang, H. L.; Ngo, S. T.; Chen, C. J.; Hu, C. K.; Li, M. S. Oligomerization of Peptides LVEALYL and RGFFYT and Their Binding Affinity to Insulin. *PLoS One* **2013**, *8* (6).

(41) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72* (5), 2002–2013.

(42) White, J. A.; Román, F. L.; González, A.; Velasco, S. Periodic Boundary Conditions and the Correct Molecular-Dynamics Ensemble. *Phys. A Stat. Mech. its Appl.* **2008**, *387* (27), 6705–6711.

(43) Flocco, M. M.; Mowbray, S. L. Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **1994**, *235* (2), 709–717.

(44) Janosi, L.; Gorfe, A. A. Simulating POPC and POPC/POPG Bilayers: Conserved Packing and Altered Surface Reactivity. *J. Chem. Theory Comput.* **2010**, *6* (10), 3267–3273.

(45) Schmidt, T. H.; Kandt, C. LAMBADA and InflateGRO2: Efficient Membrane Alignment and Insertion of Membrane Proteins for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2012**, *52* (10), 2657–2669.

(46) Makarewicz, T.; Kaźmierkiewicz, R. Molecular Dynamics Simulation by GROMACS Using GUI Plugin for PyMOL. *J. Chem. Inf. Model.* **2013**, *53* (5), 1229–1234.

(47) DeLano, W. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**.

(48) Sandeep, G.; Nagasree, K. P.; Hanisha, M.; Kumar, M. M. K. AUDocker LE: A GUI for Virtual Screening with AUTODOCK Vina. *BMC Res. Notes* **2011**, *4* (1), 445.

(49) Chen, R.; Poger, D.; Mark, A. E. Effect of High Pressure on Fully Hydrated DPPC and POPC Bilayers. *J. Phys. Chem. B* **2011**, *115* (5), 1038–1044.

(50) Dunkin, C. M.; Pokorny, A.; Almeida, P. F.; Lee, H. S. Molecular Dynamics Studies of Transportan 10 (Tp10) Interacting with a POPC Lipid Bilayer. *J. Phys. Chem. B* **2011**, *115* (5), 1188–1198.

(51) Baldwin, J. M. The Probable Arrangement of the Helices in G Protein-Coupled Receptors. *EMBO J.* **1993**, *12* (4), 1693–1703.

(52) Nugent, T.; Ward, S.; Jones, D. T. The MEMPACK Alpha-Helical Transmembrane Protein Structure Prediction Server. *Bioinformatics* **2011**, *27* (10), 1438–1439.

(53) Eberle, A. N. G Protein-Coupled Receptors (GPCRs) A Comprehensive Book Covering All. **2012**.

(54) Hollenstein, K.; De Graaf, C.; Bortolato, A.; Wang, M. W.; Marshall, F. H.; Stevens, R. C. Insights into the Structure of Class B GPCRs. *Trends in Pharmacological Sciences*, 2014, *35*, 12–22.

(55) Symposium, J. A. Structural Genomics of the Human GPCR Protein Family. **2010**.

(56) Malo, M. Selectivity of Dopamine D1 and D2 Receptor Agonists – A Combined Computational Approach; 2012.

(57) 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 A_2A Adenosine Receptor. *Science* **2011**, *332* (6027), 322–327.

(58) Pogozheva, I. D.; Przydzial, M. J.; Mosberg, H. I. Homology Modeling of Opioid Receptor-Ligand Complexes Using Experimental Constraints. *AAPS J.* **2005**, *7* (2), E434–E448.

(59) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of Two Distinct Inactive Conformations of the beta2-Adrenergic Receptor Reconciles Structural and Biochemical Observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4689–4694.

(60) Finn, R. D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; Sonnhammer, E. L. L.; Tate, J.; Punta, M. Pfam: The Protein Families Database. *Nucleic Acids Research*, 2014, *42*.

(61) Cochrane, G. R.; Galperin, M. Y. The 2010 Nucleic Acids Research Database Issue and Online Database Collection: A Community of Data Resources.

Nucleic Acids Res. **2009**, *38* (SUPPL.1).

(62) Mertens, I.; Clinckspoor, I.; Janssen, T.; Nachman, R.; Schoofs, L. FMRFamide Related Peptide Ligands Activate the *Caenorhabditis Elegans* Orphan GPCR Y59H11AL.1. *Peptides* **2006**, *27* (6), 1291–1296.

(63) Jung, J.; Lee, B. Use of Residue Pairs in Protein Sequence-Sequence and Sequence-Structure Alignments. *Protein Sci.* **2000**, *9* (8), 1576–1588.

(64) Cherezov, V.; Rosenbaum, D.; Hanson, M.; Rasmussen, S.; Thian, F.; Kobilka, T.; Choi, H.; Kuhn, P.; Weis, W.; Kobilka, B.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human beta2-Adrenergic G Protein-Coupled Receptor. *Science (80-).* **2007**, *318*, 1258–1265.

(65) Wang, C.; Wu, H.; Katritch, V.; Han, G. W.; Huang, X. P.; Liu, W.; Siu, F. Y.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the Human Smoothened Receptor Bound to an Antitumour Agent. *Nature* **2013**, *497* (7449), 338–343.

(66) Latek, D.; Pasznik, P.; Carlomagno, T.; Filipek, S. Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison. *PLoS One* **2013**, *8* (2), 1–10.

(67) Sale, K.; Faulon, J. L.; Gray, G. A; Schoeniger, J. S.; Young, M. M. Optimal Bundling of Transmembrane Helices Using Sparse Distance Constraints. *Protein Sci.* **2004**, *13* (10), 2613–2627.

(68) Perez-Aguilar, J. M. Human μ Opioid Receptor Models with Evaluation of the Accuracy Using the Crystal Structure of the Murine μ Opioid Receptor. *J. Anesth. Clin. Res.* **2012**, *03* (06), 8715–8723.

(69) Millar, R. P.; Newton, C. L. The Year in G Protein-Coupled Receptor Research. *Mol. Endocrinol.* **2010**, *24* (1), 261–274.

(70) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. II. Are Homology Models of G-Protein Coupled Receptors Suitable Targets? *Proteins Struct. Funct. Genet.* **2003**, *50* (1), 5–25.

(71) Orry, A. J.; Wallace, B. A. Modeling and Docking the Endothelin G-Protein-Coupled Receptor. *Biophys. J.* **2000**, *79* (6), 3083–3094.

(72) Huang, C. C.; Tesmer, J. J. G. Recognition in the Face of Diversity: Interactions of Heterotrimeric G Proteins and G Protein-Coupled Receptor (GPCR) Kinases with Activated GPCRs. *Journal of Biological Chemistry*, 2011, *286*, 7715–7721.

(73) Eilers, M.; Hornak, V.; Smith, S. O.; Konopka, J. B. Comparison of Class A and D G Protein-Coupled Receptors: Common Features in Structure and Activation. *Biochemistry* **2005**, *44* (25), 8959–8975.

(74) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2013**, *53*, 531–556.

(75) Kunishima, N.; Shimada, Y.; Tsuji, Y.; Sato, T.; Yamamoto, M.; Kumasaka, T.; Nakanishi, S.; Jingami, H.; Morikawa, K. Structural Basis of Glutamate Recognition by a Dimeric Metabotropic Glutamate Receptor. *Nature* **2000**, *407* (6807), 971–977.

(76) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. Modeling Activated States of GPCRs: The Rhodopsin Template. *J. Comput. Aided. Mol. Des.* **2006**, *20* (7-8), 437–448.

(77) Gil-Mast, S.; Kortagere, S.; Kota, K.; Kuzhikandathil, E. V. An Amino Acid Residue in the Second Extracellular Loop Determines the Agonist-Dependent Tolerance Property of the Human D3 Dopamine Receptor. *ACS Chem. Neurosci.* **2013**, *4* (6), 940–951.

(78) Worth, C. L.; Kleinau, G.; Krause, G. Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models. *PLoS One* **2009**, *4* (9).

(79) Reggio, P. H. Computational Methods in Drug Design: Modeling G Protein-Coupled Receptor Monomers, Dimers, and Oligomers. *AAPS J.* **2006**, *8* (2), E322–E336.

(80) Mugumbate, G.; Jackson, G. E.; Van Der Spoel, D.; Kövér, K. E.; Szilágyi, L. *Anopheles Gambiae*, Anoga-HrTH Hormone, Free and Bound Structure-A Nuclear Magnetic Resonance Experiment. *Peptides* **2013**, *41*, 94–100.

(81) Heifetz, A.; Schertler, G. F. X.; Seifert, R.; Tate, C. G.; Sexton, P. M.;

Gurevich, V. V.; Fourmy, D.; Cherezov, V.; Marshall, F. H.; Storer, R. I.; Moraes, I.; Tikhonova, I. G.; Tautermann, C. S.; Hunt, P.; Ceska, T.; Hodgson, S.; Bodkin, M. J.; Singh, S.; Law, R. J.; Biggin, P. C. GPCR Structure, Function, Drug Discovery and Crystallography: Report from Academia-Industry International Conference (UK Royal Society) Chicheley Hall, 1–2 September 2014. *Naunyn. Schmiedebergs. Arch. Pharmacol.* **2015**.

(82) Hu, L. A.; Chen, W.; Martin, N. P.; Whalen, E. J.; Premont, R. T.; Lefkowitz, R. J. GIPC Interacts with the beta1-Adrenergic Receptor and Regulates beta1-Adrenergic Receptor-Mediated ERK Activation. *J. Biol. Chem.* **2003**, *278* (28), 26295–26301.

(83) de Voux, A.; Chan, M. C.; Folefoc, A. T.; Madziva, M. T.; Flanagan, C. A. Constitutively Active CCR5 Chemokine Receptors Differ in Mediating HIV Envelope-Dependent Fusion. *PLoS One* **2013**, *8* (1).

(84) Kooistra, A. J.; Kuhne, S.; De Esch, I. J. P.; Leurs, R.; De Graaf, C. A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *Br. J. Pharmacol.* **2013**, *170* (1), 101–126.

(85) Mueller, A.; Strange, P. G. The Chemokine Receptor, CCR5. *International Journal of Biochemistry and Cell Biology*, 2004, *36*, 35–38.

(86) Tan, Q.; Zhu, Y.; Li, J.; Chen, Z.; Han, G. W.; Kufareva, I.; Li, T.; Ma, L.; Fenalti, G.; Li, J.; Zhang, W.; Xie, X.; Yang, H.; Jiang, H.; Cherezov, V.; Liu, H.; Stevens, R. C.; Zhao, Q.; Wu, B. Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex. *Science* **2013**, *341* (6152), 1387–1390.

(87) Mysinger, M. M.; Weiss, D. R.; Ziarek, J. J.; Gravel, S.; Doak, A. K.; Karpiak, J.; Heveker, N.; Shoichet, B. K.; Volkman, B. F. Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4. *Proc. Natl. Acad. Sci.* **2012**, *109* (14), 5517–5522.

(88) De Vries, A. H.; Mark, A. E.; Marrink, S. J. Molecular Dynamics Simulation of the Spontaneous Formation of a Small DPPC Vesicle in Water in Atomistic Detail. *J. Am. Chem. Soc.* **2004**, *126* (14), 4488–4489.

(89) Feller, S. E.; Venable, R. M.; Pastor, R. W. Computer Simulation of a DPPC

Phospholipid Bilayer: Structural Changes as a Function of Molecular Surface Area. *Langmuir* **1997**, *13* (24), 6555–6561.

(90) Funkhouser, T. Determination Ab Initio Methods. **2005**, 1–8.

(91) Fleishman, S. J.; Ben-Tal, N. Progress in Structure Prediction of α-Helical Membrane Proteins. *Curr. Opin. Struct. Biol.* **2006**, *16* (4), 496–504.

(92) Willard, L.; Ranjan, A.; Zhang, H.; Monzavi, H.; Boyko, R. F.; Sykes, B. D.; Wishart, D. S. VADAR: A Web Server for Quantitative Evaluation of Protein Structure Quality. *Nucleic Acids Res.* **2003**, *31* (13), 3316–3319.

(93) Lakshmi, B.; Ramakrishnan, C.; Archunan, G.; Sowdhamini, R.; Srinivasan, N. Investigations of Ramachandran Disallowed Conformations in Protein Domain Families. *Int. J. Biol. Macromol.* **2014**, *63*, 119–125.

(94) Hooft, R. W.; Sander, C.; Vriend, G. Objectively Judging the Quality of a Protein Structure from a Ramachandran Plot. *Comput. Appl. Biosci.* **1997**, *13* (4), 425–430.

(95) Berjanskii, M. V.; Wishart, D. S. The RCI Server: Rapid and Accurate Calculation of Protein Flexibility Using Chemical Shifts. *Nucleic Acids Res.* **2007**, *35* (SUPPL.2), 531–537.

(96) Schlegel, B.; Sippl, W.; Höltje, H. D. Molecular Dynamics Simulations of Bovine Rhodopsin: Influence of Protonation States and Different Membrane-Mimicking Environments. *J. Mol. Model.* **2005**, *12* (1), 49–64.

(97) Petukhov, M.; Rychkov, G.; Firsov, L.; Serrano, L. H-Bonding in Protein Hydration Revisited. *Protein Sci.* **2004**, *13* (8), 2120–2129.

(98) Friesner, R. A.; Abel, R.; Goldfeld, D. A.; Miller, E. B.; Murrett, C. S. Computational Methods for High Resolution Prediction and Refinement of Protein Structures. *Curr. Opin. Struct. Biol.* **2013**, *23* (2), 177–184.

(99) Werner, T.; Morris, M. B.; Dastmalchi, S.; Church, W. B. Structural Modelling and Dynamics of Proteins for Insights into Drug Interactions. *Adv. Drug Deliv. Rev.* **2012**, *64* (4), 323–343.

(100) Tastan, O.; Klein-Seetharaman, J.; Meirovitch, H. The Effect of Loops on the Structural Organization of α-Helical Membrane Proteins. *Biophys. J.* **2009**, *96*

(6), 2299–2312.

(101) Cho, H. C.; Tsushima, R. G.; Nguyen, T. T. T.; Guy, H. R.; Backx, P. H. Two Critical Cysteine Residues Implicated in Disulfide Bond Formation and Proper Folding of Kir2.1. *Biochemistry* **2000**, *39* (16), 4649–4657.

4 Chapter 4: Molecular docking calculations and molecular dynamic simulations of *flp18-6* peptide bound complex of nematode GPCR, FLP18R1

4.0 Introduction

Molecular docking involves estimating the augmented conformation, interactions and orientation of a ligand in a binding site of a protein target from an *in silico* target-ligand complex.[1–4] The favourable ligand position in the binding site increases the interactions while decreasing the total energy of the complex.[2,5,6] In drug discovery and development, docking is implemented to cut down on research timelines by identifying potential hit compounds.[7,8] This also reduces the cost by reducing the number of compounds analysed for wet laboratory experiments.

Docking has three stages, viz; definition of the binding site, pose generation and pose selection.[3,9] These stages are performed in two steps, which are searching and scoring.[3,7,10,11] Searching on the protein target space exploits mathematical algorithms to investigate optimal positions of the ligand within the active site and possible binding poses.[12–14] On the other hand, scoring involves evaluating the particular pose of ligands/inhibitors and ranking them according to a scoring function.[1,12] The success of a docking program relies on broad sampling methods and good scoring functions.

In cases where the binding site is not known, such as some homology models, blind docking and cavity detection programs are utilized to identify the active site.[15–17] Docking programs such as AutoDock[18,19] and SiteMap in the Glide protocol[20] of the Schrodinger package can be used to identify binding sites in blind docking. With protein-ligand interactions, for example, intermolecular hydrogen bonding and lipophilic interactions are critical for effective docking. Docking exploits the 3D structure of the target and facilitates binding affinity prediction between the ligand and the target.[21] Identifying the active site on a receptor and a ligand's binding pose is dynamic since both the protein and ligand change conformation to exploit total free energy during free association.[5,22]

Docking algorithms are able to generate a large number of possible poses and score them. Docking programs, for example, DOCK,[3] FTDOCK,[23] QSDOCK and FlexX[24] match the energy, chemistry and geometry of the active site where they position and orient the ligand.[25] These are mostly used in rigid docking where the ligand and protein are rigid.[3] Other programs such as such as GOLD,[24] AutoDock[18,26] and DARWIN[27] consider flexible ligand and rigid protein. These are used in search of the global energy minimum of the ligand position by applying genetic algorithms.[27] Optimized combinations of scoring and docking functions reduce the number of false positives and false negatives in scoring.[1,13]

An increasing number of docking protocols are beginning to incorporate protein flexibility.[1,9] Consideration of protein flexibility via energy minimization during post docking refinement helps to optimize dihedral angles. Some docking programs use geometric restraints to isolate best fit of the protein and ligand during docking studies. Docking studies have revealed that ligand binding induces diverse conformational changes in the protein such as side chain rearrangement and in some cases hinge movement of the C-domain.[9,28,29]

Estimation of relative binding affinity during post docking refinement is usually performed using Molecular Mechanics terms such as Generalised Born Surface Area (MM-GBSA).[30–33] In this study Prime MM-GBSA was applied in the Schrodinger package using an implicit solvent model with protein-ligand complex. The following equation (4.1) defines the binding affinity of the protein-ligand complex:

$$\Delta G_{bind} = G_{(PL)} - G_{(P)} - G_{(L)} \text{……..(Equation 4.1)}$$

Where ΔG_{bind} is the binding free energy, PL is the protein-ligand complex, P is the protein and L is the ligand. Relative binding affinities are calculated and are understood to relate to the free energy of binding. The more negative the binding affinity, the stronger the binding of the ligand to the receptor protein. The MM-GBSA free energy of binding is calculated according to equation 4.2:[34]

$$\Delta G_{bind} = \Delta G_{solv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \text{…….(Equation 4.2)}^{30,31}$$

Where ΔG_{solv} is the impact of the solvent due to solvent effects,

ΔG_{conf} is the induced conformational changes in the protein receptor and the ligand,

ΔG_{int} is the calculated free energy from the protein-ligand interactions,

ΔG_{rot} is the free energy lost as freezing internal rotations of the protein and ligand associate,

$\Delta G_{t/r}$ is the loss in translational and rotational free energy as protein and ligand interact to give the bound complex,

ΔG_{vib} is the vibrational mode free energy.

However, experiments have concluded that several challenges such as protein conformational changes as well as the protein flexibility and ranking of poses in a docking program still exist. These challenges are thought to result from too many degrees of freedom that are available to an active site.[17,35] As a result of this, an increasing number of binding sites exist due to these conformational changes.[9,36] Thus the quality of binding can be evaluated based on the interactions.

Where 3D structures of proteins e.g. orphan GPCRs, are not available, homology models are used in molecular docking calculations that are crucial in the discovery and design of a range of ligands presently used in pharmaceutical drug research.[37–39] In this study, the homology modeling of the nematode GPCR, FLP18R1 was presented in Chapter 3 and the 3D conformations of the ligands *af3, af4, af20 and flp18-6* were determined in chapter 2. Since this is homology modeling, the binding site is not known and thus blind docking is performed using AutoDock and SiteMap. Systematic pose generation was performed using exhaustive conformational search and a stochastic option in Glide was utilized.

4.1 Computational Methods

4.1.1 Autodock4.2 Protein preparation

The FLP18R1 receptor structure was determined in Chapter 3 and was prepared for docking calculations by adding gasteiger charges and polar hydrogen atoms using AutoDock Tools.

4.1.2 Ligand preparation

The four peptides, *af3, af4, af20* and *flp18-6,* were used as the ligands. The solution structures of these were determined and reported in Chapter 2. Polar hydrogen atoms and gasteiger charges were added. Torsional degrees of freedom and number of active torsions were set.

4.1.3 Binding Site Identification

Since the binding site is not known, blind docking (BD)[15–17] was performed to identify the binding pocket of FLP18R1 and evaluate its accessibility. BD is a technique that probes the protein receptor by the ligand to detect putative binding sites in the protein.[17,19,35] Since the ligand is known to bind outside the cell, the N-terminus domain of extracellular surface was probed. One tool that is useful in blind docking is AutoDock.[19,26,35] Hence, AutoDock 4.2 and AutoDock Tools were used to identify the binding sites of the peptides *af3, af4, af20* and *flp18-6* in FLP18R1.

The blind docking parameters reported by Hetenyi *et al* were set.[16,17] GPCR ligands are known to access the receptor and bind to the extracellular side.[40,41] Grid maps were generated to cover the whole extracellular region of FLP18R1 using the AutoGrid4 program.[18,19] Initially, *flp18-6* was docked to the receptor to determine its most preferred position in the extracellular region. The position of the determined binding site in the extracellular region was used to select the suitable grid map had grid points of 66 x 70 x 64 in *xyz,* and a spacing of 0.375nm, was used for the blind docking calculations. The grid box was centred at 32.81 x 35.5 x 69.22. The *flp18-6* ligand was flexible with 27 active torsions. Conformations of the ligand were generated within the binding pocket using Lamarckian Genetic Algorithm (LGA).

Population size of 250 was set and the maximum number of energy evaluations and generations were set at 100 x 10^6 and AutoDock4 default parameter values were not changed (Appendix IV). The *flp18-6* ligand was put through 57 trials of BD for 100 runs in order to locate the binding pocket in the FLP18R1 receptor. Cluster analysis of the protein/ligand complex was done using a cut-off of 2Å and the cluster with the

most cluster members was selected. Initially 8 trials of the docking calculations confirmed the least energy conformation. This is the pose that was selected from the most populated cluster. More focused BD calculations were performed using the identified binding site as conformation tests. Estimated free energy of binding (ΔG_b) of the protein-ligand complex at this binding site was determined. The accessibility of the binding pocket was determined for the FLPR1 receptor using grid points 72 x 64 x 62 in *xyz* with the same ligand.

Once the binding site was identified, similar docking calculations were carried out using Glide version 6.2, Schrödinger, LLC, New York, NY, 2014 docking protocol and the results were compared to those from Autodock 4.2.

4.1.4 Glide Protein Preparation

Protein Preparation Wizard[42] (Prep Wizard) from Maestro version 9.7 in Schrödinger package was used to prepare the protein receptor. The FLP18R1 homology model refined by MD simulation (Chapter 3) had its atomic clashes removed while formal charges and missing hydrogen atoms were added.[42] The protein was energy minimized for predocking refinement. The Impref utility of Maestro utilizing the Optimized Potentials for Liquid Simulations-All Atom (OPLS_2005) force field was used to energy minimize and refine the structures up to 0.3Å RMSD of non-hydrogen atoms.[42] The prepared receptor protein was retained for docking studies.

4.1.5 Glide Ligand Preparation

The Ligprep-2.9, Ligand preparation protocol,[42] in Schrödinger package was used to prepare the peptide ligand to create three-dimensional geometries. Proper bond and bond orders were assigned as well as van der Waal's parameters.[43] No tautomers were generated and initial protonation states were retained. Charges were calculated to predict pKa's for each potentially charged atom and generate a structure for each charge combination for a given pH of 7 ± 1.

4.1.6 SiteMap Identification of the binding site

SiteMap tool in Maestro version 9.7 (Schrodinger package) was used to identify the binding site. 5 binding sites were predicted using default parameters. SiteMap-3.0 defines sites using a permutation of properties computed at each site point and by cumulative properties of the entire set of points that make up each site. The SiteMap algorithm produces site points on a grid adjacent to the receptor van der Waals surface. Site 1 (Table 4.1) was selected for use in this study.

4.1.7 Receptor Grid Generation

The protein receptor grid box is the 3D space in which ligands are given freedom to move. The receptor grid generator tool of the Schrodinger package was used to generate the grid box in a ligand-free FLP18R1 receptor. The centroid of the residues of the receptor in the extracellular domain, predicted by SiteMap was well defined as the grid box. A grid box of dimesions, inner box 10 x 10 x 10 $Å^3$ and outer box, 37.8 x 37.8 x 37.8 $Å^3$ with a grid centre of 41.7, 38.3 and 35.9 in the *x, y* and *z* directions respectively was constructed. Partial charge cut-off of 0.25, using a scaling factor of 1.0, for van der Waals and OPLS_2005 force field were the default settings used for grid generation. The shape and properties of the receptor protein was mapped into the grid during the preprocessing phase. 10 poses per ligand were generated. Post-docking minimization generated 10 poses and other parameters for sampling were set to their default.

4.2 Docking calculations

Docking was performed using the homology models of FLP18R1 based on the Beta-2-adrenergic template that was elucidated in Chapter 3. Glide version 6.2, Schrödinger, LLC, New York, NY, 2014 docking protocol[20] was used for performing docking calculations[44]. Two precision choices were used; XP (extra precision) and SP (Standard Precision)-Peptide using default parameters. Grid based XP and SP-Peptide

docking was performed using the MD refined homology model (Chapter 3) and the four peptide ligands determined in Chapter 2. The top 50% docking conformations obtained from these docking calculations were refined post docking using MM-GBSA to calculate the protein-ligand binding affinity. The 64 poses generated per ligand were visually inspected.

4.2.1 Molecular Dynamics Simulations in a Mimetic Membrane

Conformational studies of the FLP18R1 receptor after docking were studied by carrying out the molecular dynamic simulations of the receptor-ligand complexes in a POPC pre-equilibrated membrane. The MD simulation protocol described in chapter 3 used 100 ns in a cubic water box.

4.3 Results and Discussion

4.3.1 Identification of binding site

The binding site was identified by applying blind docking using AutoDock4.2 and was confirmed using SiteMap from the Schrodinger's suite. The two different protocols identified similar binding sites. The binding site from SiteMap gave 5 sites that vary in size and residues. Table 4.1 shows the volumes and site score for each site.

Table 4.1 Ranking of potential binding pockets in FLP18R1 GPCR receptor found using SiteMap. From the table *EC- Extracellular # IC- Intracellular ++ TM- Transmembrane

Site Rank	Site Score	Volume	Size	Location	D-score
1	1.070	798.17	292	*EC	1.118
2	1.001	518.62	144	#IC	1.059
3	0.955	261.02	79	#IC	0.862
4	0.865	163.61	61	#IC/++TM	0.878
5	0.814	83.35	34	++TM	0.861

Careful analysis of the site map results indicates the site ranked 1 as the best site. This is large, with a volume of 798.17 and is on the extracellular domain. As previously mentioned in Chapter 3, the ligands in GPCRs approach the binding pocket from the extracellular side through the extracellular loop 2 as in the template, beta 2-adrenergic crystal structure (PDB code 2RH1).[45] Two disulphide bridges between Cys202 and Cys208 and the other one between Cys128 in Helix 3 and Cys209 (ECL2) facilitate the accessibility to this binding cavity in FLP18R1 receptor by the *flp18-6* ligand. It is also important to mention that GPCRs are transmembrane proteins and thus the membrane region is not a possible binding site, in this case, site 4 and 5. This is due to its inaccessibility and lipophilic nature. The selected binding site is shown in Figure 4.1.

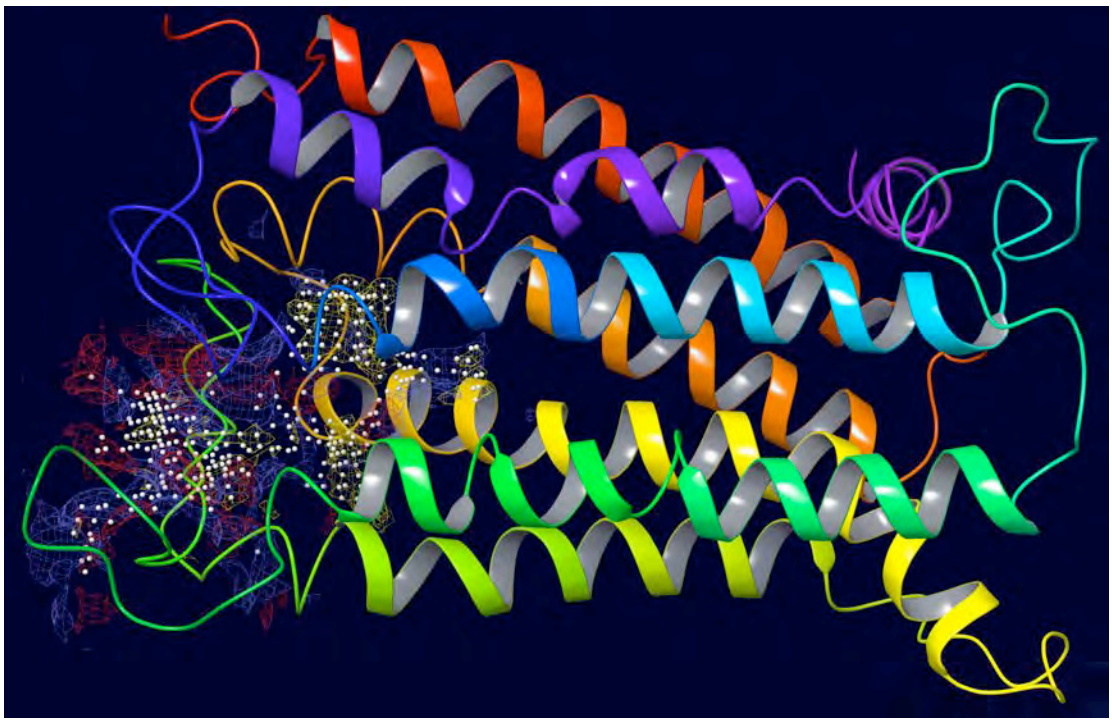


Figure 4.1: SiteMap generated from the binding site calculation in FLP18R1 defining the largest binding site (Table 4.1) shown by the wire-mesh with small site points indicating hydrophobic and hydrophilic areas in the binding cavity.

