Enhancing the Fight against Malaria*:* From Genome to Structure and Activity of a G-Protein Coupled Receptor from the Mosquito, *Anopheles Gambiae.*

Grace Chitima Mugumbate

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

March 2010

Declaration

I declare that this thesis, "*Enhancing the fight against Malaria: From genome to Structure and Activity of a G-protein Coupled Receptor from the mosquito, Anopheles Gambiae,"* is my own work that was performed under the supervision of Professor G. E. Jackson, and that it 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.

Grace Chitima Mugumbate **March 2010**

Signature_____

Dedication

This work is dedicated to my two wonderful children, Tafadzwa William,

and Channesse Rutendo, and to God.

Acknowledgements

I would like to express my sincere gratitude to the following, without whom this journey would have been difficult:

My Supervisor, **Professor G. E Jackson**, for having confidence in me, providing guidance, direction and encouragement throughout the course of this work.

Professor G. E. Jackson's research group, for the time spent working and having fun together.

The Department of Chemistry (UCT) and staff, for providing an environment that is conducive for research and learning.

Professor D. van der Spoel, and the Molecular Biophysics research group, for helping with molecular dynamic simulations and the helpful discussions during research visits to Sweden.

Dr. C. Hétényi, for helping with docking studies, binding affinity calculations and helpful discussions.

Dr Katalin Kover and **Professor Laszlo Szlargyi**, Debrecen University, Hungary, for the NMR spectra.

Family and friends, who were always available to listen, encourage and support in amazing ways.

Last but not least**, The Almighty God,** for strength, wisdom, knowledge and skill, and for providing all resources needed for the work to be carried out.

Financial Support

I would like to thank The Equity Development Programme (Department of Chemistry), National Research Fund, Department of Science and Technology and the Swedish Research Fund for providing the financial support.

Outputs From the study

1) Oral presentation at the Chemistry Equity Development Scholarship Programme Seminars, Department of Chemistry, 22 September 2009.

2) Poster presentation at the 42nd IUPAC Congress, Glasgow, Scotland, 31 July – 6 August 2009. Awarded the IUPAC poster prize.

3) Oral presentation at the Cape Organometallic Symposium COS 08, Breakwater Lodge, Waterfront, Cape Town, South Africa.

4) Oral presentation at the 39th SACI National Convention, 2008, Stellenbosch, South Africa.

5) Oral presentation at the Postgraduate Symposium 2008, University of Cape Town, South Africa.

6) Poster presentation at the Carman National Physical Chemistry Symposium, 2007, The forum, V & A Waterfront, Cape Town, South Africa.

Abstract

Enhancing the Fight against Malaria: From Genome to Structure and Activity of a G-Protein Coupled Receptor in the Mosquito, *Anopheles Gambiae*

Grace Chitima Mugumbate

Department of Chemistry, University of Cape Town, Private Bag, Rondebosch, 7701, South Africa. Submitted March 2010

G-proton coupled receptors (GPCRs) are excellent drug targets that occupy a central position in the physiology of insects and are involved in transmission of signal from the extracellular to the intracellular side of the cell. Adipokinetic hormone receptors (AKHRs) are GPCRs that mediate physiological functions of the neurohormones, adipokinetic hormones (AKHs) that regulate mobilisation of energy reserves during mosquito flight. Ligand binding to GPCRs depends on the three dimensional (3D) structures of the receptors but to date no crystal structures of insect GPCRs are available. This work focused on building molecular models of AKHR from the genome of the malaria mosquito, identifying its binding site and studying the conformational and structural changes during molecular dynamics of the active and inactive receptor.

Homology modelling was used to build the helices based on the crystal structures of rhodopsin and beta2-adrenergic receptor (β2AR). The loops were built separately and joined to their respective helices. Molecular dynamics was used for conformational search of the loops. The two resulting 3D structures of the GPCR from the malaria mosquito had similar overall structures. However, the β_2AR-based structure had an 'open' conformation in the extracellular region, whilst the rhodopsin-based model was 'closed'.

NMR restrained molecular dynamics was used to determine the solution conformation of AKH-I from *Anopheles gambiae* (Anoga_akh). Docking calculations of this peptide and the decapeptide, Del_CC, from the blister beetle showed that helices 2,3,5,6, and 7, and the extracellular domains defined the binding pocket of AKHR. The 'open' AKHR model provided easy access to the binding site and had higher affinity for the ligands than the 'closed' structure. During molecular dynamics, after binding of the agonist, the receptor binding pocket closed to protect the ligand. At the same time the intracellular region opened. Although conversion of the receptor from inactive to active state was slow with Anoga_akh, the receptor had a higher affinity for the ligand than for Del_CC as indicated by estimated free energy of binding, -47.3 kJ/mol and -38.5 kJ/mol respectively. The protein-ligand complexes were stabilised by an intense network of H-bonds, salt bridges and hydrophobic interactions. Tyr285 (H6) played an important role in binding Del_CC, whilst ILe106 (H2) was pivotal in binding Anoga_akh.

Since AKHR facilitates energy mobilisation during insect flight, knowledge of the 3D structure and binding pocket of the receptor from the malaria mosquito could lead to structure-based design of non peptide antagonists that prevent binding of AKH molecules. This would stop generation of energy for the mosquito to fly and pave the way for development of insect specific insecticides and reduction of transmission of malaria.

Keywords

Malaria

Molecular modelling

G-protein coupled receptors

Rhodopsin

Beta2-adrenergic receptors

Adipokinetic hormones

GROMACS

Docking

Homology modelling

Molecular dynamics

Abbreviations

GPCRs - G-protein coupled receptors

MD - Molecular dynamics

AKH - adipokinetic hormones

AKHR - adipokinetic hormones

β_2AR - beta2-adrenergic receptor

7TMs - seven transmembrane helices

ECL - extracellular loop

ICL - intracellular loop

Amino acids (Residues)

Name	Three letter Name	Symbol
Alanine	Ala	A
Aspartic acid	Asp	D
Asparagine	Asn	N
Arginine	Arg	R
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
IsoLeucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Trpptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Contents

CHAPTER ONE

INTRODUCTION

1.1 Summary

Insect flight is under hormonal control via adipokinetic hormones that regulate the mobilisation of energy reserves during flight. Their action occurs through binding to G-protein coupled receptors, which occupy a central position in the physiology of insects and are excellent drug targets. Binding of the hormone to a GPCR depends on the three dimensional structure (3D) of the receptor but to date no 3D structure of an insect G-protein coupled receptor (GPCR) is available.

In this thesis, the first models of the 3D structures of a GPCR from the malaria mosquito are presented. Molecular modelling of the receptor, using the primary sequence identified from the genome of *Anopheles gambiae,* was used to construct two models of its AKH receptor. Homology modelling of the 7 transmembrane helical bundles of these two models was based on the crystal structures of beta2-adrenergic receptor (β_2AR) and rhodopsin. The flexible loop regions were modelled using high temperature simulated annealing and constrained molecular dynamic simulations. The two receptor models differ in a number of critical features, the most important of which is that the rhodopsin-based model has a 'closed' structure while the β_2AR-based structure is 'open'. To mimic the receptor's true environment, extensive molecular dynamic simulations were performed in an explicit membrane to determine stability and conformational changes during molecular dynamics.

Docking calculations with insect adipokinetic hormones, showed that helices 2, 3, 5, 6 and 7, and the extracellular domain of the receptor participate in hormone binding. The binding pocket in β_2AR-based model was easily accessed by the

ligands, thus hormone binding was facilitated. Molecular dynamics of the receptor/ligand complex have revealed that the inactive state of AKH receptor has an open conformation whilst the activated form takes up a closed conformation similar to rhodopsin. Binding of the ligand involved ionic as well as hydrophobic interactions with residues in the helical bundle, for example Tyr285, and in the loop regions.

Since this receptor is involved in generation of energy during insect flight, knowledge of its 3D structure and binding pocket could lead to the design of non-peptide mimetics that can block the binding pocket and lead to the development of species specific insecticides. Therefore, by using genomic information, the fight against malaria could be enhanced.

1.2 The Mosquito, Anopheles gambiae and Malaria

The mosquito, *Anopheles gambiae,* (Figure 1.1) is the World's most efficient vector for the malaria parasite, *Plasmodium falciparum* [1,2,3].

Figure 1.1: The mosquito, *Anopheles gambiae*, the major vector of the malaria parasite, *Plasmodium falciparum*.

Malaria is endemic to large portions in Africa and is responsible for up to 500 million cases of illness and 1.5-4 million deaths each year, 90% of which occur in Sub-

Saharan Africa [1,4]. The mosquito requires multiple human blood meals for its reproductive cycle and this favours the transmission of the malaria parasite to humans [3]. For this reason, it is important to control this insect so as to reduce the spread of malaria.

Insecticide-based control measures are the primary way to kill mosquitoes that bite indoors and at night, but there is increasing resistance of *A. gambiae* to currently used insecticides like DDT, permethrin and deltamethrin [1,5,6]. These insecticides also have adverse side effects in humans, therefore there is an urgent need for continuous search for alternative insect specific insecticides. The genome of *Anopheles gambiae* has recently been published [1], hence in this thesis utilisation of genomic information in designing new insecticides and identifying alternative targets is presented.

1.3 Adipokinetic Hormones (AKHs) from Anopheles gambiae

Important in the mosquito activity is its ability to fly and flight activity is under hormonal control via the adipokinetic hormones (AKHs) [7]. The hormones are all octa-, nano- or decapeptides with N-(pyroglutamate) and C-(amidate) termini and are highly homologous, for example, Del_CC (pQLAFSPNWGNa) and Term_HrTH (pQLAFSPNWa) from the blister beetle species [7,8]. Most of these hormones contain Phe, Pro and Trp at positions 4, 6 and 8 respectively, which is postulated to be important in hormone structure and receptor binding. AKHs contribute to hemolymph sugar homeostatsis and are involved in the mobilisation of sugar and lipids from the fat body during energy-demanding activities, such as flight and locomotion [8-12].

As in most insects, two AKHs, octapeptide AKH-I (pQLTFTPAWa) and decapeptide, AKH-II (pQVTFSRDWNAa), have been identified in *A. gambiae* [9,11]. These peptides were present in the freshly hatched larvae, heads and thoraces of pupae and adult mosquitoes. AKH-I was found to be responsible for inducing mobilisation of carbohydrates and lipids during flight and therefore enhancing the flight activity of the female mosquito [9]. On the other hand AKH-II, which consists of Arg instead of Pro in position 6, had no effect on carbohydrate and lipid levels and hence did not influence flight performance [9].

Commencement of flight activity induces the release of the hormones from the *corpora cardiaca*, a neurohemal gland connected to the insect brain, into the circulatory system [12]. Binding of the AKHs to their G- protein coupled receptor (GPCRs), at the fat body, triggers a number of coordinated signal transduction processes that ultimately result in the mobilisation of carbohydrates and lipid reserves as fuel for flight activity [3,7,8,9,11,12]. It has been suggested that studies of these receptors, that play an important role in processes central to vector capacity, would contribute to the development of a new generation of mosquito repellents, attractants, insecticides, as well as other types of antimalarial drugs [1,2,3]. These drugs would reduce the spread of malaria by reducing the effectiveness with which mosquitoes locate and feed on their human victim [1]. In this work a study of the physiologically important AKH GPCR that facilitates generation of energy during mosquito flight is presented.

1.4 The G - protein Coupled receptors (GPCRs)

1.4.1 Secondary Structure and Function

GPCRs comprise the largest superfamily of membrane-bound proteins in animal genome [2,13-18] and share a similar overall tertiary structure characterised by highly conserved seven α-helices that span the cell membrane (7TMs), three extracellular (ECL) and three intracellular loops (ICL) that join the helices together, the N-terminus and the C-terminus [13,14,16,19-23]. Due to the presence of the 7TMs, this family of proteins is also known as the 7-transmembrane receptors or heptahelical receptors [24]. The structural domains commonly found in these receptors are shown in the schematic diagram (Figure 1.2) of beta2-adrenergic receptor below [23,25-28].

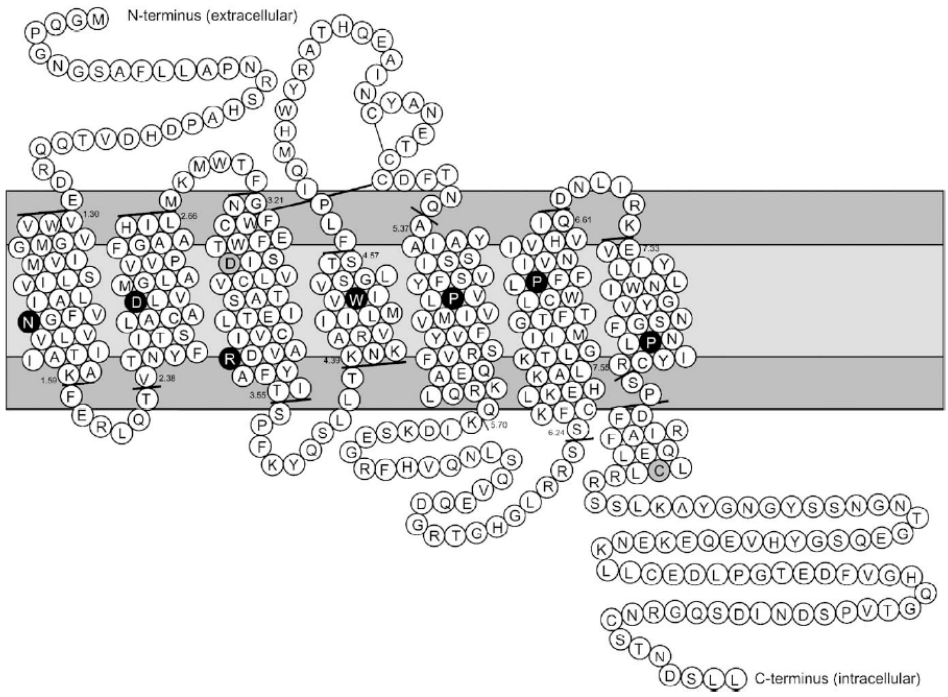


Figure 1.2: Schematic representation of the membrane topology of the human beta2-adrenergic receptor. Black lines indicate the localised transmembrane helices, light grey and dark grey show the core and the water-lipid interface regions of the lipid membrane, respectively. The black circles in the each of the 7 transmembrane helices indicate the highly conserved residues.

As shown in Figure 1.2, each helix consists of highly conserved residues, motifs and patterns for class A receptors, which are Asn (H1), Asp (H2), DRY (H3), Trp (H4), Pro (H5), CwxxY/F (H6) and NPxxY (H7), where x stands for any other residue [16,27,29,30]. Patterns are short sequence of continuous or discontinuous highly conserved amino acids in one helix and one GPCR class and motifs are described as continuous multiple-sequence regions that include conserved amino acids [30]. In addition the receptors are also characterised by disulphide bonds linking the first extracellular loop and the second extracellular loop that stabilise the helix bundles and also participate in ligand binding [19,16,23,25,26]. In β_2AR there are two such bridges between Cys106/Cys191 and Cys184/Cys190. Rhodopsin however contains only one disulphide bond between Cys 110 and Cys187. In the intracellular region, there is a palmitoylation site involving Cys341 in β_2AR, and Cys322 and Cys323 in rhodopsin [19,25,26,31-34].

GPCRs form the largest family of proteins that comprise receptors for ~80% of hormones, neurotransmitters and neuromodulators with important physiological functions [27,35]. The family includes receptors that respond to a variety of endogenous agonists and antagonists such as amines, peptides, amino acids, glycoproteins, phospholipids, fatty acids, nucleotides, Ca^{2+} ions and sensory receptors [14,15,27]. These receptors transduce extracellular stimuli into intracellular signals when agonistic binding to the extracellular receptor results in conformational changes of the 7TMs, resulting in activation of a heterotrimeric G-protein and generation of a cascade of signalling pathway [2,14,13,36]. For this reason they are good drug targets. In humans there is an estimated 948 GPCRs and about 50% of all targeted pharmaceuticals are active because of binding to GPCRs [4,15,37,38]. Therefore, AKHs receptors can provide alternative drug targets for insect specific insecticides.

1.4.2 Classification of GPCRs

Although GPCRs do not share any overall sequence homology they contain the seven transmembrane α-helices as the only common structural feature to all receptors [22]. Significant homology is however found within several sub-families which were formed based on their native ligands, phylogenetic analysis and their amino acid sequences [27]. Kolakowski, 1994 [39] first divided the receptors into six different families or classes. Each family had >20% amino acid sequence identity within the 7TMs. Three major subfamilies were divided into Classes A, B and C. Class A includes receptors related to rhodopsin and β_2-AR, Class B comprises of receptors related to glucagons and receptors related to the metabotropic neurotransmitter receptors form Class C of GPCRs [22,27]. The minor classes D, E and F comprise the yeast pheromone receptors and the cAMP receptors respectively.

Class A (rhodopsin-like) GPCRs is the largest and most studied family and is divided into six major subfamilies, which are opsins, receptors for odorants, small endogenous agonists, peptides, leucine-rich repeat motif and protease-activating receptors [22,27]. The overall homology of these receptors is low and restricted to a number of highly conserved key residues and motifs, namely Asn (H1), Asp (H2), DRY (H3), Pro (H5), CWxxY (H6) and NPxxY (H7), as shown in Figure 1.2 [22,27]. The high degree of conservation among these key residues suggests that they have an essential role for either the structural or functional features of the receptors. All Class A receptors contain the residue, arginine, in the DRY motif at the cytoplasmic side of helix 3 (H3). In addition, all Class A receptors have a disulphide bond that connects the second and third extracellular loops.

On the otherhand, Class B includes about 20 different receptors for a variety of peptide hormones and neuropeptides e.g. vasoactive intestinal peptide (VIP) and

calcitonin and glucagons [22,27]. The disulphide bond is the only structural feature also found in Class B and the important DRY motif common to all Class A receptors does not exist in this family. These receptors are characterised by a large (~ 100 residues) extracellular amino terminus with several cysteines. In contrast, Class C has an exceptionally long amino terminus of about 500-600 residues and include the metabotropic glutamate and γ-amino-butyric acid receptors, the calcium receptors and the taste receptors [22,23,27]. Class C receptors have two putative disulphide bond forming cysteines in ECL2 and ECL3 respectively. The three classes do not share any conserved residues amongst themselves. The recently published primary structure of AKH receptors in *Anopheles gambia*e has revealed that the receptor belongs to Class A GPCRs.

1.4.2.1 AKH G-protein coupled receptors

The first biochemical characterisation of AKH G-protein coupled receptors was performed for the tobacco hawk moth, *manduca sexta,* followed by cloning and identification of the receptors in the fruitfly, *Drosophila melanogaster*, the silkworm, *Bomyx mori* and the cockroach, *Periplaneta Americana* [10,19,12]. The receptors were found to be structurally and evolutionarily related to the gonadotropin-releasing hormone receptor from vertebrates [10]. AKH receptor cloned from *Bomyx mori* had 48% sequence identity and 68% conserved residues similarity with receptor-1 of *Drosophila melanogaster.* Both receptors had two potential glycosylation sites at the same position [10].

From the recently published genome for *A. gambiae* 276 genes could be interpreted as putative neuropeptide receptors. Of these, 25 were GPCRs belonging to Class A [2,3,11]. As suggested earlier, utilization of the genomic data should pave

the way for new ways of combating the spread of malaria with highly selective and environmentally friendly insecticides [4]. Thus, from orphan GPCRs of *Anopheles gambiae*, Kaufmann, C. & Brown, M. R [11] and Belmont, M. et al [4] identified the primary sequence of a putative AKH GPCR (AKHR), which belonged to Class A receptors (Figure 1.3).