This site covers the extracellular loop regions[5,6,15,46] and does not include TM1 and TM7, which, are known not to be involved in ligand binding. This preferred binding site is well defined among well-ordered side chains.[28,47,48] The binding site is made up of mostly helices 3, 5 and 6 as well as ECL2, which is typical of GPCRs.

4.3.2 Determination of binding mode

The selected site was used to generate the docking grids that were used in the docking calculations. No constraints were applied during grid generation and allowance was made for the peptide-docking grid. Docking calculations were carried out using AutoDock 4.2, XP Glide and S-peptide docking from the Schrodinger Suite.[20,49] The protocols thus used, use the empirical scoring for docking results and the force field based scoring methods of Glide and AutoDock respectively.[17,20,49] The best binding poses were rescored and filtered based on MM-GBSA free energy binding calculations using Prime.[30–32]

The rearrangement of helix 3 and helix 7 was observed during MD simulations,[36,50] that was found in Adenosine 2 Alpha (A_2A) adrenergic receptors upon ligand binding, was also realized when the peptides bound to FLP18R1.[51,52] As observed in a study by Katritch, helix 7 moves towards helix 3 in the intracellular side upon ligand binding. At the same time, helix 3 shifts upwards, along its axis, following the backbone rearrangement of NPLLF (helix 7) motif.[50,52] The movement of these two helices also result in the movement of the other helices, particularly 5 and 6, which are believed to be vital for G-protein binding,[53] subsequent GPCR stimulation and are preserved in most class A GPCRs.[9,28,54] The short helix in ECL2 of FLP18R1, a common peptide-binding motif in GPCRs such as the Opioid subfamily and in the Chemokine receptors, was also involved in ligand binding in this study.[46,55–57]

4.3.2.1 flp18-6 Ligand Binding and pharmacophore

The pharmacophore features of the interactions when *flp18-6* binds to the FLP18R1 receptor are shown in Figure 4.2. This common pharmacophore is also shown in all the protein-peptide interactions in this study, which are discussed in the following sections.

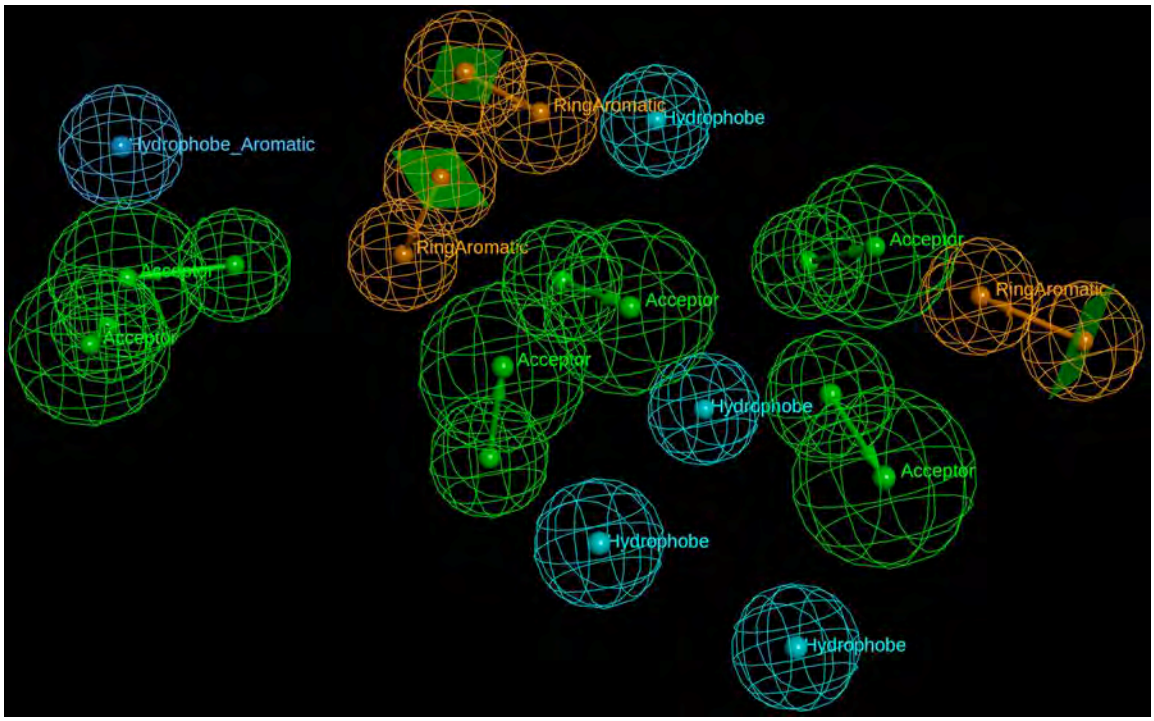


Figure 4.2: Hybrid 3D Structure and ligand based Pharmacophore features of the protein-peptide complex, which were manually created from the interactions of *flp18-6* peptide when bound to the FLP18R1 receptor using Discovery Studio 3.5 suite.

The aromatic hydrophobic (turquoise) feature on the far left in Figure 4.2 indicate the indole ring of Trp299 that is involved in a conserved π-cation interaction with the Arg (C-terminal) in all peptides. The ring aromatics indicate the aromatic ring on the Phe (shown in the middle) of the peptide as well as the aromatic residues (far right) in the receptor, in this case His194.

The bound conformation of *flp18-6* inside the binding pocket of the FLP18R1 GPCR receptor is shown in Figure 4.3 below.

(a) (b)

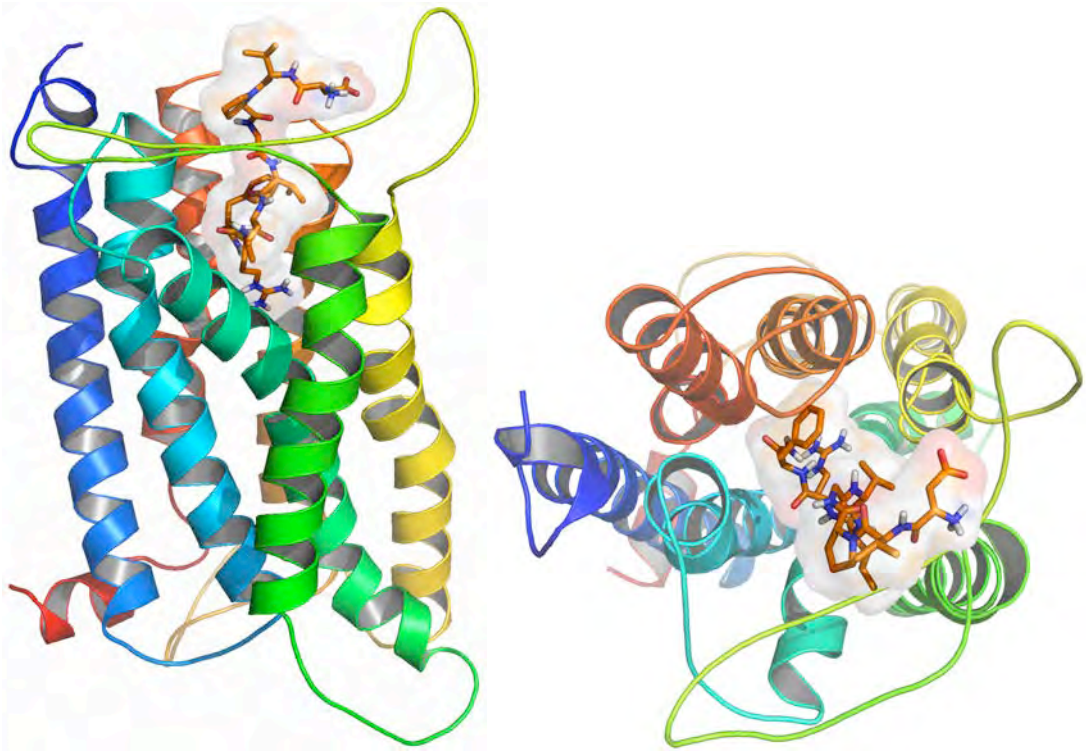


Figure 4.3: The complex of the *flp18-6* peptide bound to the FLP18R1 receptor. The 7 transmembrane helices shown in different colours with H1 (marine), H2 (cyan), H3 (lemon), H4 (green), H5 (yellow), H6 (orange), H7 (brown). (a) Side view of the *flp18-6* in the binding site in the extracellular domain. (b) Aerial view of the *flp18-6* in the binding pocket.

The *flp18-6* peptide was determined to have a cyclic conformation in its unbound form (Chapter 2). This cyclic conformation[40,58] is stabilized through a salt bridge between the carboxylate of aspartic acid (N-terminus) to the guanidinium group of arginine on the C terminal of the peptide.[59–61] The carbonyl oxygen of aspartic acid creates an intramolecular 1.91Å strong hydrogen bond with the pyrrolidine ring of proline in the cyclic conformation as well. The salt bridge cleaves (Figure 4.4) when the peptide binds to the FLP18R1 receptor.

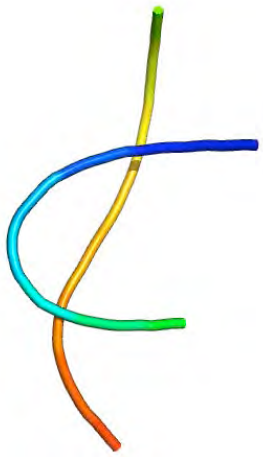
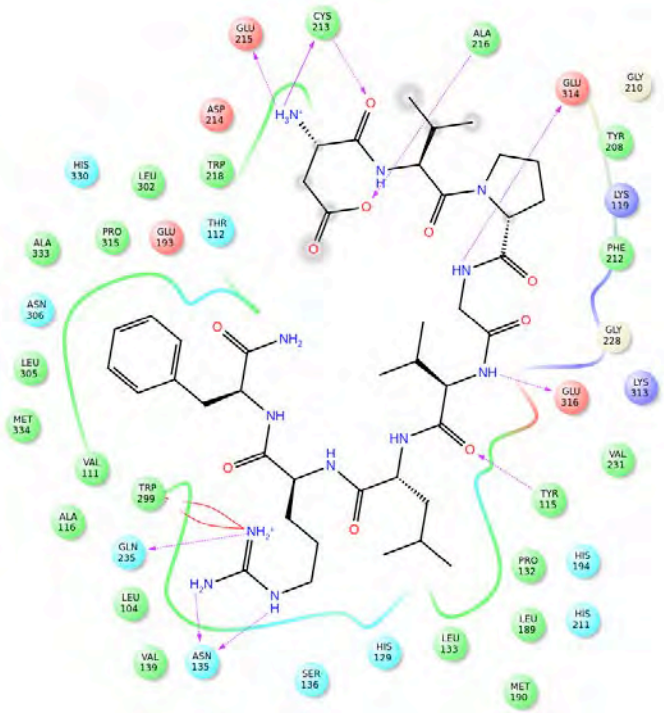


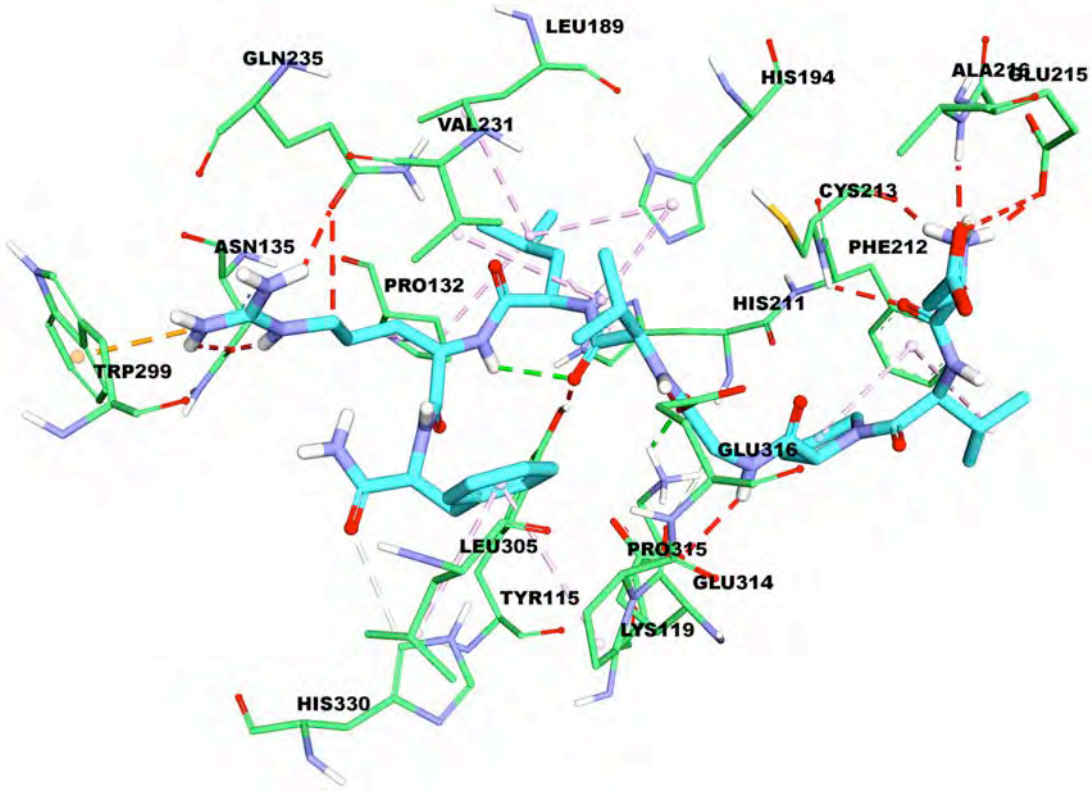
Figure 4.4: The solution ribbon structure of the *flp18-6* peptide before docking (greenish-blue) superimposed with the peptide after docking (orange yellow)

Breaking of this salt-bridge results in the guanidinium group of the peptide arginine forming a strong hydrogen bond with the OE1 of Gln235 and a π-cation interaction with the indole ring of the highly conserved[45] Trp299 (conserved in CW*x*P motif helix 6 of class A GPCRs as outlined in Chapter 3) as shown in Figure 4.5. It is understood that ligand binding to Trp299 is essential in signal transduction and results in the typical ligand-supported reorganization in the binding cavity, commonly called transmission switch.[54]

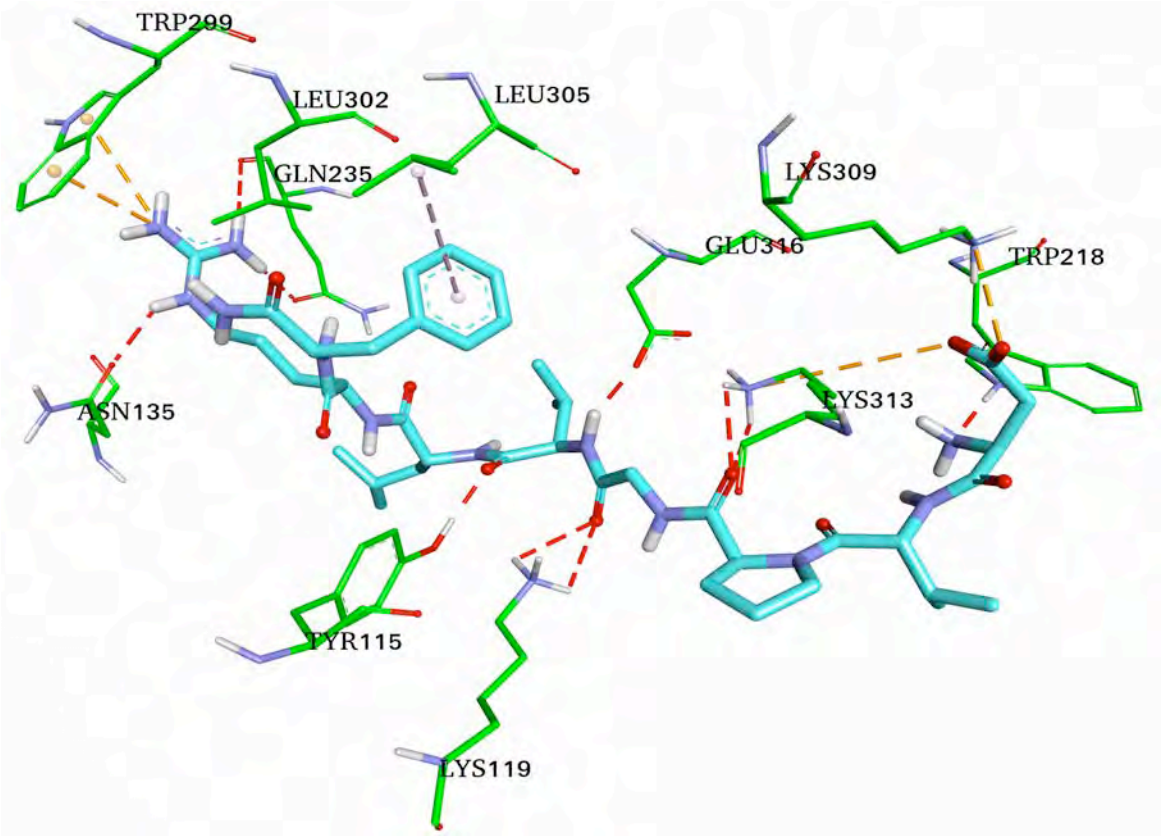
(a)



(b)


LEU189
GLN235
HIS194
ALA216
GLU215
VAL231
CYS213
ASN135
PHE212
PRO132
HIS211
TRP299
GLU316
LEU305
PRO315
GLU314
TYR115
LYS119
HIS330

(c)


TRP299
LEU302
LEU305
GLN235
LYS309
GLU316
TRP218
ASN135
LYS313
TYR115
LYS119

(d)

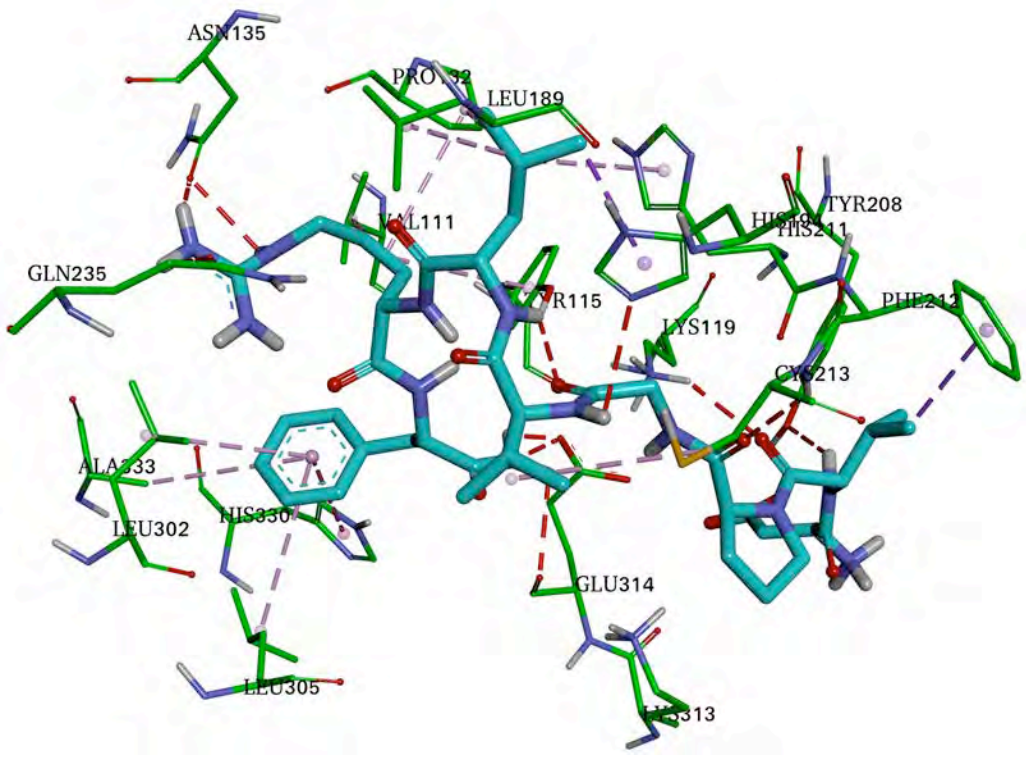


Figure 4.5: Binding mode of *flp18-6 (cyan)* to FLP18R1 (green). (a) 2D ligand interaction map of the *flp18-6* peptide in the FLP18R1 receptor. Protein-peptide binding interactions of *flp18-6* from (b) S-peptide Glide (c) XP Glide (d) AutoDock 4.2 docking interactions. Hydrogen bonds are shown as red dashed lines. Hydrophobic interactions (magenta) indicate that the C-terminus phenylalanine in the *flp18-6* is inside a hydrophobic pocket made up of Leu302, Leu305, His330 and Lys119 on the extracellular domain.

The geometric conformation of the π-cation interaction through the parallel stacking arrangement is observed to be consistent with the study by Thornton and corroborated by Flocco and Mowbray.[60,62,63] In these studies, they concluded that the guanidinium ion of arginine generally interacts with aromatic residues through the available π electron system to the cation as observed in most crystal structures in the Protein Data Bank (PDB).[62,63] The interaction is attributed to the larger arginine side chain, which is thought to interact with the aromatic rings through van der Waals interactions. Thus the indole ring of Trp299 interacts with the guanidinium ion in arginine, which concurrently forms hydrogen bonds with Gln235.[64] Figures 4.3(a) and (b) reveal the T-shaped geometry of the arginine guanidinium ion to the indole ring of Trp299.

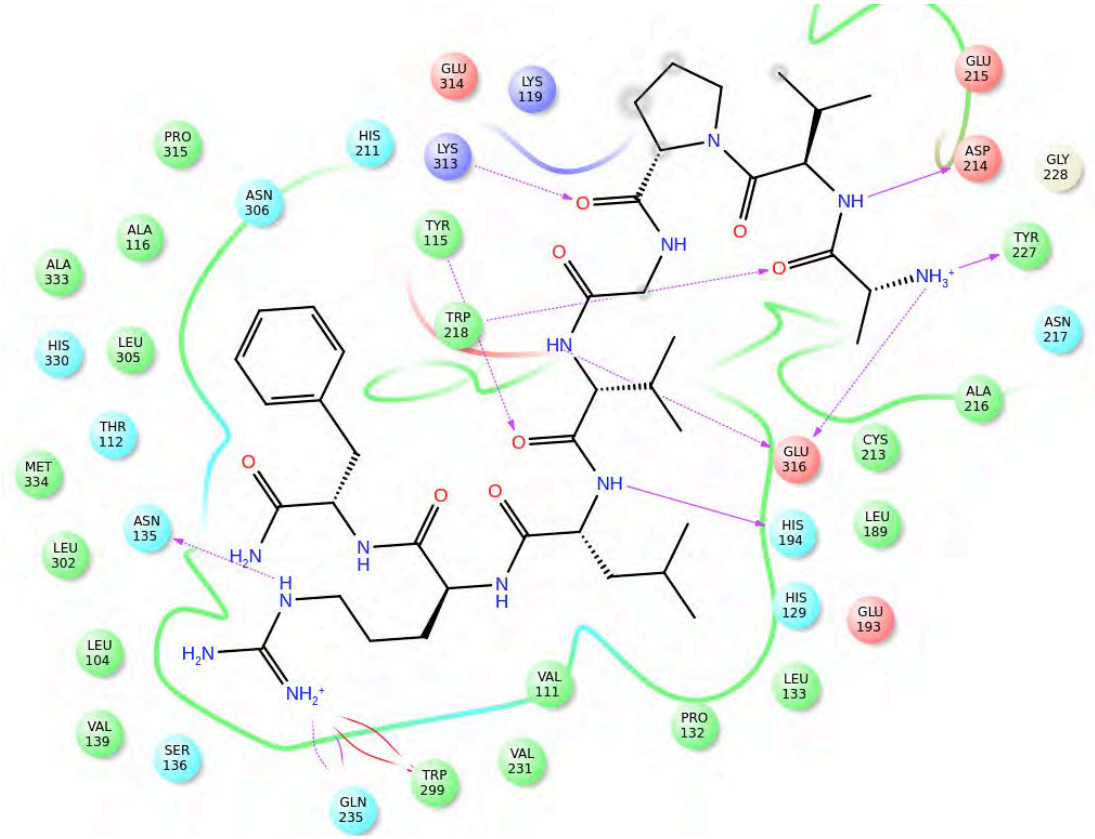
OD1 of Asn135 forms a hydrogen bond with epsilon amide (NHε) and guanidinium ion of the Arg7 in the peptide of bond lengths of 1.70 Å and 2.74 Å respectively. The OD1 has two lone pairs of electrons and thus accepts 2 hydrogen bonds from arginine that donates the hydrogen bonds. These interactions are consistent with A_2A adrenergic receptor and rhodopsin using the conserved tryptophan in helix 6 (Trp299 in this study), which render stability to the oscillating movement of helix 6 and thus restrain shifting of the tryptophan residue.[51,52,54] The Phe8 *(flp18-6)* aromatic ring is in the hydrophobic pocket comprising side chains of Leu302 (conserved in helix 6 of class A GPCRs)[65] and Leu305. However, the XP Glide docking interaction of the Asp1 carboxylate with the Trp218 was not observed using S-peptide docking and AutoDock 4.2.

ECL2 is involved in the binding to the flp18-6 through residues Cys213, Ala216 and Glu215. Strong, hydrogen bonds of between 1.91 Å and 2.14 Å are seen between Asp1, on the N-terminus of the peptide and carboxylate ion and protonated amide. Lys313 (protonated Schiff base)[66] forms a salt bridge with the carboxylate ion of the Asp1 (*flp18-6*). The XP Glide score (-8.14 kJ/ mol) is lower than the S-peptide score (-11.89 kJ/mol) and AutoDock4.2 score of -9.0 kJ/mol. Using XP Glide, similar interactions were observed using S peptide docking.

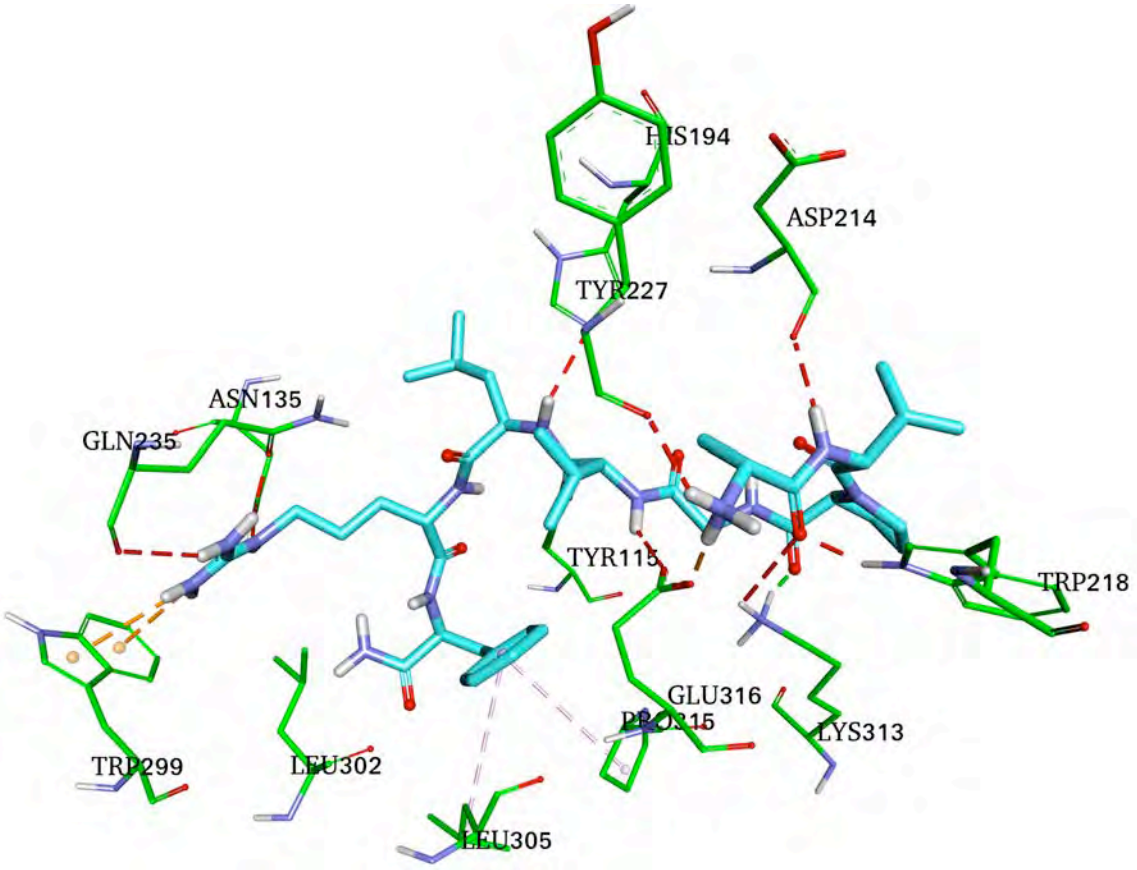
4.3.2.2 *af3 Docking*

Crucial interactions of *af3* and FLP18R1 were observed using AutoDock 4.2, S peptide and XP Glide. The docking scores were - 9.20, -10.20, and -8.27 kJ/ mol respectively. Important conserved chemoprints of class A GPCRs, Tyr115, Asn135, K313 (protonated Schiff base), Gln235 and Trp299 in FLP18R1 receptor that bind with the *af3* peptide were observed. Figure 4.6 (a) to (d) indicates the ligand interaction maps of *af3* with the GPCR, FLP18R1.

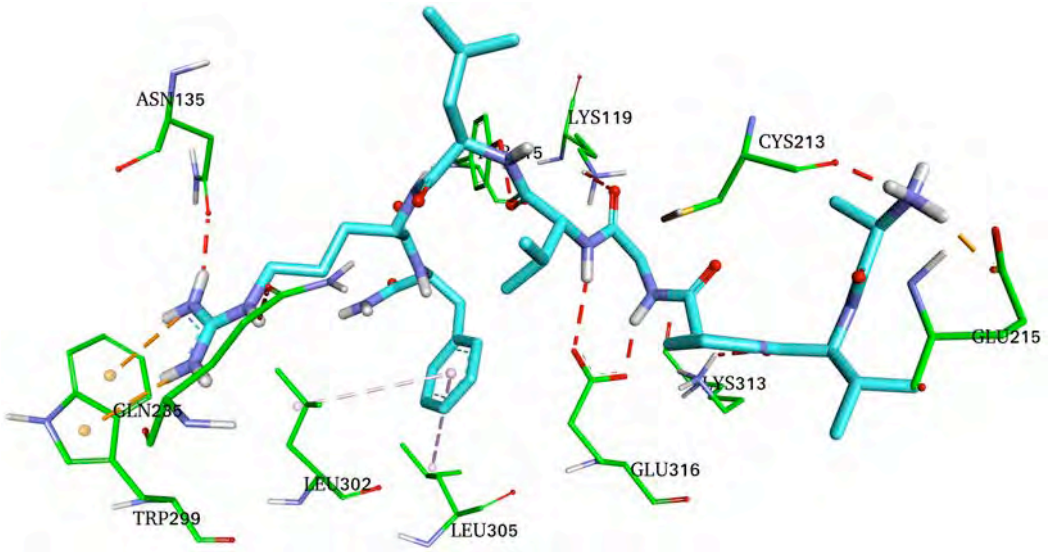
(a)


GLU 314
LYS 119
GLU 215
PRO 315
HIS 211
LYS 313
ASP 214
GLY 228
ASN 306
ALA 116
TYR 115
TYR 227
ALA 333
TRP 218
ASN 217
HIS 330
LEU 305
ALA 216
THR 112
CYS 213
MET 334
GLU 316
LEU 189
ASN 135
HIS 194
LEU 302
HIS 129
GLU 193
LEU 104
VAL 111
LEU 133
PRO 132
VAL 139
SER 136
VAL 231
TRP 299
GLN 235

(b)


HIS194
ASP214
TYR227
ASN135
GLN235
TYR115
TRP218
GLU316
LYS313
TRP299
LEU302
LEU305

(c)



(d)

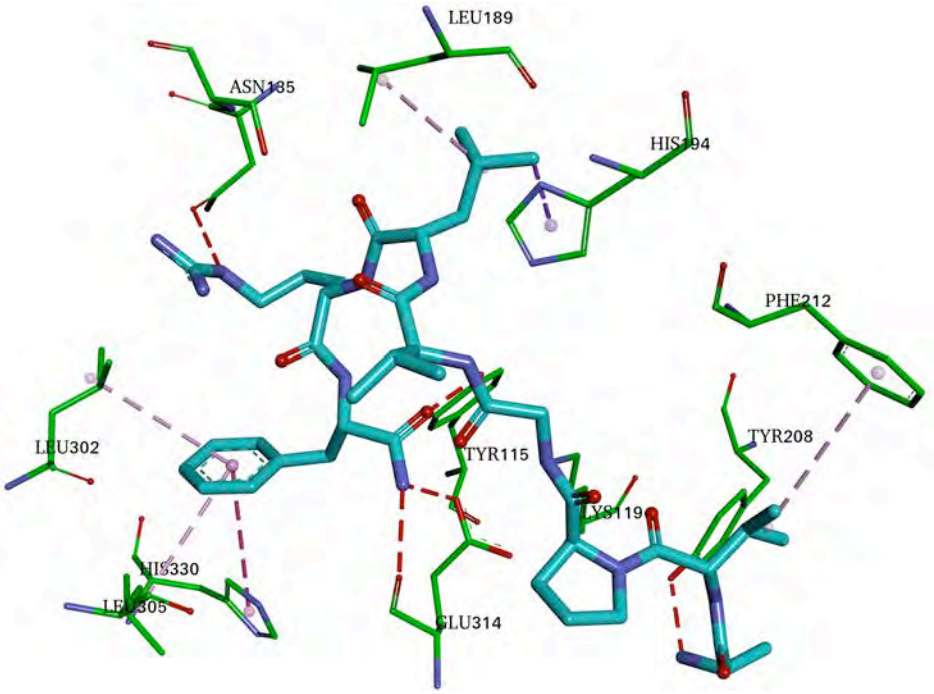


Figure 4.6: Binding mode of *af3 (cyan)* to FLP18R1 (green). Protein-peptide binding interactions (a) *af3* ligand interaction map as observed in the FLP18R1 binding pocket. Binding interactions of *af3* from (b) S-peptide Glide (c) XP Glide (d) AutoDock 4.2 docking. Hydrogen bonds are highlighted by red dashed lines. Hydrophobic interactions (magenta)

The phenolic OH on Tyr115 (Helix 2) is a hydrogen bond donor to the backbone carbonyl oxygen on Val5 on *af3*. Trp299 experiences a π-cation interaction with the guanidinium group of the Arg7 from the peptide as was observed in *flp18-6* and as

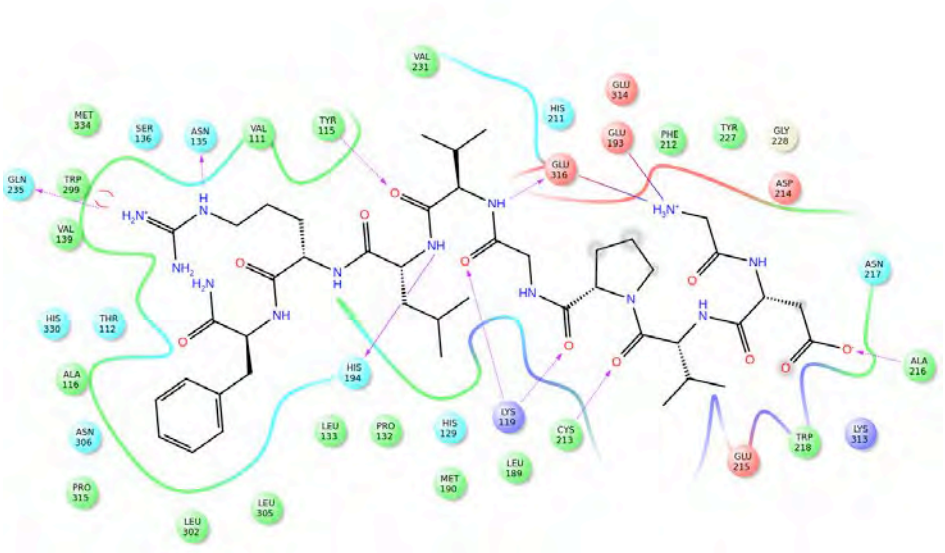
reported by Thornton.[63] Trp299 anchors the peptide in the binding pocket, preventing it from diffusing deeper into the binding pocket. Asn135 forms hydrogen bonds with the guanidinium ion and NH_2 of Arg7 from the peptide. As was postulated by Flocco and Mowbray, the π-cation interaction of 3.62 Å, is within the confines of < 6 Å length they observed.[62] Gln235 forms a hydrogen bond, which is an important interaction with NHε of the arginine of the peptide.

Cys213, on the short helix in ECL2, forms a 1.85 Å hydrogen bond with the N-terminal Ala1 (*af3*) protonated amide. Tyr227 donates a proton that potentiates the cation-π binding ability of phenol to the NH_3^+ cation on the N-terminal alanine in the peptide.[63,64] The loop residues Lys119, Phe212, Cys213, Asp214, Glu215, Glu316 and His194 are involved in ligand binding. Phe8 of the peptide has the aromatic ring hidden in the common hydrophobic pocket.

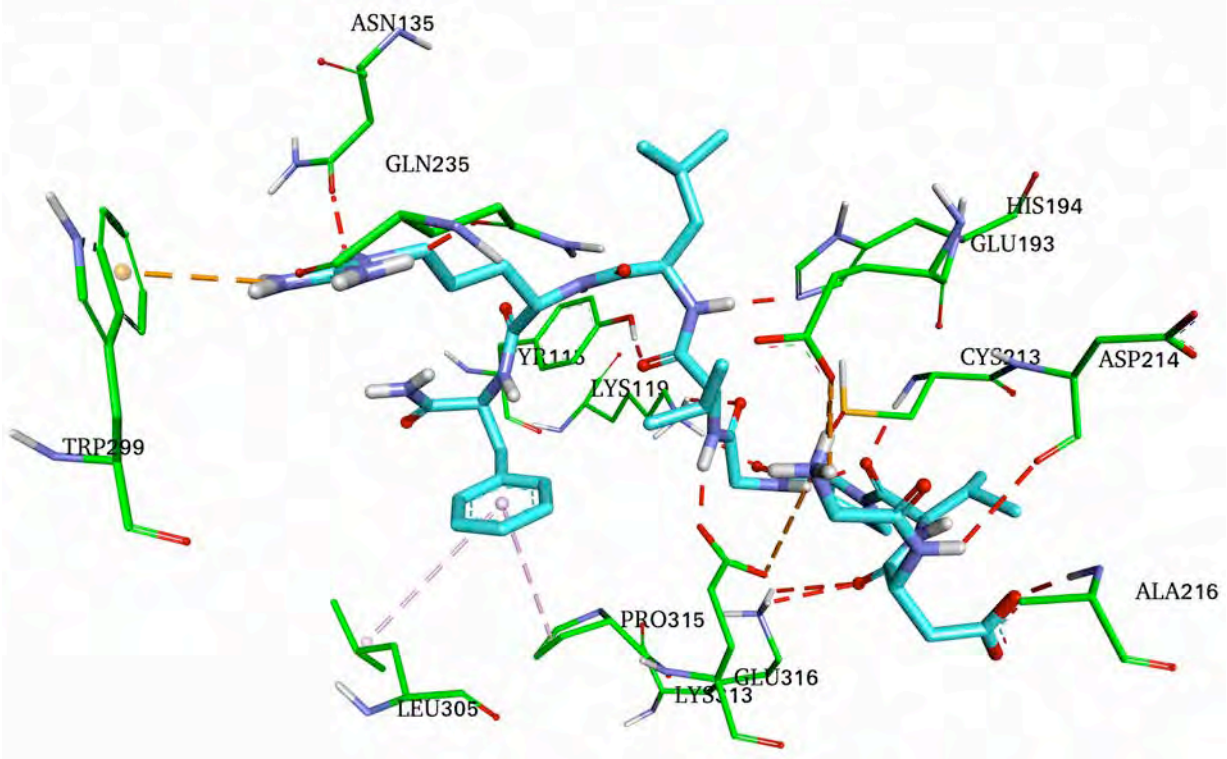
4.3.2.3 *af4 Docking*

The putative binding site residues that have been observed with *af3* and *flp18-6*, reveal similar trends with the *af4* peptide. The interaction between Arg7 and Trp299, was also observed between other peptides and *af4*. The ligand interaction maps using S-peptide, XP Glide and AutoDock4.2 of *af4* with the FLP18R1 receptor are shown in Figure 4.7 (a) to (d).

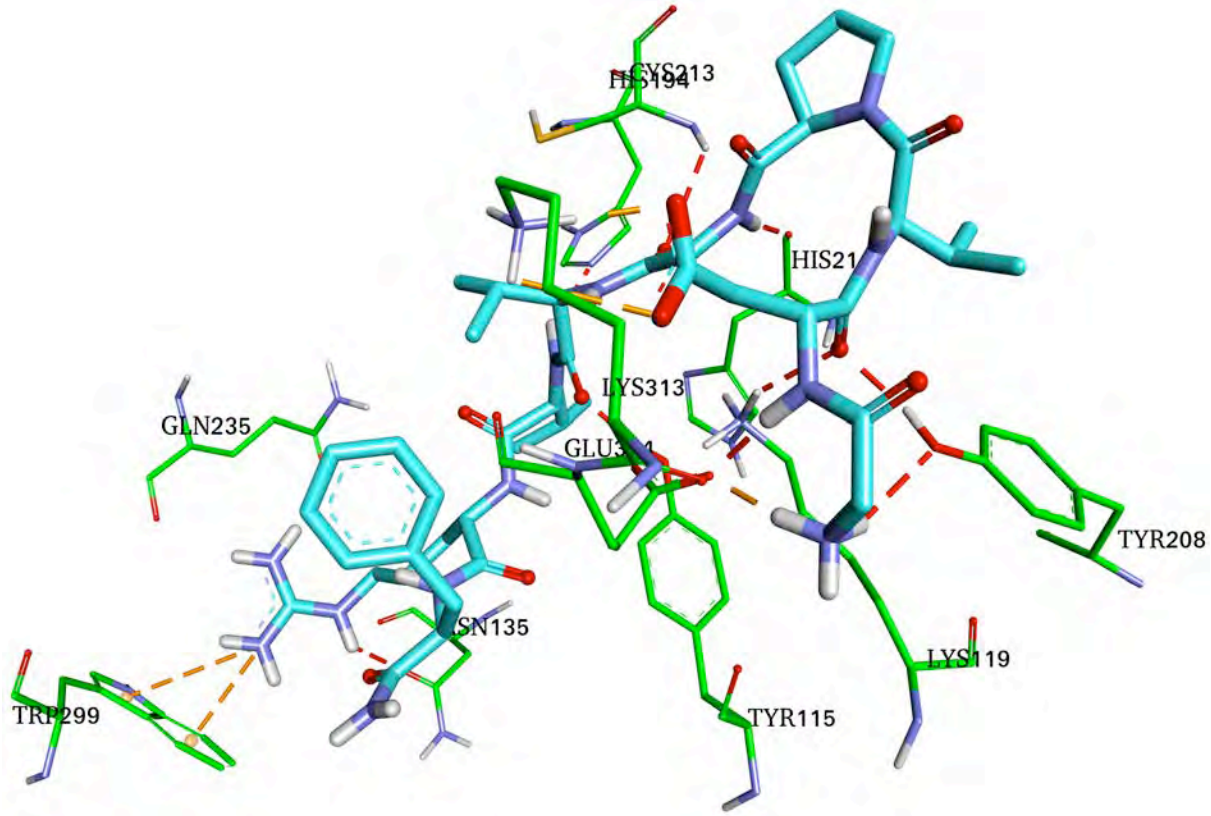
(a)



(b)


ASN135
GLN235
HIS194
GLU193
CYS213
ASP214
LYS119
TRP299
ALA216
PRO315
GLU316
LEU305

(c)


CYS213
HIS21
LYS313
GLN235
TYR208
LYS119
TYR115
TRP299

(d)

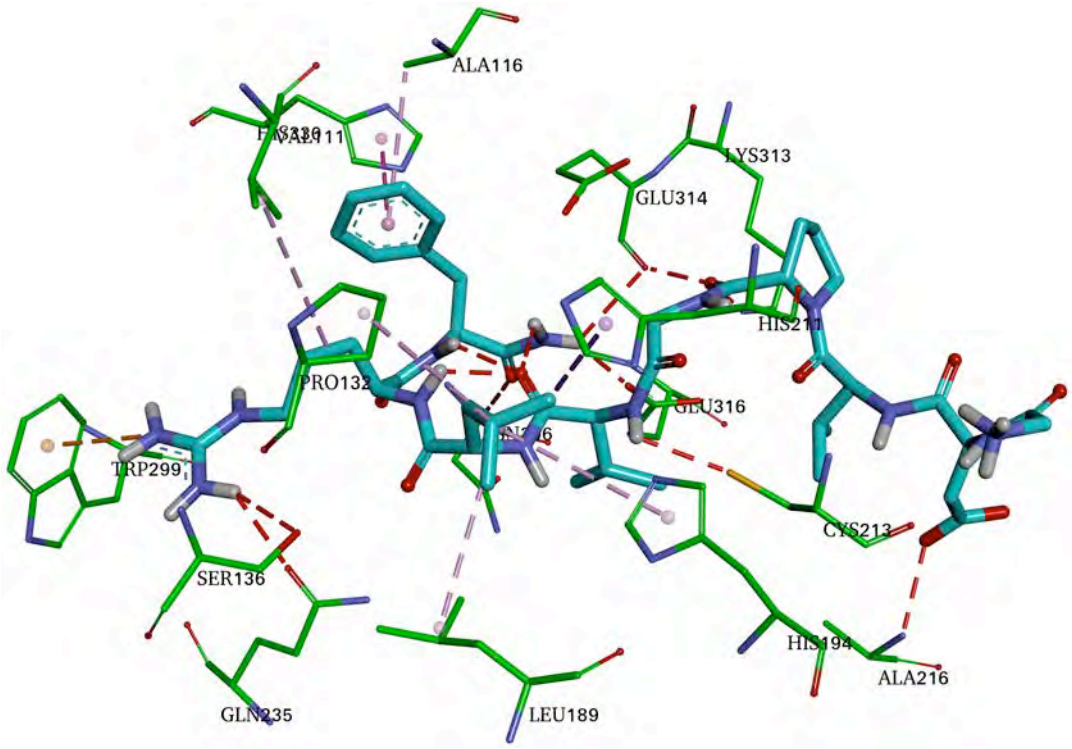


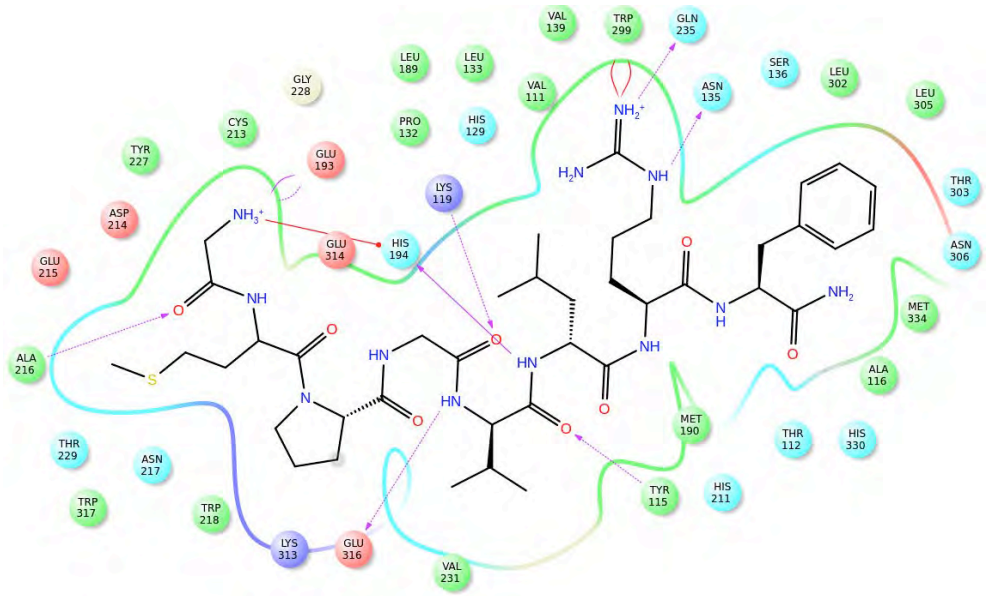
Figure 4.7: Binding mode of *af4 (cyan)* to FLP18R1 (green). (a) 2D ligand interaction map of *af4* peptide in the binding pocket. Protein-peptide binding interactions of *af4* from (b) S-peptide Glide (c) XP Glide (d) AutoDock 4.2 docking. Hydrogen bonds are shown by red dashed lines. Hydrophobic interactions (magenta)

During S-peptide docking, Glu193 and Glu316 were observed to form salt bridges with the protonated amine of Gly1 of the peptide N-terminus. A hydrophobic pocket made up of Pro132, Leu189, Leu302, Leu305 and Pro315 house the aromatic ring of the C terminal Phe9 of the peptide during S-peptide docking. When using XP Glide docking, the hydrophobic pocket of the Phe9 aromatic ring is made up of Ala116, Val 111, Thr112 and Leu 302. It is envisaged that the carbonyl oxygen of Asp2 (*af4*) interacts with the phenolic OH of Tyr208, forming a strong 1.88 Å hydrogen bond. Tyr 208 also binds to the Ala1 protonated amide. The important interactions that have so far been observed between the receptor and *af4* involve Asn135, Gln235 and Trp299 with the guanidinium ion of Arg on the C terminal of the peptide. The XP Glide score of -8.42 kJ/ mol was low compared to the S-peptide docking score of -10.50 kJ/ mol and AutoDock-4.2 score of -9.1 kJ/ mol.

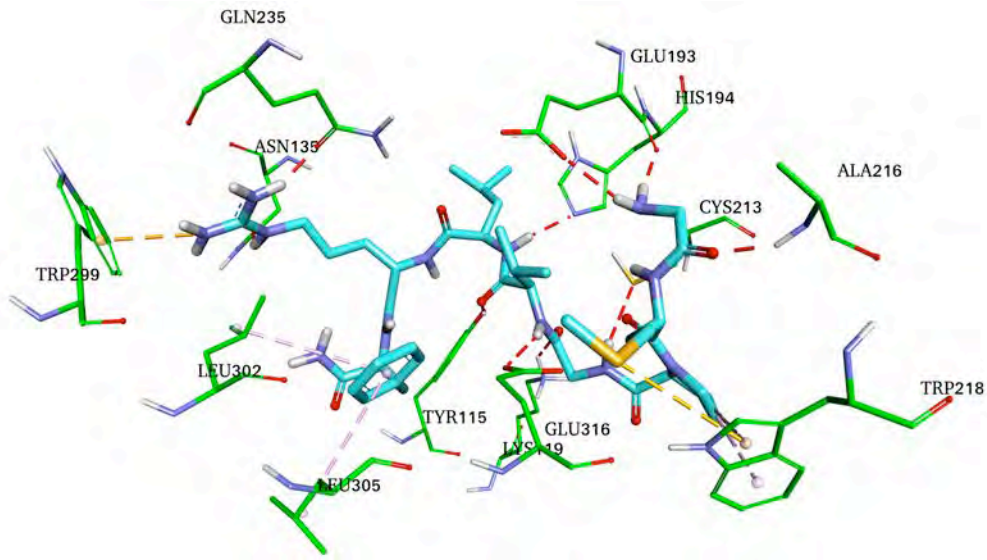
4.3.2.4 af20 docking

Docking calculations were also performed using *af20* as a ligand to the FLP18R1 receptor using a similar grid to that used for all the other docking calculations. Trp299 was also found to behave as an anchor in the binding pocket, forming a π-cation interaction with Arg7 at the C terminus of the peptide. Most interactions are similar to those shown for *flp18-6, af3* and *af4* and the ligand interaction map is shown in Figure 4.8 (a) to (d).

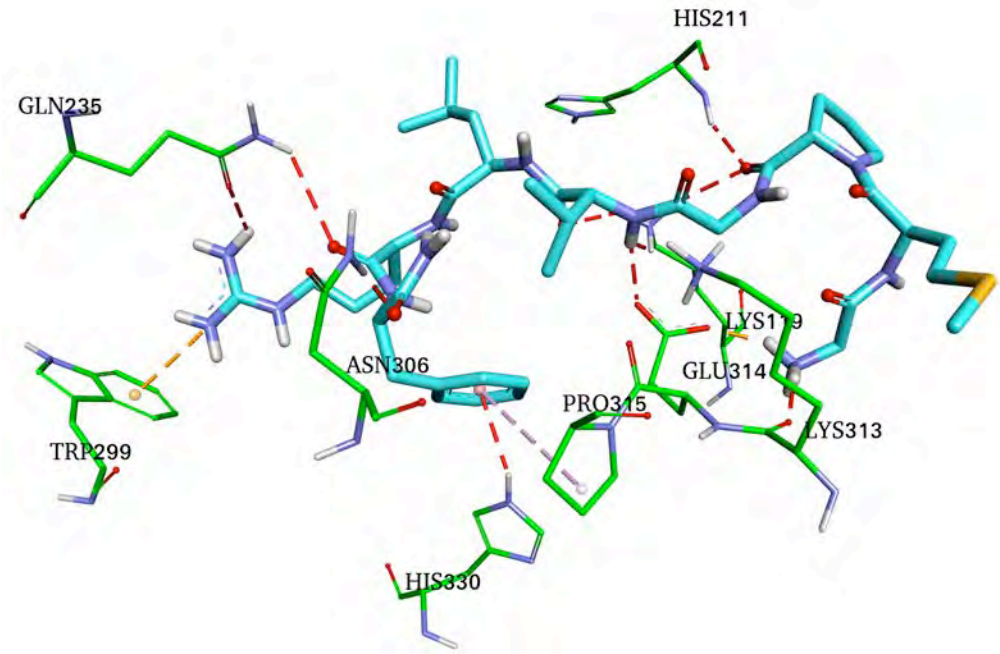
(a)



(b)



(c)



(d)

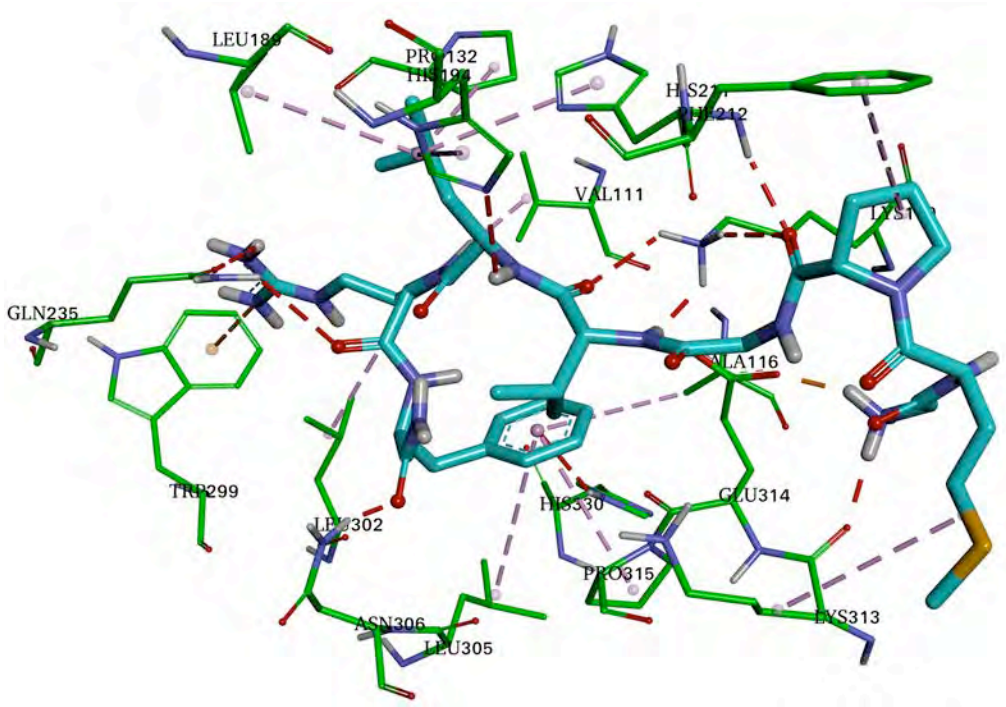


Figure 4.8: Binding mode of *af20 (cyan)* to FLP18R1 (green). (a) 2D ligand interaction map of the *af20* peptide in the binding pocket. Protein-ligand binding interactions of *af20* from (b) S-peptide Glide (c) XP Glide (d) AutoDock 4.2 docking interactions. Hydrogen bonds are indicated with red dashed lines and hydrophobic interactions (magenta)

A strong 1.83 Å hydrogen bond between helix 2 residue Tyr115 and Val5 of the peptide, and Glu316 the amide proton of Val5 lead to a stronger binding of the *af20* in the receptor during S-peptide docking (-11.18 kJ/ mol) than during XP Glide docking (-8.53 kJ/ mol) and AutoDock 4.2 (-9.0kJ/ mol). The important interactions of Asn135 (Helix 3), Gln235 (Helix 5) and Trp299 with the NHε and guanidinium ion[59,63] of the c-terminal Arg7 (af20) are conserved. The Val111 (Helix 2), Pro132 (Helix 3), Leu189 (Helix 4) and His194 (ECL2) form a hydrophobic pocket where the Leu6 (*af20*) and in all the other peptides is housed. The Met2 (*af20*) is also involved in hydrophobic interactions including a π-sulfur interaction with Cys213 in the ECL2 short helix.

4.4 Molecular Dynamics of Protein-Ligand complex

The complex of *flp18-6* with the FLP18R1 receptor was inserted in a mimetic POPC membrane and MD simulations performed over 100 ns in a water box. Since this is the main peptide of interest, further analyses were performed to understand its behavior. We performed these simulations in order to ascertain the accurate binding conformations and hydrogen bond occupancy as well as evaluate whether the peptide will sink beyond the Trp299 anchor during unrestrained MD simulations. After 1.5 ns of MD, the helices move towards [Figure 4.9(a)] each other on the extracellular[48,67] domain thus indicating that the receptor closes. At the same time the receptor opens up on the intracellular side. The ECL2 seems to be capping the binding site [Figure 4.9(b)].

(a) (b)

Figure 4.9: (a) Superimposition of the FLP18R1 receptor after docking (orange) and after 1.5 ns MD simulations (cyan) in a mimetic POPC membrane.

This correlates with theoretical and experimental findings in other GPCRs reported such as Adipokinetic hormone receptor (AKHR),[40] A_2A adrenergic receptor,[51,52] beta-2-adrenergic[54,67] among many other GPCRs. The Root Mean Square Deviation (RMSD)[9] between the starting structure and receptor structure after 1.5 ns was calculated to be 2.13 Å using the receptor backbone atoms.

However, the docking calculations reveal that, once bound in the binding site, as MD simulations were performed, the major interactions such as Asn135, Gln235 and the Trp299 were retained throughout the simulations. Tyr188 in helix 5 forms a strong hydrogen bond that was not observed after docking calculations as shown in Figure 4.10 after 1.5 ns MD simulations in a mimetic POPC membrane.