MPNTMAAHINQRIEDHRNLADWSYYANETAGEEYYEMPIDMR
FNSGHILSIMVYTTLMVFSATGNLTVLAQRKVRASSRINIMLAH
LAIADLLVTFLMMPLEIGWAYTVRWTAGDLMCRVMAFFRTFG
LYLSSFILICISVDRYFAVLKPLKVHEHRAVLMIAAAWIMSGLCS
LPQAFIFHLEGHPNTTGYQQCVTYHYFEEEIYQIIYNVLVMCLM
YTFPLIVILYCYGSIYYEIFSRTNPRNLESFRRSSIDVLGRAKRKT
LRMTIMIVIVFVVCWTPYYVMSLWYWLDKESAKNVDQRIQKG
LFLFASTNSCMNPVVYGVFNVRKKHTKKLLKTTHEKSCGSHLT
MRA

Figure 1.3: Primary sequence of adipokinetic hormone receptor (AKHR) from the malaria mosquito, *Anopheles gambiae*, extracted (Genbank database accession number AY298745). The loop regions (black), helices 1 to 7 (red) and the highly conserved residues, patterns and motifs (blue) are as predicted by Belmont, M et al, 2006 [4]. These regions were consistent with those predicted by Kaufmann and Brown, 2006 [11].

The receptor had 354 residues and was highly concentrated in the head, thorax and abdomen of the mosquito [9,11] , where the fat body, involved in nutrient storage and metabolism, is distributed [12]. The putative AKH receptor from *Anopheles gambiae* has high sequence similarity to the receptor from *Drosophila melanogaster* [4,11]. As shown in Figure 1.3, the receptor has seven transmembrane regions, an N-terminus

with 50 residues and a C-terminus of 27 residues, which is much shorter than in other AKH GPCRs [4,11]. The seven transmembrane helices contain residues, patterns and motifs commonly found in Class A GPCRs, shown in Table 1.1, where x represents any other residue.

Table 1.1: Highly Conserved residues, Patterns and motifs in AKHR and Class A GPCRs

Helix	Residue/Motif/Patterns in Class A GPCRs	Residues/Motifs/Patterns in AKHR
1	N	N
2	LAxxD	LAIAD
3	(D/E)R(Y/H)	DRY
4	W	W
5	(F/Y)xxPxxxxxxxY	YTFPLIVILYCY
6	(F/YxxxWxPYY	FVVCWTPYY
7	(N/D)PxxY	NPVVY

These residues are believed to play an important role in stabilising the helical bundle through helical interactions. In addition they are also crucial for ligand binding and receptor activation.

*1.4.3 Activation of Class A GPC*Rs

Activation of GPCRs is induced by binding of a variety of ligands to the extracellular side of the receptor. This results in diverse ligand – receptor interactions that cause changes in the conformation of the receptor molecule especially in the helix bundles [13,27]. Opsins, (rhodopsin and colour pigments), are different from other Class A receptors in that they are activated by photons of light. The ligand, a chromophore,

11-cis-retinal, is covalently bound to the protein by a protonated Schiff base bond to the e-amino group of the Lsy residue in H7 [16,32,33,40]. When exposed to light, there is a change from 11-cis to an all-trans conformation of the ligand within femtoseconds [13,16,33,32]. Activation of these receptors involves the neutralisation of the Schiff's base counter ion, and a conformational change in the intracellular side associated with protonation of two cytoplasmic acidic residues [27]. In bovine rhodopsin, it has been proved that photoactivation can induce movements of helix 3 (H3) and helix 6 (H6), relative to the rest of the 7-helix bundle, a significant rigid body movement of H6 in a clockwise direction when viewed from the extracellular side, and movement of the cytoplasmic end of the helix away from H3 that results in the disruption of the ionic bonds involving the highly conserved Arg residue in H3 [27].

Small endogenous and exogenous receptors are activated by noncovalent binding of specific agonists to the upper part of the helix bundles. Spectroscopic analysis of the fluorescent antagonist, carazolol, in β_2AR has shown that the binding sites of these receptors are buried deep inside the helices [41]. Molecular biology techniques have also illustrated that the binding sites for a number of small agonists like biogenic amines, adenosine, nucleotides and 3-peptide thyrotropin releasing hormone are between H3, H4, H5, H6 and H7 [42-46]. In addition, site-directed mutagenesis experiments have shown that some residues in the extracellular loop domains, eg ECL2 are important for binding of agonist and antagonists, [47]. The presence of the disulphide bond between ECL2 and ECL3 would covalently link the loop in close proximity to the binding site. Kim et al, 1996 [47], highlighted that the second extracellular loop may form a lid over the bound ligand and keep it in the binding site. From results of mutagenesis studies of serotonin it was suggested that

ECL1 and ECL3 may also play an important role for receptor conformation [48], or form a channel for ligand entrance to the primary binding site [27]. Ligand channelling hypothesis has also been proposed for rhodopsin [49]. The ligands may enter their binding pockets after attraction and sequential interactions with a channel between ECL1 and ECLs/TMs at the entrance to the buried binding site within the helical bundle.

Peptide receptors form the largest subfamily of Class A GPCRs and mediate important physiological roles of neurotransmitters, hormones, and paracrines [27]. Mutational studies have shown that in many peptide receptors there is critical involvement of the extracellular domains for binding of the large peptide ligands [50-54] and that these interact directly with residues in the N-terminus and the extracellular loop domains. Amongst these are receptors for angiotension II, neuropeptide Y (NPY), chemokines, thyrotropin-releasing hormone (TRH) and gonadotropin releasing hormone (GnRH) [55-61]. Experimental results have also shown that residues in the upper parts of H2, H3, H5, H6 and H7 interact directly with certain neuropeptides, for example, in human, Y_1NPY, δ opiod, ET_B endothelin [62,63,64]. Recently, docking calculations involving an insect peptide GPCR, pheromone biosynthesis activating neuropeptide receptor, have also indicated the importance of the extracellular domains in ligand binding [65,66,67]. The identified binding positions are different from the key positions believed to interact with ligands in the biogenic amine receptors [22].

GPCRs are believed to exist in equilibrium between the activated and the inactivated states. There is interconversion between the two states upon ligand binding. When bound to an agonist, the activated state is favoured leading to formation of the agonist-activated receptor-G-protein complex [27,68]. On the other

hand, inverse agonists bind to the receptor with higher affinity for the inactive state than the active state [69], whereas neutral antagonists bind with equal affinity for the activated and the inactive states and have no effect on the constitutive activity.

1.4.3.1 G-proteins and signal transduction

GPCRs act on the heterotrimeric G-proteins as guanine-nucleotide exchange factors [22,27]. There are three subunits of heterotrimeric G-proteins, α, β and γ. The receptor is bound to the GDP-ligand α-subunit and this association is greatly enhanced by association with $\beta\gamma$-subunits [27]. Upon activation, conformational changes occur in the GPCR molecule resulting in increased affinity of the associated G-protein α-subunit leading to release of GDP and the immediate binding of GTP [22,27,70]. Affinity for α-subunit for $\beta\gamma$-subunits decreases and the GTP-bound form of the α-subunit dissociates from the receptor as well as from the $\beta\gamma$- dimer. The released $\beta\gamma$-dimer can modulate several cellular signalling pathways including stimulation or inhibition of adenylate cyclases, tyrosine kinases, GPCR kinases (GRKs) and activation of phospholipases, as well as regulation of potassium and calcium channel activity resulting in cellular functions [71,72].

To quench the activation, GTP is hydrolysed to GDP, leading to resuscitation of α-GDP and $\beta\gamma$-subunits and the receptor reverts back to the inactive state [27]. The signal can also be terminated by reuptake and desensitisation of neurotransmitters or their precursor, or the degradation of neuropeptides by the cell surfaces proteinases.

1.4.4 Three dimensional structures of GPCRs

Three dimensional (3D) structures of receptors provide insight into the molecular basis of protein function for effective design of experiments like site-directed mutagenesis, disease-related mutation studies and structure-based design of specific

inhibitors [73]. Since it has been observed that small agonists and the majority of antagonists for GPCRs bind wholly to the transmembrane helices [13,16], and peptide hormones also bind to the extracellular domain [22,70], 3D structures of these receptors are crucial for ligand binding and signal transduction.

1.4.4.1 Determination of 3D structures of GPCRs

Despite their importance in physiological processes, very little information on the 3D structures of GPCRs, is available. Experimental methods that include NMR and X-ray crystallography, normally used in the determination of the 3D structures of substances, cannot be easily used to determine the 3D structure of GPCRs [15]. This is because the membrane-bound proteins are highly flexible and hence difficult to crystallise for application of X-ray crystallography [19,39,74,75]. On the other hand the large molecular size and low solubility [15,19,39,74,75] of the receptors make it impossible to apply NMR techniques. Thus, to date, the only resolved crystal structures of the GPCR, are for the mammalian rhodopsin [16,18], and the human Beta2- adrenergic receptor (β_2AR) [25,26]. However, the highly mobile third intracellular loop has been removed in both crystals. For this reason, most known pharmaceuticals that target GPCRs are based on molecular models obtained from computational methods.

1.4.4.1.1 Crystal structures of Rhodopsin and Beta2- adrenergic receptors

Rhodopsin and β_2AR have similar overall structures [14], and provide good templates for homology modelling of GPCRs [14,19,75]. The two receptors belong to Class A GPCRs [76], in which rhodopsin is a member of the opsin subfamily [19,26] and β_2AR belongs to the amine family [14,19,26].

Rhodopsin

Rhodopsin is found in the retina of the eye and is structurally stabilised by a covalently bound ligand, 11–cis retinal [19, 28,29,36,75]. The first crystal structure of this receptor was resolved at 2.8 Å by Palczewski *et al,* [16] and deposited in the Protein Databank (PDB) as 1F88A (Figure 1.3a &b). The receptor had a total of 348 residues with 194 of the residues making up the 7TMs.

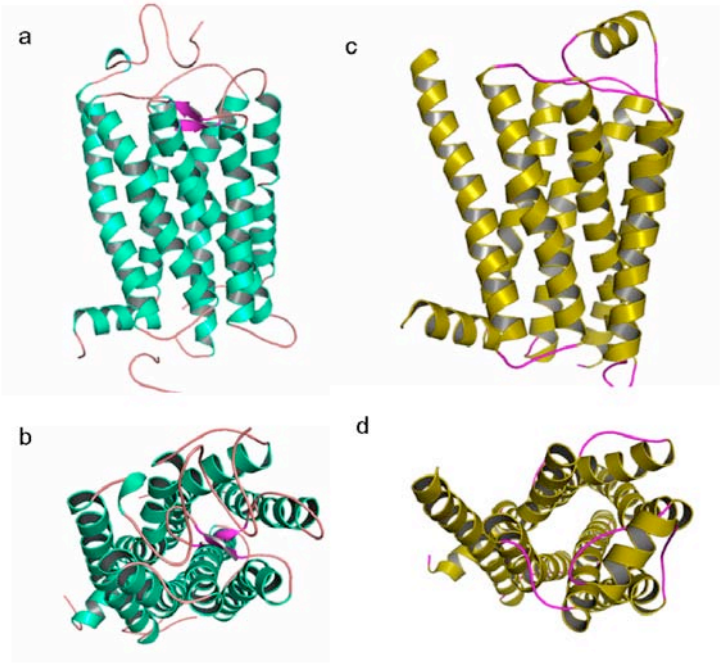


Figure 1.3: Ribbon structures of the two crystal structures of GPCRs. a) Side view of the structure of rhodopsin (PDB, 1F88A), showing the helices (cyan), the loops (brown) and two β-sheets (pink) in the second extracellular loop. The loop forms a lid over the binding site to protect the bound ligand (not shown). b) Extracellular view of rhodopsin indicating the arrangement of the helices. c) Side view of β_2AR (PDB, 2rh1A) indicating the helices (gold) and the loops (pink). The second extracellular loop has a helix that keeps away from the binding pocket and d) shows the extracellular view of the molecule which illustrates the open conformation of the helix bundle.

The crystal structure showed the packing of the seven transmembrane helices across the membrane, an eighth helix parallel to the membrane inside the cell, and the positions of the termini and the loop domains. The binding site is covered by a β–sheet lid formed by the second extracellular loop, (ECL2) and a cap–like structure formed by the N-terminus and the helix domain [19,20,26]. The disulphide bond is highly conserved in most GPCRs and is formed by Cys110 at the end of helix 3 and Cys187 in the second extracellular loop.

As mentioned earlier this receptor is activated only by a photon of light that induces conformational changes in the bound ligand. Crystallographic refinement produced the model of rhodopsin at resolution of 2.8Å (PDB code, 1HZX) [33]. Okada et al refined the ground-state of rhodopsin to 2.6Å at which water molecules could be assigned to the helix bundle [32]. Up until 2007, this has been the only known crystal structure of a GPCR, that has been used as a template for building 3D structures of all most all GPCRs.

Beta2-adrenergic receptor

The crystal structure of β_2AR was obtained at a resolution of 2.4Å, by replacing part of the third extracellular loop of the receptor with T4 lysozyme to reduce its flexibility. The structure was published in 2007 [19,25,26] (Figure 1.3c & d) and deposited in PDB as 2rh1A. The receptor has a broader range of signalling behaviour, couples to more than one G-protein and responds to a spectrum of many diffusible ligands. The binding site was located in a similar position to that of rhodopsin [19,20]. In comparison to rhodopsin, it had a more open structure in the extracellular region of the helix bundle, with the second extracellular loop constrained by two disulphide bonds so that there is open access to the ligand-binding pocket [19].

Both β_2AR and rhodopsin, however, undergo similar conformational changes in the process of activation [19]. The structural differences between β_2AR and rhodopsin helices clearly show that the two structures would predict quite different locations for the same residues when used as templates in homology modelling of the same GPCRs and hence give models with different structural features.

Modelling results of large protein molecules like AKHR are largely dependent on the starting structure and may not be identical with the structure of AKHR or the AKHR/AKH determined by x-ray crystallography or by NMR spectroscopic methods. However, the crystal structures of rhodopsin and β_2AR gave valuable information on the structure and activity of Class A GPCRs, and hence, facilitated structure-based drug design. Therefore, the 3D structure of the adipokinetic hormone receptor from *Anopheles gambiae,* could give essential information for the design of non-peptide mimetics which can block the receptor binding pocket, reduce the amount of energy available for the mosquito to fly and hence prevent the transmission of malaria.

1.5 Problem Statement and Research Motivation

Despite their importance in physiological processes, very little is known about the 3D structures of insect GPCRs. In fact, no crystal structures of these crucial receptors are known. That is, to date the 3D structures of GPCRs for adipokinetic hormones from *Anopheles gambiae* and knowledge of their binding pockets are not available. This hinders the use of these receptors in structure-based drug design that could lead to production of alternative insecticides and identification of alternative targets.

Since only two crystal structures of GPCRs, rhodopsin and Beta2-adrenergic receptor, have been resolved [16,19,25,26,28,32], in humans, ~50% of all known pharmaceuticals target GPCRs have been designed using homology models of these

receptors [36,65,67,75-84]. Therefore, from an accurate model of AKHR, the binding site can be identified and this should facilitate structure-based design of non-peptide ligands that can act with high potency at the receptor as antagonists and agonists [22]. These will lead to production of insect specific insecticides that reduce the vector capacity of the mosquito by reducing the amount of energy released during flight, thus enhancing the eradication of malaria. To this reason, this thesis focuses on the determination of the 3D structure and location of binding site of adipokinetic hormones G-protein coupled receptors from the mosquito, *Anopheles gambiae*.

1.6 Aims and Objectives

Aims

To determine the three dimensional structure of a GPCR, adipokinetic hormone receptor, from the malaria mosquito, *Anopheles gambiae* and to use this in docking studies to identify the hormone binding pocket.

Objectives

- Use the primary sequence of adipokinetic hormone receptor to build the three dimensional structure of the helix bundles using homology modelling based on the crystal structures of rhodopsin and beta2-adrenergic receptor.
- Separately build the loop domains using molecular dynamic simulations.
- Join the loops to their respective helices and optimise the final models of the adipokinetic hormone receptor in vacuum and then in an explicit membrane/water system.
- Identify the binding pocket of the adipokinetic hormone receptor by performing docking calculations using available insect hormones and carry out

conformational search of the bound receptor by molecular dynamic simulations.

- Use NMR restrained molecular dynamics to determine the solution structure of an adipokinetic hormone from *anopheles gambiae*.
- Perform docking calculations using the adipokinetic hormone receptor and hormone from *anopheles gambiae*.

1.7 Questions addressed in the Thesis

1. Are the three dimensional structures of adipokinetic hormone receptor based on rhodopsin and Beta2-adrenergic receptor molecules the same? If not is there any relationship between the structures?

2. What conformational and structural changes occur during molecular dynamic simulations of both models in their true, membrane environment? How well do these changes correlate with experimental results on other GPCRs?

3. How does the location of the binding site in adipokinetic hormone receptor compare to those in rhodopsin and Beta2-adrenergic receptor? How does this position agree with experimental results from other peptide GPCRs?

4. What are the free energies of binding of AKHs to the receptor and is the accessibility of the binding pocket the same for both the rhodopsin-based and the beta2-adrenergic-based structures? If not, to what extent are they different?

5. What conformational and structural changes occur during molecular dynamics of bound beta2-adenergic-based model?

6. Is there any relationship between the inactive state of AKHR and the Beta2-adrenergic-based structure, and between the active state and the rhodopsin-based structure?

1.8 Brief Overview of Chapters

This thesis is made up of seven chapters.

- Chapter one gives the introduction including review of AKHs and their GPCRs, problem statement, aims and objectives and thesis statement.
- Chapter two gives a description of molecular modelling methods, docking and NMR experiments used in this thesis.

Chapters three to six contain the bulk of the thesis. Each chapter gives a summary, experimental procedures, equipments, results and discussion, and conclusions for the major achievements.

- Chapter three focuses on molecular modelling of adipokinetic hormone receptor. The results highlight the similarities and differences between the rhodopsin-based and the Beta2-adrenegic receptor-based models.
- Chapter four describes molecular dynamic studies of the models in a membrane model. Stability, conformational changes and role of the membrane during molecular dynamics of the receptor are given.
- Chapter five concentrates on docking studies and molecular dynamic simulations of the bound receptor. The results present the position of the binding pocket of the receptor as identified by an active hormone, Del_CC, accessibility of the site in both models and the change from open to closed conformation upon activation of the receptor.
- Chapter six gives NMR experiments, conformational search and docking calculations of an active adipokinetic hormone from *Anopheles gambiae*. The structure of the adipokinetic hormone from malaria mosquito and results from the docking calculations are presented.
- Chapter seven gives an overall discussion, conclusion and future work.

1.9 References

1. Holt, R. *et al*. The genome sequence of the malaria mosquito, *Anopheles gambiae*. *Science* **298**, 129 – 149 (2002).
2. Hill, C. A. & Brown, M. R. G – protein coupled receptors in *Anopheles Gambiae. Science* **298**, 176 – 178 (2002).
3. Riehle, M. A., Garczynski, S. F., Crim, J., W., Hill, C. A. & Brown, M. R. Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* **298**, 172-175 (2002).
4. Belmont, M., Cazzamali, G., Williamson, M., Hauser, F. & Grimmelikhuijzen, C. J. P. Identification of four evolutionary related G – protein coupled receptors from the malaria mosquito *Anopheles gambiae*. *Biochem. and Res. Comm.* **344**, 160 – 165 (2006).
5. Ramphul, U., Boase, T., Bass, C., Okedi, L. M., Donnelly, M. J., and Müller, P. Insectiside resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. *Trans. Royal Society of Tropical Medicine and Hygiene* **103**, 1121-1126 (2009).
6. Nwane, P. et al. Trends in DDT and pyrethroid resistance in Anopheles gambiae s.s populations from urban and agro-industrial settings in southern Cameroon. *BMC Infectious Diseases*, **9**, 163 (2009).
7. Nair, M., Jackson, G. E. & Gade, G. Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics. *J. Comp. – aided Mol. Des.***15**, 259 – 270 (2001)**.**
8. Gade, G. & Auerswold, L. Flight substrates and their regulation by a member of the AKH/RPCH family of neuropeptides in cerambycidae. *J. Ins. Phys.* **46**, 1575 – 1584 (2000).
9. Kaufmann, C. & Brown, M. R. Regulation of carbohydrates metabolism and flight performance by a hypertrehalosaemic hormone in the mosquito *Anopheles gambiae*. *J. Ins. Phys.* **54**, 367-377 (2008).
10. Staubli, F. *et al* Molecular identification of the insect adipokinetic hormone receptors. *Proc. Nat. Acad. Sci.* **99**, 6, 3446 – 3451 (2002).
11. Kaufmann, C. & Brown, M. R. Adipokinetic hormones in the malaria mosquito, *Anopheles Gambiae*: Identification and expression of genes for two

peptides and a putative receptor. *Ins. Biochem. and Mol. Biol.* **36**, 466 – 481 (2006).

12. van der Horst, D. J. Insect adipokinetic hormones: release and integration of flight energy metabolisation. *Comp. Biochem. and Phys. Part B* **136**, 217 – 226 (2003)**.**
13. Kobilka, B. G-protein coupled receptor structure and activation. *Biochim Biophys. Acta* **1768**(4), 794-807 (2007).
14. Kroeze, W. K., Sheffler, D., J. & Roth, B. L. G – proteins coupled receptors at a glance. *J. Cell Sci*. **116** (24), 4867 – 4869 (2003)**.**
15. Zhang, Y., DeVries, M. E. & Skolnick, J. Structure modeling of all identified G – protein couple receptors in the human genome. *Comp. Biol* **2**(2), 88-99 (2006).
16. Palcezewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science* **289** (5480), 739 – 45 (2000).
17. Beeley, N. R. A. & Sage, C. GPCRs: An update on structural approaches to drug discovery, *Targets* **2**(1),19-25 (2003).
18. Hansen, K. K., Hauser, F., Cazzamali, G., Williamson, M. & Grimmelikhuijzen C. J. P. Cloning and characterization of the adipokinetic hormone receptor from the cockroach *Periplacenta americana*. *Biochem. Biophys. Res. Comm.* **343**, 638-643 (2006).
19. Kobilka, B. & Schertler, G. F. X. New G-protein coupled receptor crystal structures: Insights and limitations. *Trends Pharmacol Sci* **29**(2), 79-83 (2008).
20. Chelikani, P. *et al.* Role of group conserved residues in the helical core of β2-adrenergic receptor. *PNAS **104***(17), 7027 – 7032 (2007).
21. Ballesterro, J. A., and Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function in G-protein coupled receptors, *Methods Nuerosci,* 25, 366-428 (1994).
22. Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Reviews*, **21**, 90-113 (2000).
23. Fanelli, F. and De Benedetti P. G. Computational Modelling Approaches to Structure-Function Analysis of G Protein-Coupled Receptors. *Chem Rev*, **105**, 3297-3351 (2005).

24. Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell*. Biol, 3, 639-650 (2002).
25. Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2 adrenergic G-protein coupled receptor. *Nature* **450**(7168), 383-387 (2007).
26. Cherezov, V. *et al.* High – resolution crystal structure of an engineered human β2- adrenergic G- Protein coupled Receptor. *Science* **318**, 1358-1264 (2007).
27. Kristiansen, K. Molecular mechanism of ligand binding, signalling, and regulation within the superfamily of G-protein-coupled receptors: molecular modelling and mutagenesis approaches to receptor structure and function. *Pharmac Therapeutics* **103**, 21-80 (2004).
28. Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**(5854), 1266- 1273 (2007).
29. Lundstrom, K. Structural genomics of GPCRs. *Trends Biotech*. **23**(2), 103-108 (2005).
30. Bissantz, C., Logcan, A., and Rognan, D. High-Throughput Modelling of Human G-Protein Coupled Receptors: Amino acid Sequence-alignment, Three Model building and Receptor Library Screening, *J. Chem,. Inf. Comput. Sci.* 44, 1162-1176 (2004).
31. Mouillac, B., Caron, M., Bonin, H., Dennis, M., and Bouvier, M. Agonist-modulated palmitoylation of B2-adrenergic receptor in sf9 cells. *j. Biol Chem,* 267, 21733-21737 (1992).
32. Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., and Shichida, Y. Functional role of internal water molecules in rhodopsin revealed ny X-ray crstallogarphy. *Proc Natl Acad Sci USA,*99, 5982-5987 (2002).
33. Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., and Skenkamp, R. E. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein coupled receptors (GPCRs), *Biochemistry*, 40, 7761-7772 (2001).
34. Stenkemp, R. E., Teller, D. C. & Palczewski, P. Crystal structure of Rhodopsin: A G-protein coupled receptor, *Chem. Biochem*. **3**, 963 – 967 (2002).
35. Birnbaumer, L., Abramowitz, J., and Brown, A. M. Receptor-effector coupling by G-proteins. *Biochem. Biophys Acta,* **1031**, 163-224 (1990).

36. Bhattacharya, S., Hall, S. E., Li, H., & Vaidehi, N. Ligand stabilization states of human β2 - adrenergic receptor: Insight into G- protein coupled receptor activation. *Biophys.* **94**, 2027-2042 (2008).
37. Drews, J. Drug discovery: a historical perspective. *Science*, 287, 1960-1964 (2000).
38. Hopkins, A. L., and Groom, C. R. The druggable genome. *Nat Rev Drug Discov,* **1**, 727-730 (2002).
39. Koloakowski Jr., L. F. GCRDb: a G protein-coupled receptor data-base. *Recept Channel*, **2**, 1-7 (1994).
40. Filipek, S., Stenkamp, R. E., Teller, D. C., Palczewski, K. G protein-coupled receptor rhodopsin: a prospectus. *Annu Rev Physiol*, 65, 851-879 (2003).
41. Tota, M. R., and Strader, C. D. Characterisation of the binding domain of the beta-adrenergic receptor with the fluorescent antagonist carazolol. Evidence for a buried ligand binding site, *J. Biol Chem*, **265**, 16891-16897 (1990).
42. Strader, C. D., Gaffney, T., Sugg, E. E., Rios Candelore, M., Keys, R., Patchett, A. A. and Dixon, R. A. F. Allele-specific activation of genetically engineered receptors. *J. Biol Chem*, **266**, 5-8 (1997).
43. Liapakis, G., Ballesteros, J. A., Papachristou, S., Chan, W. C., Chen, X., and Javitch, J. A. The forgotten serine. A critical role for ser-203 5.42 in ligand binding to and activation of the beta 2-adrenergic receptor, *J. Biol Chem,* **275**, 37779-37788 (2000).
44. Jiang, Q., Guo, D., Lee, B. X., Van Rhee, A. M., Kim, Y. C., Nicholas, R. A., Schachter, J. B., Harden, T. K., and Jacobson, K. A. A mutational analysis of residues essential for ligand recognition at the human P2Y1 receptor. *Mol Pharmacol*, **52**, 499-507 (1997a).
45. Jiang, Q., Lee, B. X., Glashofer, M., Van Rhee, A. M., and Jacobson, K. A. Mutagenesis reveals structure-activity parallels between human A2A adenosine receptors and biogenic amine G protein-coupled receptors, *J Med Chem*, **40**, 2588-2595 (1997b).
46. Perlman, J. H., Laakkonen, L.J., Guarnieri, F., Osman, R., and Gershengorn, M. C. A refined model of the thyrotropin-releasing hormone (TRH) receptor binding pocket. Exprimental analysis and energy minimisation of the complex betweem TRH and TRH receptor, *Biochemistry*, **35**, 7643-7650 (1996).

47. Kim, J. H., Jiang, Q. L., Glashofer, M., Yehle, S., Wess, J., and Jacobson, K. A. Glutamate residues in the second extracellular loop of the human A_{2a} adenosine receptor are required for ligand recognition. *Mol Pharmacol*, **49**, 683-691 (1996).
48. Kroeze, W. K., Kristiansen, K., and Roth, B. L. Molecular biology of serotonin receptors structure and function at the molecular level, *Curr Top Med Chem*, **2**, 507-528 (2002).
49. Schadel, S. A., Heck, M., Maretzki, D., Filipek, S., Teller, D. C., Palczewski, K., and Hofmann, K. P. Ligand channelling within a G protein-coupled receptor. The entry and exit of retinal's in native opsin. *J Biol chem.* **278**, 24896-24903 (2003)
50. Yokota, Y., Akazawa, C., Ohkubo, H., and Nakanishi, S., Delineation of structural domains involved in subtype specificity of tachykinin receptors through chimeric formation of substance P/substance K receptors, *EMBO J*, **11**, 3585-3591 (1992).
51. Gether, U., Johansen, T. E., and Schwartz, T. W. Chimeric NK-1 (substance P)/NK-3 (Neurokinin-B) receptors: identification of domains determining the binding specificity of tachykinin agonists, *J Biol Chem*, **268**, 7893-7898 (1993).
52. Fong, T. M., Yu, H., Huang, R. R. C., and Strader, C. D. The extracelluar domain of neurokinin-1 receptor is required for high-affinity binding of peptides. *Biochem*, **31**, 11806-11811 (1992).
53. Huabg, R. R., Vcario, P. P., Strader, C. D., and Fong, T. M. Identification of residues involved in ligand binding to the neurokinin-2 receptor, *Biochem*, **34**, 10048-10055 (1995).
54. Bhogal, N., Donnelly, D., and Findlay, J. B. The ligand binding site of neurokinin 2 receptor. Site-directed mutagenesis and identification of neurokinin A binding residues in the human neurokinin2 receptor, *J Biol Chem*, **269**, 27269-27274 (1994).
55. Hjorth, S. A., Schambye, H. T., Greenlee, W. L., and Schwartz, T. W. Identification of peptide binding residues in the extracellular domains of the AT1 receptor, *J Biol Chem,* **269**, 30953-30959 (1994).

56. Walker, P., Munoz, M., Martinez, R., and Peitsch, M. C. Acidic residues in extracellular loops of the human-y1 neuropeptide-y receptor are essentila for ligand binding, *J Biol Chem,* **269**, 2863-2869 (1994).
57. Heerding, J. N., Yee, D. K., Jacobs, S. L., and Fluharty, S. J. Mutational analysisi of the angiotensin II type 2 receptor: contribution of conserved extracellula amino acids, *regul pept*, **72**, 92-103 (1997).
58. Leong, S. sr, Kabakoff, R. C., and Hebert, C. Complete mutagenesis of the extracellular domain of interleukin-8 (IL-8) type A receptor identifies charged residues mediating IL-8 binding and signal transduction, *J Biol Chem,* **269**, 19343-19348 (1994).
59. Cotte, N., Balestre, M. N., Phalipou, S., Hibert, M., Manning, M., Barberis, C., and Mouillac, B. Identification of residues responsible for selective binding of peptide anatagonists and agonists in the V2 vasopressin receptor, *J Biol Chem,* **273**, 29462-29468 (1998).
60. Davidson, J. S., Assafa, D., Pawson, A., Hapgood, J., Becker, I., Flanagan, C., Roeske, R., and Millar, R. Irreversible activation of gonadotropin-releasing hormone receptor by photoaffinity cross-linking: localisation of attachment site to Cys residues in N-terminal segment, *Biochem*, **36**, 12881-12889 (1997).
61. Han, B., and Tashjian Jr., A. H., Importance of the extracellular domains for ligand binding in the thyrotropin-releasing hormone receptor, *Mol Endocrinol*, **9**, 1708-1719 (1995).
62. Kannoa, T., *et al,* Different binding sites for the neuropeptide Y Y 1 antagonists 1229U91 and J-194870 on human Y1 receptors, *Peptides,* **22**, 405-413 (2001).
63. Valiquette, M., Vu, H. K., Yue, S. Y., Wahlestedt, C., and Walker, P. Involvement of Trp-284, Val-297 and Val-297 of the human delta-opioid receptor in binidnig of delta-selective ligands. *J Biol Chem*, **271**, 18789-18796 (1996).
64. Lee, J. A., *et al*. Lysine 182 of endothelin B receptor modulates agonist selectivity and antagonis affinity: evidence for the overlap of peptide and non-peptide ligand binding sites, *Biochem*, **33**, 14543-14549 (1994).

65. Stern, P. S., Yu, L., Choi, M-Y., Russell, J. A., Becker, L., and Rafaeli, A. Molecular modeling of the binding of pheromone biosynthesis activating neuropeptide to its receptor, *J Insec Physiol*, **53**, 803-818 (2007).
66. Choi, M-Y., Fuerst, E-J., Rafaeli, A., and Jurenka, R. Role of extracellular domains in PBAN/pyrokinin GPCRs from insects using chimera receptors, *Insec Biochem and Mol Biol*, **37**, 296-306 (2007).
67. Rafaeli, A. Pheromone biosynthesis activating neuropeptide (PBAN): Regulatory role and mode of action. *Gen. Comp. Endocri.* **162**, 69-78 (2009).
68. Samama, P ., Cotecchia, S., Costa, T., and Lefkowitz, R. J. A mutation-induced activated state of the β_2-adrenergic receptor. *J Biol Chem*, **268**, 4625-4636 (1993).
69. Seifert, R., and Wenzel-Seifert, K. Constitutive activity of G-protein-coupled receptors: cause of disease and commom property of wild-type receptors, *Nauny-Schmiedeberg's Arch Pharmacol*, **366**, 381-416 (2002).
70. Broeck, J. V., Insect G Protein-Coupled Receptors and signal Transduction. *Arch. Insect Biochem. Physiol*, **48**, 1-12 (2001).
71. Pierce, K. L., Premont, R. T., and Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell*. Biol, **3**, 639-650 (2002).
72. Cabrera-Vera, T. M., Vanhauwe, J., Thomas, T. O., Medkova, M., Preininger, A., Mazzoni, M. R., and Hamm, H. E. Insights into G protein structure, function, and regulation. *Endocri Rev,* **24**, 765-781 (2003).
73. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C., SWISS-MODEL: an au**t**oma**t**ed pro**t**ein homology-modeling server. *Nucleic Acids Research*, **31**(13), 3381-3385 (2003).
74. Trabanino, R. J. *et al.* First principles prediction of the structure and function of G- protein coupled receptors: Validation for bovine rhodopsin. *J. Biophys.* **86**, 194 – 1921 (2004).
75. Orry, A.J.W. and Wallace, B.A. Modeling and docking the endothelin G-protein coupled receptor. *J. Biophys.* **79**, 3083 – 3094 (2000)**.**
76. Krystek, S. R., Jr. *et al.* Modeling and active Site refinement for G-protein coupled receptors: Application to β2- adrenergic receptors. *J.Comp. Aided Mol. Design* **20**, 463 – 470 (2006).

77. Du, P., *et al*. Modeling the G-protein-coupled neuropeptide Y Y1 receptor agonist and antagonist binding sites. *Protein Engineering*, **10**, 109-117 (1997).
78. Singh, S., Malik, B.K. & Sharma, D. K. Targeting HIV-1 through molecular modeling and docking Studies of CXCR4: Leads for therapeutic development. *Chem Biol Drug Design* **69**, 191-203 (2007)**.**
79. Pogozheva, I. D., Przydzial, M. J. & Mosberg, H.I. Homology modeling of opioid receptor – ligand complexes using experimental constraints. *AAPS* **7**(2), 43, E434 –E448 (2005).
80. Baker, D. and Sali, A. Protein structure prediction and structure genomics. *Science* **294**, 93 – 96 (2001).
81. Elling, C. E. and Schwartz T.W., Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.* **15**, 6213-6219 (1996).
82. Swaminath, G., Deupi, X., Lee, T. W., Zhu, W., Thian, F. S., Kobilka, T. S., and Kobilka, B. Probing the B2 adrenoceptor binding site with catechol reveals differences in binding and activation by agonist and partial agonists. *J. Biol. Chem.* **280**, 22165-22171 (2005).
83. Ravina, C. G., Seda, M. A., Pinto, F. M., Fernandez-Sanchez, M., Pintado, O. C., and Candenas, M. L. Characterization of tachykinin receptors in human sperm. *Fertility and Sterility*, **88**, 362 (2007).
84. Springael, J-Y., de Poorter, C., Deupi, X., Durme, J. V, Pardo, L., and Parmentier, M. The activation mechanism of chemokine receptor CCR5 involves common structural changes but a different network of interhelical interactions relative to rhodopsin**.** *Cellular Signalling* **19** 1446-1456, (2007).

CHAPTER TWO

Molecular Modelling and Docking for GPCRs

2.1 Summary

Since experimental determination of the three dimensional structures of GPCRs has proved to be very difficult, molecular modeling has been a very powerful tool in designing drugs that target this family of proteins. Chemical properties of molecules are obtained by using quantum and molecular mechanics. In molecular mechanics, homology modelling and molecular dynamics are commonly used to produce good quality models of GPCRs that have been used and are currently used in structure-based drug design. In conjunction with the models, information from docking calculations has opened a way for the production of most used pharmaceuticals that target the GPCRs. In this work, homology modelling, molecular dynamics and docking studies have been used to determine and characterise the 3D structure of an adipokinetic hormone receptor from the malaria mosquito. An overview of these methods is presented in this chapter. Details of experimental procedures and materials used are given in chapter three to six.

2.2 Molecular Modelling

Molecular modelling gives numerical representations of molecular structures and simulates their behaviour with equations of quantum and classical physics [1]. Therefore, programmes used in molecular modelling generate and present molecular data ranging from bond lengths, bond angles, torsion angles, energies, electronic properties, volumes, viscosity and spectroscopic properties [1,2]. Thus, the method is used to design new materials for which an accurate prediction of physical properties of a realistic system is required [3]. Using *ab in*itio protein structure prediction

methods and molecular mechanics, in molecular modeling, chemical properties of molecular structures are studied [2,3,4].

Ab initio methods require a long simulation period and a large amount of disk space to store intermediate data files, it is limited to compounds with less than twenty atoms [1,2,3]. When applied to modelling of large GPCRs molecules, such as D2 dopamine, β_2-adrenergic and 5-HTs serotonin receptors, *ab initio* methods gave results inconsistent with the rhodopsin structure [4]. The helices were predicted to be having a clockwise or nonsequential arrangement, whereas Class A receptor helices have an anticlockwise and sequential arrangement from helix one to seven [5,6,7,8].

In contrast, molecular mechanics calculations are fast and efficient, and are used to simulate systems containing thousand of atoms [1,2,4]. The method is used to compute molecular properties that include geometry, rotational barriers, vibrational spectra, heats of formation and the relative stability of conformers [1,2]. Molecular mechanics relies on the laws of Newtonian physics and experimentally derived parameters to calculate geometry as a function of potential energy based on a force field (equation 2.1).

$$\mathbf{E}_{pot} = \sum \mathbf{E}_{bnd} + \sum \mathbf{E}_{ang} + \sum \mathbf{E}_{tor} + \sum \mathbf{E}_{oop} + \sum \mathbf{E}_{vdw} + \sum \mathbf{E}_{el} \qquad 2.1$$

Where $\mathbf{E}_{pot}$ = total steric energy

$\mathbf{E}_{bnd}$ = the energy resulting from deforming a bond length from its natural value.

$\mathbf{E}_{ang}$ = the energy due to deformation of an angle from its natural value

$\mathbf{E}_{tor}$ = energy from deforming the torsion or dihedral angle

$\mathbf{E}_{oop}$ = the out-of-plane bending component of the steric energy

$\mathbf{E}_{vdw}$ = energy arising from van der Waals non-bonded interactions, and

$\mathbf{E}_{el}$ = the energy arising from coulombic forces.

When simulations based on physical properties of atomic interactions fail due to complexity of the system, homology or comparative molecular modelling methods, based on a similarity analysis of known structural and chemical data is performed [3]. These methods have been widely applied in molecular modelling of complex and highly flexible GPCRs based on known crystal structures.

2. 2. 1 Computer Simulation Methods

Molecular dynamic (MD) and Monte Carlo simulations are the most popular computer simulation methods that can be employed in molecular modelling when measurements are made on large samples that contain extremely large numbers of atoms or molecules [2,3]. The methods enable prediction of their properties using techniques that consider small replications of the macroscopic system with manageable numbers of atoms or molecules. During simulation, representative configurations are generated in such a way that accurate values of structural and thermodynamic properties can be obtained with a feasible amount of computation force [2].

In Monte Carlo simulations, there is no temporal relationship between successive configurations and the outcome of each trial depends only upon its immediate antecedent [2]. Total energy is determined directly from the potential energy and simulations do not require computation forces. In comparison to MD, the simulation method does not yield significant better statistics within a specified amount of computer time [3].