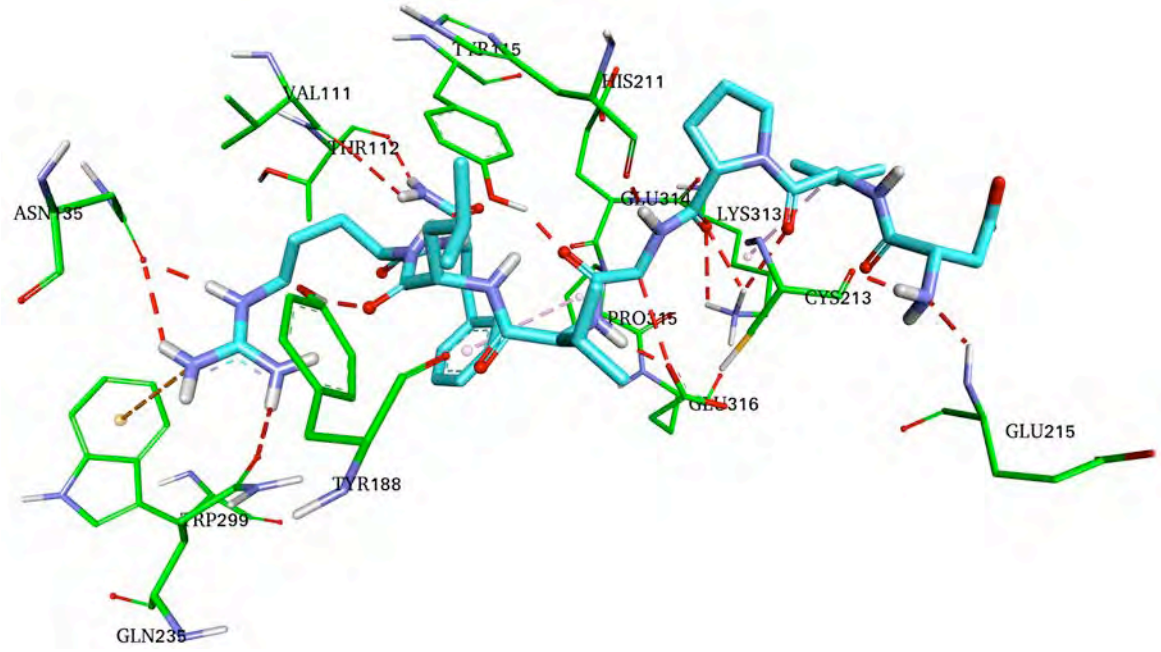


Figure 4.10: Protein-peptide interactions during MD simulations after 1.5 ns in a mimetic membrane. New interactions form such as the hydrogen bonds from the phenolic OH of Tyr188 to the carbonyl oxygen of Leu6 in *flp18-6*, Val 111, and Lys313 to the carbonyl oxygen of Val2 in *flp18-6*.

The binding of Tyr188 in helix 5 was found to be a result of the clockwise movement of helix 6 towards helix 3, and helix 7 towards helix 4 discussed earlier. The hydrophobic pocket where the leucine of the peptide was hidden was made up of Pro132, His211 and Thr112. The Val111 forms new hydrogen bonds with the peptide. The Leu302, Leu305 and Pro315 maintain the hydrophobic pocket where the aromatic ring of the C-terminus phenylalanine in the peptide.

The protein-ligand complex shows that the extracellular side closes as MD simulations continue. Cluster analysis of the complex after 100 ns MD simulations in a mimetic membrane show that the ligand is bound strongly and the intracellular side opens up. This is believed to be the activated complex and the new interactions formed are shown in Figure 4.11. The transmission switches and ionic locks in the intracellular domain are broken after ligand binding as a result of helical rearrangement.

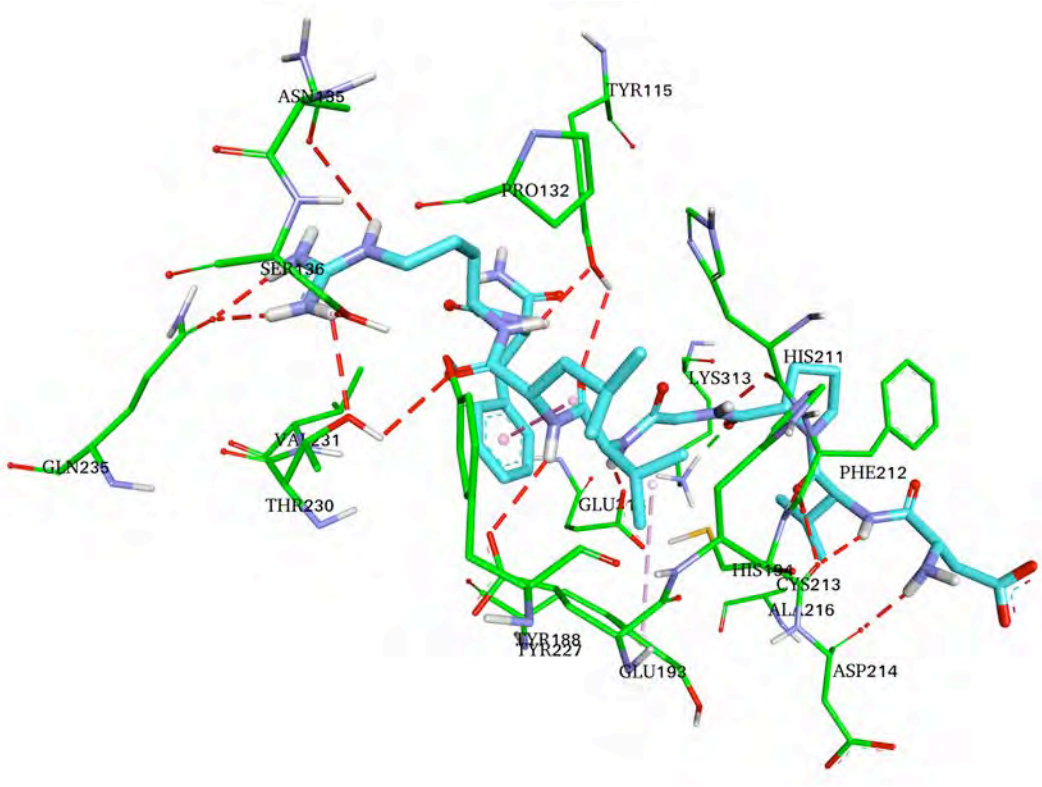


Figure 4.11: Protein-*flp18-6* complex after 75 ns showing the least energy bound conformation from MD simulations.

Cluster analysis of the FLP18R1 bound to *flp18-6* after 75 ns indicates the stability of the interactions of Tyr115 (Helix 2), Pro132, Asn135 and Ser136 (Helix 2), Tyr188 (Helix 4), Glu193, His194, Phe212, Cys213, Asp214 and Ala216 (ECL2), Thr230, Val 231 and Gln235 (Helix 5), Lys313, and Glu314 in ECL3. Most of these interactions were not observed after docking calculations since the binding pocket is large. However, the studies revealed that the bound conformation further close up at the top as the structure stabilizes further.

The carboxylate oxygen of Glu193 (ECL2), is involved in a hydrogen bond with the *flp18-6* peptide. This interaction is also observed in the co-crystal structure of A2A bound to ZM241385.[51] The side chain hydroxyl of Thr230 (Helix 5) binds to the carbonyl oxygen on the Phe8 of the peptide C-terminus. The role played by Phe168 in ECL2 of A_2A (Phe212 of FLP18R1) has been reported by Sabbadin in studies of the A_2A adenosine receptor revealing loss of binding affinity to ligands when this residue is not available.[20,65] Therefore, in FLP18R1, the interactions involving this residue are

observed during molecular dynamics. The orientations of the side chain of this residue changes and thus result in this important interaction.

4.5 Conclusions

FLP18R1, a GPCR was found to have a large binding cavity. There were limited steric clashes in the interactions observed. Shape recognition[68,69] of GPCR peptides is a common phenomenon with the general beta turn, that is critical for diffusion into the binding site, was identified.[70] The study revealed that during docking calculations, the cyclic conformation of the peptides was not maintained as the hydrogen bonds and the salt bridges cleave during the docking. When these break, the peptides assume an L-shaped conformation from the C or U-shaped conformation.[40]

The guanidinium ion of peptide arginine was involved in a conserved π-cation interaction with the Trp299 in all peptides binding to FLP18R1. Charge complementarity of the guanidinium ion in the arginine is important for this interaction.[68,71] ECL2 is important for ligand binding as is common in most class A GPCRs known to date such as beta-2- adrenergic and adenosine 2 adrenergic (A_2A) receptors.[15,45,51,52,65] The conserved hydrogen bonds were between the Arg (C-terminus of each peptide) Asn135 and Gln235 of the receptor.

The hydrophobic pockets made up of val111, Pro132, Leu189 and His194 and the other made up of Leu302, Leu305, Pro315 and His330 was used to house the Leucine and phenylalanine residues on the C-terminus of the peptides. Lys119 and Glu316 were the more conserved loop residues on the receptor involved in ligand binding. The local similarities in the binding pockets for ligands were detected in FLP18R1. The functional selectivity[72,73] of the hydrophobic and hydrophilic interactions between ligands (peptides) and the FLP18R1 receptor are vital for determining signal transduction pathways to give the required therapeutic effect.[74]

The peptides investigated in this study have shown that there may be pharmacological similarity between the four peptides, *af3, af4, af20* and *flp18-6* based on their

interactions with the receptor. According to Johnson and Maggiora (1990), the structural similarities[10,75] in related compounds, which may be true in the case of these peptides which share a common C-terminus motif, may as well illustrate similarities in the biological activities of these compounds.[76]

The binding modes of the peptides were similar using the three protocols: AutoDock4.2, XP Glide and S-peptide docking. Molecular Mechanics (MM-GBSA)[30–32] was used to select the best pose with the lowest binding energy. The poses that were selected had MM-GBSA energies of -134.19kcal/mol for *flp18-6*, -148.18 kcal/mol for *af3*, -163.42kcal/ mol for *af4* and -143.21kcal/mol for *af20* using S-peptide docking conformations. Generally, it can be concluded that peptide introduces small conformational changes in their binding partner, in this case the FLP18R1 receptor.

4.6 References

(1) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins Struct. Funct. Genet.* **2002**, *47* (4), 409–443.

(2) Onay, A.; Demiray, Y. S.; Hiçsönmez, S. Principles of Docking : An Overview of Docking Algorithms. *Survey Paper Review*, Turkey

(3) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and Validation of a Modular, Extensible Docking Program: DOCK 5. *J. Comput. Aided. Mol. Des.* **2006**, *20* (10-11), 601–619.

(4) Studi, D.; Milano, D. I.; Facolt, B.; Matematiche, S. Development of Methodologies for Molecular Docking and Their Applications. **2009**.

(5) Skolnick, J.; Zhou, H.; Gao, M. Are Predicted Protein Structures of Any Value for Binding Site Prediction and Virtual Ligand Screening? *Curr. Opin. Struct. Biol.* **2013**, *23* (2), 191–197.

(6) Pérot, S.; Sperandio, O.; Miteva, M. A.; Camproux, A. C.; Villoutreix, B. O. Druggable Pockets and Binding Site Centric Chemical Space: A Paradigm Shift in Drug Discovery. *Drug Discov. Today* **2010**, *15* (15-16), 656–667.

(7) Kalyaanamoorthy, S.; Chen, Y. P. P. Structure-Based Drug Design to Augment Hit Discovery. *Drug Discov. Today* **2011**, *16* (17-18), 831–839.

(8) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9* (1), 71.

(9) Dalton, J. a R.; Jackson, R. M. Homology-Modelling Protein-Ligand Interactions: Allowing for Ligand-Induced Conformational Change. *J. Mol. Biol.* **2010**, *399* (4), 645–661.

(10) Pearson, W. R. Selecting the Right Similarity-Scoring Matrix. *Curr. Protoc. Bioinforma.* **2013**, No. SUPL.43.

(11) Hein, M.; Zilian, D.; Sotriffer, C. a. Docking Compared to 3D-Pharmacophores: The Scoring Function Challenge. *Drug Discov. Today Technol.* **2010**, *7* (4), e229–e236.

(12) Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M. High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5* (10).

(13) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935–949.

(14) Vyas, V.; Jain, A.; Jain, A.; Gupta, A. Virtual Screening: A Fast Tool for Drug Design. *Sci. Pharm.* **2008**, *76* (3), 333–360.

(15) Ghersi, D.; Sanchez, R. Improving Accuracy and Efficiency of Blind Protein-Ligand Docking by Focusing on Predicted Binding Sites. *Proteins Struct. Funct. Bioinforma.* **2009**, *74* (2), 417–424.

(16) Hetényi, C.; Van Der Spoel, D. Toward Prediction of Functional Protein Pockets Using Blind Docking and Pocket Search Algorithms. *Protein Sci.* **2011**, *20* (5), 880–893.

(17) Hetényi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2002**, *11* (7), 1729–1737.

(18) Lindstrom, W.; Morris, G. M.; Weber, C.; Huey, R. Using AutoDock 4 for Virtual Screening. *Screening* **2008**, No. January, 1–37.

(19) Ma, J.; Chen, B. Identification of Falcipain-2 Binding Sites Using Docking Methods AUTODOCK Methods. *Report for I619 final project*.

(20) Beuming, T.; Sherman, W. Current Assessment of Docking into GPCR Crystal Structures and Homology Models: Successes, Challenges, and Guidelines. *J. Chem. Inf. Model.* **2012**, *52* (12), 3263–3277.

(21) Huang, S. Y.; Zou, X. Advances and Challenges in Protein-Ligand Docking. *Int. J. Mol. Sci.* **2010**, *11* (8), 3016–3034.

(22) Adams, R.; Worth, C. L.; Guenther, S.; Dunkel, M.; Lehmann, R.; Preissner, R. Binding Sites in Membrane Proteins - Diversity, Druggability and Prospects. *Eur. J. Cell Biol.* **2012**, *91* (4), 326–339.

(23) Orry, A. J.; Wallace, B. A. Modeling and Docking the Endothelin G-Protein-Coupled Receptor. *Biophys. J.* **2000**, *79* (6), 3083–3094.

(24) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. II. Are Homology Models of G-Protein Coupled Receptors Suitable Targets? *Proteins Struct. Funct. Genet.* **2003**, *50* (1), 5–25.

(25) Schneider, G.; Baringhaus, K. H. Molecular Design. Concepts and Applications; 2008; Vol. 3.

(26) Huey, R.; Morris, G. M. Using AutoDock with AutoDockTools: A Tutorial. *Scripps Res. Institute, Mol. Graph. Lab. La Jolla, CA* **2003**, 1–56.

(27) Taylor, J. S.; Burnett, R. M. DARWIN: A Program for Docking Flexible Molecules. *Proteins Struct. Funct. Genet.* **2000**, *41* (2), 173–191.

(28) Isin, B.; Estiu, G.; Wiest, O.; Oltvai, Z. N. Identifying Ligand Binding Conformations of the β2-Adrenergic Receptor by Using Its Agonists as Computational Probes. *PLoS One* **2012**, *7* (12).

(29) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494* (7436), 185–194.

(30) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, 1–13.

(31) Mulakala, C.; Viswanadhan, V. N. Could MM-GBSA Be Accurate Enough for Calculation of Absolute Protein/ligand Binding Free Energies? *J. Mol. Graph. Model.* **2013**, *46*, 41–51.

(32) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31* (4), 797–810.

(33) Godschalk, F.; Genheden, S.; Söderhjelm, P.; Ryde, U. Comparison of MM/GBSA Calculations Based on Explicit and Implicit Solvent Simulations. *Phys. Chem. Chem. Phys.* **2013**, *15* (20), 7731–7739.

(34) Leach, A. R. Molecular Modelling: Principles and Applications; **2001**.

(35) Goodsell, D. S. Computational Docking of Biomolecular Complexes with Autodock. *Cold Spring Harb. Protoc.* **2009**, *4* (5).

(36) Niv, M. Y.; Skrabanek, L.; Filizola, M.; Weinstein, H. Modeling Activated States of GPCRs: The Rhodopsin Template. *J. Comput. Aided. Mol. Des.* **2006**, *20* (7-8), 437–448.

(37) Reddy, C. S.; Vijayasarathy, K.; Srinivas, E.; Sastry, G. M.; Sastry, G. N. Homology Modeling of Membrane Proteins: A Critical Assessment. *Comput. Biol. Chem.* **2006**, *30* (2), 120–126.

(38) Pierri, C. L.; Parisi, G.; Porcelli, V. Computational Approaches for Protein Function Prediction: A Combined Strategy from Multiple Sequence Alignment to Molecular Docking-Based Virtual Screening. *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804* (9), 1695–1712.

(39) Yarnitzky, T.; Levit, A.; Niv, M. Y. Homology Modeling of G-Protein-Coupled Receptors with X-Ray Structures on the Rise. *Curr. Opin. Drug Discov. Devel.* **2010**, *13* (3), 317–325.

(40) Mugumbate, G.; Jackson, G. E.; Van Der Spoel, D.; Kövér, K. E.; Szilágyi, L. Anopheles Gambiae, Anoga-HrTH Hormone, Free and Bound Structure-A Nuclear Magnetic Resonance Experiment. *Peptides* **2013**, *41*, 94–100.

(41) Gloriam, D. E.; Foord, S. M.; Blaney, F. E.; Garland, S. L. Definition of the G Protein-Coupled Receptor Transmembrane Bundle Binding Pocket and Calculation of Receptor Similarities for Drug Design. *J. Med. Chem.* **2009**, *52* (14), 4429–4442.

(42) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234.

(43) Bharatham, N.; Bharatham, K.; Shelat, A. A.; Bashford, D. Ligand Binding Mode Prediction by Docking: Mdm2/Mdmx Inhibitors as a Case Study. *J. Chem. Inf. Model.* **2014**, *54* (2), 648–659.

(44) Priya, S. S. L.; Devi, R. P.; Madeswaran, A. *In Silico* Docking Studies of RP2 (X-Linked Retinitis Pigmentosa) Protein Using Anthocyanins as Potential Inhibitors. *Bangladesh J. Pharmacol.* **2013**, *8* (3), 292–299.

(45) Cherezov, V.; Rosenbaum, D.; Hanson, M.; Rasmussen, S.; Thian, F.; Kobilka, T.; Choi, H.; Kuhn, P.; Weis, W.; Kobilka, B.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human beta2-Adrenergic G Protein-Coupled Receptor. *Science (80-).* **2007**, *318*, 1258–1265.

(46) Schiedel, A. C.; Hinz, S.; Thimm, D.; Sherbiny, F.; Borrmann, T.; Maass, A.; Müller, C. E. The Four Cysteine Residues in the Second Extracellular Loop of the Human Adenosine A2B Receptor: Role in Ligand Binding and Receptor Function. *Biochem. Pharmacol.* **2011**, *82* (4), 389–399.

(47) Vilar, S.; Ferino, G.; Phatak, S. S.; Berk, B.; Cavasotto, C. N.; Costanzi, S. Docking-Based Virtual Screening for Ligands of G Protein-Coupled Receptors: Not Only Crystal Structures but Also *in Silico* Models. *J. Mol. Graph. Model.* **2011**, *29* (5), 614–623.

(48) Schneider, M.; Wolf, S.; Schlitter, J.; Gerwert, K. The Structure of Active Opsin as a Basis for Identification of GPCR Agonists by Dynamic Homology Modelling and Virtual Screening Assays. *FEBS Lett.* **2011**, *585* (22), 3587–3592.

(49) Sandeep, G.; Nagasree, K. P.; Hanisha, M.; Kumar, M. M. K. AUDocker LE: A GUI for Virtual Screening with AUTODOCK Vina. *BMC Res. Notes* **2011**, *4* (1), 445.

(50) Werner, T.; Morris, M. B.; Dastmalchi, S.; Church, W. B. Structural Modelling and Dynamics of Proteins for Insights into Drug Interactions. *Adv. Drug Deliv. Rev.* **2012**, *64* (4), 323–343.

(51) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6-Angstrom Crystal Structure of a Human A_2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (5905), 1211–1217.

(52) 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 A_2A Adenosine Receptor. *Science* **2011**, *332* (6027), 322–327.

(53) Vohra, S.; Taddese, B.; Conner, A. C.; Poyner, D. R.; Hay, D. L.; Barwell, J.; Reeves, P. J.; Upton, G. J.; Reynolds, C. A. Similarity between Class A and Class B G-Protein-Coupled Receptors Exemplified through Calcitonin Gene-Related Peptide Receptor Modelling and Mutagenesis Studies. *J R Soc Interface* **2013**, *10* (79), 20120846.

(54) Katritch, V.; Cherezov, V.; Stevens, R. C. Structure-Function of the G Protein-Coupled Receptor Superfamily. *Annu Rev Pharmacol Toxicol* **2013**, *53*, 531–556.

(55) Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K. Crystal Structure of the M -Opioid Receptor Bound to a Morphinan Antagonist. **2012**, 0–6.

(56) Mueller, A.; Strange, P. G. The Chemokine Receptor, CCR5. *International Journal of Biochemistry and Cell Biology*, 2004, *36*, 35–38.

(57) de Voux, A.; Chan, M. C.; Folefoc, A. T.; Madziva, M. T.; Flanagan, C. a. Constitutively Active CCR5 Chemokine Receptors Differ in Mediating HIV Envelope-Dependent Fusion. *PLoS One* **2013**, *8* (1).

(58) a., L.; C.D., G. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.

(59) Flocco, M. M.; Mowbray, S. L. Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **1994**, *235* (2), 709–717.

(60) Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J. A Survey of Aspartate-Phenylalanine and Glutamate-Phenylalanine Interactions in the Protein Data Bank: Searching for Anion-π Pairs. *Biochemistry* **2011**, *50* (14), 2939–2950.

(61) Dossey, A. T.; Reale, V.; Chatwin, H.; Zachariah, C.; DeBono, M.; Evans, P. D.; Edison, A. S. NMR Analysis of Caenorhabditis Elegans FLP-18 Neuropeptides: Implications for NPR-1 Activation. *Biochemistry* **2006**, *45* (24), 7586–7597.

(62) Flocco, M. M.; Mowbray, S. L. Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **1994**, *235* (2), 709–717.

(63) Mitchell, J. B.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. Amino/aromatic Interactions in Proteins: Is the Evidence Stacked against Hydrogen Bonding? *J. Mol. Biol.* **1994**, *239* (2), 315–331.

(64) Gallivan, J. P.; Dougherty, D. a. Cation-Pi Interactions in Structural Biology. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (17), 9459–9464.

(65) Sabbadin, D.; Ciancetta, A.; Moro, S. Bridging Molecular Docking to Membrane Molecular Dynamics to Investigate GPCR-Ligand Recognition: The Human A2A Adenosine Receptor as a Key Study. *J. Chem. Inf. Model.* **2014**, *54* (1), 169–183.

(66) Grossfield, A. Recent Progress in the Study of G Protein-Coupled Receptors with Molecular Dynamics Computer Simulations. *Biochim. Biophys. Acta* **2011**, *1808* (7), 1868–1878.

(67) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of Two Distinct Inactive Conformations of the beta2-Adrenergic Receptor Reconciles Structural and Biochemical Observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4689–4694.

(68) Sulea, T.; Purisima, E. O. Profiling Charge Complementarity and Selectivity for Binding at the Protein Surface. *Biophys. J.* **2003**, *84* (5), 2883–2896.

(69) Hiller, C.; Kühhorn, J.; Gmeiner, P. Class A G-Protein-Coupled Receptor (GPCR) Dimers and Bivalent Ligands. *J. Med. Chem.* **2013**, *56* (17), 6542–6559.

(70) Erguner, B.; Hattori, M.; Kanehisa, M. Characterizing Common Substructures of Ligands for GPCR Protein Subfamilies, **2010**.

(71) Chau, P. L.; Dean, P. M. Electrostatic Complementarity between Proteins and Ligands. 1. Charge Disposition, Dielectric and Interface Effects. *J. Comput. Aided. Mol. Des.* **1994**, *8* (5), 513–525.

(72) Pogozheva, I. D.; Przydzial, M. J.; Mosberg, H. I. Homology Modeling of Opioid Receptor-Ligand Complexes Using Experimental Constraints. *AAPS J.*

2005, *7* (2), E434–E448.

(73) Bissantz, C.; Logean, A.; Rognan, D. High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building, and Receptor Library Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1162–1176.

(74) Rognan, D. Chemogenomic Approaches to Rational Drug Design. *Br. J. Pharmacol.* **2007**, *152* (1), 38–52.

(75) Kooistra, A. J.; Kuhne, S.; De Esch, I. J. P.; Leurs, R.; De Graaf, C. A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *Br. J. Pharmacol.* **2013**, *170* (1), 101–126.

(76) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57* (8), 3186–32

5 Chapter 5: Structure-Based Virtual Screening of NCI database and *in vitro* evaluation for the identification of FLP18R1 antagonists

5.0 Introduction

The core aim in drug discovery is to identify new chemical entities that are effective with enhanced potency, and possess desirable physicochemical and pharmacokinetic properties.[1–3] In order to identify hit and lead compounds in the drug discovery and development cascade, structure-based and ligand-based drug design approaches have been executed.[4,5] Structure-based drug design strategies rely on a knowledge or availability of 3D structure of the target protein determined either through NMR, X-ray crystallography or comparative modeling.[6–8] When the 3D structures are known, compound libraries and fragment databases can be docked into the ligand-binding cavity of the target to identify potent ligands.[9,10] If the 3D structures are not available, comparative modeling can be performed against suitable templates that are homologous to the protein of interest.[11,12]

In cases where active and inactive compounds are known, ligand-based approaches are employed to identify compounds with similar chemical features to the known ligands.[5,13,14] Ligand-based approaches may involve 2D or 3D similarity searches and as QSAR-pharmacophore models used to screen compound databases.[15–18] The compounds obtained from both approaches can then be tested for *in vitro* and *in vivo* efficacy.

In this study, comparative (homology) modeling of the FLP18R1 receptor was performed and reported in Chapters 3 and 4, and the structure of the peptides that are known to bind to the receptor were determined in Chapter 2. The binding cavity and binding site residues of the receptor were discussed in Chapter 4. This chapter describes structure-based virtual screening (docking-based) taking into account the protein chemoprints from the docking calculations in Chapter 4. ECL2 and ECL3 are involved in the binding of the ligands through the conserved Glu193 and Glu314 as well as helices 2 (Pro132, Asn135, Ser136), 5 (His194) and 6 (Trp299, Asn306). The

disulfide bonds that were reported in Chapter 3 facilitate access of the ligands from the extracellular side.

5.1 Virtual screening of the National Cancer Institute (NCI) Database

The FLP18R1 GPCR receptor bound to the peptide ligands, *af3*, *af4*, *af20* and *flp18-6* was used to perform virtual screening. It was observed in Chapter 3 that the homology model opened up in the extracellular domain during 100 ns molecular dynamic simulations in a mimetic POPC membrane. This may have been caused by the relaxation and stabilization of the receptor's helices during MD simulations and the extracellular side of FLP18R1 open up. FLP18R1 was modeled against beta-2-adrenergic receptor bound to carazolol, a partial inverse agonist. The bound form has a closed extracellular domain. As reported in Chapter 4, the two-disulfide bridges in the extracellular domain enable accessibility of the binding side through the extracellular domain.

Docking calculations on the peptide ligands in Chapter 4 reveal that when the peptides bind to the receptor, and unrestrained 100 ns MD simulations of the complex are performed, the extracellular side close up[19] and the peptides sink deeper into the binding pocket as the receptor stabilizes. In the intracellular region, the ionic lock between Helices 3 and 6 breaks after peptide binding when the receptor is activated. As such, the intracellular side opens up. The binding site residues of the receptor to the peptide ligands are used as chemoprints for the small molecules database screening.

Docking-based virtual screening of the NCI database on the FLP18R1 GPCR receptor was performed. This database is easily accessible and compounds are available free of charge. The schematic cascade of the virtual screening workflow is given in Figure 5.1. An analysis of class A GPCR chemoprints reveal that conserved residues in the GPCRs are responsible for most of the ligand binding.[20,21] The combination of the structure-based approaches and examination of peptide ligand interactions in the

conserved residue region was envisaged to overcome some of the characteristic limitations of computational modelling.[22,23]

5.2 Materials and Methods

5.2.1 Preparation of the FLP18R1 Protein receptor

The optimized homology model of FLP18R1 generated and reported in Chapter 3 and used in docking calculations in Chapter 4 was employed in virtual screening. The protein structure was prepared for docking using the Protein Preparation Wizard from the Schrodinger package.[24] The preparation entailed correcting bond orders, charges and adding polar hydrogen atoms. The OPLS_2005 all-atom force field was employed for restrained energy minimization. Hydrogen bond assignment and protonation states of histidine, and the orientation of the hydroxyl, asparagine and glutamine were optimized.