On the other hand, MD simulations predict the configuration of the system at any time in the future or at any time in the past and give information about the time dependence of properties of the system such as bond lengths, bond angles, proper and

improper dihedral angles and non-bonded interactions [2,3]. MD has a kinetic contribution to the total energy and is performed under conditions of constant number of particles (N), volume (V), and energy (E). The method can be adapted to simulate at constant N, pressure (P) and temperature (T), the NPT ensemble. MD has commonly been used in simulations of GPCR models and remains an important tool for studying their stability and flexibility in water [9] and for visualisation of conformational changes in a membrane [10,11].

2.2.2 Molecular dynamics simulations

MD simulations involve the integration of Newton's equations of motion for a number of atoms interacting under the influence of a force (F) (equations 2.2 and 2.3) [2,3].

$$\mathbf{F} = \mathbf{ma} \qquad 2.2$$

Where **m** is mass of an atom and **a** is acceleration of the atom,

$$\mathbf{d}^2x_i/\mathbf{d}t^2 = \mathbf{F}x_i/\mathbf{m}_i \qquad 2.3$$

Where, $\mathbf{m}_i$ = mass of an atom along one coordinate x_i

$\mathbf{F}x_i$ = Force on the atom in that direction.

Integration of the equations of motion yields a trajectory that describes how the positions, velocities and accelerations of the atom in the system vary with time [2,3].

Verlet Algorithm

The Verlet algorithm [12] is the most widely used method for integrating the equation of motion in MD simulations [2]. The method uses the positions (**r**) and accelerations, (**a**), at time, t, and the positions from the previous step, $\mathbf{r}(t - \delta t)$, to calculate the new positions at $t + \delta t$, $\mathbf{r}(t + \delta t)$, where δt is the time step. In the final

Verlet integration algorithm, equation 2.4, velocities (**v**) do not appear but can be calculated from either equation 2.5 or estimated at half-step (t + ½δ*t*) (equation 2.6).

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t) \quad 2.4$$

$$\mathbf{v}(t) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t)]/2\,\delta t \quad 2.5$$

$$\mathbf{v}(t + \tfrac{1}{2}\delta t) = [\mathbf{r}(t + \delta t) - \mathbf{r}(t)]/\delta t \quad 2.6$$

Leap-frog algorithms

The Leap-frog algorithm [13], is a variation of the Verlet algorithm that uses the relationships shown in equations 2.7 and 2.8 [2].

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\, \mathbf{v}(t + \tfrac{1}{2}\delta t) \quad 2.7$$

$$\mathbf{v}(t + \tfrac{1}{2}\delta t) = \mathbf{r}(t - \tfrac{1}{2}\delta t) + \delta t \boldsymbol{a}(t) \quad 2.8$$

The velocities at time, *t* can be calculated from

$$\mathbf{v}(t) = \tfrac{1}{2}[\mathbf{v}(t + \tfrac{1}{2}\delta t) + \mathbf{v}(t - \tfrac{1}{2}\delta t)] \quad 2.9$$

There is a 'leap-frog' of velocities over the positions to give their value at $t + \frac{1}{2}\delta t$. The positions then leap over the velocities to give their new values at $t + \delta t$, ready for the velocities at $t + 3/2\delta t$, and so on [2,3]. The inclusion of the velocity and lack of calculation of the differences in large numbers make the leap-frog algorithm advantageous over the Verlet algorithm. However, it has the disadvantage that the positions and velocities are not synchronised. The integration of equations in GROMACS is performed by the leap-frog algorithm [3].

Choice of time step

The time step used during MD simulations is important and if an inappropriate value is used this leads to artefacts [2]. Specifically, with very small time steps the trajectory will cover only a limited proportion of phase space but too large a time step

results in instability in the integration algorithm due to high energy overlaps between atoms [2,3]. The stabilities would lead to a violation of energy and linear momentum conservation resulting in a failure because of numerical overflow. Longer time step can be used when the properties of the bonds are constrained to their equilibrium values while still allowing the rest of the degrees of freedom to vary under the intramolecular and intermolecular forces available [2]. In GROMACS, a timestep of 1fs or 2fs is normally used [3].

Periodic boundary conditions

Periodic boundary conditions (PBC) help to minimise edge effects in a finite system [2,3,14]. The conditions are applied by putting the system to be simulated into a space-filling box surrounded by translated copies of itself (Figure 2.1)

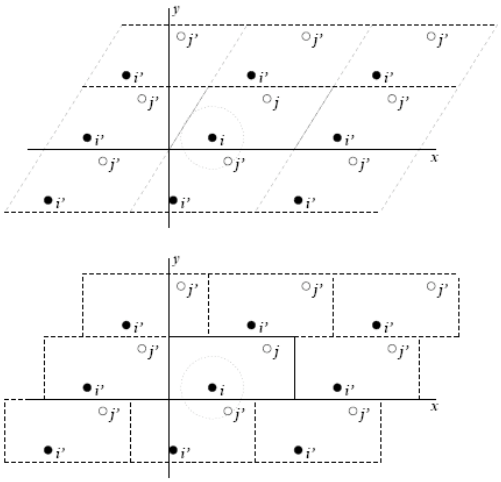


Figure 2.1: Periodic boundary conditions in two dimensions [3].

Therefore, the error of the periodic conditions replaces the error caused by unwanted boundaries in an isolated cluster. Hypothetically, any cell shape such as, cube, hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the elongated dodecahedron, can be used provided it fills all the space by translation operations of the central box in three dimensions. Typically, simulation using a

spherical cell requires fewer solvent particles than a cubic cell [2, 14,15,16]. The triclinic unit cell is the most general space-filling unit cell for it comprises all possible space-filling shapes [14,16].

Simulated Annealing

To solve the problem of local minimum in predicting the 3D structure of AKHR, simulated annealing was used during conformational space search [2,3,17,18]. Simulated annealing is a process in which the temperature of a system is reduced gradually with the aim of obtaining a conformation that would correspond to the global minimum of the potential energy [2]. Using Molecular dynamics, the system is allowed to reach 'thermal equilibrium' at a given temperature. At high temperatures, the molecules can overcome high-energy barriers [2]. As the temperature falls gradually, the conformations of low energy become more probable according to Boltzmann distribution and at absolute zero, the system should occupy the lowest-energy state giving the global minimum energy conformation. However, practically this is difficult to achieve, so a collection of runs is used for the global minimum to correspond to the true minimum [2].

2.3 Molecular modelling of AKHR

To facilitate drug design using GPCRs, homology (comparative) modelling, and molecular dynamics (MD) simulations have been increasingly successful in predicting the 3D structures of this important protein family[19,20,21,22]. Almost all known drugs that target GPCRs have been designed from molecular models of their receptors, for example, Neuropeptide Y receptors (NPY) [23], CXCR4 [22], endothelin receptors [20], opioid receptors [21], beta2-adrenergic receptors [24, 25,26], gonadotropin releasing hormone (GnRH) receptor [27], tachykinin receptors [7,8,28], chemokine receptors [29] and thyrotropin-releasing hormone receptors [30].

Recently, molecular models have been successfully used in characterisations of insect peptide GPCRs like pheromone biosynthesis activating neuropeptide (PBAN) receptors [31,32].

In this thesis, homology modelling was used to build the 7TMs of AKHR. For optimisation of the loops and the whole AKHR molecule, MD simulations were performed in vacuum, water and membrane using the Groningen Machine for Chemical Simulation (GROMACS) software package [3] available at http//www.gromacs.org.

2.3.1 Homology Modelling of 7TMs

Homology modelling produces an atomic resolution model of a protein from its amino acid sequence (target), based on its alignment to one or more related known protein structures (templates) [33,34]. The success of homology modelling is largely dependent on the evolutionary relationship between the target and the template proteins. It has been shown that GPCR homology models are accurate enough to identify known antagonists seeded in any 'drug-like' library [34,35].

Before 2007, the only available templates for building 3D models of structurally diverse GPCRs, were the crystal structures of rhodopsin [5,36,37]. Rhodopsin is homologous to most GPCRs and its helices have an anticlockwise arrangement [5]. The arrangement of helices of Class A receptors was previously evidenced by application of protein engineering [38,39] and confirmed by the crystal structures of rhodopsin [5,36,37]. Therefore, most GPCR models have been constructed with the 7TMs in an anticlockwise arrangement when viewed from the extracellular side [40]. The recently published crystal structures of the human β_2AR

[41], provided an alternative templates that could be used in homology modeling of Class A GPCRs for diffusible ligands such as AKH [42].

In this work, the recently published primary sequence of AKHR [43,44] was used to build the 3D structure of the receptor. Since the helical region is the core of GPCRs and plays an important role in ligand binding and receptor activation, homology modelling was applied to model the transmembrane region of AKHR, using crystal structures of both rhodopsin and β_2AR receptor as templates [4-6,19,23,36,40,41,45]. To predict the helical regions, programmes such as PSIPRED and the MEMSAT3 [46,47] were used.

2.3.1.1 Homology modelling protocol

The first step in homology modelling involves identifying templates whose sequence is closely related to the sequence of the target, from the PDB [2,33,34,35]. Choice of template is important given that models produced are usually more similar to the template used rather than the native structure of the sequence of interest [19]. Since AKHR belongs to the same receptor family as rhodopsin and β_2AR, both crystal structures were used as templates for building its 7TMs.

Secondly, pairwise sequence alignments of the target sequence and the template is performed using programmes such as the Swiss –Model workshop [34,35] and ClustalW available on http://www2.edi.ac.uk/clustalW [48]. Sequence alignment identifies and computes the matches between template and target and is the most crucial step in homology modelling [2,19,49]. In the case of GPCRs, the highly conserved residues found in the core of the receptors are matched. Percent sequence identities and similarities between the target and the templates can be calculated by dividing the total number of identical/similar residues by the number of residues in

the shorter sequence [47,49,50]. Accuracy of homology modelling is related to the percent sequence identity [33], with 50% sequence identity of target to template giving highly accurate models, 30-50% producing medium accuracy and low comparable models are obtained from < 30% identity. However, it has been shown that proteins with similar sequences tend to have similar 3D structures [2]. In general, GPCRs have below 40% sequence identity with their templates, so modelling of the 7TMs is based on the high percent sequence similarity and the similar 3D structure of this domain [2,45].

Having obtained an acceptable sequence alignment, the known 3D structure and the alignment are used to build the model using tools like, MODELLER [51], WHATIF [52] and Swiss-Model workshop [34]. The final stage involves analysis of the model by considering, for example, the Ramachandran plots for psi and phi angles of each residue [3,55,56,57], in order to determine whether the amino acid residues occupy energetically favourable regions [2]. Where available the results are compared with experimental results.

2.3.2 MD simulations of AKHR

As with most GPCR models, the loop regions of AKHR were built separately, joined to their respective helices and then optimised by performing molecular dynamic (MD) simulations [23,20,56,57]. To search for membrane associated conformations of the free and unbound AKHR molecules MD simulation was performed in a membrane model.

2.3.2.1 MD simulation protocol

To set up and run MD simulation, firstly, the initial configuration for the system should be established either from experimental data, theoretical model or from a

combination of the two [1,2,3]. The configuration chosen should be close to the state that it is desired to simulate. A description of the force field for the system is given in the topology file [3]. A variety of force fields such as the GROMACS, GROMOS and OPLS/AA [3] could be applied in GROMACS. The OPLS/AA all atom-force field that includes the amino acid dihedrals [58], was used in molecular dynamic simulations of AKHR. The force could include contributions of the various terms in the force filed such as bonds, angles, torsion and non-bonded interactions.

The potential functions used in most force fields are divided into non-bonded, bonded interactions and restraints [3]. The non-bonded (the Lennard-Jones and Coulomb) interactions are computed based on a neighbor list in which exclusions are already removed [3]. On the otherhand, the bonded interactions (covalent bond-stretching, angle-bending, improper and proper dihedrals) and restraints such as (position, angle distance and orientation restraints) are calculated based on fixed lists. Initial velocities are assigned to the atoms by random selection from a Maxwell-Boltzmann distribution at a temperature of interest [2,3]. The initial velocities are adjusted so that the total momentum of the system was zero. Differentiation of the potential function calculated the force acting on each atom at each step during the simulation.

Secondly, energy minimization of the system is performed to ensure that the initial configuration did not contain high-energy interactions as these may cause instabilities [2,3]. In GROMACS, this is achieved by using the steepest descent, conjugate gradients or the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) algorithms [3]. Steepest descent is robust and easy to implement whilst conjugate gradient is slower in the beginning of minimization but becomes more efficient closer to the energy minimum [3]. On the other hand, l-bfgs algorithm

convergs faster and was used in most energy minimizations of AKHR.

Thirdly, the system is equilibrated to enable it to evolve from the initial configuration to reach equilibrium [2,3]. During these stages thermodynamic and structural properties like kinetic, potential and total energies, velocities, temperatures and pressure, are monitored until stability is achieved.

The production stage involves calculation of simple properties of the system [2,3]. For this stage molecular dynamics parameter file containing all parameters to be used during dynamics are the input file [3] and at regular intervals, the atomic coordinates of the system are recorded. To include knowledge from experimental data and for imposing restraints on the motion of the system, distance and position restraints were used in MD simulation of AKHR [3].

Position restraints

Position restraints are used to restrain particles to fixed reference positions, $\mathbf{R}_i$. The solvent and any counter ions are allowed to evolve whilst the protein atoms remained at fixed positions, [2,3]. Position restrained MD should be sufficient to allow the solvent to completely readjust to the potential of the solute and soak the protein. The length of this solvent equilibration stage should be longer than the relaxation time of the solvent, for example, 10ps for water [2,3]. In GROMACS, position restraint potential ($\mathbf{V}_{pr}$) is calculated from equation 2.9.

$$\mathbf{V}_{pr}(\mathbf{r}_i) = 1/2\ k_{pr} |\mathbf{r}_i - \mathbf{R}_i|^2 \qquad 2.9$$

Distance restraints

Distance restraints add a penalty to the potential when the distance between specified pairs of atoms exceeds a threshold value [3]. The restraints are used to impose experimental restraints, as from nuclear magnetic resonance (NMR) experiments, on the motion of the system. Thus, MD simulations of AKHR were used for structure

refinement using NMR data [3] and measured interhelical distances. The potential form for distance restraints is normally quadratic below a specified lower bound and between two specified upper bounds and linear beyond the largest bound [3] as shown in figure 2.2.

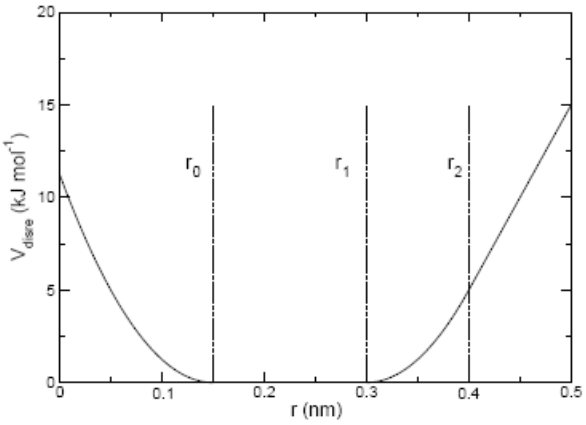


Figure 2.2: Distance restraints potentials [3]

2.3.2.2 MD simulations of AKHR in a membrane

GPCRs exhibit very complex and dynamic structural properties due to their high flexibility [10]. Vacuum MD simulations of most models of GPCRs have generated models currently used in drug design but this does not represent the true environment of the receptor. To represent the physiological conditions of these GPCRs, a number of membrane-embedded human GPCRs models have recently been published [10,11,59-61]. Phospholipid bilayers represent the fundamental structures of most biomembranes [62]. Periodic lipid bilayers consisting of, for example, 1-palmitoyl-2-oleoyl-sn-glcero-3-phosphocholine (POPC) and dipalmitoyl phosphatidyl choline molecules (DPPC), solvated with water, have been simulated and equilibrated [59]. POPC and DPPC are commonly used in MD simulation of GPCRs [61]. To that effect, MD simulations of both free and bound AKHR embedded in a pre-equilibrated POPC membrane model were performed.

2.3.2.3 Analysis of the MD simulation trajectory

In this thesis, analysis of the trajectory obtained from simulations of the protein, using tools set in GROMACS software package [3] was the last stage. These included cluster analysis to find similar conformation within a given root mean square deviation [2]. From the most popular cluster, the lowest energy conformation gave the final structure of the molecule. The root mean square deviations (rmsd), root mean square fluctuations (rmsf), radius of gyration (rg), secondary structure analysis [3,9-11,63] were applied to determine the stability, structural and conformational changes of the receptor molecules. Molecular visualisation was carried out using PYMOL [64], MOLMOL [65], VMD [66] and PMV [67].

Root mean square deviation

The root mean square deviation of the backbone atoms (N, Ca, C) in a molecule with respect to a reference structure, e. g. the initial structure or the crystal structure, was calculated by the least-square fitting of the backbone atom positions to the reference molecule [3,9]. Subsequent calculation of rmsd was performed using the equation 2.10,

$$\mathrm{RMSD}(t_1,t_2)=[1/M\sum_{i=1}^{N} m_i\|\boldsymbol{r}_i(t_1)-\boldsymbol{r}_i\|^2]^{1/2} \qquad 2.10$$

Where M is the sum of the masses of the atoms, m_i, and $\boldsymbol{r}_i(t)$ is the position of the atomi at time t.

Root mean square fluctuation

The root mean square fluctuations (rmsf), is the standard deviation of the Ca atoms of the protein during simulations, after fitting to the initial structure [3]. Rmsf values give an indication of the degree of mobility and flexibility of the Cα atoms with time.

Radius of gyration

The radius of gyration (Rg) is a measure of the compactness of the molecule during simulation [3,68,69] and is calculated as follows,

$$Rg = (\sum_i ||\boldsymbol{r}_i||^2 m_i / \sum_i m_i)^{1/2} \qquad 2.11$$

Where m_i is the mass of the atom, i and $\boldsymbol{r}_i$ the position of the atom with respect to the centre of mass of the molecule.

Secondary structure analysis

To look at the structural changes during dynamics, the programme DSSP [70] is used to determine the secondary structure variations of each residue. The DSSP programme analyses the secondary structure of the protein by a pattern-recognition process of hydrogen-bonding patterns and geometrical features extracted from atomic coordinates. [3,70]. For each residue, secondary structure variations are plotted against time as shown in Figure 2.3, where each colour represents a specific structural feature.

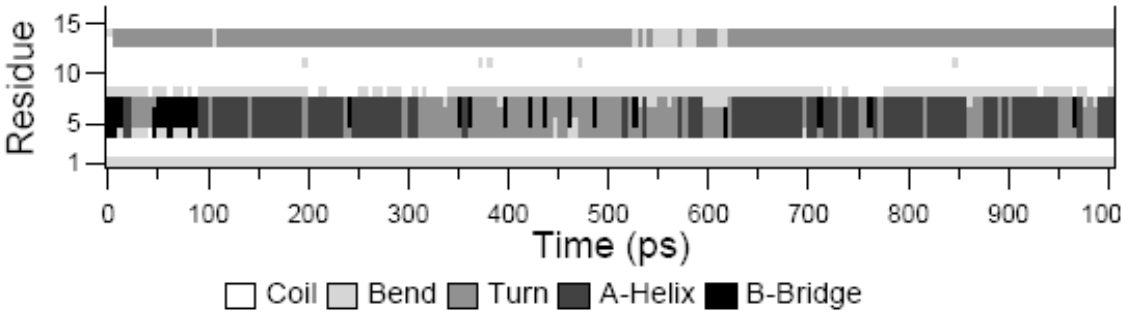


Figure 2.3: Variation of the secondary structure elements of a peptide with time [3].

2.4 Molecular Docking

In the absence of a crystal structure of ligand/receptor complexes, molecular docking is used to provide information on the position, interactions and preferred interactions between ligands and their receptors [71]. Molecular docking methods place (dock) a ligand in the binding pocket of a receptor [71-73] and therefore the structure of the ligand/receptor complex and binding modes are predicted [2]. Structural information obtained proved to be very important to the Medicinal Chemist and has been developed to implement structure-based drug design [2,71-74]. Since the 3D structures of most GPCRs are not known, their molecular models are used in molecular docking leading to identification of a variety of ligands currently used in pharmaceutical industry [20-22,75-77]. Therefore, application of molecular docking to adipokinetic hormone receptor model could lead to identification of ligands with a high affinity for the receptor. These compounds can be used in the production of agonist and antagonists with high specificity for the protein that could block its binding pocket and hinder generation of energy during mosquito flight.

Several docking methods have been developed and these differ mainly in how they position the ligand, explore conformational space, represent ligand/receptor interactions and in scoring (binding affinity estimation) methods [71,78-80]. The first group of algorithms included DOCK [81,82], FTDOCK [83], QSDOCK [84], FlexX [85,86] which try to fit the ligand into the binding pocket by geometrically, chemically and energetically matching [79]. On the other hand, GOLD [87,88], DARWIN [89] and AUTODOCK [90] apply genetic algorithms to search for the global minimum of the energy in the continous conformational space of the ligand [2,78-80].

2.4.1 Docking protocol

Docking procedures involve firstly, identification of the binding site in a target molecule. For previously defined binding sites the search space is restricted to a relatively small part of the protein. When no information on the binding site is available, putative sites could be identified by cavity detection programmes [91,92] and blind docking [93]. Blind docking allows the ligand to scan the protein surface to identify accessible cavities. By using Autodock, blind docking was successfully perfomed to identify binding pockets of a variety of proteins [93-97].

Secondly, docking calculations involve choosing the most suitable search algorithm that effectively samples the set of possible ligand positions and conformations on the protein molecule [78,79]. A good algorithm should be able to sample regions of the search space close to the correct solution without a complete search of all conformations and orientations [78]. Random search involved generation of a large set of random ligand positions and conformations and the ligands are ranked with a suitable scoring function. The algorithm was mainly used as a starting point for other algorithms [78]. Rigid body docking involved rigid ligand and protein molecules and is commonly used in DOCK [81], FTDOCK [82] and QSDOCK [84] methods. The third search algorithm is simulated annealing which is based on a model of systematic reduction of temperature [98,99], which was commonly used by first versions of Autodock [78]. GOLD [87], Autodock [90] and DARWIN [89] are examples of docking methods that applied genetic algorithms when searching for the position and conformations of the ligand [78]. In genetic algorithms, a population of solutions undergoes mutation and higher order transfomations. The generated solutions experience selection biased towards the most fit among them [78,79].

The third and last component of docking procedures is scoring that estimates the binding free energy (ΔG_b) for the ligand when bound to the receptor [2]. The more negative the ΔG_b, the stronger the intractions between the ligand and the receptor and the higher the binding affinity. Simulation techniques used considered that ΔG_b was written as an additive equation of various components to reflect the various contributions to binding (equation 2.12) [2].

$$\Delta G_b = \Delta G_{solv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \qquad 2.12$$

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

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

ΔG_{int} is the free energy due to specific protein-ligand interactions,

ΔG_{rot} is the free energy loss associated with freezing internal rotations of the ligand and the protein,

$\Delta G_{t/r}$ is the loss in translational and rotational free energy caused by association of two bodies to give the bound complex,

ΔG_{vib} is the free enrgy changes due to vibrational modes.

In addition to identifying the 'true' docking mode of ligands the scoring functions should also be able to rank the ligands [2,78].

2.4.2 Molecular docking of AKHR

Since Autodock has been successfully used in blind docking of ligands to receptors whose binding sites were unknown [93,96,97], Autodock4 and Autogrid4 [90] were used to identify the binding site of AKHR and the binding modes of two active adipokinetic hormones, anoga_akh from *Anopheles gambiae* and Del_CC from the blister beetle [100].

Autodock uses a hybrid of genetic algorithm, the Lamarckian genetic algorithm (LGA), to search for conformational space of the ligand [78,79,90]. The ligand's state is given as a chromosome with strings of real-valued genes describing its position, orientation and conformation [78,79,90]. Placement of the ligand begins by creating a random population of individuals, then specifying a number of generation cycles. Each generation cycle is followed by local search and the solutions are scored using the energy-based scoring function [79,90]. This includes terms accounting for short-range van der Waals and electrostatic interactions, loss of entropy upon ligand binding, hydrogen bonding and solvation [78,79,90].

Identifying the binding pocket and binding of a ligand to a receptor is dynamic with both ligand and protein changing conformation to maximise total free energy during free association [71,101]. However, practically this is not easily achieved, so docking calculations are carried out with rigid protein molecules [2,71,78,79]. To further refine the docking calculations, in this thesis, MD simulations of AKHR/ligand complexes were performed with both ligand and protein flexible.

To circumvent the slow estimation of ΔG_b for large and flexible bioactive ligands like AKHs, Hetenyi et al, 2006 [102] designed a fast and accurate calculation method for binding affinities of such ligands. The method makes use of an ensemble of structures produced by MD simulation and a modified scoring function of Autodock 3.0 [90]. Scoring is an important stage in drug discovery for it helps with identification of the conformation of the ligand closest to the true structure and binding affinities of the ligand to its target [103,104]. Therefore, to efficiently calculate the binding affinities and determine the binding modes of AKHs to AKHR, in this work, the improved, fast scoring protocol described by Hetenyi, 2006 [102] was followed. The results obtained should be vital in structure-based design of high

affinity non-peptides that block the binding pocket of the mosquito adipokinetic hormone receptor and thus lead to production of effective insecticides that would enhance the fight against malaria.

2.5 References

1 Richon, A. B. An Introduction to Molecular modeling. *Mathematech*, **1**, 83-94 (1994).

2 Leach, A. R., Molecular *Modelling: Principles and Applications*, 2nd Edition, Person Education Limited, 522-549 (2001).

3 van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., Berendsen, H. J. C. *Gromacs User Manual, Versions 3.3*. www.gromacs.org (2005).

4 Fanelli, F. and De Benedetti P. G. Computational Modelling Approaches to Structure-Function Analysis of G Protein-Coupled Receptors. *Chem Rev.*, **105**, 3297-3351 (2005).

5 Palczewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science,* **289** (5480), 739 – 45 (2000).

6 Cherezov, V. *et al.* High – resolution crystal structure of an engineered human β2- adrenergic G- Protein coupled Receptor. *Science,* **318**, 1358-1264 (2007).

7 Elling, C. E. and Schwartz T.W., Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.,* **15**, 6213-6219 (1996a).

8 Elling, C. E., and Schwartz T.W. Connectivity and rotation of transmembrane helices in &TM receptors probed by metal-ion site engineering. Regul Pept [abstaract]. **64**, 40 (1996b).

9 van der Spoel , Vogel H. J and Berendsen H.J.C. Molecular Dynamic Simulations of N-terminal Peptides from a Nucleodite Binding Protein, *PROTEINS: Structure, Function and Genetics*, **24**, 450-466 (1996).

10 Dror, R. O., *et al*. Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Nat. Acad. Sci.*, **106**(12), 4689-4694 (2009).

11 Ivetac, A and Sansom, M. S. P. Molecular dynamics simulations and membrane protein structure quality. *J. Eur. Biophys.*, **37**, 403-409 (2008).

12 Verlet, L. Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.,* **159**(1), 98-103 (1967).

13 Hockney, R. W., Potential Calculation and Some Applications. *Methods Comput. Phys.*, **9**, 136-211 (1970).

14 Dzwinel, W., Kitowski, J., and Moscinski, J. 'Checker Board' Periodic Boundary Conditions in Molecular Dynamics Codes. *Molec Simul*, **7**, 171-179 (1991).

15 Adams, D. J., Adams, E. M., Hills, G. J. The computer simulation of polar liquids. *Mol. Phys*., **38**, 387–400 (1979).

16 Bekker, H., Dijkstra, E. J., Renardus, M. K. R., Berendsen, H. J. C. An efficient, box shape independent non-bonded force and virial algorithm for molecular dynamics. *Mol. Sim*., **14**,137–152 (1995).

17 Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. Optimisation by Simulated Annealing. *Science*, **220**, 671-680 (1983).

18 Bertsimas, D., and Tsitsiklis, J. Simulated Annealing. *Stat Sci.*, **8**(1), 10-15 (1993).

19 Zhang, Y., DeVries, M. E. & Skolnick, J. Structure modeling of all identified G – protein couple receptors in the human genome. *Comp. Biol.,* **2**(2), 88-99 (2006).

20 Orry, A.J.W. & Wallace, B.A. Modeling and docking the endothelin G-protein coupled receptor. *J. Biophys.,* **79**, 3083 – 3094 (2000)**.**

21 Pogozheva, I. D., Przydzial, M. J. & Mosberg, H.I. Homology modeling of opioid receptor – ligand complexes using experimental constraints. *AAPS,* **7**(2), 43, E434 –E448 (2005).

22 Singh, S., Malik, B.K. & Sharma, D. K. Targeting HIV-1 through molecular modeling and docking Studies of CXCR4: Leads for therapeutic development. *Chem Biol Drug Design,* **69**, 191-203 (2007)**.**

23 Du, P, *et al*. Modeling the G-protein-coupled neuropeptide Y Y1 receptor agonist and antagonist binding sites. *Protein Engineering*, **10**, 109-117 (1997).

24 Krystek, S. R., Jr. *et al.* Modeling and active Site refinement for G-protein coupled receptors: Application to β2 - adrenergic receptors. *J.Comp. Aided Mol. Design,* **20**, 463 – 470 (2006).

25 Swaminath, G., Deupi, X., Lee, T. W., Zhu, W., Thian, F. S., Kobilka, T. S., and Kobilka, B. Probing the B2 adrenoceptor binding site with catechol reveals differences in binding and activation by agonist and partial agonists. *J. Biol. Chem.,* **280**, 22165-22171 (2005).

26 Yao. X., Prnot, C., Deupi, X., Ratnala, V. R., Swaminath, G., Farrens, D and Kobilka, B. Couplng ligand structure to specific conformational switches in the β2-adrenoceptor. *Nat. Chem. Biol.,* **2**, 417-422 (2006).

27 Sealfon, S., C., Weinstein, H. & Millar, R., P. Molecular mechanisms of ligand interaction with the gonadotropin releasing hormone receptor. *Endocr. Rev.,* **18,** 2, 180-205 (1997).

28 Ravina, C. G., Seda, M. A., Pinto, F. M., Fernandez-Sanchez, M., Pintado, O. C., and Candenas, M. L. Characterization of tachykinin receptors in human sperm. *Fertility and Sterility*, **88**, 362 (2007).

29 Jean-Yves Springael, Cédric de Poorter, Xavier Deupi, Joost Van Durme, Leonardo Pardo and Marc Parmentier. The activation mechanism of chemokine receptor CCR5 involves common structural changes but a different network of interhelical interactions relative to rhodopsin. *Cellular Signalling* **19,** 1446-1456 (2007).

30 Perlman, J. H. *et al*. Interactions between Conserved Residues in Transmembrane Helices 1, 2, and 7 of the Thyrotropin-releasing Hormone Receptor. *J. Bio. Chem.,* **272**(18), 11937-11942 (1997).

31 Rafaeli, A. Pheromone biosynthesis activating neuropeptide (PBAN): Regulatory role and mode of action. *Gen. Comp. Endocri.* **162**, 69-78 (2009).

32 Choi, M-Y., Fuerst, E-J., Rafaeli, A. and Jurenka, R. Role of extracellular domains in PBAN/pyrokinin GPCRs from insects using chimer receptors. *Insec. Biochem and Mol. Bio.,* **37,** 296-306 (2007).

33 Baker, D. & Sali, A. Protein structure prediction and structure genomics. *Science,* **294**, 93 – 96 (2001).

34 Arnold, K., et al 'The Swiss-Model Workspace' a web-based environment for protein homology modelling. *Bioinf.,* **22**(2), 195-201 (2006)

35 Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C., SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31**(13), 3381-3385 (2003).

36 Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K., and Skenkamp, R. E. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein coupled receptors (GPCRs), *Biochemistry*, **40**, 7761-7772 (2001).

37 Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., and Shichida, Y. Functional role of internal water molecules in rhodopsin revealed ny X-ray crstallogarphy. *Proc Natl Acad Sci USA,* **99**, 5982-5987 (2002).

38 Liu, J., Schoneberg, T., Vanrhee, M., and Wess, J. Mutational analysis of the relative orientation of transmembramne helices I and VII in G protein-coupled receptors. *J Biol Chem.*, **270**, 19532-19539 (1995).

39 Mizobe, T., Maze, M., Lam, V., Suryanarayana, S., and Kobilka, B. K. Arrangement of transmembrane domains in adrenergic receptors-Similarity to bacteriorhodopsin. *J Biol Chem.*, **271**, 2387-2389 (1996).

40 Kristiansen, K. Molecular mechanism of ligand binding, signalling, and regulation within the superfamily of G-protein-coupled receptors: molecular modelling and mutagenesis approaches to receptor structure and function. *Pharmac Therapeutics,* **103**, 21-80 (2004).

41 Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2-adrenergic G-protein coupled receptor. *Nature*, **450**(7168), 383-387 (2007).

42 Kobilka, B. G-protein coupled receptor structure and activation. *Biochim Biophys. ACTA,* **1768**(4), 794-807 (2007).

43 Kaufmann, C. & Brown, M. R. Adipokinetic hormones in the malaria mosquito, *Anopheles Gambiae*: Identification and expression of genes for two peptides and a putative receptor. *Ins. Biochem., and Mol. Biol.,* **36**, 466 – 481 (2006).

44 Belmont, M., Cazzamali, G., Williamson, M., Hauser, F. & Grimmelikhuijzen, C. J. P. Identification of four evolutionary related G – protein coupled receptors from the malaria mosquito *Anopheles gambiae*. *Biochem. and Res. Comm.,* **344**, 160 – 165 (2006).

45 Reggio, P. H. Computational Methods in Drug Design: Modellling G Protein-Coupled Receptor Monomer, Dimers and Oligomers. *J AAPS*, **8**(2) 37, pages, (2006).

46 McGuffin, L. T., Bryson, K. & David, J. T. The PSIPRED protein structure prediction server. *Bioinform. Appl. Note,* **16**(4), 404 – 405 (2000)**.**

47 David, T. J. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics,* **23**(5), 538-544 (2007).

48 Thompson, J.D., Higgins, D.G., and Gibson, T.J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res*. **22**, 4673-4680 (1994).

49 Bissantz, C., Logcan, A., and Rognan, D. High-Throughput Modelling of Human G-Protein Coupled Receptors: Amino acid Sequence-alignment, Three Model building and Receptor Library Screening, *J. Chem,. Inf. Comput. Sci.* **44**, 1162-1176 (2004).

50 Schlessinger, A and Rost, B., Protein Flexibility and Rigidity Predicted From Sequence. *Proteins: Structure, Functions, and Bioinformatics,* **61**, 115-126 (2005).

51 Sali, A., and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol*. **234**, 779-815 (1993).

52 Vriend, G. WHATIF: A molecular modeling and drug design program., *J. Mol. Graph*., **8**, 52-56 (1990).

53 Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V, Stereochemistry of polypeptide chain configurations, *J. Mol. Biol*., **7**, 95-99 (1963).

54 Gunasekaran, K., Ramakrishnan, C. and Balaram, P. Disallowed Ramachandran Conformations of Amino Acid Residues in Protein Structures. *J Mol Biol.*, **264**, 191-198 (1996).

55 Prakash, T., Ramakrishnan, C., Dash, D., and Brahmachari, S. K. Conformational Analysis of Invariant Peptide Sequences in Bacteria Genome. *J Mol Biol.*, **345**, 937-955 (2005).

56 Lundstrom, K. Structural genomics of GPCRs. *Trends Biotech*., **23**(2), 103-108 (2005).

57 Bhattacharya, S., Hall, S. E., Li, H., & Vaidehi, N. Ligand stabilization states of human β2 - adrenergic receptor: Insight into G- protein coupled receptor activation. *Biophys.,* **94**, 2027-2042 (2008).

58 Kaminski, G., A., Friesner, R., A., Tirado-Rives, J., and Jorgensen, W., L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, **105**, 6474-6487 (2001)

59 Bissantz, C., Logcan, A., and Rognan, D. High-Throughput Modelling of Human G-Protein Coupled Receptors: Amino acid Sequence-alignment, Three Model building and Receptor Library Screening, *J. Chem,. Inf. Comput. Sci.* **44**, 1162-1176 (2004).

60 Rivail, L., Chipot, C., Maigret, B., Bestel, I. Sicsic, S., and Tarek, M. Large – scale molecular dynamics of a G protein coupled receptor, the human 5-HT4-serotonin receptor, in a lipid bilayer. *J. Molec. Structure (THEOCHEM),* **817**, 19-26 (2007).

61 Spijker, P., Vaidehi, N., Freddolino, P. L., Hilber, P. A. J. and Goddard III, W. A. Dynamic behaviour of fully solvated B2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. *PNAS*, **103**, (130), 4882-4887 (2006).

62 Anézo, C., de Vries, A. H., Höltje, H-D., Tieleman, D. P. and Marrink S-J. Methodology Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B* **107***,* 9424-9433, (2003)

63 Shrivastava I. H., Capener, C. E., Forrest, L. R. & Sansom, M. S. P. Structure and dynamics of K Channel Pore-Lining Helices: A Comparative Simulation Study. *J. Biophys*. **78**, 79-92 (2000).

64 Delano, W., L. "The PyMol Molecular graphics system." Delano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.

65 Koradi, R., Billeter, M., and Wűthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics,* **14**, 41-55 (1996).

66 Humphrey, W., Dalke, A. and Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33-38 (1996).

67 Sanner, M. F. Python: A Programming Language for Software Integration and Development. *J. Mol. Graphics Mod*., **17**, 57-61 (1999).

68 Lobanov M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compatness, *Molec. Bio*, **42**, 623-628 (2008).

69 Izumi, Y., Kuwamoto, S., Jinbo, Y., and Yoshino, H. Increase in the molecular weight and radius of gyration of apocalmodulin induce by binding of target peptide: evidence for complex formation. *FEBS letters,* **495**, 126-130 (2001).

70 Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637 (1983).

71 Erickson, J. A., Jalaie, M., Robertson, D. H., Lewis, R. A., and Vieth, M. Lessons in Molecular Recognition: The effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.,* **47**, 45-55 (2004).

72 Gschwend, D. A., Good, A. C., and Kuntz, I. D. Molecular Docking Towards Drug Discovery. *J. Molec. Recogn.,* **9**, 175-186 (1996).

73 Shoichet, B. K., McGovern, S. L., Wei, B., and Irwin, J. J. Lead discovery using molecular docking. *Curr. Opion Chem. Biol,* **6**, 439-446 (2002).

74 Whittle, P. J., and Blundell, T. L. Protein structure-based drug design. *Annu. Rev. Biophys. Struct.,* **23**, 349-375 (1994).

75 Manetti, F., Tintori, C., Armmand-Ugon, M., Clotet-Codina, I., Massa, S., Ragno, R., Este, J. A., and Botta, Maurizio. A Combination of Molecular Dynamics and Docking Calculations to Explore the Binding Mode of ADS-J1,

a Polyanionic Compound Endowed with Anti-HIV-1 Activity. *J. Chem. Inf. Model.,* **46**, 1344-1351, (2006).

76 Liu, X., Kai, M., and Wang, R. Molecular modeling studies to predict the possible binding modes of endomorphin analogs in μ opioid receptor. *Bioorg. Med. Chem. Lett.,* **19**, 5387-5391 (2009).

77 Huber, T., Menon, S., and Sakmar, T. P. Structural Basis for Ligand Binding and Specificity in Adrenergic Receptors: Implications for GPCR-Targetd Grug Discovery. *Biochem.,* **47**(42), 11014-11023 (2008).

78 McConkey, B. J., Sobolev, V., and Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Sci.* **83** (7), 845-856 (2002).

79 Bursulaya, B. D., Totrov, M., Abagyan, R., and Brooks III, C. L. Comparative stud of several algorithms for flexible ligand docking. *J. Comp Aided Molec Des.,* **17**, 755-763 (2003).

80 Kontoyianni, M., McClellan, L. M., and Sokol, G. Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.,* **47**, 558-565 (2004).

81 Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. A Geometric Approach to Macromolecular-ligand Interactions. *J. Molec. Biol,* **161**, 269-288 (1982).

82 Ewing, T. J., and Kuntz, I. D. Critical Evaluation of Search Algorithms for Automated Molecular Docking and Database Screening. *J. Comp. Chem.,* **18**(9), 1175-1189 (1996).