5.2.2 Database preparation

The ligand dataset containing 265,262 molecules was obtained from National Cancer Institute database[25] available on http://cactus.nci.nih.gov/download/nci/. Ligand database preparation was performed using filters of drug-like properties and the dataset was filtered to collect 244,672 unique drug-like molecules using Discovery Studio 4.0 (Accelerys Inc., San Diego, CA). Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) descriptors[26] (Blood Brain Barrier (BBB), aqueous solubility, hepatotoxicity, plasma protein binding, Cyp2D6 binding and intestinal absorption properties) were employed to further filter the dataset. Lipinski's "rule of five"[27] guidelines observed were hydrogen bond donors < 5, hydrogen bond acceptors <10, molecular weight < 500, C logP <5. Functional group filters were applied to eliminate reactive functional groups such as reactive alkyl halide peroxide and certain unsuitable leads such as crown ethers as well as unacceptable natural products such as quinones and cycloheximidine derivatives [28–31]returned 230,557 molecules.[32–34] These ADMET descriptors help to remove molecules that may otherwise fail or have issues

down the drug discovery pipeline and thus ensure that some molecules are eliminated prior to synthesis.[26,35]

The ligand preparation protocol Ligprep-2.9,[36] in Maestro version 9.7 was used to generate three-dimensional coordinates of the 230,557 ligands. Proper bond and bond orders were assigned as well as van der Waal's parameters.[37] Specified chiralities were retained as a default, accessible tautomers and protonation states were generated using Epik version 2.7.[36] Charges were calculated to predict pKa's for each potential protonation centre and a structure for each charge combination in the given pH range of 7 ± 2 was generated.

5.2.3 Docking calculations

Glide,[8,37] implemented in the Schrodinger package was used to perform the molecular docking experiments. The ligands were given full torsional freedom[38] while the receptor was kept rigid. The grid was generated considering the SiteMap[36,39] cavity dimensions of 34 x 38 x 40 Å with the ligand-midpoint box of (34.17, 38.48, 38.49) Å with dimensions sufficient to accommodate compounds of a length ≤ 25A. A scaling factor of 1.0 was applied to the van der Waals (vdW) radii to soften potential for nonpolar parts of the receptor, with a partial charge cut-off of 0.25e.

Standard Precision (SP) and Extra Precision (XP) scoring functions of Glide were implemented.[8] Bias sampling of torsions for nonpolar conformation of amides only was employed, and Epik state penalties were added to docking scores. A post-docking minimization for the flexible ligands was executed on the output complexes to reduce the number of poses initially collected from 25 to 5 per ligand. 10 % of SP docking top ranking poses were selected for rescoring with the XP scoring function of Glide. Prime[40] MM-GBSA[40,41] algorithm was used to predict the binding free energy estimations and to rescore the selected poses. Visual inspection of the binding poses and orientation was done using Maestro and PyMol.[42]

5.2.4 Biological assays

Selected compounds from the NCI database[25] were assessed for antagonist activity against three invertebrate G protein-coupled receptors (GPCRs) functionally expressed in the yeast *Saccharomyces cerevisiae*;[43,44] two receptors from *Caenorhabditis elegans* namely Ce4,[45] a FLP18 receptor and Ce50, a FLP2 receptor and one from *Drosophila melanogaster*,[46] namely DM4, an Allatostatin IV receptor.

Three *S. cerevisiae* strains were obtained under license from Cadus Corp.[44,45] Each strain was transfected with the vector Cp4258 cloned with cDNA encoding either a FLP-18 receptor, Ce4, a FLP-2 receptor, Ce50, both from *C. elegans* or DM4, an Allatostatin IV receptor from *D. melanogaster*.[47,48] In the Cadus system, the yeast GPCR is deleted and the Gα subunit (GPA1) is replaced by one of several yeast-mammalian Gα chimeras that enhance recognition of heterologous GPCRs.[49,50] Also, receptor activation has been phenotypically linked to induction of an enzyme that repairs an auxotrophy: HIS3 to reverse a his^- phenotype (Wang et al., 2006; Kimber et al., 2009).[45]

Positive transformants were selected on agar plates with Complete Minimal (CM) dropout medium (Sigma–Aldrich, St. Louis, MO) lacking leucine (CM/Leu^-) and used for receptor activation assays in CM liquid medium (CM/Leu^-/His^-), supplemented with 0.05 M MOPS, pH 6.8. Briefly, a colony was grown overnight in CM/Leu^- at 250 rpm and 30 ˚ C, cells were harvested by centrifugation at 4000 *g* for 1 minute and rinsed thoroughly in CM/Leu^-/His^- to eliminate histidine. Cell density was established by absorbance at 600 nm and 3000 cells/well were seeded in CM/Leu^-/His^- in a 96-well microtiter plate with test compound (100 µM), Alamar blue (10%) (Sigma–Aldrich) and respective peptide ligands at 10 µM or histidine.

The respective peptide ligands were used as controls, i.e. FLP18, FLP2 and Allatostatin IV, respectively.[45,51] After 44 hrs incubation at 30˚C, cell growth was measured by Alamar blue fluorescence in a Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek) set at 560 nm excitation/590 nm emission. Experiments were performed in two replicates and absence of test compound was used as positive control and absence of peptide ligand as negative control. Only assays with a Z-factor of 0.5-1 were used for further analysis. A test compound was considered a hit with a

z-score[52] (Equation 2) of -4.0 or less. A hit compound inhibits growth of yeast in the presence of peptide ligand but not in the presence of histidine.

Equation 1

$$Z\text{-}factor = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

Equation 2

$$Z = \frac{x - \mu}{\sigma}$$

Where;

σ_p= standard deviation of positive controls

σ_n= standard deviation of negative controls

μ_p= mean of positive controls

μ_n= mean of negative controls

μ = mean of the population

σ = standard deviation of the population

x = sample score

5.3 Results and Discussion

5.3.1 Structure-based Virtual Screening

The virtual screening protocol (Figure 5.1) using SP Glide docking of these 230,557 compounds returned 7403 compounds with a glide score cut-off of -7.5 kcal/mol. The top 10% (1920) of the compounds post docking were selected for further docking in XP docking with a cut-off XP Glide score of -7.5 kcal/mol. This yielded 415 compounds. Visual inspection of these compounds was then undertaken. The interactions of the receptor residues in the binding cavity that were involved in ligand binding with the peptide were observed and considered. Post-docking refinement and rescoring was performed with Prime MM-GBSA[41,40,53] and cluster analysis using Extended Connectivity Fingerprint (ECFP4)[54] and 2D[16] similarity to the peptide ligands. 49 compounds were selected and observed to have similar interactions with conserved residues as was observed with the peptide ligands (Chapter 4) in the putative binding site. Nine of these compounds were not available for purchase from

the NCI; hence 40 hit compounds were submitted for biological evaluation. The virtual screening workflow applied is given in Figure 5.1.

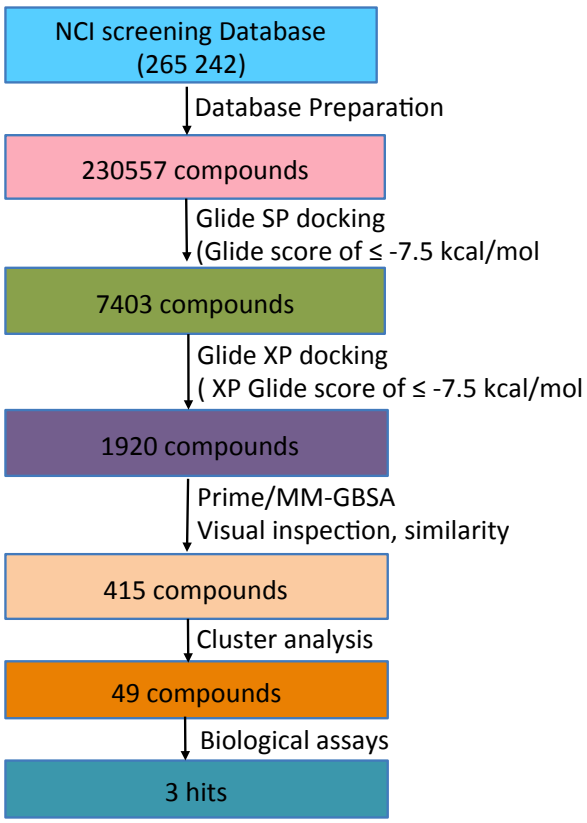
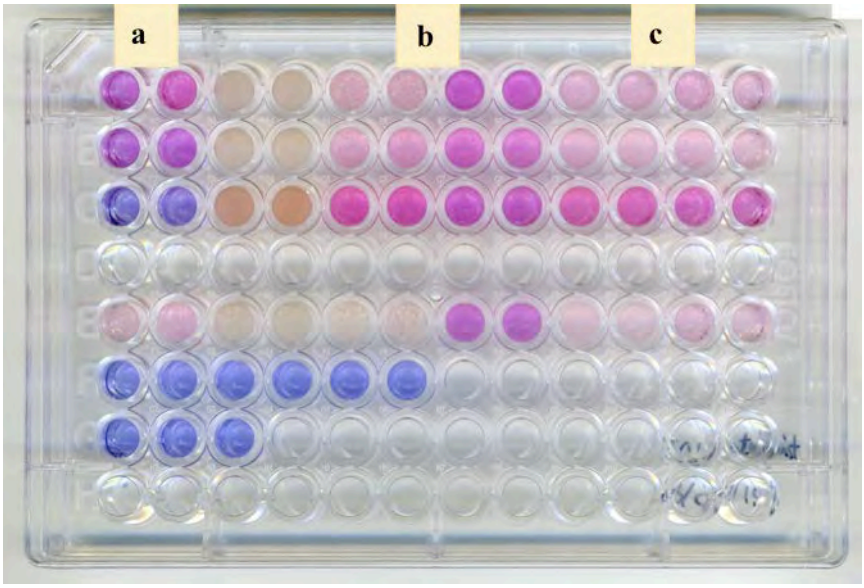


Figure 5.1: Schematic representation of virtual screening strategy adopted in this study

5.3.2 In vitro assays

Using *in vitro* assays, three compounds were found to have antagonist activity at 100 µM against Ce4 FLP-18 receptor of *C. elegans* heterologously expressed in *S. cerevisiae* (Figure 5.2). There was no observed activity against Ce50 (FLP2) and DM4 (*Drosophila melanogaster*)[46,55] receptors. Results are shown in Appendix 5. HIS3 was used as a reporter gene and Alamar blue used to monitor the growth. The active compounds are NCI 327396, NCI 303243 and NCI 15228 which are given in Table 5.1. NCI 327396 had the strongest activity among tested samples. Assays in the agonist mode did not yield any actives (data not shown).

A



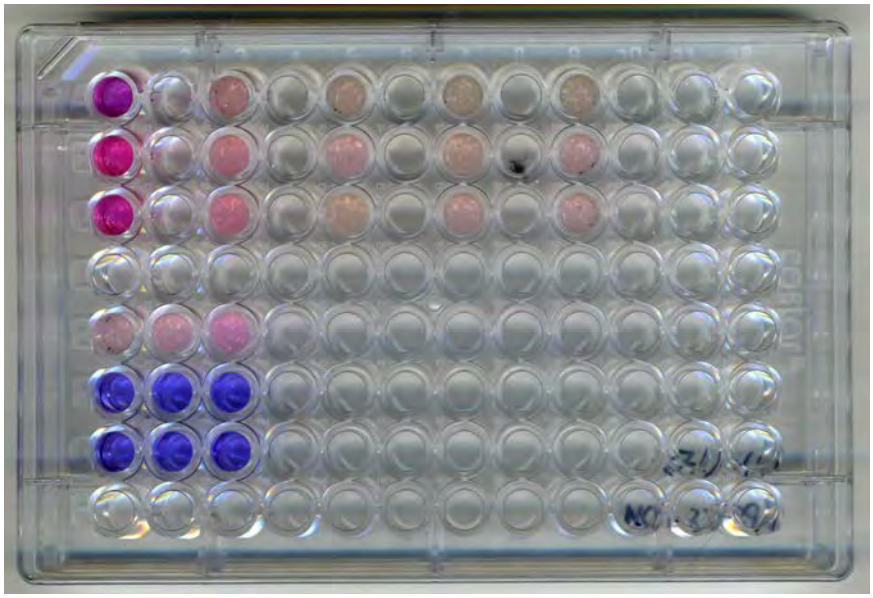
B

A	Ce4 + flp18 + 15228	Ce4 + flp18 + 15228	Ce50 + flp2 + 15228	Ce50 + flp2 + 15228	DM4+ *A+ 15228	DM4 + A++ 15228	Ce4 + his + 15228	Ce4 + his + 15228	Ce50 + his + 15228	Ce50 + his + 15228	DM4 + his + 15228	DM4 + his + 15228
B	Ce4 + flp18 + 303243	Ce4 + flp18 + 303243	Ce50 + flp2 + 303243	Ce50 + flp2 + 303243	DM4 + *A+ 303243	DM4 + A+ 303243	Ce4 + his + 303243	Ce4 + his + 303243	Ce50 + his + 303243	Ce50 + his + 303243	DM4 + his + 303243	DM4 + his + 303243
C	Ce4 + flp18 + 327396	Ce4 + flp18 + 327396	Ce50 + flp2 + 327396	Ce50 + flp2 + 327396	DM4 + *A+ 327396	DM4+ *A + 327396	Ce4 + his + 327396	Ce4 + his + 327396	Ce50 + his + 327396	Ce50 + his + 327396	DM4 + his + 327396	DM4 + his + 327396
D												
E	Positive	positive	positive	positive	Positive	positive	positiv	positiv	positive	positive		
F	Negative	negative	negative	negative								
G	Blank	blank	blank									
H												

Figure 5.2. Three NCI compounds have antagonist activity against Ce4 FLP18 receptor heterologously expressed in *S. cerevisiae* at 100 µM (A). Alamar blue colour change used as growth indicator. Blue indicates no growth, purple-pink to white indicates growth: (a) 15228, (b) 303243 and (c) 327396 (B) Assay plate layout. **(*A+ : Allatostatin +)***

Titration of compound NCI 327396 (10 µM - 1nM), Figure 5.3, showed that it was inhibiting the yeast growth at 10 µM but appears to have little/no activity at lower concentrations. The other two compounds did not have activity at 10 µM (data not shown). Appendix 5 contains the fluorescence data and the plate layout.

A



B

Ce4	327396				
A	10µM	1 µM	0.1 µM	10 nM	1 nM
B	10µM	1 µM	0.1 µM	10 nM	1 nM
C	10µM	1 µM	0.1 µM	10 nM	1 nM
D					
E	positive	positive	positive		
F	negative	negative	negative		
G	blank	blank	blank		
H					

Figure 5.3: (A) Titration of NCI 327396 (B) Assay plate layout

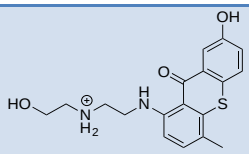
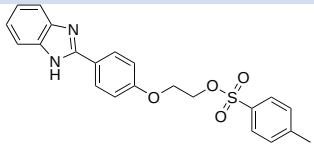
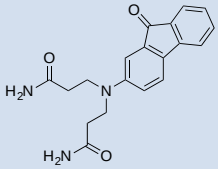
Figure 5.3a indicate that there is growth inhibition by NCI 327396 at 10 µM but there was no or very minimal inhibition of growth at lower concentrations.

5.4 Binding Modes of Compound NCI 327396

Having found 3 actives from the biological screening assays, detailed docking calculations were undertaken to study their binding modes. Docking calculations were carried out using the same grid and docking parameters used during peptide docking in Chapter 4.

The protein-ligands interactions observed from protein-peptide complexes were used as the benchmark interactions with the inhibitor. Thus, the binding site residues (Tyr 115, Lys119, His129, Pro132, Asn135, Ser136, Glu193, Gln235, Glu314, Pro315, Glu316) in earlier docking calculations (Chapter 4) were anticipated; geometric and charge complementarity was considered. The XP Docking scores for the three active compounds NCI 327396, NCI 303243, and NCI 15228 were -9.45 kcal/mol, -8.32 kcal/mol and -8.32 kcal/mol respectively (Table 5.1).

Table 5.1 The structures and physicochemical properties of the three compounds that inhibited yeast growth below 100 μM and the docking scores.

Compound (NCI)	Structure	MW (g/mol)	PSA	XP Glide (kcal/mol)	IC_{50} (μM)
327396		486.49	156.76	-9.45	10
303243		408.47	89.66	-8.32	100
15228		337.38	106.49	-8.32	100

The protein-ligand interactions were visualized using PyMOL (Figure 5.4 to 5.6). The ligands occupied the same binding pocket. Compound 1 interacted with the FLP18R1 protein through a less common π-anion interaction (salt bridge) between the aromatic ring on NCI 327396 and the anionic carboxylate ion of Glu193 (Figure 5.5 and 5.6). This interaction was suggested by Howell and colleagues who highlighted that carboxylate moieties of glutamic and aspartic acids are preferentially positioned on the ring edges of aromatic compounds and reported as the aromatic-anion interaction.[56–58]

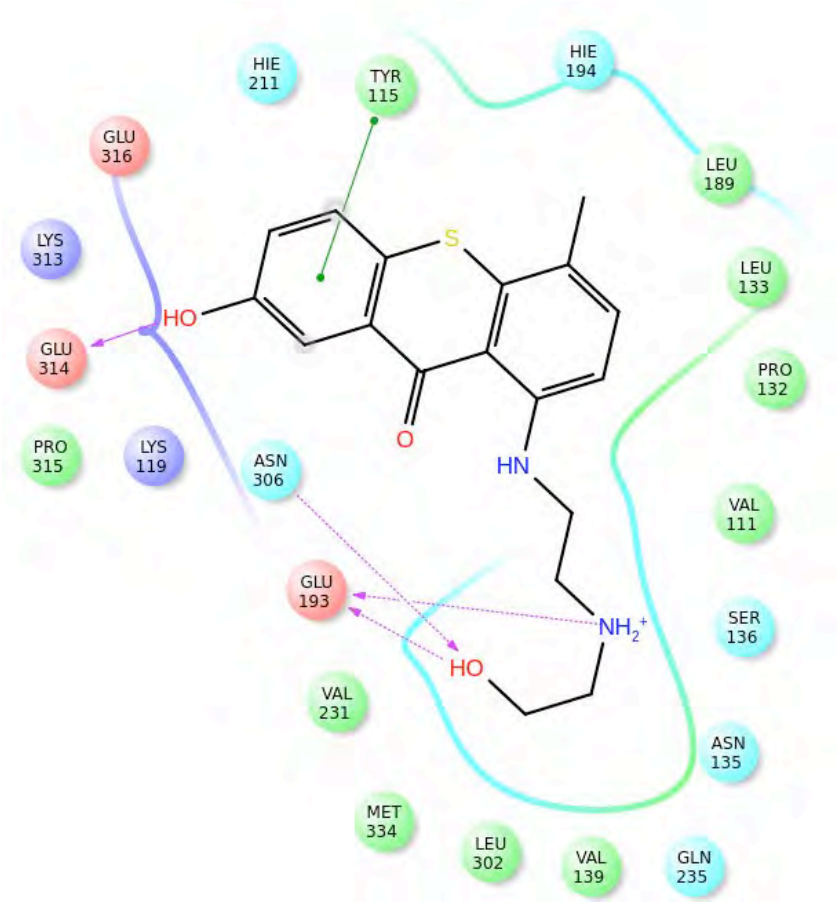


Figure 5.4: 2D interaction map of compound 1 (NCI 327396) in the binding pocket of the FLP18R1 receptor. Tyr115 is involved in a π- π stacking with the aromatic ring of ligand indicated by the green solid line. The dashed magenta lines indicate hydrogen bonds to side chain residues while a solid magenta line indicates a hydrogen bond to the backbone of the protein. Green circled residues are hydrophobic i.e. Pro132, Leu133, Leu189, Leu 302, Met334.

The Gln235 in the binding pocket is indicated to have a very weak hydrogen bond to the ligand. The strong hydrogen bond (1.61 Å) between the hydroxyl of compound 1 (NCI 327396) to the carboxylate of Glu193 helps position the ligand in the binding pocket (Figure 5.4). Tertiary amine of compound 1 also forms a salt bridge with this residue. The glutamine residue is important as it is conserved in ECL2 of Class A GPCRs such as A_2A adenosine and is involved in ligand binding.

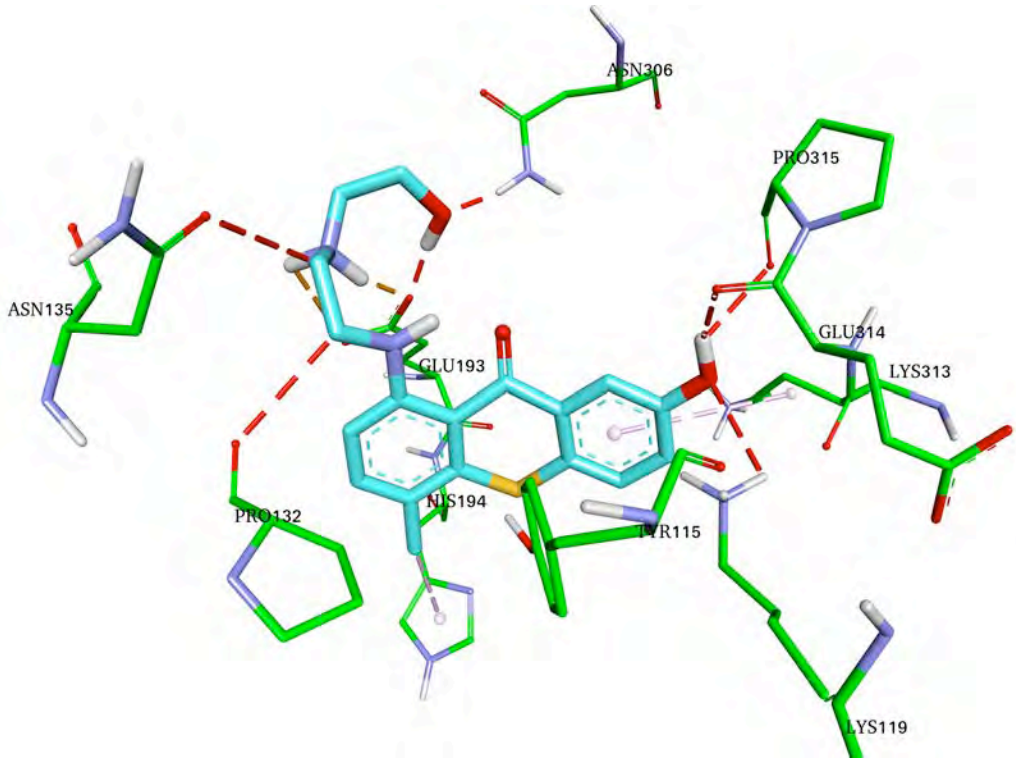


Figure 5.5: The 3D interaction map of NCI 327396 with FLP18R1 showing the bifurcated hydrogen bond from the hydroxyl moiety on Compound NCI 327396 to the Glu314 (2.07Å) and Pro315 (2.61Å) carbonyl oxygen atoms in the receptor. This moiety (OH) accepts hydrogen from the NHz of Lys119 (2.89 Å)

The aromatic ring of Tyr115 and pyrrolidine ring of Pro132 (Helix 2) create a hydrophobic pocket around the inhibitor. The conserved Asn135 (Helix 3) bind to the inhibitor through a transient (3.2 Å) hydrogen bond on the tertiary amine substituent on the 2 position of the aromatic ring (Figure 5.5). This is an important interaction that was also observed with the peptides. Ligand binding with the extracellular loop 3 (ECL3) is also observed through Glu314 and Pro315.

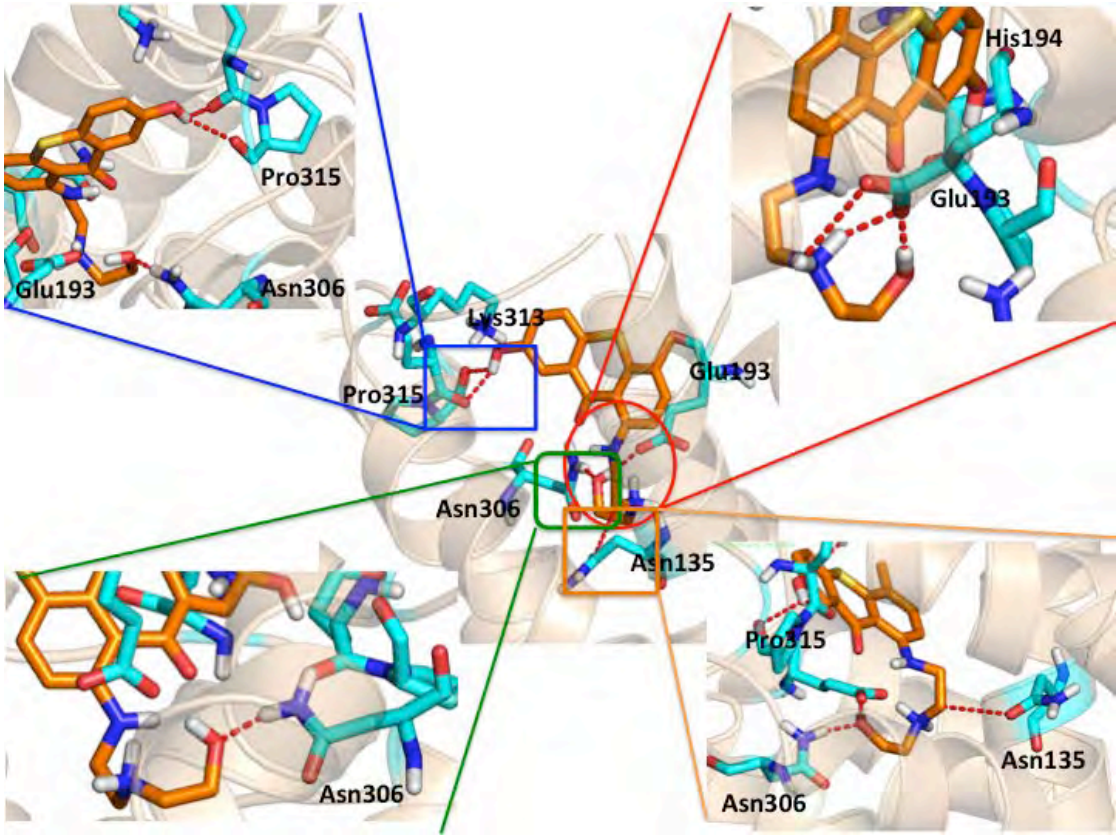


Figure 5.6: The Protein-ligand interactions of NCI 327396 to FLP18R1, highlighting the important interactions that represent the reference chemoprints in Class A GPCRs that include Asn135, Glu193, His194, Asn306, Lys313 and Pro315. The interaction maps of the NCI 15228 and NCI 303243 are given in Appendix 5B

The Asn306 in Helix 6 binds to the inhibitor through a strong 1.79Å hydrogen bond from the Hδ to the hydroxyl on the inhibitor.

5.5 Conclusions

Structure based virtual screening was successfully employed to screen the National Cancer Institute (NCI) database.[25,59] 49 Hit compounds were obtained from the virtual screening exercise of the database, of which 9 compounds were not available for *in vitro* studies. Of the 40 compounds that were screened *in vitro* for activity against FLP18R1, as indicated by the yeast growth inhibition[44] in the antagonist mode, three compounds were observed to be active below 100 μM. Of these 3, in further titrations, NCI 327396 was observed to have activity at 10 μM.

Binding pocket residues in the FLP18R1 receptor were involved in peptide binding were found to bind similarly to small molecules. The residues are conserved in class A GPCR. These include loop residues Tyr115, Lys119 and Glu193, helical residues His129, Pro132, Asn135, Ser136, and ECL3 residues Glu314, Pro315 and Glu316 as well as Gln235, which is a helical residue. GPCRs in class A that have been experimentally determined to date reveal that ECL2 and helical regions are involved in ligand binding.[60,61] This was the observed trend in the study. The Glide scores of compound NCI 327396, NCI 303243 and NCI 15228 were -9.45 kcal/mol, -8.32 and -8.32 kcal/mol respectively using XP Glide scores correlate with the activities. Post docking refinement with Prime/MM-GBSA gave binding scores of -99.13, -72.98 and -89.08 kcal/ mol for the three compounds NCI 327396, NCI 303243 and NCI 15228. The salt bridge from Glu193 (ECL2) to the ligand was observed in other class A GPCR such as A_2A adenosine receptor [60,62]and Beta-1 adrenergic receptors.[61,63]

5.6 References

(1) Mysinger, M. M.; Weiss, D. R.; Ziarek, J. J.; Gravel, S.; Doak, A. K.; Karpiak, J.; Heveker, N.; Shoichet, B. K.; Volkman, B. F. Structure-Based Ligand Discovery for the Protein-Protein Interface of Chemokine Receptor CXCR4. *Proc. Natl. Acad. Sci.* **2012**, *109* (14), 5517–5522.

(2) Kalyaanamoorthy, S.; Chen, Y. P. P. Structure-Based Drug Design to Augment Hit Discovery. *Drug Discov. Today* **2011**, *16* (17-18), 831–839.

(3) Mugumbate, G.; Newton, A. S.; Rosenthal, P. J.; Gut, J.; Moreira, R.; Chibale, K.; Guedes, R. C. Novel Anti-Plasmodial Hits Identified by Virtual Screening of the ZINC Database. *J. Comput. Aided. Mol. Des.* **2013**, *27* (10), 859–871.

(4) Rognan, D. Development and Virtual Screening of Target Libraries. *J. Physiol. Paris* **2006**, *99* (2-3), 232–244.

(5) Schneider, M.; Wolf, S.; Schlitter, J.; Gerwert, K. The Structure of Active Opsin as a Basis for Identification of GPCR Agonists by Dynamic Homology Modelling and Virtual Screening Assays. *FEBS Lett.* **2011**, *585* (22), 3587–3592.

(6) Herrmann, T.; Güntert, P.; Wüthrich, K. Protein NMR Structure Determination with Automated NOE-Identification in the NOESY Spectra Using the New Software ATNOS. *Journal of Biomolecular NMR*, **2002**, *24*, 171–189.

(7) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494* (7436), 185–194.

(8) Beuming, T.; Sherman, W. Current Assessment of Docking into GPCR Crystal Structures and Homology Models: Successes, Challenges, and Guidelines. *J. Chem. Inf. Model.* **2012**, *52* (12), 3263–3277.

(9) Skolnick, J.; Zhou, H.; Gao, M. Are Predicted Protein Structures of Any Value for Binding Site Prediction and Virtual Ligand Screening? *Curr. Opin. Struct. Biol.* **2013**, *23* (2), 191–197.

(10) Levit, A.; Beuming, T.; Krilov, G.; Sherman, W.; Niv, M. Y. Predicting GPCR Promiscuity Using Binding Site Features. *J. Chem. Inf. Model.* **2014**, *54* (1), 184–194.

(11) Latek, D.; Pasznik, P.; Carlomagno, T.; Filipek, S. Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison. *PLoS One* **2013**, *8* (2), 1–10.

(12) Pogozheva, I. D.; Przydzial, M. J.; Mosberg, H. I. Homology Modeling of Opioid Receptor-Ligand Complexes Using Experimental Constraints. *AAPS J.* **2005**, *7* (2), E434–E448.

(13) Vyas, V.; Jain, A.; Jain, A.; Gupta, A. Virtual Screening: A Fast Tool for Drug Design. *Sci. Pharm.* **2008**, *76* (3), 333–360.

(14) Ghose, A. K.; Herbertz, T.; Salvino, J. M.; Mallamo, J. P. Knowledge-Based Chemoinformatic Approaches to Drug Discovery. *Drug Discov. Today* **2006**, *11* (23-24), 1107–1114.

(15) Sakkiah, S.; Thangapandian, S.; Kim, Y. S.; Lee, K. W. Pharmacophore Modeling and Molecular Dynamics Simulation to Find the Potent Leads for Aurora Kinase B. *Bull. Korean Chem. Soc.* **2012**, *33* (3), 869–880.

(16) Dobi, K.; Hajdú, I.; Flachner, B.; Fabó, G.; Szaszkó, M.; Bognár, M.; Magyar, C.; Simon, I.; Szisz, D.; Lõrincz, Z.; Cseh, S.; Dormán, G. Combination of 2D/3D Ligand-Based Similarity Search in Rapid Virtual Screening from Multimillion Compound Repositories. Selection and Biological Evaluation of Potential PDE4 and PDE5 Inhibitors. *Molecules* **2014**, *19* (6), 7008–7039.

(17) Seidel, T.; Ibis, G.; Bendix, F.; Wolber, G. Strategies for 3D Pharmacophore-Based Virtual Screening. *Drug Discov. Today Technol.* **2010**, *7* (4).

(18) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J. Med. Chem.* **2005**, *48* (17), 5448–5465.

(19) Grace Chitima Mugumbate, *Dissertation*, Department of Chemistry, University of Cape Town, Enhancing the Fight against Malaria: From Genome to

Structure and Activity of a G-Protein Coupled Receptor from the Mosquito, *Anopheles Gambiae*, March, **2010**

(20) Isin, B.; Estiu, G.; Wiest, O.; Oltvai, Z. N. Identifying Ligand Binding Conformations of the β2-Adrenergic Receptor by Using Its Agonists as Computational Probes. *PLoS One* **2012**, *7* (12).

(21) Huang, C. C.; Tesmer, J. J. G. Recognition in the Face of Diversity: Interactions of Heterotrimeric G Proteins and G Protein-Coupled Receptor (GPCR) Kinases with Activated GPCRs. *Journal of Biological Chemistry*, **2011**, *286*, 7715–7721.

(22) Leach, A. R. *Molecular Modelling: Principles and Applications*; **2001**.

(23) Kooistra, A. J.; Kuhne, S.; De Esch, I. J. P.; Leurs, R.; De Graaf, C. A Structural Chemogenomics Analysis of Aminergic GPCRs: Lessons for Histamine Receptor Ligand Design. *Br. J. Pharmacol.* **2013**, *170* (1), 101–126.

(24) Schrödinger Workshop 2013 Structure Based Virtual Screening - Various Approaches. **2013**.

(25) Milne, G. W.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (5), 1219–1224.

(26) Moroy, G.; Martiny, V. Y.; Vayer, P.; Villoutreix, B. O.; Miteva, M. a. Toward in Silico Structure-Based ADMET Prediction in Drug Discovery. *Drug Discov. Today* **2012**, *17* (1-2), 44–55.

(27) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1-3), 3–26.

(28) Blagg, J. Structural Alerts for Toxicity. In *Burger's Medicinal Chemistry and Drug Discovery*; **2010**; pp 301–334.

(29) Métivier, J. P.; Lepailleur, A.; Buzmakov, A.; Poezevara, G.; Crémilleux, B.; Kuznetsov, S. O.; Goff, J. Le; Napoli, A.; Bureau, R.; Cuissart, B. Discovering Structural Alerts for Mutagenicity Using Stable Emerging Molecular Patterns.

J. Chem. Inf. Model. **2015**, *55* (5), 925–940.

(30) Edwards, P. J.; Sturino, C. Managing the Liabilities Arising from Structural Alerts: A Safe Philosophy for Medicinal Chemists. *Curr. Med. Chem.* **2011**, *18*, 3116–3135.

(31) Von Der Ohe, P. C.; Kühne, R.; Ebert, R. U.; Altenburger, R.; Liess, M.; Schüürmann, G. Structural Alerts - A New Classification Model to Discriminate Excess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay. *Chem. Res. Toxicol.* **2005**, *18*, 536–555.

(32) Lavecchia, A.; Di Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20* (23), 2839–2860.

(33) Schmidt, T.; Bergner, A.; Schwede, T. Modelling Three-Dimensional Protein Structures for Applications in Drug Design. *Drug Discov. Today* **2014**, *19* (7), 890–897.

(34) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40* (D1).

(35) Clark, D. E.; Pickett, S. D. Prediction of " Drug-Likeness." **2000**, *5* (2), 49–58.

(36) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234.

(37) Bharatham, N.; Bharatham, K.; Shelat, A. A.; Bashford, D. Ligand Binding Mode Prediction by Docking: Mdm2/Mdmx Inhibitors as a Case Study. *J. Chem. Inf. Model.* **2014**, *54* (2), 648–659.

(38) Hetényi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2002**, *11* (7), 1729–1737.

(39) Chen, Z.; Li, H.; Zhang, Q.; Bao, X.; Yu, K.; Luo, X.; Zhu, W.; Jiang, H.

Pharmacophore-Based Virtual Screening versus Docking-Based Virtual Screening: A Benchmark Comparison against Eight Targets. *Acta Pharmacol. Sin.* **2009**, *30* (12), 1694–1708.

(40) Mulakala, C.; Viswanadhan, V. N. Could MM-GBSA Be Accurate Enough for Calculation of Absolute Protein/ligand Binding Free Energies? *J. Mol. Graph. Model.* **2013**, *46*, 41–51.

(41) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, 1–13.

(42) DeLano, W. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**.

(43) Pausch, M. H. G-Protein-Coupled Receptors in Saccharomyces Cerevisiae: High-Throughput Screening Assays for Drug Discovery. *Trends Biotechnol.* **1997**, *15* (12), 487–494.

(44) Dowell; Brown. Yeast Assays for G Protein-Coupled Receptors. *Methods Mol. Biol. (Clifton, NJ)* **2009**, *552*, 213–229.

(45) Larsen, M. J.; Lancheros, E. R.; Williams, T.; Lowery, D. E.; Geary, T. G.; Kubiak, T. M. Functional Expression and Characterization of the C. Elegans G-Protein-Coupled FLP-2 Receptor (T19F4.1) in Mammalian Cells and Yeast. *Int. J. Parasitol. Drugs drug Resist.* **2013**, *3*, 1–7.

(46) Feng, G.; Reale, V.; Chatwin, H.; Kennedy, K.; Venard, R.; Ericsson, C.; Yu, K.; Evans, P. D.; Hall, L. M. Functional Characterization of a Neuropeptide F-like Receptor from *Drosophila Melanogaster*. *Eur. J. Neurosci.* **2003**, *18* (October 2002), 227–238.

(47) Nathoo, A. N.; Moeller, R. A.; Westlund, B. A.; Hart, A. C. Identification of Neuropeptide-like Protein Gene Families in *Caenorhabditis elegans* and Other Species. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (24), 14000–14005.

(48) Joshua Trueheart, Jeremy I. Paul, Hans A. Fuernkranz, Debra Nathan, S. H. *United States Patent. US 6,159,705*. **2000**.

(49) Sarramegna, V.; Talmont, F.; Demange, P.; Milon, A. Heterologous Expression of G-Protein-Coupled Receptors: Comparison of Expression Systems from the

Standpoint of Large-Scale Production and Purification. *Cell. Mol. Life Sci.* **2003**, *60* (8), 1529–1546.

(50) Tate, C. G.; Grisshammer, R. Heterologous Expression of G-Protein-Coupled Receptors. *Trends Biotechnol.* **1996**, *14* (11), 426–430.

(51) Kubiak, T. M.; Larsen, M. J.; Bowman, J. W.; Geary, T. G.; Lowery, D. F. FMRFamide-like Peptides Encoded on the Flp-18 Precursor Gene Activate Two Isoforms of the Orphan *Caenorhabditis Elegans* G-Protein-Coupled Receptor Y58G8A.4 Heterologously Expressed in Mammalian Cells. *Biopolym. - Pept. Sci. Sect.* **2008**, *90* (3), 339–348.

(52) Hooft, R. W.; Sander, C.; Vriend, G. Objectively Judging the Quality of a Protein Structure from a Ramachandran Plot. *Comput. Appl. Biosci.* **1997**, *13* (4), 425–430.

(53) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31* (4), 797–810.

(54) Heikamp, K.; Bajorath, J. Large-Scale Similarity Search Profiling of ChEMBL Compound Data Sets. *J. Chem. Inf. Model.* **2011**, *51* (8), 1831–1839.

(55) Larsen, M. J.; Burton, K. J.; Zantello, M. R.; Smith, V. G.; Lowery, D. L.; Kubiak, T. M. Type A Allatostatins from *Drosophila Melanogaster* and *Diplotera Puncata* Activate Two *Drosophila* Allatostatin Receptors, DAR-1 and DAR-2, Expressed in CHO Cells. *Biochem. Biophys. Res. Commun.* **2001**, *286* (5), 895–901.

(56) Jackson, M. R.; Beahm, R.; Duvvuru, S.; Narasimhan, C.; Wu, J.; Wang, H. N.; Philip, V. M.; Hinde, R. J.; Howell, E. E. A Preference for Edgewise Interactions between Aromatic Rings and Carboxylate Anions: The Biological Relevance of Anion-Quadrupole Interactions. *J. Phys. Chem. B* **2007**, *111* (28), 8242–8249.

(57) Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J. A Survey of Aspartate-Phenylalanine and Glutamate-Phenylalanine Interactions in the Protein Data Bank: Searching for Anion-π

Pairs. *Biochemistry* **2011**, *50* (14), 2939–2950.

(58) Schwans, J. P.; Sunden, F.; Lassila, J. K.; Gonzalez, A.; Tsai, Y.; Herschlag, D. Use of Anion-Aromatic Interactions to Position the General Base in the Ketosteroid Isomerase Active Site. *Proc. Natl. Acad. Sci.* **2013**, *110* (28), 11308–11313.

(59) Díaz Nafría, J. M.; Zimmermann, R. E. Emergence and Evolution of Meaning. *TripleC* **2013**, *11* (1), 13–35.

(60) 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.* **2011**, *332* (6027), 322–327.

(61) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of Two Distinct Inactive Conformations of the beta2-Adrenergic Receptor Reconciles Structural and Biochemical Observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4689–4694.

(62) Sabbadin, D.; Ciancetta, A.; Moro, S. Bridging Molecular Docking to Membrane Molecular Dynamics to Investigate GPCR-Ligand Recognition: The Human A2A Adenosine Receptor as a Key Study. *J. Chem. Inf. Model.* **2014**, *54* (1), 169–183.

(63) Vaidehi, N.; Floriano, W. B.; Trabanino, R.; Hall, S. E.; Freddolino, P.; Choi, E. J.; Zamanakos, G.; Goddard, W. A. Prediction of Structure and Function of G Protein-Coupled Receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12622–12627

6 Chapter 6: Conclusions and recommendations

6.0 Summary

Nematode infections are one of the leading causes of morbidity affecting nearly 2 billion people in the developing world.[1,2] Most current anthelmintics exploit the function of the nematode neurotransmitter receptors.[3,4] Some human nematode pathogens have developed resistance to these anthelmintics, resulting in relatively poor efficacy.[4,5] There have also been concerns surrounding the sustainability of periodic deworming using the current available drugs.[6] Hence, identification of new targets for anthelmintic drug discovery is warranted in order to combat nematode infections.[4,7] G-Protein Coupled Receptors (GPCRs) are central to the signaling pathway of FMRFamide-like Peptides (FLPs) of nematodes.[8] The recent publication of the primary sequence of FLP18R1 GPCR has offered a potential alternative target[9] that can be used in target-based virtual screening for new anthelmintics.[4] Coupled to this important development is the availability of matched neuropeptidic agonists coded for the precursor gene FLP18 in nematode *Caenorhabditis elegans.*[9]

However, the lack of sufficient information and 3D structures of most GPCRs make target-based virtual screening challenging.[10,11] In some cases, ligand-based strategies can be used if there is ligand information available.[12,13] In this study, the primary sequence of FLP18R1 was used to elucidate its 3D homology model using the crystal structure of beta-2-adrenegic receptor (2RH1)[14] as a template. The homology model was refined and optimized using unrestrained molecular dynamic simulation in a POPC membrane.[15–17] The matched peptidic agonist[18] solution structures were determined using NMR-restrained molecular dynamic simulations in DPC micelle solution.[19] The agonists that include, four neuropeptides, *af3*, *af4*, *af20* and *flp18-6* coded by the FLP18 gene were investigated and their structures determined.[9,20] Blind docking[21] of the peptidic agonists to the GPCR was performed to identify the binding cavity chemoprints. The selected docked complex of the receptor was optimized with molecular dynamic (MD) simulations for 100 ns in a mimetic POPC membrane. Virtual screening of the National Cancer Institute (NCI) database[22] was carried out using the same grid generated during blind docking.

6.1 Determination of the 3D structure of FLP18R1

A target-based (Structure-Based Drug Design) approach was used in this study to identify potential anthelmintics. This followed the sequencing of the DNA of *Caenorhabditis Elegans,*[20,23,24] which led to the availability of primary sequences of the FLP18R1 nematode GPCR.[9] The primary sequence was used to predict the secondary structure of the GPCR using PSI-PRED[25,26] and MEMSAT3.[25] The predicted structure had 7 transmembrane helices,[27,28] with each helix connected to the next by a loop. The PSI BLAST tool available on http://blast.ncbi.nlm.nih.gov/ was used to identify templates from the Protein Data Bank (PDB). The template selected was beta-2 adrenergic receptor (2RH1) which had sequence identity of 24.87% to the target, FLP18R1. The schematic flow diagram of the homology modeling is shown in Figure 6.1.

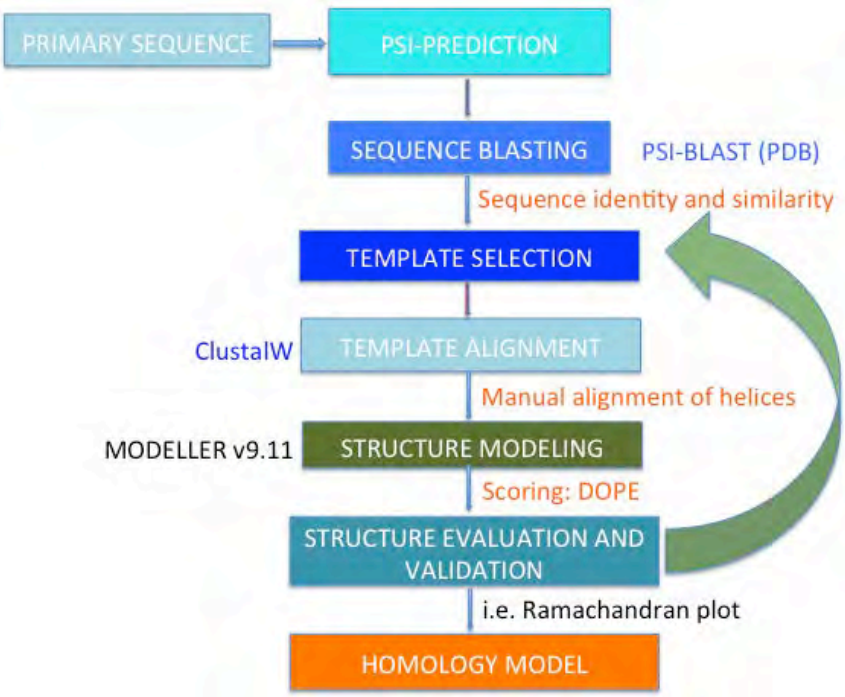


Figure 6.1: Schematic flow diagram of the homology modeling of the FLP18R1 receptor

Manual adjustment of the sequence alignment was performed to remove gaps in the helical regions. Upon obtaining a satisfactory alignment, Modeller version 9.11 was used to perform homology modeling and the best model was selected. The homology model (Figure 6.2a) had a closed structure in the extracellular region, which is

believed to be inherited from the beta-2 adrenergic receptor template that was bound to carazolol, an inverse agonist. There were two observed disulfide bonds between Cys128-Cys208 and Cys203-Cys209. Helix 3 was linked to helix 6 through an ionic lock that is conserved in most class A GPCR receptors.

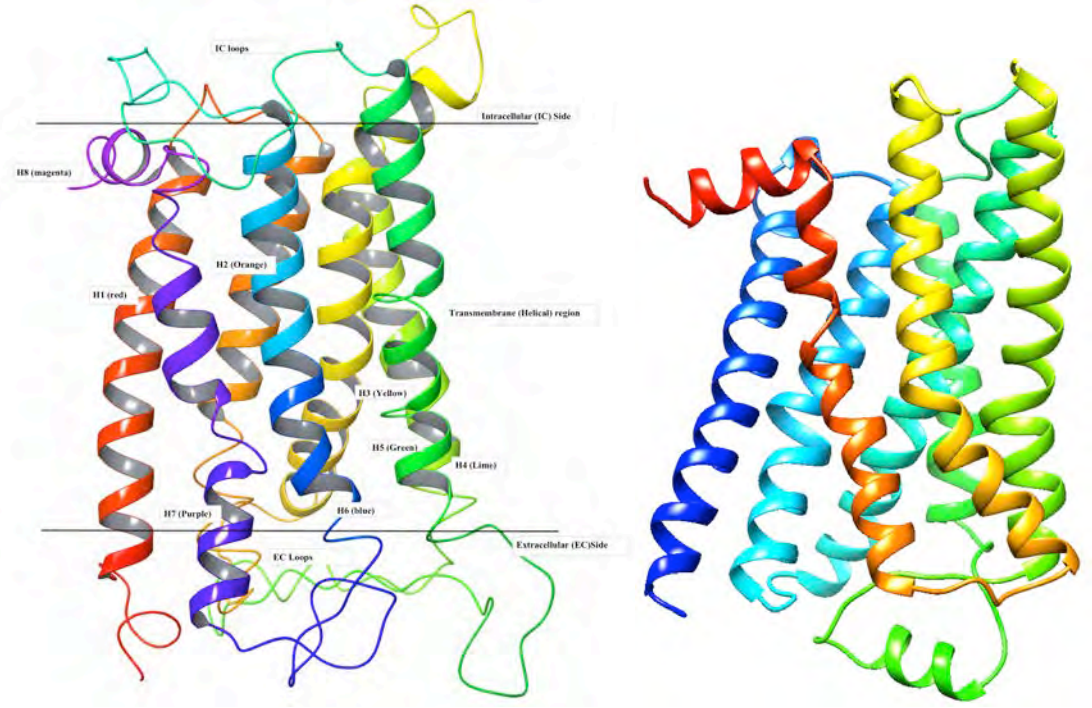


Figure 6.2(a) Homology model from Modeller (b) Optimized and refined homology model after 100 ns Molecular Dynamic Simulations

The homology model was placed in a mimetic POPC membrane and unrestrained molecular dynamic simulations were carried out using GROMACS[29,30] version 4.6.5 in a water box. The loops were optimized and refined during the 100 ns MD simulations. Over the course of the MD simulations, after 1.5 ns, the extracellular region opened up as the helical regions moved apart and the POPC membrane stabilized the homology model, maintaining its overall structure.[30,31] It is believed this resulted from atomic packing and the relaxation of the structure when removed from the template. The ionic lock[32,33] between Helix 3 and 6 was disrupted during the MD simulations. Extracelular Loop 2 (ECL2) was observed to contain a short alpha helix after 40 ns MD simulations, which was conserved throughout the MD simulations. The disulfide bridges ensure accessibility to the binding pocket by the diffusible ligands. The opening up of the extracellular side was concluded to reveal that the refined structure represented the unbound FLP18R1 structure.[34,35]

6.2 Solution structure of the peptide agonists

Ca^{2+} mobilization assays in Fluorescence Imaging Plate Reader (FLIPR) format[18] identified FMRF-amide peptides with a putative C-terminus, -PVGLRF-NH_2 motif,[8] to be agonists for FLP18R1, a nematode GPCR. Using the published primary sequences of the four peptides *af3, af4, af20* and *flp18-6*[9,20] the peptides were synthesized, and analysed using 500MHz, 600MHz and 700 MHz NMR spectroscopy. The data was processed using NMRPipe and Sparky[36] was used to assign NOESY and TOCSY spectra. NMR interproton distances[37] were used during distance restrained molecular Dynamic Simulations of the peptides *in vacuo*, water and DPC micelle solutions.

Analysis of the solution structures of the four peptides indicate that the peptides are flexible and mobile as depicted by the S^2 order parameters[38] (Figure 2.2, 2.9, 2.17 and 2.23). The RCI values corroborate this observation from the NMR experiments which indicate that the proteins are unstructured on the N-terminus. The four peptides possess a type II beta turn[39] in their structures. The amide proton chemical shifts[40] decrease with temperature, and thus there is an up field signal shift with temperature. The Molecular dynamic experiments reproduced the NMR experimental results as highlighted in Chapter 2. The peptides have a U-shaped or C-shaped conformation, consistent with the reports that mammalian peptides of over 100 GPCR ligands have a cyclic conformation.[41,42] The 3D structures of these peptides were given in Figures 2.5, 2.12, 2.19 and 2.25. The RMSD fluctuations during the MD simulations indicate that the peptides have an invariably stable structure, that is not rigid and not too compact, and but still consistent with NMR experiments.

During the MD simulations in DPC-micelle solution, the hydrophobicity of the peptide residues forms hydrophobic interactions with the DPC surface. The DPC solution has an effect on the structure of the peptides. The corresponding π-alkyl interactions and the salt bridges formed mostly with the conserved arginine on the C-terminus provided stability in the peptides. Thus, despite the large conformational space, MD simulations successfully revealed the structural similarities in the peptides, and reproduced the NMR chemical shift data.[38,43]

6.3 Docking calculations and MD simulations

Understanding the protein-small molecule interactions is crucial to any drug discovery program. The binding pose of a ligand in a binding pocket increases favourable interactions and reduces the total energy of the bound complex. The binding residues are believed to be conserved residues in class A GPCRs.[44] These are known as chemo prints and thus the docking calculations investigated this concept. Since the binding cavity on the FLP18R1 was not known, blind docking[45] was performed using AutoDock4.2 and Glide docking. The binding pocket was identified to be extracellular region of the FLP18R1 (Figure 6.3). The loop and helical regions were shown to be involved in ligand binding. This observation is consistent with findings from the docking calculations of A_2A adenosine receptor.[46,47]

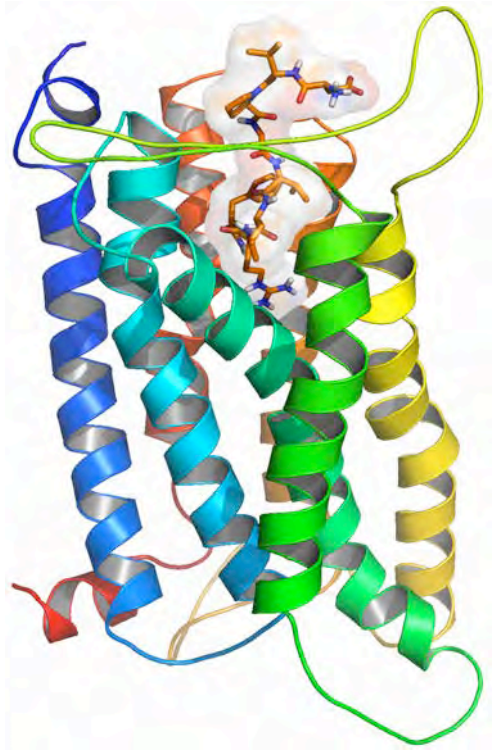


Figure 6.3: The *flp18-6* peptide in a binding pocket of FLP18R1 identified using blind docking with AutoDock Vina and similarly with Glide docking.

The 2 disulfide bonds (Cys128-Cys209 and Cys202-Cys208) ensure accessibility of the binding pocket by the diffusible ligand. The bound complex was inserted in a mimetic POPC membrane and molecular dynamic simulations carried out for 100 ns. When the peptides (ligands) bind to the protein (FLP18R1), the receptor closed up in

the extracellular region and opened up in the intracellular region. The cyclic conformation of the GPCR ligands with a beta turn is vital for the diffusion of the peptides into the binding pocket. However, several studies have shown that the salt bridges[32] and the hydrogen bonds responsible for the conformation are broken during docking calculations, thus assuming an L-shaped conformation.[48] The four peptides occupied the same binding pocket and bound to nearly similar residues. The hydrophobic pockets of the FLP18R1, defined by Val111, Pro132, Leu189, His194, Leu302, Leu305, Pro315 and His330, housed the leucine (Leu7) and phenylalanine (Phe9) residues in the peptides. ECL2 (Tyr115, Glu193 and ECL3 residues (Lys313, Glu314, Pro315, Glu316) was also involved in ligand binding. (Fig 4.5, 4.6, 4.7 and 4.8).

Therefore, the peptides stabilized and bonds were conserved throughout the simulations and interactions with Asn135, Tyr115, Gln235 and Trp299 were maintained (Figure 4.9 and Figure 4.10).

6.4 Structure-Based Virtual screening and biological assays

From the docking calculations favourable interactions were identified and used during virtual screening of the NCI database (265,262 compounds) against FLP18R1. Database preparation was carried out by filtering the database using ADMET descriptors[49] and Lipinski guidelines[50] highlighted in section 5.2.2. Structural alerts were also filtered out. Docking calculations were performed using SP and XP Glide docking and a Glide score cut-off of -7.5 was applied. Prime MM-GBSA[51,52] was used to perform post docking refinement. The virtual screening cascade in Figure 5.1 was successfully employed. 49 Hit compounds were obtained, of which 9 were not available for *in vitro* studies.

Of the 40 compounds that were screened for activity against FLP18R1, as indicated by the yeast (*Saccharomyces Cerevisiae*)[18,53] growth inhibition in the antagonist mode, three compounds were identified as active at concentrations below 100 μM (Figure 5.2 and Figure 5.3). *In vitro* studies in the agonist mode yielded no results. The three compounds were NCI 327396, NCI 303243 and NCI 15228 and the

docking scores are given in Table 5.1. Of these three, only NCI 327396 was observed to have activity at 10 μM.

The protein-ligand interactions observed from protein-peptide complexes were used as benchmark interactions with the inhibitor. The binding site residues found during docking calculations with the peptide ligands were observed to be similar to those with the antagonists. The protein-ligand interactions of the hit compounds, highlighted in Figure 5.4, 5.5 and 5.6, indicate the protein ligand interactions of NCI 327396 with the FLP18R1 receptor. The interactions of the other two compounds and the structures of the rest of the hit compounds are given in Appendix 5A and 5B.

The IC_{50} values that were obtained from the *in-vitro* assays of 10-100 μM for the three potential anthelmintic agents are within the IC50 values of other GPCR antagonists. In a study that was carried out using Structure-based virtual screening on the chemokine receptor (CXCR7), 21 novel ligands that have IC_{50} values between 1.29-11.4 μM were identified that had different scaffolds. The CHEMBL[54] database of bioactivity data reveals that most reported GPCR ligands have IC_{50} values $\leq$ 50 μM. In this regard it shows that our leads are promising compounds as the optimization and further SAR studies will improve the potency of the compounds. Carazolol that is an beta-blocker in beta 2 adrenergic receptor (homologous to FLP18R1 receptor) has IC_{50} of 0.114 nM. Mosapride targets the Alpha 2 adrenergic receptor (A_2A) and is a registered drug with an IC_{50} of 3.8 μM.

6.5 Biological and Biomedical significance of the study

The research identifies new potential leads that promise to be effective anthelmintic agents. The lead compounds will however require being refined to optimize the therapeutic potency so as to develop new drugs that have good efficacy and which the helminths strains have not developed resistance. It is also hoped this will do away with periodic deworming and thus offer alternative control mechanisms to helminth infections. This will aid to improve the health and reduce child mortality and improve health of mankind. A hybrid approach to drug design is believed to offer alternative

anthelmintic agents and thus effective remedy in poverty stricken areas as well as control mechanisms against new incidences.

6.6 Future Work

Longer MD simulations (500 ns) using bigger computer clusters of protein-ligand complexes are envisaged to enable comprehensive understanding of the protein-ligand interactions. These will perhaps help to understand the dynamics at the micro scale and thus aid in designing agonists and antagonists with sufficient activity against FLP18R1. This would in turn facilitate drug discovery initiatives for potential anthelmintic agents. The bigger computer clusters will also speed up the time taken during these complex simulations.