83 Gabb, H. A., Jackson, R. M., and Sternberg, M. J., Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. *J. Mol. Biol.,* **272**, 106-120 (1997).

84 Goldman, B. B., and Wipke, W. T., QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock). *Proteins*, **38**(1), 79-94 (2000).

85 Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Molec. Biol.,* **261**, 470-489 (1996).

86 Kramer, B. M., Rarey, M., and Lengauer, T. Evaluation of the FLEXX Incremental Construction Agorithm for Protein-Ligand Docking. *Proteins*, **37**, 228-241 (1999).

87 Jones, G., Willett, P., and Glen, R. C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Molec. Biol*., **245**, 43-53 (1995).

88 Jones, G., Willett, P., Glen, R. C., and Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Molec. Biol.,* **267**, 727-748 (1997).

89 Taylor, J. S., and Burnett, R. M. DARWIN: A program for docking flexible molecules. *Proteins*, **41**, 173-191 (2000).

90 Morris, G., M. *et al*. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **19(**14) 1639-1662 (1998).

91 Peters, K. P., Fauck, J., and Frommel, C. The automatic search for ligand binding sites in protein of known three-dimensional structures using only geometric criteria. *J. Molec. Biol.,* **256**, 201-213 (1996).

92 Brady, G. P. Jr., and Stouten, P. F. W. Fast prediction and visualization of protein binding pockets with PASS**.** *J. Comput-Aided Mol. Des.,* **14**, 383-401 (2000).

93 Hetényi, C., & van der Spoel, D. Efficient docking of peptides without prior knowledge of the binding site. *Prot. Sci.* **11**, 1729-1737 (2002).

94 Hetenyi, C. and van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Letters*, **580**(5), 1447-1450 (2006).

95 Kovacs, M., Toth, J., Hetenyi, C., Malnasi-Csizmadia, A., and Sellers, J.R. Mechanism of blebbistatin inhibition of myosin II. *J. Biol Chem.*, **279**(34): 35557-35563 (2004).

96 Bikadi, Z., Hazai, E., Zsila, F., and Lockwood, S.F. Molecular modeling of non-covalent binding of homochiral (3S,3 ' S)-astaxanthin to matrix metalloproteinase-13 (MMP-13). *Bioorg & Med. Chem.*, **14**(16): 5451-5458 (2006).

97 Iorga, B., Herlem, D., Barre, E., and Guillou, C. Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the "blind docking" approach. *J. Mol. Mod.*, **12**(3): 366-372 (2006).

98 Cummings, M. D., Hart, T. N., and Read, R. J., Monte Carlo docking with ubiquitin. *Protein Sci.,* **4**(5), 885-899 (1995).

99 Trosset, J. Y., and Scheraga, H. A. Flexible docking simulations: scaled collective variable Monte Carlo minimization approach using Bezier splines, and comparison with a standard Monte Carlo algorithm. *J. Comput. Chem*. **20**(2), 244-252 (1999).

100 Nair, M., Jackson, G. E. & Gade, G. Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics. *J. Comp. – aided Mol. Des.***15**, 259 – 270 (2001).

101 Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. Complexity and simplicity of ligand-macromolecule interactions: The energy landscape perspective. *Curr. Opin. Struct. Biol.* **12**, 197-203 (2002).

102 Hetényi, C., Paragi, G., Maran, U., Timár, Z., Karelson, M., and Penke, B. Combination of a Modified Scoring Function with Two-Dimensional Descriptors for Calculation of Binding Affinities of Bulky, Flexible Ligands to Proteins. *J. Am. Chem. Soc.* **128**(4), 1232-1239 (2006).

103 Leach, A. R., Scoichet, B. K., and Peishoff, C. E. Prediction of Protein-Ligand Interactions. Docking and Scoring : Successes and Gaps. *J. Med. Chem.,* **49**(20), 5851-5855 (2006).

104 Michel, J., Verdonk, M. L., and Essex, J. W. Protein-Ligand Complexes: Computation of the Relative Free Energy of Different Scaffolds and Binding Modes. *J. Chem.Theory Comp.,* **3**, 1645-1655 (2007).

CHAPTER THREE

Building the 3D structure of AKHR: Homology modelling and Molecular Dynamics

3.1 Summary

To facilitate structure based drug design using insect GPCRs, the recently published primary sequence of AKHR was used to determine its 3D structure. Homology modelling [1], which produces an all atom model of a sequence, was used to build the transmembrane helices based on their alignments to the helix bundles in rhodopsin and beta2-adrenergic receptor (β_2AR). Rhodopsin has successfully been used as a template for most GPCR models but it has a closed structure that protects its ligand. The recently published beta2-adrenergic receptor has an open structure, which allows ligands to diffuse in and out of the binding pocket [2,3].

The loops in AKHR are long and flexible so molecular dynamics (MD) simulations were used to determine their conformations [4,5] while tethered to their respective helices. By joining each loop to the helices, two 3D structures of AKHR were built. The resulting molecule was energy minimised and MD simulations, in vacuum, performed to optimise the loops. Movement of the helices was restrained so that the general conformations of the helix bundles were maintained. The minimum energy conformation of each molecule gave the final, vacuum, 3D structures of AKHR. These will be referred to as the β_2AR-based and the rhodopsin-based structures. Although the two models have similar overall structure, they differ in, for example, the conformation of the helix bundle, helical lengths, interhelical distances and interactions.

A comparison of the two structures should help in determining the most suitable model for use in docking experiments.

3.2 Experimental Methods

3.2.1 Modelling the α-Helices

3.2.1.1 Prediction of secondary structure and transmembrane regions

The primary sequence of AKHR, (Genbank database accession number AY298745), was used to determine the 3D structure of the receptor. PSIPRED and the MEMSAT3 programmes available on http://globin.bio.warwick.ac.uk/psipred/, predicted the secondary structure and the transmembrane regions of the receptor. PSIPRED makes use of two feed-forward neural networks, which perform analysis on output obtained from PSI_BLAST [6] and has shown to be capable of achieving an average score of 76.5%. MEMSAT3 uses neural networks to determine reliable topogenic scores from the evolutionary information in the PSI-BLAST- derived sequence profiles [7]. Based on the prediction results, the primary sequence of AKHR was divided into 7TMs, six loops, N-terminus and C-terminus segments.

3.2.1.2 Sequence Alignment and homology modelling

AKHR belongs to the same Class A GPCR family as rhodopsin and beta2-adrenergic receptor and has the highly conserved residues commonly found in the helices of proteins in this family. Therefore the crystal structures of rhodopsin, (1F88A) and beta2-adrenergic receptor, (2RH1A) from the Protein Data Bank [8] were used as templates for homology modelling of its helix bundles. Pairwise sequence alignments of the AKHR helix residues with those extracted from the templates PDB files were performed using ClustalW available on http://www2.ebi.ac.uk/clustalW. To determine the ends of the helices accurately, residues from the loop regions were

included on both ends of each helix. The BLOSUM scoring matrix was used for the alignments, with the gap penalty set at 10 and the penalty for extending a gap and separation of a gap set at 0.05. The alignments were manually adjusted to remove gaps from the helix regions. Percent sequence identities and similarities between our receptor and the templates were calculated by dividing the total number of identical/similar residues by the number of residues in the shorter sequence.

Upon obtaining an acceptable alignment, homology modelling of the 7TMs was carried out using Modeller 9 version 3. Modeller implements comparative protein structure modeling by satisfying of spatial restraint [9]. This gives the rhodopsin-based and the β_2AR-based helical bundles of AKHR.

3.2.2 Modelling loops, N- and C-terminus

The loops, N- and C-termini were built separately using InsightII 2005. Conformational search for each segment was performed by subjecting them to molecular dynamic simulations using GROMACS version 3.3 [10]. All simulations were performed at a time step of 2 fs, using the OPLS-AA/L all – atom [11] force field and constant temperature, volume and number of particles (NVT). A cut-off 1.0nm was used for van der Waals interactions and electrostatic interactions for real space calculations.

3.2.2.1 Molecular dynamics simulations

In vacuum

Each loop was put into a simulation box and energy minimised using the low–memory BFGS minimiser for 500 steps to remove bad contacts. For each loop, the terminal residues were distance restrained to the interhelical distances of their respective helices during 0.1 ns MD simulations at 600K. The anchor residues were

position restrained at the interhelical distances and a further 10ns MD simulation was performed at 600K. Since only one starting structure was used, conformational search was at high temperature so that the molecule could overcome energy barriers [12]. During the production stage, 100 structures collected during the 10 ns high temperature simulation, were subjected to 0.1ns simulated annealing to 300K. The loops were energy minimised and cluster analysis performed with a cut-off distance of <0.2.

In water

The minimum energy conformation from the simulations in vacuum was used as the initial structure of the loop for 10ns MD simulations in water. The system was solvated with SPC water. A force constant of 1000 $kJmol^{-1}nm^{-2}$ was used during position-restrained simulations for 0.1 ns to equilibrate the system and allow the water to soak the protein. All simulations were performed with periodic boundary conditions. The protein and solvent were coupled separately using the Berendsen coupling method to a temperature of 300K using a coupling constant, tau (τ)_t, of 1 x 10^{-4} ns.

The anchor residues were distance restrained to their respective interhelical distances during 0.1ns MD simulations. In the data collection stage, 100 structures were collected for analysis during 10 ns molecular dynamic simulations. Cluster analysis of the structures was performed based on a cut-off value of <0.2 nm for superimposing backbone atoms. The minimum energy conformation was identified from the largest cluster and used as a starting structure of the loop. The process was repeated for all loops, N-terminus and C-terminus.

3.2.3 Constructing the 3D structure of AKHR

The N-terminus, C-terminus and the loops were independently joined to the β_2AR-

based and rhodopsin based AKHR helical bundles. The protein molecules obtained were energy minimised and subjected to 10ns MD simulations in vacuum to relax and optimise loops. The helices were fixed to maintain their conformation and allow free movement of the loops. The minimum energy structures, from a 10ns MD simulation, gave two models of AKHR. The models were viewed and analysed using the programs PYMOL [13] and MolMol [14].

3.3 Results and Discussion

The recently published primary sequence [15,16] of the adipokinetic hormone receptor (AKHR), (Genbank database accession number AY298745), from the malaria mosquito, *Anopheles gambiae*, was used to predict and model its 3D structure. Secondary structure and transmembrane region predictions of AKHR, by the PSIPRED programme and the MEMSAT3 programs [6,7], indicate that 218 residues make up the seven transmembrane alpha helices (7TMs) (Figure 3.1, Appendix A).

Figure 3.1: Predicted Secondary structure of AKHR. Pred: Predicted secondary structure (H=helix, E=strand, C=coil), AA: Target sequence. Red indicates residues in the

seven transmembrane helices, black gives residues in loops, N- and C- termini and green shows the highly conserved residues.

These helices consist of the highly conserved residues and patterns commonly found in Class A GPCRs [1,17,18], namely, N in helix 1, D in helix 2, D/ERY in helix 3, CWTPY in helix 6 and NPVVY in helix 7.

Sequence alignments of the transmembrane residues of AKHR with those of beta2-adrenergic receptor (β_2AR) [19] (Protein Data Bank (PDB) identity 2RH1A, Appendix B) and rhodopsin [2] (PDB identity 1F88A), in Appendix B and C showed relatively low sequence identity but high sequence similarity (Table 3.1). The low percent sequence identity values between AKHR helices and the two templates are typical of GPCR superfamily proteins. These proteins are known to have low sequence identity but similar structural and functional features due to high sequence similarity [20].

Table 3.1: Comparison of sequences of AKHR, Rhodopsin and β_2AR helices

Helix	Percent sequence (%) identity with		Percent sequence (%) similarity with	
	β_2AR	rhodopsin	β_2AR	rhodopsin
1	14	23	54	50
2	25	14	58	52
3	29	30	68	60
4	28	40	76	80
5	23	17	65	60
6	34	23	83	74
7	33	35	71	71

The helix segments in AKHR have high sequence similarity with both rhodopsin and β_2AR indicating a high structural similarity between them [21]. The alignments were used to build two homology models of the 7TM bundles of AKHR using MODELLER 9 version 3 [9] (Figure 3.2).

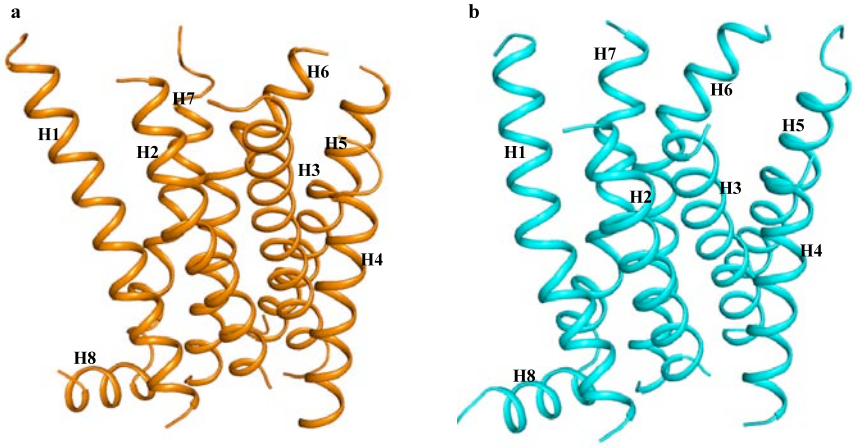


Figure 3.2: The seven transmembrane helices of adipokinetic hormone receptor. a) β_2AR-based seven transmembrane helices of AKHR and b) Rhodopsin-based seven transmembrane helices of AKHR. The two helical bundles have the same overall structure and an anticlockwise arrangement of helices. Helix3 (H3) is tilted and helix 4 (H4) is the shortest. The bundles have the eighth intracellular helix that is parallel to the cell membrane.

Accurate loop modelling determines the usefulness of a protein model in recognising ligand-binding sites and for ligand docking experiments [22]. The N-terminus, extracellular loops (ECL), intracellular loops (ICL) and the C-terminus in AKHR are longer and more flexible than those in both β_2AR and rhodopsin. Therefore, molecular dynamics (MD) simulations were used to search conformational space for each structural domain. Cluster analysis and an overlay of the 100 structures obtained from a 10ns MD simulation in water, showed that each loop has a strongly preferred conformation (Figure 3.3).

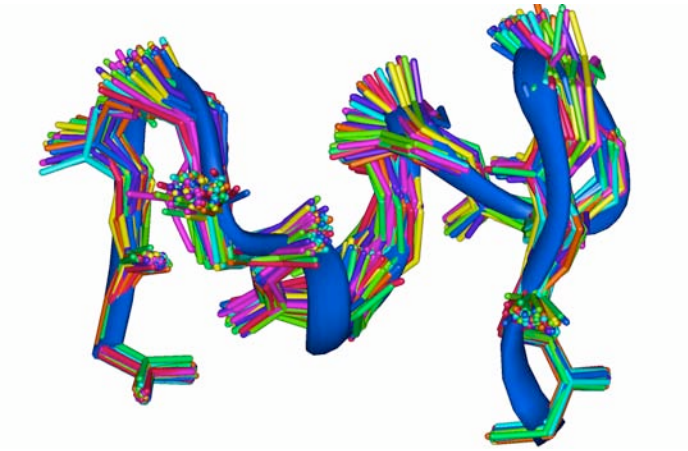
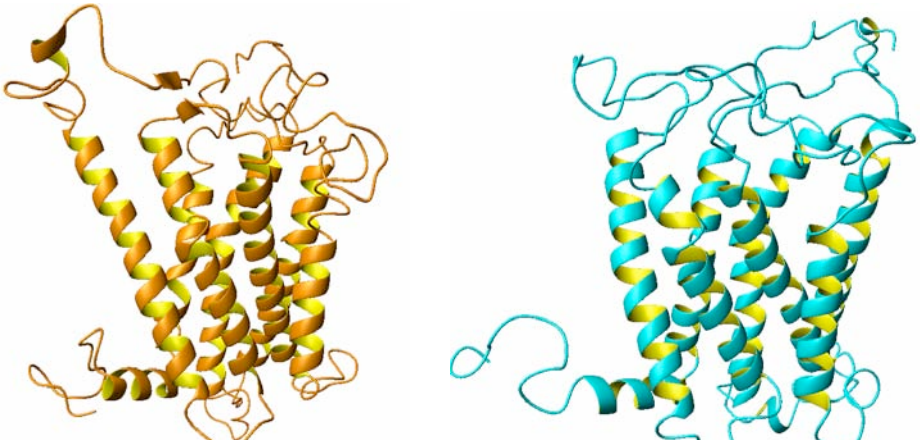


Figure 3.3: Cluster of 100 structures from MD simulation of the second extracellular loop.

The conformation of minimum energy of each loop was added to the two 3D structures of the helices to give starting structures for molecular dynamics of the complete receptor.

The cysteine – cysteine disulphide bond between the second extracellular loop (ECL2) and helix 3, which plays an important role in receptor stability and function, has been found in ~90% of the identified Class A GPCRs [23,24]. In this receptor, Cys121 at the extracellular end of helix 3 and Cys197 in ECL2 have a high probability of forming a disulphide bond. A disulphide bond between Cys121 at the extracellular end of helix 3 and Cys197 in ECL2 was therefore built. A further 10ns MD simulation of the complete receptor to optimise the loops gave two 3D structures of AKHR (Figure 3.4).



(a) (b)

Figure 3.4: a) The β_2AR-based structure and b) the rhodopsin-based structure. In the extracellular region (region outside the cell) of the β_2AR-based AKHR, there are β-strands in the ECL1 and the N-terminus, and a helix in the N-terminus, not found in the rhodopsin-based structure. Both molecules have the eighth intracellular (inside the cell) helix at right angles to the helical bundle.

Ramchandran plots [25,26,27] of the two structures show that, 89% of residues in β_2AR-based AKHR and 86% of residues in rhodopsin based AKHR have psi (ψ) and phi (Φ) angles in favoured regions (Figure 3.5), indicating that the models are of acceptable quality. Most residues in the disallowed region are found in the ECL and the ICL regions and residues in the helix bundle are found in the highly favoured region.

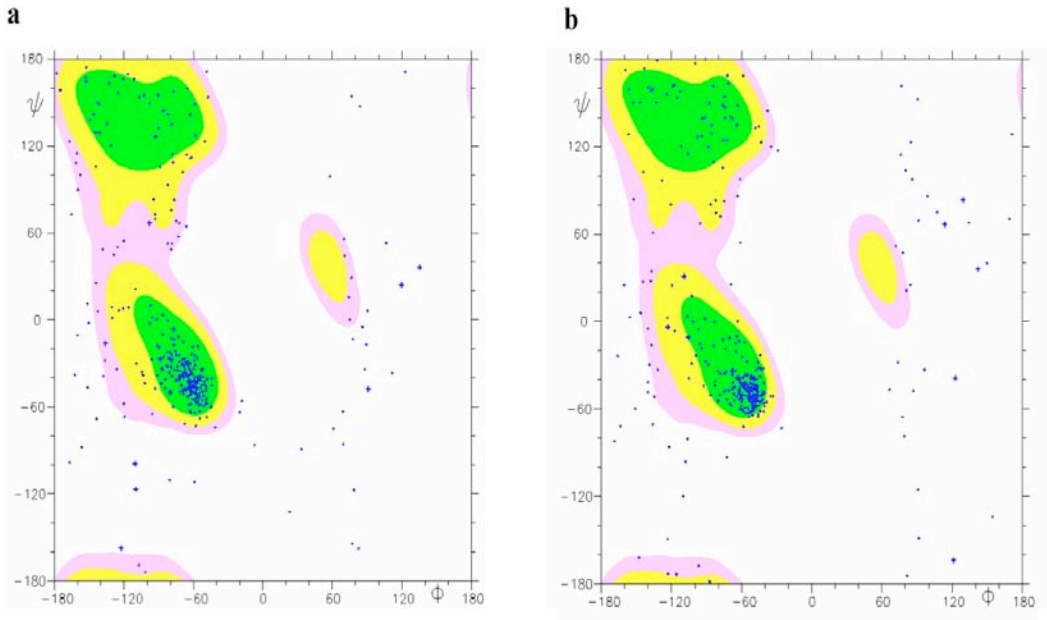


Figure 3.5: Ramchandran plots of psi (ψ) and phi (Φ) angles of residues in a) the β_2AR-based and b) rhodopsin-based 3D structures of AKHR. The highly favoured, favoured, least favoured and disallowed regions for values of ψ and Φ angles are indicated as green, yellow, pink and white respectively. Each blue dot represents a residue and the blue cross represents glycine. About 89% and 86% of residues in open AKHR and closed AKHR, respectively, have ψ and Φ angles in the favoured regions.