Further studies of the binding mode of the other three peptides *af3, af4* and *af20* is proposed in order to understand the variation in the binding and assessing the possible reasons for the differences in the conformational space.

Some other molecular modeling techniques can be employed to identify other potential anthelmintic agents. These techniques include pharmacophore modeling[55,56] and fragment-based drug design.[57,58]

6.7 References

(1) Barry, M. A; Simon, G. G.; Mistry, N.; Hotez, P. J. Global Trends in Neglected Tropical Disease Control and Elimination: Impact on Child Health. *Arch. Dis. Child.* **2013**, *98* (8), 635–641.

(2) Chan, M. S. The Global Burden of Intestinal Nematode Infections - Fifty Years on. *Parasitology Today*, **1997**, *13*, 438–443.

(3) Ezenwa, V. O.; Etienne, R. S.; Luikart, G.; Beja-Pereira, A.; Jolles, A. E. Hidden Consequences of Living in a Wormy World: Nematode-induced Immune Suppression Facilitates Tuberculosis Invasion in African Buffalo. *Am. Nat.* **2010**, *176* (5), 613–624.

(4) Martin, R. J. Modes of Action of Anthelmintic Drugs. *Vet. J.* **1997**, *154* (1), 11–34.

(5) Lustigman, S.; Prichard, R. K.; Gazzinelli, A.; Grant, W. N.; Boatin, B. a.; McCarthy, J. S.; Basáñez, M. G. A Research Agenda for Helminth Diseases of Humans: The Problem of Helminthiases. *PLoS Negl. Trop. Dis.* **2012**, *6* (4).

(6) Bethony, J.; Brooker, S.; Albonico, M.; Geiger, S. M.; Loukas, A.; Diemert, D.; Hotez, P. J. Soil-Transmitted Helminth Infections: Ascariasis, Trichuriasis, and Hookworm. *Lancet*, **2006**, *367*, 1521–1532.

(7) McCoy, C. J.; Atkinson, L. E.; Zamanian, M.; McVeigh, P.; Day, T. A.; Kimber, M. J.; Marks, N. J.; Maule, A. G.; Mousley, A. New Insights into the FLPergic Complements of Parasitic Nematodes: Informing Deorphanisation Approaches. *EuPA Open Proteomics* **2014**, *3*, 262–272.

(8) McVeigh, P.; Leech, S.; Mair, G. R.; Marks, N. J.; Geary, T. G.; Maule, A. G. Analysis of FMRFamide-like Peptide (FLP) Diversity in Phylum Nematoda. *Int. J. Parasitol.* **2005**, *35* (10), 1043–1060.

(9) Kubiak, T. M.; Larsen, M. J.; Bowman, J. W.; Geary, T. G.; Lowery, D. F. FMRFamide-like Peptides Encoded on the Flp-18 Precursor Gene Activate Two Isoforms of the Orphan *Caenorhabditis Elegans* G-Protein-Coupled Receptor Y58G8A.4 Heterologously Expressed in Mammalian Cells. *Biopolym. - Pept. Sci. Sect.* **2008**, *90* (3), 339–348.

(10) Reddy, C. S.; Vijayasarathy, K.; Srinivas, E.; Sastry, G. M.; Sastry, G. N. Homology Modeling of Membrane Proteins: A Critical Assessment. *Comput. Biol. Chem.* **2006**, *30* (2), 120–126.

(11) Sarramegna, V.; Talmont, F.; Demange, P.; Milon, A. Heterologous Expression of G-Protein-Coupled Receptors: Comparison of Expression Systems from the Standpoint of Large-Scale Production and Purification. *Cell. Mol. Life Sci.* **2003**, *60* (8), 1529–1546.

(12) Bissantz, C.; Logean, A.; Rognan, D. High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional Model Building, and Receptor Library Screening. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 1162–1176.

(13) Tramontano, A.. Homology Modeling with Low Sequence Identity. *Methods* **1998**, *14* (3), 293–300.

(14) Cherezov, V.; Rosenbaum, D.; Hanson, M.; Rasmussen, S.; Thian, F.; Kobilka, T.; Choi, H.; Kuhn, P.; Weis, W.; Kobilka, B.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human beta2-Adrenergic G Protein-Coupled Receptor. *Science (80-.).* **2007**, *318*, 1258–1265.

(15) Janosi, L.; Gorfe, A. A. Simulating POPC and POPC/POPG Bilayers: Conserved Packing and Altered Surface Reactivity. *J. Chem. Theory Comput.* **2010**, *6* (10), 3267–3273.

(16) Dunkin, C. M.; Pokorny, A.; Almeida, P. F.; Lee, H. S. Molecular Dynamics Studies of Transportan 10 (Tp10) Interacting with a POPC Lipid Bilayer. *J. Phys. Chem. B* **2011**, *115* (5), 1188–1198.

(17) Mondal, S.; Johnston, J. M.; Wang, H.; Khelashvili, G.; Filizola, M.; Weinstein, H. Membrane Driven Spatial Organization of GPCRs. *Sci. Rep.* **2013**, *3*, 2909.

(18) Larsen, M. J.; Lancheros, E. R.; Williams, T.; Lowery, D. E.; Geary, T. G.; Kubiak, T. M. Functional Expression and Characterization of the *C. Elegans* G-Protein-Coupled FLP-2 Receptor (T19F4.1) in Mammalian Cells and Yeast. *Int. J. Parasitol. Drugs drug Resist.* **2013**, *3*, 1–7.

(19) Cross, T. A.; Opella, S. J. Solid-State NMR Structural Studies of Peptides and Proteins in Membranes. *Current Opinion in Structural Biology*, **1994**, *4*, 574–581.

(20) Geary, T. G.; Kubiak, T. M. Neuropeptide G-Protein-Coupled Receptors, Their Cognate Ligands and Behavior in *Caenorhabditis Elegans*. *Trends Pharmacol. Sci.* **2005**, *26* (2), 56–58.

(21) Hetényi, C.; Van Der Spoel, D. Toward Prediction of Functional Protein Pockets Using Blind Docking and Pocket Search Algorithms. *Protein Sci.* **2011**, *20* (5), 880–893.

(22) Milne, G. W.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. National Cancer Institute Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (5), 1219–1224.

(23) Janssen, T.; Husson, S. J.; Lindemans, M.; Mertens, I.; Rademakers, S.; Ver Donck, K.; Geysen, J.; Jansen, G.; Schoofs, L. Functional Characterization of Three G Protein-Coupled Receptors for Pigment Dispersing Factors in *Caenorhabditis Elegans*. *J. Biol. Chem.* **2008**, *283* (22), 15241–15249.

(24) Blaxter, M. *Caenorhabditis Elegans* Is a Nematode. *Science* **1998**, *282* (5396), 2041–2046.

(25) McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000**, *16* (4), 404–405.

(26) Nugent, T.; Ward, S.; Jones, D. T. The MEMPACK Alpha-Helical Transmembrane Protein Structure Prediction Server. *Bioinformatics* **2011**, *27* (10), 1438–1439.

(27) Sale, K.; Faulon, J. L.; Gray, G. a; Schoeniger, J. S.; Young, M. M. Optimal Bundling of Transmembrane Helices Using Sparse Distance Constraints. *Protein Sci.* **2004**, *13* (10), 2613–2627.

(28) Wilson, S.; Bergsma, D. J.; Chambers, J. K.; Muir, A. I.; Fantom, K. G.; Ellis, C.; Murdock, P. R.; Herrity, N. C.; Stadel, J. M. Orphan G-Protein-Coupled Receptors: The next Generation of Drug Targets? *Br. J. Pharmacol.* **1998**, *125* (7), 1387–1392.

(29) Schlegel, B.; Sippl, W.; Höltje, H. D. Molecular Dynamics Simulations of Bovine Rhodopsin: Influence of Protonation States and Different Membrane-Mimicking Environments. *J. Mol. Model.* **2005**, *12* (1), 49–64.

(30) Mugumbate, G.; Jackson, G. E.; Van Der Spoel, D.; Kövér, K. E.; Szilágyi, L. Anopheles Gambiae, Anoga-HrTH Hormone, Free and Bound Structure-A Nuclear Magnetic Resonance Experiment. *Peptides* **2013**, *41*, 94–100.

(31) Gardner, M. R.; Kattenhorn, L. M.; Kondur, H. R.; Von Schaewen, M.; Dorfman, T.; Chiang, J. J.; Haworth, K. G.; Decker, J. M.; Alpert, M. D.; Bailey, C. C.; Neale Jr, E. S.; Fellinger, C. H.; Joshi, V. R.; Fuchs, S. P.; Martinez-Navio, J. M.; Quinlan, B. D.; Yao, A. Y.; Mouquet, H.; Gorman, J.; Zhang, B.; Poignard, P.; Nussenzweig, M. C.; Burton, D. R.; Kwong, P. D.; Piatak Jr, M.; Lifson, J. D.; Gao, G.; Desrosiers, R. C.; Evans, D. T.; Hahn, B. H.; Ploss, A.; Cannon, P. M.; Seaman, M. S.; Farzan, M. AAV-Expressed eCD4-Ig Provides Durable Protection from Multiple SHIV Challenges. *Nature* **2015**.

(32) Schneider, E. H.; Schnell, D.; Strasser, A.; Dove, S.; Seifert, R. Impact of the DRY Motif and the Missing “Ionic Lock” on Constitutive Activity and G-Protein Coupling of the Human Histamine H4 Receptor. *J. Pharmacol. Exp. Ther.* **2010**, *333* (2), 382–392.

(33) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of Two Distinct Inactive Conformations of the beta2-Adrenergic Receptor Reconciles Structural and Biochemical Observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (12), 4689–4694.

(34) Bierzyński, A. Methods of Peptide Conformation Studies. *Acta Biochim. Pol.* **2001**, *48* (4), 1091–1099.

(35) Li, C.; Kim, K. Family of FLP Peptides in *Caenorhabditis Elegans* and Related Nematodes. *Front. Endocrinol. (Lausanne).* **2014**, *5* (October), 1–16.

(36) Assigning Proteins Using Sparky. http://www.cgl.ucsf.edu/home/sparky/manual/

(37) Thomas, P. D.; Basus, V. J.; James, T. L. Protein Solution Structure

Determination Using Distances from Two-Dimensional Nuclear Overhauser Effect Experiments: Effect of Approximations on the Accuracy of Derived Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (4), 1237–1241.

(38) Berjanskii, M.; Wishart, D. S. NMR: Prediction of Protein Flexibility. *Nat. Protoc.* **2006**, *1* (2), 683–688.

(39) Fuchs, P. F. J.; Alix, A. J. P. High Accuracy Prediction of ß-Turns and Their Types Using Propensities and Multiple Alignments. *Proteins Struct. Funct. Genet.* **2005**, *59* (4), 828–839.

(40) Cierpicki, T.; Otlewski, J. Amide Proton Temperature Coefficients as Hydrogen Bond Indicators in Proteins. *J. Biomol. NMR* **2001**, *21* (3), 249–261.

(41) Tyndall, J. D. a; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2005**, *105* (3), 793–826.

(42) Ruiz-Gómez, G.; Tyndall, J. D. a; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Update 1 of: Over One Hundred Peptide-Activated G Protein-Coupled Receptors Recognize Ligands with Turn Structure. *Chem. Rev.* **2010**, *110* (4), PR1–PR41.

(43) Baxter, N. J.; Williamson, M. P. Temperature Dependence of 1H Chemical Shifts in Proteins. *J. Biomol. NMR* **1997**, *9* (4), 359–369.

(44) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular Signatures of G-Protein-Coupled Receptors. *Nature* **2013**, *494* (7436), 185–194.

(45) Hetényi, C.; van der Spoel, D. Efficient Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Protein Sci.* **2002**, *11* (7), 1729–1737.

(46) 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* **2011**, *332* (6027), 322–327.

(47) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6-Angstrom Crystal Structure

of a Human A_2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322* (5905), 1211–1217.

(48) Teodoro, M. L.; Phillips, J.; Kavraki, L. E. Molecular Docking: A Problem with Thousands of Degrees of Freedom. *Proc. - IEEE Int. Conf. Robot. Autom.* **2001**, *1*, 960–965.

(49) Moroy, G.; Martiny, V. Y.; Vayer, P.; Villoutreix, B. O.; Miteva, M. a. Toward in Silico Structure-Based ADMET Prediction in Drug Discovery. *Drug Discov. Today* **2012**, *17* (1-2), 44–55.

(50) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1-3), 3–26.

(51) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, 1–13.

(52) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31* (4), 797–810.

(53) Pausch, M. H. G-Protein-Coupled Receptors in *Saccharomyces Cerevisiae*: High-Throughput Screening Assays for Drug Discovery. *Trends Biotechnol.* **1997**, *15* (12), 487–494.

(54) Chen, Z.; Li, H.; Zhang, Q.; Bao, X.; Yu, K.; Luo, X.; Zhu, W.; Jiang, H. Pharmacophore-Based Virtual Screening versus Docking-Based Virtual Screening: A Benchmark Comparison against Eight Targets. *Acta Pharmacol. Sin.* **2009**, *30* (12), 1694–1708.

(55) Warszycki, D.; Mordalski, S.; Kristiansen, K.; Kafel, R.; Sylte, I.; Chilmonczyk, Z.; Bojarski, A. J. A Linear Combination of Pharmacophore Hypotheses as a New Tool in Search of New Active Compounds - An Application for 5-HT1A Receptor Ligands. *PLoS One* **2013**, *8* (12), 1–13.

(56) Vass, M.; Schmidt, É.; Horti, F.; Keseru, G. M. Virtual Fragment Screening on GPCRs: A Case Study on Dopamine D3 and Histamine H4 Receptors. *Eur. J.*

Med. Chem. **2014**, *77*, 38–46.

(57) Bondensgaard, K.; Ankersen, M.; Thøgersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of Privileged Structures by G-Protein Coupled Receptors. *J. Med. Chem.* **2004**, *47* (4), 888–899.

APPENDICES

APPENDIX 2A

Interproton distances *flp18-6* peptide

[distance_restraints]

a_i	a_j	type	index	type'	strong	medium	weak	fac
5	7	1	0	1	0.189	0.300	0.350	1.0
5	8	1	0	1	0.189	0.300	0.350	1.0
7	15	1	1	1	0.220	0.400	0.450	1.0
8	15	1	1	1	0.220	0.400	0.450	1.0
5	15	1	2	1	0.258	0.500	0.550	1.0
7	15	1	3	1	0.263	0.500	0.550	1.0
8	15	1	3	1	0.263	0.500	0.550	1.0
17	40	1	4	1	0.190	0.300	0.350	1.0
17	41	1	4	1	0.190	0.300	0.350	1.0
17	40	1	5	1	0.206	0.400	0.450	1.0
17	41	1	5	1	0.206	0.400	0.450	1.0
19	21	1	6	1	0.180	0.300	0.350	1.0
19	22	1	6	1	0.180	0.300	0.350	1.0
19	23	1	6	1	0.180	0.300	0.350	1.0
21	40	1	7	1	0.217	0.400	0.450	1.0
22	40	1	7	1	0.217	0.400	0.450	1.0
23	40	1	7	1	0.217	0.400	0.450	1.0
21	41	1	7	1	0.217	0.400	0.450	1.0
22	41	1	7	1	0.217	0.400	0.450	1.0
23	41	1	7	1	0.217	0.400	0.450	1.0
32	45	1	8	1	0.180	0.300	0.350	1.0
40	19	1	9	1	0.245	0.400	0.450	1.0
41	19	1	9	1	0.245	0.400	0.450	1.0
40	17	1	10	1	0.274	0.500	0.550	1.0
41	17	1	10	1	0.274	0.500	0.550	1.0

37	45	1	11	1	0.256	0.500	0.550	1.0
38	45	1	11	1	0.256	0.500	0.550	1.0
47	52	1	12	1	0.177	0.300	0.350	1.0
48	52	1	12	1	0.177	0.300	0.350	1.0
54	68	1	13	1	0.180	0.300	0.350	1.0
56	68	1	14	1	0.226	0.400	0.450	1.0
70	87	1	15	1	0.180	0.300	0.350	1.0
72	21	1	15	1	0.180	0.300	0.350	1.0
72	22	1	16	1	0.180	0.500	0.550	1.0
72	23	1	16	1	0.180	0.500	0.550	1.0
73	21	1	16	1	0.180	0.500	0.550	1.0
73	22	1	16	1	0.180	0.500	0.550	1.0
73	23	1	16	1	0.180	0.500	0.550	1.0
72	25	1	16	1	0.180	0.500	0.550	1.0
72	26	1	16	1	0.180	0.500	0.550	1.0
72	27	1	16	1	0.180	0.500	0.550	1.0
73	25	1	16	1	0.180	0.500	0.550	1.0
73	26	1	16	1	0.180	0.500	0.550	1.0
73	27	1	16	1	0.180	0.500	0.550	1.0
77	118	1	17	1	0.380	0.550	0.700	1.0
78	118	1	17	1	0.380	0.550	0.700	1.0
79	118	1	17	1	0.380	0.550	0.700	1.0
77	120	1	17	1	0.380	0.550	0.700	1.0
78	120	1	17	1	0.380	0.550	0.700	1.0
79	120	1	17	1	0.380	0.550	0.700	1.0
76	118	1	18	1	0.380	0.550	0.700	1.0
80	118	1	18	1	0.380	0.550	0.700	1.0
76	120	1	18	1	0.380	0.550	0.700	1.0
80	120	1	18	1	0.380	0.550	0.700	1.0
87	77	1	19	1	0.209	0.400	0.450	1.0
87	78	1	19	1	0.209	0.400	0.450	1.0
87	79	1	19	1	0.209	0.400	0.450	1.0
89	111	1	20	1	0.180	0.300	0.350	1.0
113	103	1	21	1	0.195	0.500	0.550	1.0
113	104	1	21	1	0.195	0.500	0.550	1.0
113	106	1	21	1	0.195	0.500	0.550	1.0

113	107	1	21	1	0.195	0.500	0.550	1.0
113	87	1	22	1	0.251	0.500	0.550	1.0
115	103	1	23	1	0.264	0.550	0.600	1.0
115	104	1	23	1	0.264	0.550	0.600	1.0
123	21	1	24	1	0.380	0.650	0.700	1.0
123	22	1	24	1	0.380	0.650	0.700	1.0
125	21	1	24	1	0.380	0.650	0.700	1.0
125	22	1	24	1	0.380	0.650	0.700	1.0
123	25	1	24	1	0.380	0.650	0.700	1.0
123	26	1	24	1	0.380	0.650	0.700	1.0
123	27	1	24	1	0.380	0.650	0.700	1.0
125	25	1	24	1	0.380	0.650	0.700	1.0
125	26	1	24	1	0.380	0.650	0.700	1.0
125	27	1	24	1	0.380	0.650	0.700	1.0

APPENDIX 2B

Interproton distances *af3* peptide

[distance_restraints]

; a_i	**a_j**	**type**	**index**	**type'**	**strong**	**medium**	**weak**	**fac**
6	14	1	0	1	0.167	0.300	0.350	1.0
16	18	1	1	1	0.143	0.300	0.350	1.0
16	20	1	2	1	0.139	0.300	0.350	1.0
16	21	1	2	1	0.139	0.300	0.350	1.0
16	22	1	2	1	0.139	0.300	0.350	1.0
16	24	1	2	1	0.139	0.300	0.350	1.0
16	25	1	2	1	0.139	0.300	0.350	1.0
16	26	1	2	1	0.139	0.300	0.350	1.0
16	20	1	3	1	0.151	0.300	0.350	1.0
16	21	1	3	1	0.151	0.300	0.350	1.0
16	22	1	3	1	0.151	0.300	0.350	1.0
16	24	1	3	1	0.151	0.300	0.350	1.0
16	25	1	3	1	0.151	0.300	0.350	1.0
16	26	1	3	1	0.151	0.300	0.350	1.0
16	14	1	4	1	0.138	0.300	0.350	1.0
16	39	1	5	1	0.161	0.300	0.350	1.0
16	40	1	5	1	0.161	0.300	0.350	1.0
16	39	1	6	1	0.164	0.300	0.350	1.0
16	40	1	6	1	0.164	0.300	0.350	1.0
18	14	1	7	1	0.158	0.300	0.350	1.0
20	18	1	8	1	0.189	0.300	0.350	1.0
21	18	1	8	1	0.189	0.300	0.350	1.0
22	18	1	8	1	0.189	0.300	0.350	1.0
24	18	1	8	1	0.189	0.300	0.350	1.0
25	18	1	8	1	0.189	0.300	0.350	1.0
26	18	1	8	1	0.189	0.300	0.350	1.0
20	14	1	9	1	0.156	0.300	0.350	1.0
21	14	1	9	1	0.156	0.300	0.350	1.0

22	14	1	9	1	0.156	0.300	0.350	1.0
24	14	1	9	1	0.156	0.300	0.350	1.0
25	14	1	9	1	0.156	0.300	0.350	1.0
26	14	1	9	1	0.156	0.300	0.350	1.0
20	16	1	10	1	0.177	0.300	0.350	1.0
21	16	1	10	1	0.177	0.300	0.350	1.0
22	16	1	10	1	0.177	0.300	0.350	1.0
24	16	1	10	1	0.177	0.300	0.350	1.0
25	16	1	10	1	0.177	0.300	0.350	1.0
26	16	1	10	1	0.177	0.300	0.350	1.0
20	39	1	11	1	0.112	0.300	0.350	1.0
21	39	1	11	1	0.112	0.300	0.350	1.0
22	39	1	11	1	0.112	0.300	0.350	1.0
24	39	1	11	1	0.112	0.300	0.350	1.0
25	39	1	11	1	0.112	0.300	0.350	1.0
26	39	1	11	1	0.112	0.300	0.350	1.0
20	40	1	11	1	0.112	0.300	0.350	1.0
21	40	1	11	1	0.112	0.300	0.350	1.0
22	40	1	11	1	0.112	0.300	0.350	1.0
24	40	1	11	1	0.112	0.300	0.350	1.0
25	40	1	11	1	0.112	0.300	0.350	1.0
26	40	1	11	1	0.112	0.300	0.350	1.0
20	39	1	12	1	0.136	0.300	0.350	1.0
21	39	1	12	1	0.136	0.300	0.350	1.0
22	39	1	12	1	0.136	0.300	0.350	1.0
24	39	1	12	1	0.136	0.300	0.350	1.0
25	39	1	12	1	0.136	0.300	0.350	1.0
26	39	1	12	1	0.136	0.300	0.350	1.0
20	40	1	12	1	0.136	0.300	0.350	1.0
21	40	1	12	1	0.136	0.300	0.350	1.0
22	40	1	12	1	0.136	0.300	0.350	1.0
24	40	1	12	1	0.136	0.300	0.350	1.0
25	40	1	12	1	0.136	0.300	0.350	1.0
26	40	1	12	1	0.136	0.300	0.350	1.0
33	44	1	13	1	0.138	0.400	0.450	1.0
34	44	1	13	1	0.138	0.400	0.450	1.0

33	44	1	14	1	0.118	0.400	0.450	1.0
34	44	1	14	1	0.118	0.400	0.450	1.0
33	67	1	15	1	0.184	0.780	0.800	1.0
34	67	1	15	1	0.184	0.780	0.800	1.0
39	16	1	16	1	0.184	0.300	0.350	1.0
40	16	1	16	1	0.184	0.300	0.350	1.0
39	14	1	17	1	0.130	0.430	0.500	1.0
40	14	1	17	1	0.130	0.430	0.500	1.0
39	16	1	18	1	0.177	0.300	0.350	1.0
40	16	1	18	1	0.177	0.300	0.350	1.0
39	20	1	19	1	0.135	0.300	0.350	1.0
39	21	1	19	1	0.135	0.300	0.350	1.0
39	22	1	19	1	0.135	0.300	0.350	1.0
39	24	1	19	1	0.135	0.300	0.350	1.0
39	25	1	19	1	0.135	0.300	0.350	1.0
39	26	1	19	1	0.135	0.300	0.350	1.0
40	20	1	19	1	0.135	0.300	0.350	1.0
40	21	1	19	1	0.135	0.300	0.350	1.0
40	22	1	19	1	0.135	0.300	0.350	1.0
40	24	1	19	1	0.135	0.300	0.350	1.0
40	25	1	19	1	0.135	0.300	0.350	1.0
40	26	1	19	1	0.135	0.300	0.350	1.0
46	20	1	20	1	0.114	0.550	0.650	1.0
47	20	1	20	1	0.114	0.550	0.650	1.0
46	21	1	21	1	0.114	0.400	0.450	1.0
47	21	1	21	1	0.114	0.400	0.450	1.0
46	22	1	22	1	0.114	0.550	0.600	1.0
47	22	1	22	1	0.114	0.550	0.600	1.0
46	24	1	23	1	0.114	0.550	0.600	1.0
47	24	1	23	1	0.114	0.550	0.600	1.0
46	20	1	24	1	0.124	0.400	0.450	1.0
47	20	1	24	1	0.124	0.400	0.450	1.0
46	21	1	25	1	0.124	0.300	0.350	1.0
47	21	1	25	1	0.124	0.300	0.350	1.0
46	22	1	26	1	0.124	0.450	0.550	1.0
47	22	1	26	1	0.124	0.450	0.550	1.0

46	24	1	27	1	0.124	0.550	0.650	1.0
47	24	1	27	1	0.124	0.550	0.650	1.0
46	25	1	28	1	0.124	0.600	0.700	1.0
47	25	1	28	1	0.124	0.600	0.700	1.0
46	26	1	29	1	0.124	0.600	0.700	1.0
47	26	1	29	1	0.124	0.600	0.700	1.0
46	39	1	30	1	0.230	0.500	0.600	1.0
46	40	1	30	1	0.230	0.500	0.600	1.0
47	39	1	30	1	0.230	0.500	0.600	1.0
47	40	1	30	1	0.230	0.500	0.600	1.0
46	67	1	31	1	0.177	0.400	0.450	1.0
47	67	1	31	1	0.177	0.400	0.450	1.0
46	86	1	32	1	0.094	0.700	0.750	1.0
47	86	1	32	1	0.094	0.700	0.750	1.0
53	44	1	33	1	0.116	0.500	0.550	1.0
57	70	1	34	1	0.182	0.500	0.550	1.0
58	70	1	34	1	0.182	0.500	0.550	1.0
59	70	1	34	1	0.182	0.500	0.550	1.0
61	70	1	34	1	0.182	0.500	0.550	1.0
62	70	1	34	1	0.182	0.500	0.550	1.0
63	70	1	34	1	0.182	0.500	0.550	1.0
57	71	1	34	1	0.182	0.500	0.550	1.0
58	71	1	34	1	0.182	0.500	0.550	1.0
59	71	1	34	1	0.182	0.500	0.550	1.0
61	71	1	34	1	0.182	0.500	0.550	1.0
62	71	1	34	1	0.182	0.500	0.550	1.0
63	71	1	34	1	0.182	0.500	0.550	1.0
69	86	1	35	1	0.146	0.300	0.350	1.0
71	86	1	36	1	0.168	0.400	0.450	1.0
72	86	1	36	1	0.168	0.400	0.450	1.0
71	110	1	37	1	0.138	0.550	0.600	1.0
72	110	1	37	1	0.138	0.550	0.600	1.0
76	86	1	38	1	0.096	0.450	0.550	1.0
77	86	1	38	1	0.096	0.450	0.550	1.0
78	86	1	38	1	0.096	0.450	0.550	1.0
80	86	1	38	1	0.096	0.450	0.550	1.0

81	86	1	38	1	0.096	0.450	0.550	1.0
82	86	1	38	1	0.096	0.450	0.550	1.0
76	51	1	39	1	0.098	0.400	0.450	1.0
77	51	1	39	1	0.098	0.400	0.450	1.0
78	51	1	39	1	0.098	0.400	0.450	1.0
80	51	1	39	1	0.098	0.400	0.450	1.0
81	51	1	39	1	0.098	0.400	0.450	1.0
82	51	1	39	1	0.098	0.400	0.450	1.0
74	86	1	40	1	0.114	0.500	0.550	1.0
88	67	1	41	1	0.162	0.600	0.700	1.0
88	110	1	42	1	0.156	0.300	0.350	1.0
90	44	1	43	1	0.133	0.730	0.750	1.0
91	44	1	43	1	0.133	0.730	0.750	1.0
90	67	1	44	1	0.126	0.450	0.550	1.0
91	67	1	44	1	0.126	0.450	0.550	1.0
99	110	1	45	1	0.095	0.400	0.450	1.0
93	67	1	46	1	0.104	0.650	0.700	1.0
94	67	1	46	1	0.104	0.650	0.700	1.0
93	110	1	47	1	0.133	0.300	0.350	1.0
94	110	1	47	1	0.133	0.300	0.350	1.0
114	93	1	48	1	0.177	0.350	0.450	1.0
115	93	1	48	1	0.177	0.350	0.450	1.0
114	94	1	48	1	0.177	0.350	0.450	1.0
115	94	1	48	1	0.177	0.350	0.450	1.0

APPENDIX 2C

Interproton distances *af4* peptide

[distance_restraints]