3.3.1 Analysis of the helix bundles

The rhodopsin and β_2AR-based 3D structures of AKHR have similar overall structures. When viewed from the extracellular region, the AKHR helix bundles have an anticlockwise arrangement of the 7TMs from helix 1 to helix 7 and a tilted helix 3, which is common in all Class A members of the GPCR family [2,7]. The anticlockwise arrangement of helices in GPCRs is supported by the distance constraints from the engineered Zinc (II) binding sites in neurokinin 1 (NK-1) receptor and the **κ**-opiod receptor [28, 29].

The seven transmembrane helices are stabilized by interhelical interactions involving the highly conserved residues [17]. Class A GPCRs contain a conserved aspartic acid in helix 2 (AspII.10 according to the Schwartz numbering scheme [30]) and an asparagines in helix 7 (AsnVII.16), involved in intramolecular interactions due to the close proximity of the two helices [30]. In contrast, in AKHR, helix 2 has the aspartic acid, Asp94 (AspII.11) and Asn319 (Asn VII.12) in helix 7 at different positions and with their side chains orientated away from each other. This reduces the chances of interactions between them in both the β_2AR-based and the rhodopsin-based structures. Instead, in the β_2AR structure Asn94 forms a salt-bridge with Ser135 in helix 3. Although these residues are both orientated towards the centre of the helix bundle in the rhodopsin-based structure, the distance between them, 0.42 nm, hinders any intramolecular interactions. However, helix 2 and helix 7 interact through formation of a H-bond between Asn84 and Tyr324 found in the highly conserved NPxxY motif of helix 7. There is also a polar interaction between Asn65 (helix 1) and Ser317 (helix 7) which help in stabilising the structure. These interactions involving helices 1, 2 and 7 are structurally important for all Class A GPCRs [31].

Experimental work supports the presence of kinks that cause helices to twist (bend) [32]. The helices in the two AKHR models have proline-induced kinks and bends [33] at approximately the same positions. Where Met, Thr or Tyr precedes the proline, the kinks result in the formation of a bend or a turn, *viz* helix 2 at Met101-Met102, helix 5 at Tyr221-Thr222, and helix 7 at Thr314. In helix 6 the kink causes a twist towards helix 5 in the extracellular region. It was observed that although helices 5 and 6 in our models differ in length, they both extend into the cytoplasm by

4-3 residues. Evidence of the helices extending two to three turns in GPCRs, was provided by studying the accessibility of attached nitroxide labels to collisions with paramagnetic probes in solution in the third intracellular loop that joins helices 5 and 6 during electromagnetic spectroscopy of rhodopsin [34]. In general, both models have shorter helices than their respective templates, 179 residues in the β_2AR-based model and 174 residues in the rhodopsin-based model compared to 204 and 194 residues in β_2AR [19] and rhodopsin respectively [35].

An overlay of the two 3D structures gave a root mean square deviation (rmsd) of 0.79 nm, for the optimal superimposition of C- alpha (Cα) atoms [36] of the whole molecule and 0.45 nm for the helix bundles. Although the overall structures are similar, the two have different structural features that cannot be ignored. The β_2AR-based AKHR model has an open structure in the extracellular region as shown in Figure 3.6.

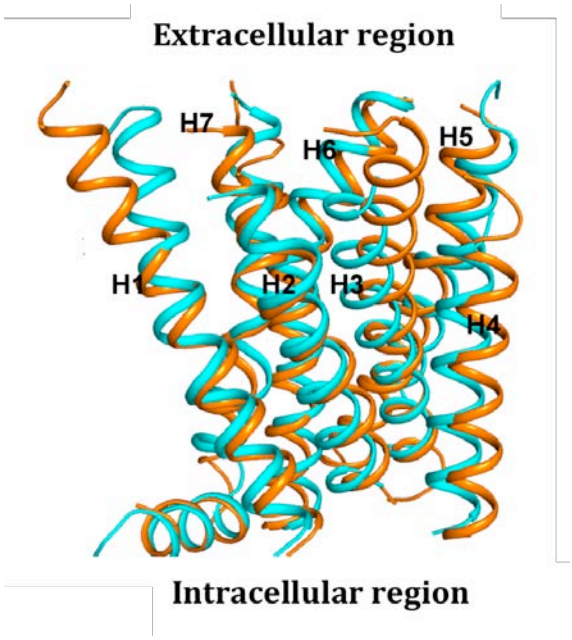


Figure 3.6: Side view of an overlay of helices of β_2AR-based (orange) and rhodopsin-based (blue) 3D structures of AKHR. There is an anticlockwise arrangement of

helices 1-7 (H1-7) and in the extracellular region, β_2AR-based AKHR helices are drawn away from the centre of the helix bundle. This gives it an open conformation compared to the conformation in the rhodopsin-based AKHR.

Longer interhelical distances and an outward deflection of helices 1, 3, 4 and 6 away from the centre of the helical bundle in β_2AR evidence the open conformation. The angle of deflection between helix 1 in the β_2AR and rhodopsin based AKHRs is 44 $\pm$ 0.25^o and the distance between the ends of the helices is 10 $\pm$ 0.20 Å (Figure 3.7a). The rhodopsin based AKHR has a closed conformation as in the crystal structure of rhodopsin. A closed conformation in rhodopsin protects the covalently bound cis-retinal from hydrolysis [19]. Since binding of peptide hormones often involves the transmembrane domain and the amino terminus [29], the closed conformation would limit the accessibility of the binding pocket by diffusible agonists.

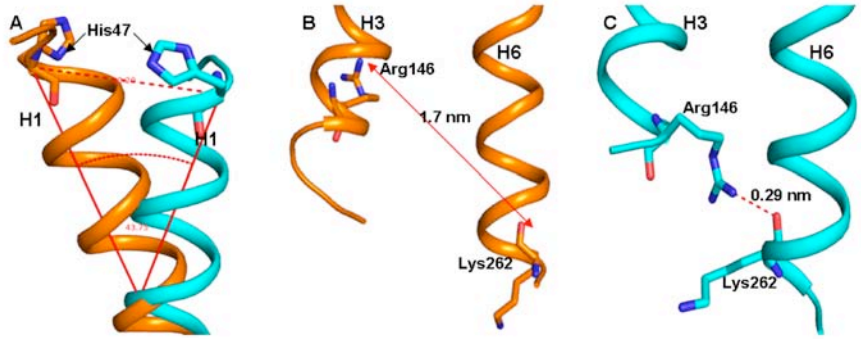


Figure 3.7: Comparisons between helices in the AKHR models

A) The deflection of helix 1 (H1) in β_2AR-based AKHR (orange) from helix 1 (H1) in rhodopsin-based AKHR (blue). The distance between the Cα atoms of Histidine at the extracellular ends of the helices is 1 $\pm$ 0.2 nm and angle of deflection is 44 $\pm$ 0.25^o. Different arrangements of the side chains of the helix residues result in different interhelical interaction. B) There is no ionic interaction between the highly conserved Arg146 at the intracellular end of helix 3 (H3) and residues in helix 6 (H6) of β_2AR-based AKHR. C) An ionic interaction (broken line in red) exists between Arg146 in H3 and Lys262 in H6 of the rhodopsin based AKHR. These residues are 0.29 nm apart.

In the β_2AR-based structure, the distance between these residues (1.7 nm) is too large for the two to interact.

Rhodopsin and β_2AR yield models in which the same residues occupy different positions in the helical bundle. The side chains of these residues have different orientations, resulting in different interhelical interactions. This structural difference is clearly shown around the highly conserved DRY sequence at the cytoplasmic end of helix 3 [3]. In the β_2AR-based AKHR there is no ionic interaction involving the positively charged side chain nitrogen atom of Arg146 in helix 3 and the backbone C=O of Lys262 in helix 6, which has a distance of 1.7 nm between them (Figure 3.7b). This open structure is therefore very flexible with high agonist-independent activity inside the cell (basal activity). Instead, Arg146 forms a salt bridge with Ile142 in helix 3. To stabilise the molecule, helices 5 and 6 form salt bridges in the intracellular side, involving Ser241, Arg242, Tyr236 in helix 5 and Lys265, Arg268 in helix 6. Dror R. O et al (2009), found that the inactive β_2AR exists in equilibrium between conformations with the ionic lock 'on and off' [37]. This could be the case with the β_2AR-based AKHR structure, which does not have the ionic lock between helices 3 and 6 but between helices 5 and 6. In contrast, the rhodopsin based AKHR has an ionic interaction between Arg146 and Lys263, since they are 0.29 nm apart (Figure 3.7c). These intramolecular interactions reduce conformation flexibility, basal activity and play an important role in receptor activation [30].

3.3.2 Analysis of the loop regions

The open and closed conformations of the 7TMs result in different conformations and structural features in the extracellular and intracellular domains. In the closed conformation, the second extracellular loop and the N-terminus stretch over the

helices covering all potential binding pockets (Figure 3.2). There are neither beta (β) sheets nor helices in the loop regions except for a small-predicted helical turn from Asn3–Met5 in the N-terminus.

In contrast, the open structure has extracellular loops drawn away from the centre of the helical bundle. ECL1 has a β-strand (Trp114-Gly117), which is not found in the rhodopsin-based structure. The N-terminus does not form a cap over the helices. A β-strand involving Tyr24 and Tyr25, and a small helical turn from Glu32-Tyr35 keep it in an upright position above the helical bundle (Figure 3.2a). Although the N-terminus of rhodopsin was not used to build the AKHR N-terminus, a similar helical turn is in the crystal structure of rhodopsin involving Tyr29-Leu31. In the extracellular side of most GPCRs, the helices and the extracellular domains open up to form a cavity that serves as a binding pocket for the ligand [30]. As such, the open conformation of the beta2 based structure leaves a deep cavity and hence a potential binding pocket is created in the extracellular region between helix 7, helix 1, ECL3 and the N-terminus. The potentially N-glycosylated Asn27 in the N-terminus of AKHR could play an important role in stability of the receptor and in ligand binding [38].

The intracellular side of AKHR is tightly packed with the second loop drawn away from the helix bundle. The two models contain the eighth helix which is parallel to the membrane surface in most Class A GPCRs and is important in G-protein binding and activation [30]. Palmitoylation plays an important role in stabilising the intracellular region of GPCRs. Thus, the presence of Cys346 after the

eighth helix indicates that the C-terminus of AKHR will be stabilised through palmytolation of the residue.

The third intracellular loop, normally not included in GPCR crystal structure determination, contains 20 residues in the β_2AR-based model and 24 residues in the rhodopsin based model. The rhodopsin-based structure has a helical turn in this loop from Ser250 – Ser 254 that is absent in the β_2AR-based structure. This structural feature was also found in the third intracellular loop of a β_2AR model after molecular dynamic simulations in a membrane [37]. Since GPCRs are highly flexible, the different structural features especially in the extra and intracellular domains of the two models could represent features in different conformations during molecular dynamics and/or in the active and inactive state of the receptor.

3.4 Conclusion

In conclusion, the first open and closed 3D structures of an adipokinetic hormone receptor from the malaria mosquito were successfully modelled. The models display structural features and intramolecular interactions similar to those found in other Class A GPCRs. The structural features and the general conformation of our models are in agreement with most mutational data and could give an insight into the 3D structures of insect GPCRs. They are therefore good starting structures for further studies involving the AKH receptor and similar insect GPCRs. In the following chapters, the open and closed AKHR molecules were used in molecular dynamics and ligand binding studies.

3.5 References

1. Trabanino, R. J. *et al.* First principles prediction of the structure and function of G- protein coupled receptors: Validation for bovine rhodopsin. *J. Biophys.* **86**, 194 – 1921 (2004).
2. Palcezewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science* **289** (5480), 739 – 45 (2000).
3. Kobilka, B. and Schertler, G. F. X. New G-protein coupled receptor crystal structures: Insights and limitations. *Trends Pharmacol Sci* **29**(2), 79-83 (2008).
4. Orry, A.J.W. and Wallace, B.A. Modeling and docking the endothelin G-protein coupled receptor. *J. Biophys.* **79**, 3083 – 3094 (2000)**.**
5. Nair, M., Jackson, G. E. and Gade, G. Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics. *J. Comp. – aided Mol. Des.***15**, 259 – 270 (2001)
6. McGuffin, L. T., Bryson, K. and David, J. T. The PSIPRED protein structure prediction server. *Bioinform. Appl. Note* **16**(4), 404 – 405 (2000)**.**
7. David, T. J. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics,* **23**(5), 538-544 (2007).
8. Berman, H. M. *et al*. The protein data bank and the challenge of structural genomics. *Nature Struct*. *Biol*. 957 – 959 (2000).
9. A. Sali and T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol*. 234, 779-815 (1993).
10. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., E. and Berendsen, H., J., C. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.* **26,** 1701-1719 (2005)
11. Kaminski, G., A., Friesner, R., A., Tirado-Rives, J., and Jorgensen, W., L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, **105**, 6474-6487 (2001)
12. Jackson, G. E., Solution conformations of an insect neuropeptide: Crustacean cardioactive peptide (CCAP), *Peptides,* **30**, 557-564 (2009).

13. Delano, W., L. "The PyMol Molecular graphics system." Delano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org date accessed.
14. Koradi, R., Billeter, M., and Wűthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics,* **14**, 41-55 (1996).
15. Belmont, M., Cazzamali, G., Williamson, M., Hauser, F. and Grimmelikhuijzen, C. J. P. Identification of four evolutionary related G – protein coupled receptors from the malaria mosquito *Anopheles gambiae*. *Biochem. and Res. Comm.* **344**, 160 – 165 (2006).
16. Kaufmann, C. and Brown, M. R. Adipokinetic hormones in the malaria mosquito, *Anopheles Gambiae*: Identification and expression of genes for two peptides and a putative receptor. *Ins. Biochem. and Mol. Biol.* **36**, 466 – 481 (2006).
17. Chelikani, P. *et al.* Role of group conserved residues in the helical core of β2- adrenergic receptor. *PNAS* ***104***(17), 7027 – 7032 (2007).
18. Singh, S., Malik, B.K. and Sharma, D. K. Targeting HIV-1 through molecular modeling and docking Studies of CXCR4: Leads for therapeutic development. *Chem Biol Drug Design* **69**, 191-203 (2007)**.**
19. Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2 adrenergic G-protein coupled receptor**.** *Nature* **450**(7168), 383-387 (2007)**.**
20. Leach, A. R., Molecular Modelling: Principles and Applications, 2nd Edition, Person Education Limited, 522-549 (2001).
21. Martí-Renam, M. A. et al. Comparative protein structure modeling of genes and genomes. *Annual Rev.Biophys. Biomol. Struct*. **29**, 291-325 (2000).
22. Fiser, A., Do, R., K.,G and Š ali, A. Modeling of loops in protein structures. *Prot. Sci*. **9**, 1753-1773 (2000).
23. Massotte, D and Kieffer, B. L., The second extracellular loop : a damper for G protein-coupled receptors ? *Nat. Struct. Mol. Biol.* **12,** 287-288, (2005).

24. Baldwin, J. M. Structure and function of receptors coupled to G-proteins. *Curr Opin Cell Biol,* **6**, 180-190 (1994).
25. Tulika, P. C., Ramakrishnan, D. D. and Samir, K. B. Conformational Analysis of Invariant Peptide Sequences in Bacterial Genomes. *J. Mol. Biol*. **345**, 937-955 (2005).
26. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V, Stereochemistry of polypeptide chain configurations, *J. Mol. Biol*. **7**, 95-99 (1963).
27. Ramachandran, G. N. and Sasisekharan, V. Conformation of polypeptides and proteins. *Advan. Proptein Chem*. **23**, 283-437 (1968).
28. Elling, C. E. and Schwartz T.W., Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.* **15**, 6213-6219 (1996).
29. Elling, C. E., Nielsen, S. M. and Shwartz T. W. Conversion of antagonist-binding site to metal-ion site in the tachykini NK-1 receptor. *Nature*, **375,** 74-77 (1995).
30. Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Reviews*, **21**, 90-113 (2000).
31. Perlman, J. H. *et al*. Interactions between Conserved Residues in Transmembrane Helices 1, 2, and 7 of the Thyrotropin-releasing Hormone Receptor. *J. Bio. Chem,* **272**(18), 11937-11942 (1997).
32. Fu, D., Ballesteros, J. A., Weinstein, H., Chen, J. and Javitch, J. A. Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding site crevice. *Biochem,* **35**, 11278-11285 (1996).
33. Deupi, X., Olivella, M., Govaets, C., Ballesteros, J. A., Campillo, M. and Pardo, L. Ser and Thr residues modulated the conformation of Pro-kinked transmembrane α-helices. *Biophys. J.* **86,** 105-115 (2004).
34. Farahbakhsh, Z. T., Ridge, K. D., Khorana, H. G. and Hubbell, W. L. Mapping light-dependent structural changes in the cytoplasmic loop connecting helices C and D in rhodopsin: a site-directed spin labeling study. *Biochem.* **34**, 8812-8819 (1995).

35. Stenkemp, R. E., Teller, D. C. and Palczewski, P. Crystal structure of Rhodopsin: A G-protein coupled receptor, *Chem. Biochem*. **3**, 963 – 967 (2002).
36. Godzik, A. The structural alignment between two proteins: Is there a unique answer? *Prot. Sci*. **5**, 1325-1338 (1996).
37. Dror, R. O., *et al*. Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Nat. Acad. Sci*, **106**(12), 4689-4694 (2009).
38. Wheatly, M. and Hawtin, S. R., Glycosylation of G-protein-coupled receptors for hormones central to normal reproductive functioning: its occurrence and role. *Hum. Rprod. Update,* **5**, 356-364 (1999).

CHAPTER FOUR

Molecular Dynamic Simulations of Adipokinetic Hormone Receptor in a Membrane

4.1 Summary

GPCRs exhibit very complex and dynamic structural properties due to their high flexibility [1]. MD simulations remain an important tool for studying the stability and flexibility of proteins in water [2] and for visualisation of conformational changes in a membrane [1,3]. Vacuum MD simulations of GPCRs have generated models currently used in drug design but this does not represent the true environment of the receptor. To represent the physiological conditions of GPCRs, to date, a number of membrane-embedded human GPCRs models have been recently published [1,3,4,5,6]. Phospholipid bilayers represent the fundamental structures of most biomembranes [7]. Periodic lipid bilayers consisting of, for example, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and dipalmitoyl phosphatidyl choline (DPPC) molecules, solvated with water, have been simulated and equilibrated [4]. POPC and DPPC are commonly used in MD simulation of GPCRs [6].

In chapter three, the β_2AR-based with an 'open' conformation, and the rhodopsin-based AKHR with a 'closed' conformation in the extracellular region, were built. In this Chapter, the 'open' β_2AR-based and 'closed' rhodopsin-based AKHR models were used to study the stability and changes in conformation of the receptor molecules during MD simulations in a membrane. Despite changes in conformation of AKHR molecules, it was observed that the overall structure and position of the helix bundles were maintained throughout the MD simulations. Movement of the helices, stability and structural features of the molecules, were determined by how

closely packed the lipid molecules were around the protein. In the β_2AR-based AKHR the helices moved away from each other and the extracellular side widened to give the receptor a more open conformation, whilst the rhodopsin-based structure maintained the closed conformation. The N-terminus and the third intracellular loop, which are highly flexible and not usually included in crystal structure determination of GPCRs, have well defined structures and maintain stable structural features during simulations.