; ai	**aj**	**type**	**index**	**type'**	**strong**	**medium**	**weak**	**fac**
13	15	1	0	1	0.147	0.300	0.350	1.0
13	16	1	0	1	0.147	0.300	0.350	1.0
13	15	1	1	1	0.138	0.300	0.350	1.0
13	16	1	1	1	0.138	0.300	0.350	1.0
13	11	1	2	1	0.139	0.300	0.350	1.0
13	27	1	3	1	0.091	0.500	0.550	1.0
13	23	1	4	1	0.160	0.300	0.350	1.0
13	53	1	5	1	0.077	0.700	0.800	1.0
121	119	1	6	1	0.143	0.300	0.350	1.0
15	23	1	7	1	0.124	0.350	0.400	1.0
16	23	1	7	1	0.124	0.350	0.400	1.0
25	11	1	8	1	0.092	0.450	0.500	1.0
25	48	1	9	1	0.168	0.300	0.350	1.0
25	48	1	9	1	0.168	0.300	0.350	1.0
25	48	1	10	1	0.161	0.300	0.350	1.0
35	48	1	10	1	0.161	0.300	0.350	1.0
27	11	1	11	1	0.095	0.630	0.650	1.0
27	23	1	12	1	0.158	0.300	0.350	1.0
27	99	1	13	1	0.091	0.500	0.550	1.0
27	100	1	13	1	0.091	0.500	0.550	1.0
40	42	1	14	1	0.159	0.300	0.350	1.0
40	43	1	14	1	0.159	0.300	0.350	1.0
42	53	1	15	1	0.124	0.400	0.450	1.0
43	53	1	15	1	0.124	0.400	0.450	1.0
42	76	1	16	1	0.109	0.350	0.400	1.0
43	76	1	16	1	0.109	0.350	0.400	1.0
42	95	1	17	1	0.079	0.550	0.600	1.0
43	95	1	17	1	0.079	0.550	0.600	1.0

42	53	1	18	1	0.130	0.400	0.450	1.0
43	53	1	18	1	0.130	0.400	0.450	1.0
42	76	1	19	1	0.118	0.350	0.400	1.0
43	76	1	19	1	0.118	0.350	0.400	1.0
48	25	1	20	1	0.171	0.300	0.350	1.0
49	25	1	20	1	0.171	0.300	0.350	1.0
48	23	1	21	1	0.116	0.450	0.500	1.0
49	23	1	21	1	0.116	0.450	0.500	1.0
48	49	1	22	1	0.178	0.300	0.350	1.0
48	53	1	23	1	0.069	0.450	0.500	1.0
49	53	1	23	1	0.069	0.450	0.500	1.0
45	53	1	24	1	0.125	0.300	0.350	1.0
46	53	1	24	1	0.125	0.300	0.350	1.0
55	11	1	25	1	0.111	0.750	0.800	1.0
56	11	1	25	1	0.111	0.750	0.800	1.0
55	23	1	26	1	0.096	0.500	0.550	1.0
56	23	1	26	1	0.096	0.500	0.550	1.0
55	53	1	27	1	0.178	0.300	0.350	1.0
56	53	1	27	1	0.178	0.300	0.350	1.0
55	60	1	28	1	0.165	0.300	0.350	1.0
56	60	1	28	1	0.165	0.300	0.350	1.0
55	99	1	29	1	0.092	0.450	0.500	1.0
55	100	1	29	1	0.092	0.450	0.500	1.0
56	99	1	29	1	0.092	0.450	0.500	1.0
56	100	1	29	1	0.092	0.450	0.500	1.0
55	108	1	30	1	0.118	0.450	0.500	1.0
56	108	1	30	1	0.118	0.450	0.500	1.0
55	95	1	31	1	0.095	0.500	0.550	1.0
56	95	1	31	1	0.095	0.500	0.550	1.0
55	56	1	32	1	0.189	0.300	0.350	1.0
55	66	1	33	1	0.144	0.350	0.400	1.0
56	66	1	33	1	0.144	0.350	0.400	1.0
55	67	1	33	1	0.144	0.350	0.400	1.0
56	67	1	33	1	0.144	0.350	0.400	1.0
55	68	1	33	1	0.144	0.350	0.400	1.0
56	68	1	33	1	0.144	0.350	0.400	1.0

55	70	1	33	1	0.144	0.350	0.400	1.0
56	70	1	33	1	0.144	0.350	0.400	1.0
55	71	1	33	1	0.144	0.350	0.400	1.0
56	71	1	33	1	0.144	0.350	0.400	1.0
55	72	1	33	1	0.144	0.350	0.400	1.0
56	72	1	33	1	0.144	0.350	0.400	1.0
55	102	1	34	1	0.106	0.600	0.700	1.0
56	103	1	34	1	0.106	0.600	0.700	1.0
55	66	1	35	1	0.133	0.360	0.400	1.0
56	66	1	35	1	0.133	0.360	0.400	1.0
55	67	1	35	1	0.133	0.360	0.400	1.0
56	67	1	35	1	0.133	0.360	0.400	1.0
55	68	1	35	1	0.133	0.360	0.400	1.0
56	68	1	35	1	0.133	0.360	0.400	1.0
55	70	1	35	1	0.133	0.360	0.400	1.0
56	70	1	35	1	0.133	0.360	0.400	1.0
55	71	1	35	1	0.133	0.360	0.400	1.0
56	71	1	35	1	0.133	0.360	0.400	1.0
55	72	1	35	1	0.133	0.360	0.400	1.0
56	72	1	35	1	0.133	0.360	0.400	1.0
62	53	1	36	1	0.104	0.600	0.650	1.0
62	95	1	37	1	0.103	0.400	0.450	1.0
62	119	1	38	1	0.084	0.750	0.800	1.0
78	53	1	39	1	0.098	0.750	0.800	1.0
78	76	1	40	1	0.141	0.300	0.350	1.0
78	95	1	41	1	0.146	0.350	0.400	1.0
78	119	1	42	1	0.124	0.400	0.450	1.0
80	119	1	43	1	0.141	0.500	0.550	1.0
81	119	1	43	1	0.141	0.500	0.550	1.0
97	53	1	44	1	0.083	0.550	0.600	1.0
97	60	1	45	1	0.110	0.600	0.700	1.0
97	95	1	46	1	0.148	0.300	0.350	1.0
97	119	1	47	1	0.155	0.350	0.400	1.0
99	95	1	48	1	0.167	0.300	0.350	1.0
100	95	1	48	1	0.167	0.300	0.350	1.0
121	123	1	49	1	0.164	0.300	0.350	1.0

121	124	1	49	1	0.164	0.300	0.350	1.0
102	76	1	50	1	0.106	0.650	0.700	1.0
103	76	1	50	1	0.106	0.650	0.700	1.0
102	119	1	51	1	0.126	0.450	0.500	1.0
103	119	1	51	1	0.126	0.450	0.500	1.0
121	99	1	52	1	0.106	0.470	0.500	1.0
121	100	1	52	1	0.106	0.470	0.500	1.0
121	102	1	53	1	0.097	0.600	0.700	1.0
121	103	1	53	1	0.097	0.600	0.700	1.0

APPENDIX 2D

Interproton distances *af20* peptide

[distance_restraints]

; a_i	**a_j**	**type**	**index**	**type'**	**strong**	**medium**	**weak**	**fac**
6	10	1	0	1	0.248	0.400	0.450	1.0
7	10	1	0	1	0.248	0.400	0.450	1.0
13	44	1	1	1	0.183	0.600	0.700	1.0
13	45	1	1	1	0.183	0.600	0.700	1.0
18	44	1	2	1	0.144	0.550	0.600	1.0
18	45	1	2	1	0.144	0.550	0.600	1.0
18	45	1	3	1	0.170	0.600	0.700	1.0
19	44	1	4	1	0.224	0.600	0.700	1.0
19	45	1	4	1	0.224	0.600	0.700	1.0
29	34	1	5	1	0.218	0.500	0.550	1.0
29	35	1	6	1	0.230	0.500	0.550	1.0
29	38	1	7	1	0.220	0.500	0.550	1.0
32	38	1	8	1	0.175	0.500	0.550	1.0
112	113	1	9	1	0.178	0.400	0.450	1.0
34	37	1	10	1	0.262	0.500	0.550	1.0
35	37	1	11	1	0.184	0.300	0.350	1.0
35	38	1	12	1	0.203	0.400	0.450	1.0
34	38	1	13	1	0.262	0.500	0.550	1.0
120	86	1	14	1	0.185	0.600	0.700	1.0
122	86	1	14	1	0.185	0.600	0.700	1.0
120	110	1	15	1	0.152	0.450	0.550	1.0
122	110	1	15	1	0.152	0.450	0.550	1.0
37	34	1	16	1	0.252	0.500	0.550	1.0
37	35	1	17	1	0.175	0.300	0.350	1.0
38	34	1	18	1	0.241	0.400	0.450	1.0
38	35	1	19	1	0.165	0.400	0.450	1.0
118	110	1	20	1	0.128	0.450	0.500	1.0

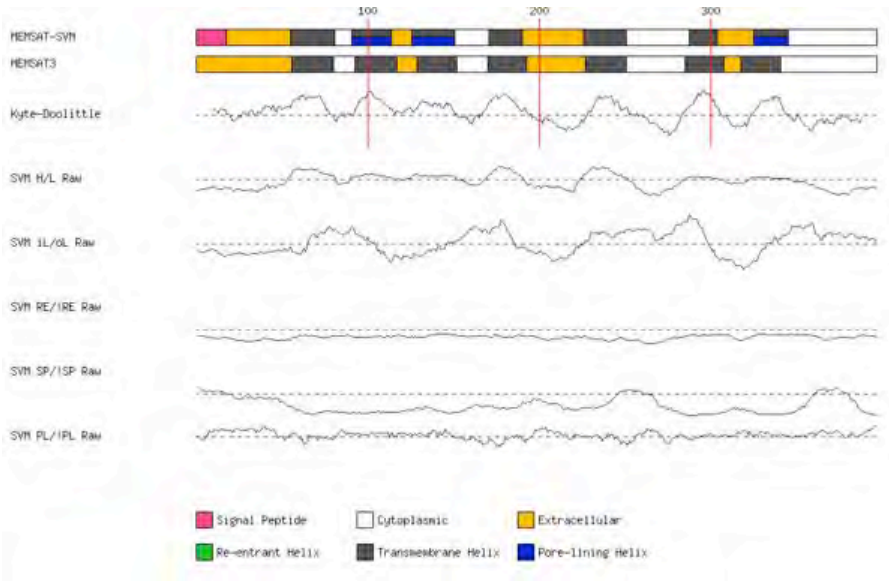
41	35	1	21	1	0.177	0.500	0.550	1.0
41	44	1	22	1	0.256	0.500	0.550	1.0
41	51	1	23	1	0.312	0.500	0.550	1.0
44	45	1	24	1	0.351	0.550	0.700	1.0
44	53	1	25	1	0.143	0.550	0.700	1.0
44	55	1	26	1	0.118	0.500	0.550	1.0
45	44	1	27	1	0.267	0.500	0.550	1.0
48	44	1	28	1	0.230	0.400	0.450	1.0
48	51	1	29	1	0.211	0.400	0.450	1.0
48	53	1	30	1	0.216	0.400	0.450	1.0
48	55	1	31	1	0.162	0.300	0.350	1.0
51	53	1	32	1	0.232	0.400	0.450	1.0
51	55	1	33	1	0.251	0.400	0.450	1.0
53	55	1	34	1	0.184	0.300	0.350	1.0
53	51	1	35	1	0.235	0.400	0.450	1.0
53	59	1	36	1	0.217	0.400	0.450	1.0
113	110	1	37	1	0.233	0.400	0.450	1.0
59	51	1	38	1	0.176	0.400	0.450	1.0
64	44	1	39	1	0.148	0.500	0.550	1.0
64	51	1	40	1	0.300	0.550	0.700	1.0
64	53	1	41	1	0.123	0.400	0.450	1.0
64	55	1	42	1	0.200	0.450	0.550	1.0
64	67	1	43	1	0.202	0.400	0.450	1.0
64	69	1	44	1	0.347	0.500	0.550	1.0
64	70	1	45	1	0.187	0.400	0.450	1.0
67	70	1	46	1	0.154	0.400	0.450	1.0
67	79	1	47	1	0.176	0.350	0.400	1.0
67	88	1	48	1	0.241	0.500	0.550	1.0
69	67	1	49	1	0.169	0.300	0.350	1.0
70	94	1	50	1	0.239	0.600	0.700	1.0
70	95	1	50	1	0.239	0.600	0.700	1.0
70	67	1	51	1	0.215	0.400	0.450	1.0
74	67	1	52	1	0.290	0.500	0.550	1.0
74	67	1	53	1	0.171	0.400	0.450	1.0
75	67	1	53	1	0.171	0.400	0.450	1.0
76	67	1	53	1	0.171	0.400	0.450	1.0

78	67	1	53	1	0.171	0.400	0.450	1.0
79	67	1	53	1	0.171	0.400	0.450	1.0
80	67	1	53	1	0.171	0.400	0.450	1.0
83	44	1	54	1	0.136	0.600	0.700	1.0
83	51	1	55	1	0.202	0.500	0.550	1.0
83	67	1	56	1	0.209	0.500	0.550	1.0
83	72	1	57	1	0.117	0.500	0.550	1.0
83	86	1	58	1	0.309	0.500	0.550	1.0
83	92	1	59	1	0.277	0.500	0.550	1.0
83	88	1	60	1	0.360	0.500	0.550	1.0
83	94	1	61	1	0.176	0.400	0.450	1.0
83	95	1	61	1	0.176	0.400	0.450	1.0
86	95	1	62	1	0.277	0.500	0.550	1.0
86	88	1	63	1	0.172	0.300	0.350	1.0
86	89	1	63	1	0.172	0.300	0.350	1.0
86	91	1	64	1	0.220	0.400	0.450	1.0
88	86	1	65	1	0.250	0.400	0.450	1.0
88	74	1	66	1	0.193	0.500	0.550	1.0
89	74	1	66	1	0.193	0.500	0.550	1.0
88	75	1	66	1	0.193	0.500	0.550	1.0
89	75	1	66	1	0.193	0.500	0.550	1.0
88	78	1	66	1	0.193	0.500	0.550	1.0
89	78	1	66	1	0.193	0.500	0.550	1.0
88	79	1	66	1	0.193	0.500	0.550	1.0
89	79	1	66	1	0.193	0.500	0.550	1.0
88	80	1	66	1	0.193	0.500	0.550	1.0
89	80	1	66	1	0.193	0.500	0.550	1.0
89	91	1	67	1	0.120	0.300	0.350	1.0
91	86	1	68	1	0.156	0.300	0.350	1.0
91	86	1	69	1	0.156	0.300	0.350	1.0
92	94	1	70	1	0.180	0.300	0.350	1.0
94	86	1	71	1	0.199	0.400	0.450	1.0
94	88	1	72	1	0.242	0.400	0.450	1.0
94	89	1	73	1	0.242	0.400	0.450	1.0
95	88	1	73	1	0.242	0.400	0.450	1.0
95	89	1	73	1	0.242	0.400	0.450	1.0

95	92	1	74	1	0.177	0.300	0.350	1.0
95	92	1	74	1	0.177	0.300	0.350	1.0
116	86	1	75	1	0.340	0.550	0.700	1.0
118	86	1	75	1	0.340	0.550	0.700	1.0
107	86	1	76	1	0.148	0.550	0.700	1.0
107	88	1	77	1	0.203	0.550	0.700	1.0
107	89	1	77	1	0.203	0.550	0.700	1.0
107	112	1	78	1	0.221	0.500	0.550	1.0
107	113	1	78	1	0.221	0.500	0.550	1.0
107	112	1	79	1	0.304	0.500	0.550	1.0
107	113	1	79	1	0.304	0.500	0.550	1.0
107	92	1	80	1	0.185	0.400	0.450	1.0
107	91	1	80	1	0.185	0.400	0.450	1.0
107	110	1	81	1	0.308	0.500	0.550	1.0
110	112	1	82	1	0.238	0.400	0.450	1.0
110	113	1	82	1	0.238	0.400	0.450	1.0
110	112	1	83	1	0.226	0.400	0.450	1.0
110	113	1	83	1	0.226	0.400	0.450	1.0
112	113	1	84	1	0.183	0.400	0.450	1.0
112	110	1	85	1	0.329	0.500	0.550	1.0
113	110	1	85	1	0.329	0.500	0.550	1.0
116	110	1	86	1	0.128	0.400	0.450	1.0
116	112	1	87	1	0.171	0.400	0.450	1.0
116	113	1	87	1	0.171	0.400	0.450	1.0
118	112	1	87	1	0.171	0.400	0.450	1.0
118	113	1	87	1	0.171	0.400	0.450	1.0

APPENDIX 3

MEMSAT3 secondary Structure prediction



APPENDIX 4A

Blind docking parameters of *flp18-6* to FLP18R1

Parameter and quantity	**Description**
outlev 1	
intelec	# calculate internal electrostatics
seed pid time	# seeds for random generator
ligand_types A C HD OA N	# atoms types in ligand
fld flp18_1.maps.fld	# grid_data_file
map flp18_1.A.map	# atom-specific affinity map
map flp18_1.C.map	# atom-specific affinity map
map flp18_1.HD.map	# atom-specific affinity map
map flp18_1.OA.map	# atom-specific affinity map
map flp18_1.N.map	# atom-specific affinity map
elecmap flp18_1.e.map	# electrostatics map
desolvmap flp18_1.d.map	# desolvation map
flp18-6.pdbqt	# small molecule
about 35.17 42.02 41.03	# small molecule center
tran0 35.17 42.02 41.03	# initial coordinates/A or random
grid_box 62, 60, 48	# box dimensions *xyz*
quat0 random	# initial quaternion
ndihe 25	# number of active torsions
dihe0 random	# initial dihedrals (relative) or random
tstep 2.0	# translation step/A
qstep 40.0	# quaternion step/deg
dstep 40.0	# torsion step/deg
torsdof 27 0.274000	# torsional degrees of freedom and coefficient
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 250	# number of individuals in population
ga_num_evals 250000000	# maximum number of energy evaluations

ga_num_generations 250000000	# 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_sw1	# set the above Solis & Wets parameters
compute_unbound_extended	# compute extended ligand energy
ga_run 10	# do this many hybrid GA-LS runs
Analysis	# perform a ranked cluster analysis

Blind docking of peptides into the FLP18R1 GPCR receptor was performed using the above parameters

APPENDIX 4B

Refined docking parameters for FLP18R1 complex with *af3* (af3.pdbqt)

Parameter and quantity	**Description**
af3.pdbqt	# small molecule
ndihe 25	# number of active torsions
torsdof 27 0.274000	# torsional degrees of freedom and coefficient
ga_pop_size 300	# number of individuals in population
ga_num_evals 250000000	# maximum number of energy evaluations
ga_num_generations 250000000	# maximum number of generations
ga_run 100	# do this many hybrid GA-LS runs
Analysis	# perform a ranked cluster analysis

AutoDock4.2 default parameters were refined as shown

APPENDIX 4C

Refined docking parameters for FLP18R1 complex with *af4*

Parameter and quantity	**Description**
Af4.pdbqt	# small molecule
ndihe 30	# number of active torsions
torsdof 32 0.274000	# torsional degrees of freedom and coefficient
ga_pop_size300	# number of individuals in population
ga_num_evals 250000000	# maximum number of energy evaluations
ga_num_generations 250000000	# maximum number of generations
ga_run 100	# do this many hybrid GA-LS runs
Analysis	# perform a ranked cluster analysis

AutoDock4.2 default parameters were refined as shown

APPENDIX 4D

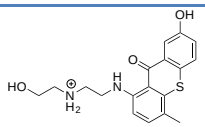
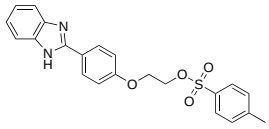
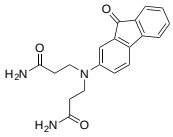
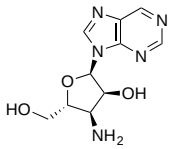
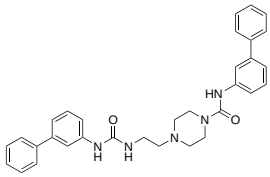
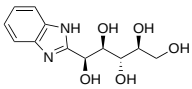
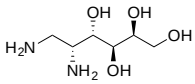
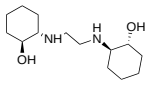
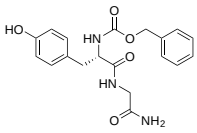
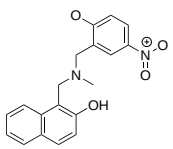
Refined docking parameters for FLP18R1 complex with *af20*

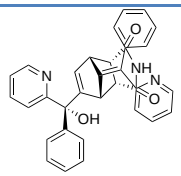
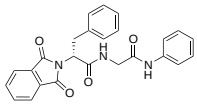
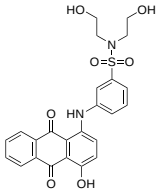
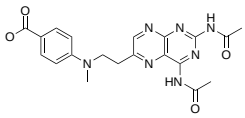
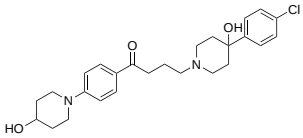
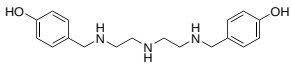
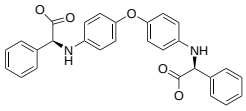
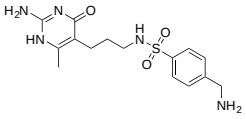
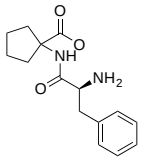
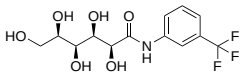
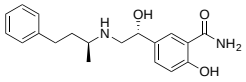
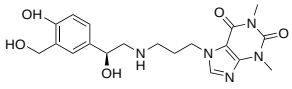
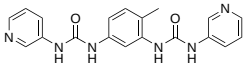
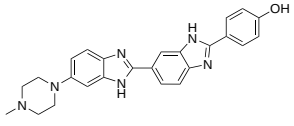
Parameter and quantity	**Description**
af20.pdbqt	# small molecule
ndihe 24	# number of active torsions
torsdof 27 0.274000	# torsional degrees of freedom and coefficient
ga_pop_size 250	# number of individuals in population
ga_num_evals 250000000	# maximum number of energy evaluations
ga_num_generations 250000000	# maximum number of generations
ga_run 100	# do this many hybrid GA-LS runs
Analysis	# perform a ranked cluster analysis

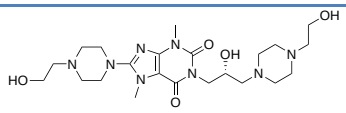
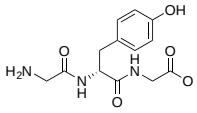
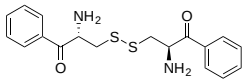
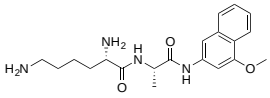
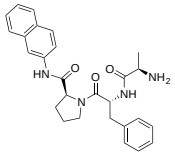
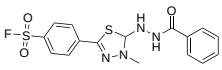
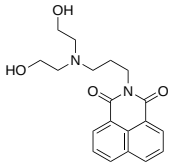
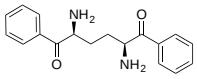
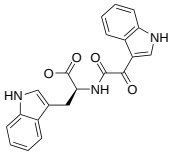
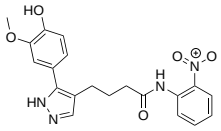
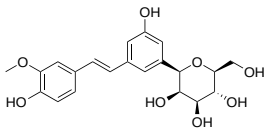
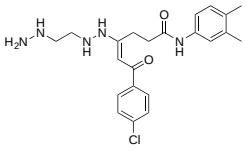
AutoDock4.2 default parameters were refined as shown

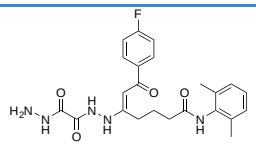
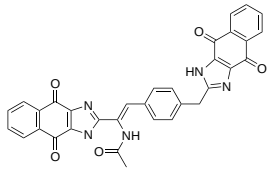
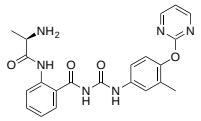
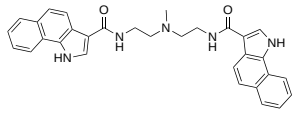
APPENDIX 5A

Compound Data from virtual screening exercise reported in Chapter 5

Cpd (NCI)	Structure	MW (g/mol)	PSA	XP Glide (kcal/mol)
327396		486.49	156.76	-9.45
303243		408.47	89.66	-8.32
15228		337.38	106.49	-8.32
18192		266.26	145.33	-9.41
55441		519.65	76.71	-8.32
56088		268.27	129.83	-9.25
75359		216.66	132.96	-9.82
87004		256.39	64.52	-9.14
89644		371.39	130.75	-9.82
98802		338.36	89.52	-9.29

101316		511.58	92.18	-8.10
117593		427.46	95.58	-8.71
135516		482.51	152.61	-8.52
136026		423.43	150.59	-8.38
170475		457.01	64.01	-8.42
197206		351.88	76.55	-11.90
201733		468.51	107.89	-8.69
210413		387.88	148.05	-8.68
241263		312.80	92.42	-8.50
254212		339.27	130.25	-9.92
290312		364.87	95.58	-9.37
293892		403.44	131.16	-10.56
319994		362.39	108.04	-8.223
322921		460.97	84.07	-8.94

326676		494.59	132.09	-8.57
333493		295.29	141.75	-8.41
337813		360.49	136.78	-9.46
339925		486.49	156.76	-9.45
339934		495.02	104.53	-9.55
372143		392.42	124.88	-9.23
377608		342.39	81.08	-8.89
603632		296.37	86.18	-10.39
608049		375.38	115.04	-9.29
622396		396.40	133.06	-8.38
622472		404.42	139.84	-9.14
625919		457.92	142.75	-8.75

632022		455.49	142.75	-9.64
637201		567.56	154.73	-9.39
654260		549.81	148.33	-9.70
704622		503.60	93.02	-9.85

APPENDIX 5B

Compounds obtained from virtual screening exercise

Well ID	**Well**	**560, 590**	**Blank 560,590**	**Count**	**Mean**	**Std Dev**	**CV (%)**	**Blank 560,590 his**	**Blank 560,590 flp18**	**fold changes**
BLK	G11	13100	-179	4	0	281	????	-348	-179	
	H11	13539	260					170	260	
	G12	12980	-299					-254	-299	
	H12	13497	218					433	218	
CTL1	A11	30540	17261	4	11024	4913	44,566	20768	17261	
	B11	25344	12065					22545	12065	
	A12	18940	5661					19605	5661	
	B12	22388	9109					26663	9109	
CTL2	D11	13001	-278	4	-312	49	-15,812	-698	-278	
	E11	12946	-333					-486	-333	
	D12	12907	-372					-503	-372	
	E12	13013	-266					-434	-266	
SPL1	A1	25329	12050	1	12050	????	????	29244	12050	0,412050335
SPL2	B1	43951	30672	1	30672	????	????	26401	30672	1,161774175
SPL3	C1	37097	23818	1	23818	????	????	12005	23818	1,984006664
SPL4	D1	37381	24102	1	24102	????	????	13260	24102	1,817647059
SPL5	E1	35516	22237	1	22237	????	????	9002	22237	2,470228838
SPL6	F1	40689	27410	1	27410	????	????	13208	27410	2,07525742
SPL7	G1	35418	22139	1	22139	????	???	9901	22139	2,236036

							??			764
SPL8	H1	30439	17160	1	17160	????	????	18027	17160	0,951905475
SPL9	A2	37705	24426	1	24426	????	????	13196	24426	1,851015459
SPL10	B2	41427	28148	1	28148	????	????	8186	28148	3,438553628
SPL11	C2	40800	27521	1	27521	????	????	23303	27521	1,181006737
SPL12	D2	44358	31079	1	31079	????	????	13957	31079	2,22676793
SPL13	E2	42539	29260	1	29260	????	????	16703	29260	1,751781117
SPL14	F2	41369	28090	1	28090	????	????	8909	28090	3,152991357
SPL15	G2	32973	19694	1	19694	????	????	9779	19694	2,013907352
SPL16	H2	37931	24652	1	24652	????	????	11951	24652	2,062756255
SPL17	A3	46603	33324	1	33324	????	????	12964	33324	2,570502931
SPL18	B3	45506	32227	1	32227	????	????	9389	32227	3,432420918
SPL19	C3	41629	28350	1	28350	????	????	8134	28350	3,485370052
SPL20	D3	44194	30915	1	30915	????	????	9640	30915	3,206950207
SPL21	E3	21459	8180	1	8180	????	????	22953	8180	0,35638043
SPL22	F3	44644	31365	1	31365	????	????	19637	31365	1,597239904
SPL23	G3	11111	-2168	1	-2168	????	????	-1634	-2168	1,326805386
SPL24	H3	28274	14995	1	14995	????	???	11546	14995	1,298718

							??			171
SPL25	A4	16142	2863	1	2863	????	????	32851	2863	0,087151076
SPL26	B4	43418	30139	1	30139	????	????	13878	30139	2,171710621
SPL27	C4	30622	17343	1	17343	????	????	11604	17343	1,494570838
SPL28	D4	45470	32191	1	32191	????	????	9911	32191	3,248007265
SPL29	E4	38143	24864	1	24864	????	????	12036	24864	2,065802592
SPL30	F4	26997	13718	1	13718	????	????	18611	13718	0,737090968
SPL31	G4	38656	25377	1	25377	????	????	6819	25377	3,721513418
SPL32	H4	40095	26816	1	26816	????	????	11895	26816	2,254392602
SPL33	A5	40356	27077	1	27077	????	????	14449	27077	1,873970517
SPL34	B5	28882	15603	1	15603	????	????	14646	15603	1,065342073
SPL35	C5	39797	26518	1	26518	????	????	22163	26518	1,196498669
SPL36	D5	36071	22792	1	22792	????	????	10612	22792	2,147757256
SPL37	E5	47295	34016	1	34016	????	????	14200	34016	2,395492958
SPL38	F5	36008	22729	1	22729	????	????	33194	22729	0,684732181
SPL39	G5	39398	26119	1	26119	????	????	15203	26119	1,718016181
SPL40	H5	15948	2669	1	2669	????	????	3457	2669	0,772056697