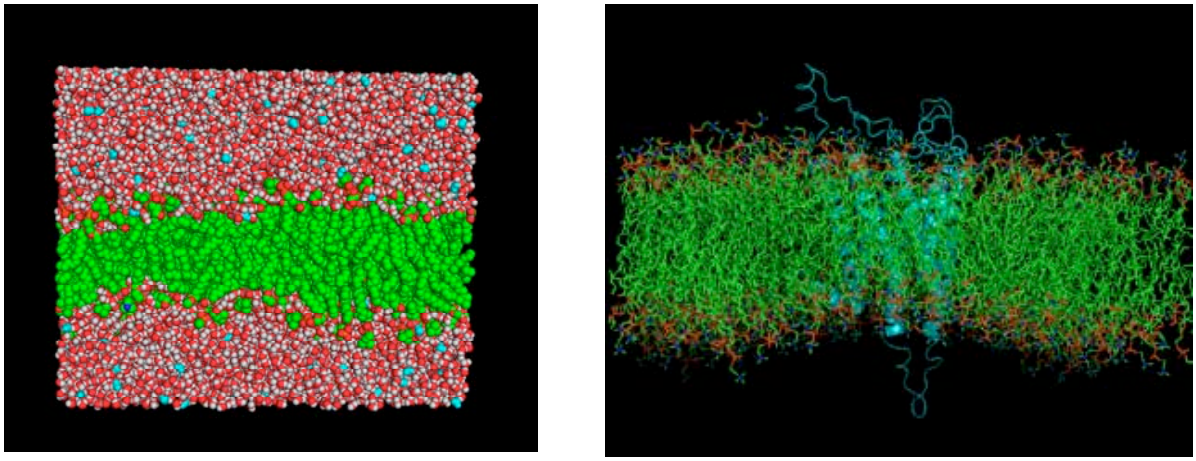
4.2 Experimental Methods

4.2.1 Membrane preparation and Insertion of the Protein

GPCRs are transmembrane receptors and to mimic closely the actual environment of the receptor, rhodopsin-based and β_2AR-based AKHR models were used as starting structures in MD simulations in a membrane. Teilman's pre-equilibrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane model [8] containing 128 lipids molecules and 2460 water molecules from http://moose.bio.ucalgary.ca/, was used. The number of water molecules was increased from 2460 to 10241 in the *z*-direction. The water model used was SPC, which behaves well in lipid bilayer/water simulations [9]. The lipid/water system was increased by a factor of four (x4) in the *xy* planes. Water molecules that occupied the lipid region were removed manually using PYMOL [10] (Figure 4.1a).

The AKHR molecules were inserted into the membrane, and aligned with the lipid bilayer so that they had the same co-ordinate space as the lipid, using VMD [11]. The new protein coordinates were transferred so that they were within the same box as the lipid. Lipid molecules that overlay the protein molecules were removed using

the program genbox from GROMACS to give two protein/lipid complexes, (Figure 4.1b), consisting of protein, 438 POPC molecules and 37020 water molecules.



(a) (b)

Figure 4.1: (a) POPC (green) membrane box filled with water (red and white sphere). (b) AKHR (blue ribbon) inserted in the POPC bilayer (green). For clarity, the water molecules are not shown.

4.2.2 Molecular Dynamics Simulations of AKHR in a Membrane

Molecular dynamics simulations of each protein/membrane complex were carried out using the development version of GROMACS 4.0.4 [12]. Energy minimisation was performed for 1000 steps using the low – memory BFGS minimiser. The system was neutralised by replacing eleven water molecules with chloride ions (Cl^-). A force constant of 1000 $KJmol^{-1}nm^{-2}$ was used during position-restrained MD simulations for 1ns to equilibrate the system and allow the membrane and water to soak the protein. MD simulation with fixed helices was performed for 100ns and since the systems reached steady a state within 10ns, structures were extracted at 10ns, 15ns and 20ns to give three starting protein/membrane systems from both β_2AR- and rhodopsin-based complexes. The six structures were separately subjected to 100ns of free MD simulations. All simulations were performed with periodic boundary conditions.

The simulations were carried out with a time step of 2 fs, using the OPLS-AA/L all – atom [13] force field and at constant temperature, pressure and number of particles (NPT). A cut-off of 1.15 nm was used for van der Waals interactions and the particle–mesh Ewald [14,15] method was used to compute the electrostatic interactions using a grid spacing of 0.1 nm and a distance from the coulomb cut-off of 1.15 nm. The protein, POPC and solvent/Cl^- were coupled separately using the Berendsen coupling method to a temperature of 300K using a coupling constant, τ_T, of 1 x 10^{-4} ns. The pressure coupling constant τ_p was 5.0ps and a compressibility of 0.000045 (1/bar) was used. Neighbourlists were updated every fifth integration time step. The runs were done on a twelve Intel Xeon cluster, each with eight cores running at a clock frequency of 2.33GHz.

4.3 Results and Discussion

Six 100ns MD simulations of AKHR in a membrane were performed. To have an insight of the changes in structural features during dynamics of the flexible GPCR, changes in conformational and the secondary structure were analysed. Estimation of the stability of the receptor molecules, was performed by finding the time dependence of the root mean square deviation (rmsd) from the starting structure, the radius of gyration (Rg) and the root mean square fluctuations (rmsf). These properties have been used to give a relative estimate of the stability of proteins in water and in membrane [1,2,3,16].

4.3.1 Molecular Dynamics simulations of β_2AR-based AKHR model

The starting structures, o_akhr10, o_akhr15 and o_akhr20, for the 100ns MD simulations, were extracted from a trajectory generated from MD simulation of AKHR with fixed helices at 10, 15 and 20ns respectively. During the MD

simulations in a membrane, the protein molecules maintained the overall structure and an open conformation in the extracellular region. However, there were significant fluctuations in different regions of the three molecules. Due to the presence of interhelical interactions and the disulphide bond between Cys197 and Cys212, the helix bundles were fairly rigid while the intra and extracellular loop regions were flexible. Figure 4.2 shows the root mean square fluctuations (rmsf) of the Ca atoms of the protein during the three simulations. Although the results for the different simulations are similar, the protein in simulation o_akhr10 has the highest variability while the movement of o_akhr20 is more restricted. This may be due to tighter packing of the membrane lipid molecules around the protein.

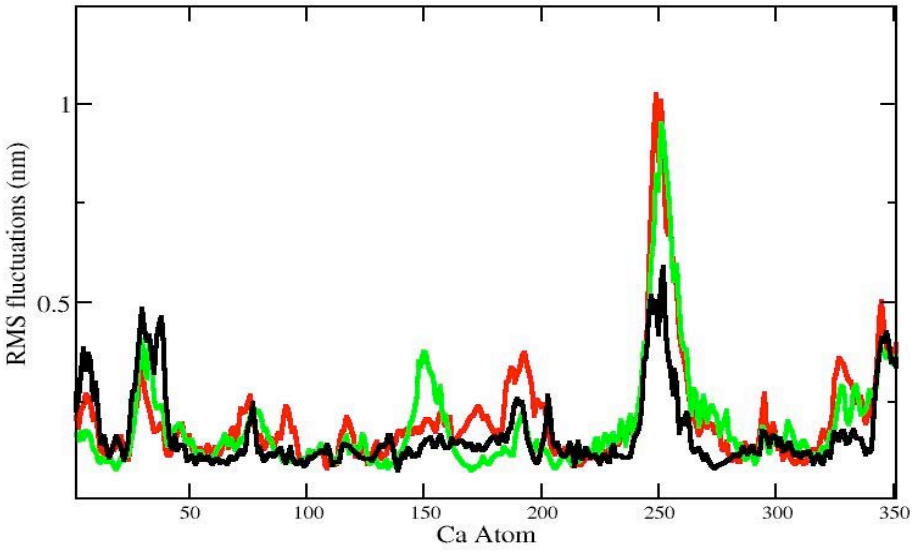


Figure 4.2: The rms fluctuations of o_akhr10 (red), o_akhr15 (green) and o_akhr20 (black) as a function of time during the 100ns MD simulations.

In all three simulations, ICL3 (243-264), has the highest fluctuations with an rmsf of ~0.5-1.0nm. In addition, as expected, there is substantial molecular motion in the N-

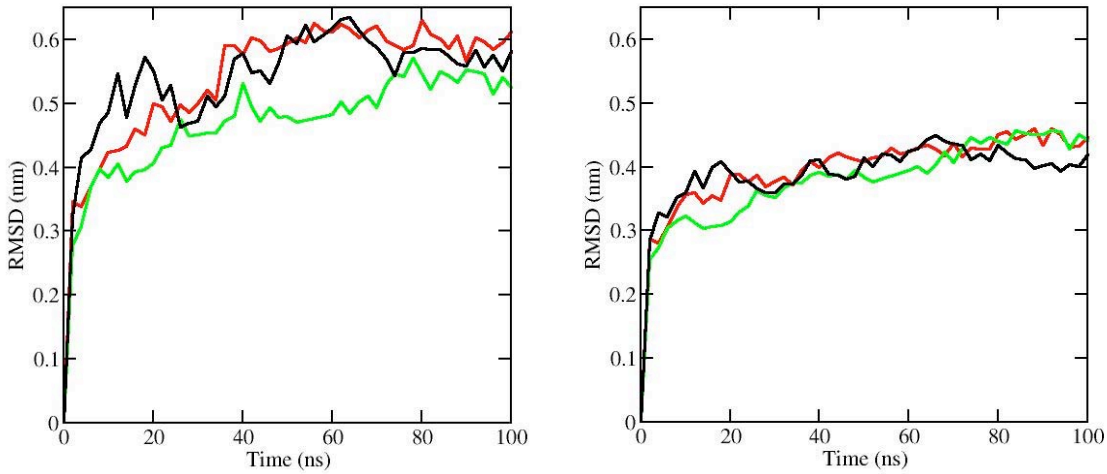
and C-termini (339-354). In the N-terminus (1-46), residues 13-25 are much more rigid. This region is stabilised by the formation of a salt bridge between Tyr24 and Thr115 and the subsequent formation of β-sheets involving Tyr24-Tyr25 and Trp114-Thr115 in ECL1. This is similar to the interactions between the N-terminus and ECL2 in rhodopsin that results in the formation of β-sheets in the two domains [17].

4.3.1.1 Deviations of molecular structures from initial structures

Structural and conformational changes of AKHR during the three simulations in a membrane account for high deviations of the structures from their initial structures as shown by the root mean square deviations (rmsd) in Figure 4.3. The rmsd of the backbone atoms were calculated by least-square fitting of the atom positions of our model to the starting structure and subsequently calculating the rmsd according to equation 1,

$$\mathrm{RMSD}(t_1,t_2) = [1/\mathrm{M}\sum_{i=1}^{\mathrm{N}} m_i \|\boldsymbol{r}_i(\mathrm{t}_1)-\boldsymbol{r}_i\|^2]^{1/2} \qquad (1)$$

Where M is the sum of all the atom masses (m_i) and $\boldsymbol{r}_i$ is the position of atom i at time t [18].



(a) (b)

Figure 4.3: Variation of the rmsd values during 100ns MD simulations. (a) Rmsd for AKHR structures initially increased to 0.3–0.4 nm. There was steady increase afterwards and the systems reached steady state after about 50ns for the o_akhr10 (red) and o_akhr20 (black) and at 70ns for o_akhr15 (green) b) Rmsd for the helix bundles only showing that steady state was attained after 30ns.

As indicated by the sudden rise in initial rmsd to ~0.3-0.4nm, there were marked changes in the AKHR molecules in all simulations. Highest deviations were observed in akhr20 (black), with initial rmsd of ~0.4 nm and fluctuations from 2-30ns and 50-70 ns were mainly due to movements in the extracellular and intracellular domains, since the helices reached steady state after 30ns (Figure 4.3b). There was a steady increase in rmsd from ~0.3nm to ~0.44nm for o_akhr10 (red) and o_akhr15 (green). In o_akhr10 and o_akhr20 equilibratium was reached after about 50ns but o_akhr15 only attained this state after 70ns. However, the three systems approached the same rmsd of ~0.57-0.59nm indicating that the structures accomplished stable conformations after 70ns with approximately the same deviations from the starting structure.

4.3.1.2 Compactness of molecules during MD simulations

The radius of gyration (Rg) indicates the compactness of the protein structures [2,19,20,] during the 100ns MD simulations. This is shown for o_akhr10, o_akhr15 and o_akhr20 in Figure 4.4. Although the initial structures had similar compactness, simulation o_akhr20 (black) maintained an average Rg of ~2.46 nm from the start, indicating that the protein molecule was tightly packed from the beginning of the simulation. However, the drop in Rg at 48 ns could be due to the formation of β-sheets in the N-terminus (Tyr24-Tyr25) and the second extracellular loop (Trp114-Gly117). Structure o_akhr15 (green) was the least compact with a Rg as high as ~2.58nm due mainly to pronounced changes from α- helicity to bends and turns in the secondary structure of the helices. In o_akhr10 (red), Rg fluctuated during the first 50ns, and then there was a 3% decrease to 2.43nm, which was maintained until the end of the simulation indicating a very small change in the positions of the atoms, during the simulation.

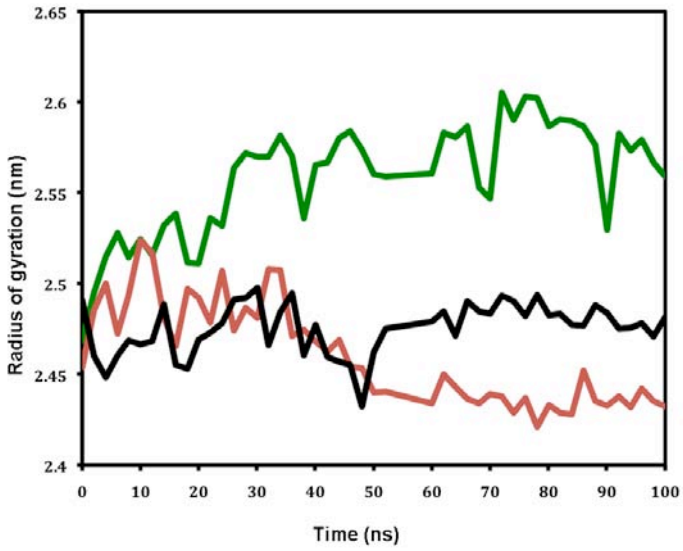


Figure 4.4: Radius of gyration (Rg) as a function of time of AKHR molecules, o_akhr10 (red), o_akhr15 (green), and o_akhr20 (black) during 100ns MD simulations.

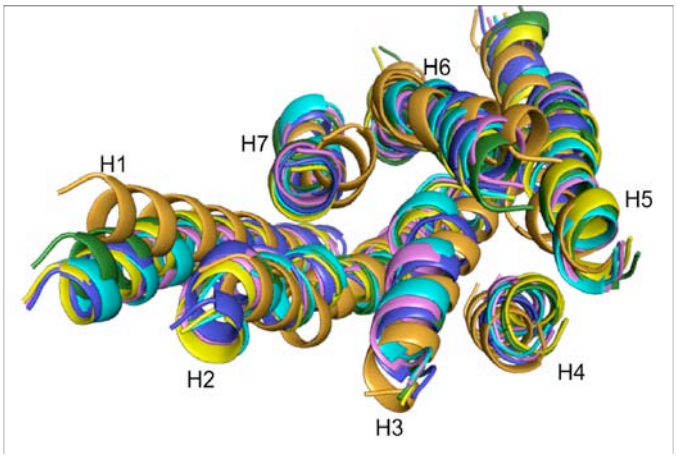
In all three AKHR simulations, movement of atoms was mainly along the y-axis and least along the plane of the membrane.

4.3.1.3 Conformation Changes during simulations

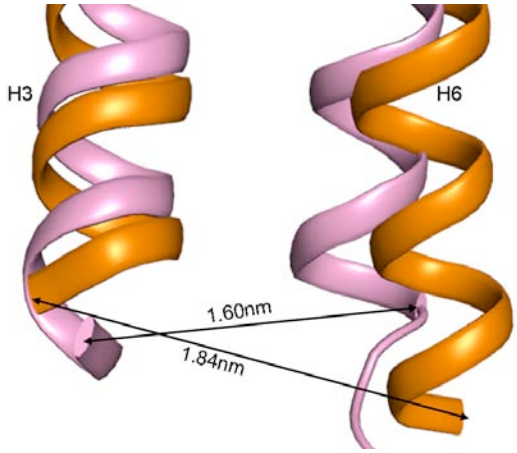
4.3.1.3.1 Conformational Changes in the Helix bundles

In all three simulations, the extracellular side of the helix bundle, moved outward, as illustrated by o_akhr20 helices in Figure 4.5. Although the overall structure of the helix bundle was preserved, an overlay of the helix bundles with those in the vacuum structure gave an rmsd of ~3.1 nm. The movement gave the molecule a more open conformation compared to the vacuum structure. This open conformation, as well as the anticlockwise arrangement of the helices, was consistent throughout the 100ns simulations. Experimental results have supported the anticlockwise arrangement in GPCRs [21,17], which is also found in the crystal structures of rhodopsin and β_2AR [22,23]. In support of the similarity between AKHR helical bundles and those in β_2AR and rhodopsin, it has been revealed that most members of Class A GPCRs share a similar folding of the 7TM domain [24,25,26].

An 'open' extracellular and a 'closed' intracellular side of β_2AR-based AKHR are typical of inactive GPCRs, wherein an extracellular cavity that serves as a binding pocket for ligands [27], is formed. This conformation is found in the recently published crystal structures of β_2AR that contain the inverse agonist, (-)-carazolol [22,28] and the antagonist (-)-timolol [29]. The inverse agonist and the antagonists stabilise the receptor in its inactive state [6]. Adipokinetic hormones are large octa- to decapeptides [30], therefore opening up of the helix bundle creates enough binding space for the diffusible AKH agonist.



(a)



(b)

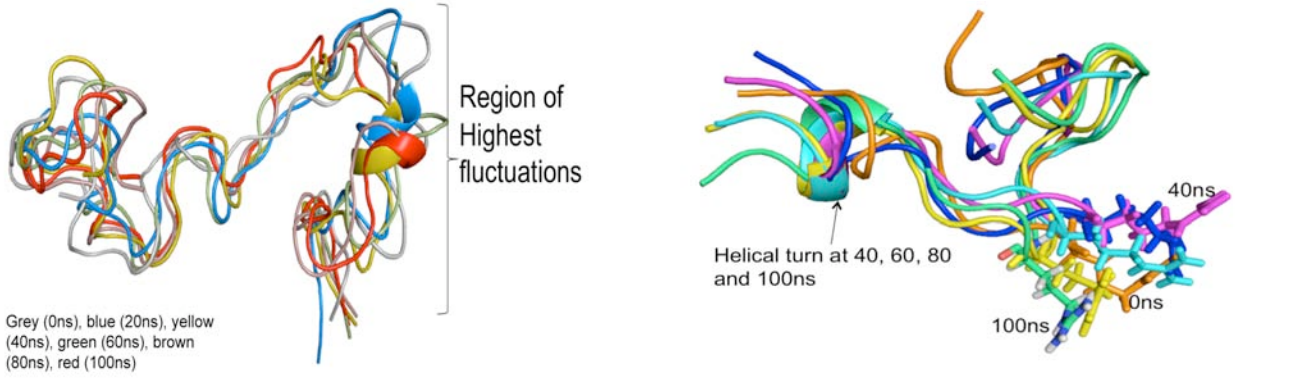
Figure 4.5: Extracellular view of an overlay of o_akhr20 helices at 0ns (brown), 20ns (blue), 40ns (pink), 60ns (cyan), 80ns (yellow) and 100ns (green) from trajectory. There is good agreement in conformation of all helices with a more open conformation than starting structure. b) Position of H3 and H6 after MD simulation in o_akhr20 (pink) and initial AKHR structure (orange). There was an inward shift of H6 in the intracellular region reducing its distance from H3 from 1.84 nm to 1.60 nm during simulation of free AKHR.

As in the β_2AR crystal structure, Helix 1 (H1) is inclined to the plane of the receptor. The angle of inclination is greater than in the β_2AR crystal structure due to the outward divergence of the helix from the starting structure. The divergence is evidenced by the shift of the Cα atom in His47 at the extracellular end of the helix by 0.6 ± 0.02 nm with respect to the vacuum structure. The helices in the vacuum structure have similar inclination to β_2AR since the crystal structure was used as a template to build the helix bundles.

Helices 3 and 6 were drawn inwards, reducing the distance between them by about 0.2nm in the intracellular region (Figure 4.5b). There was an alternating anticlockwise-clockwise rotation and an inward movement of helix 6 (H6). Such movements are commonly found in inactive states of GPCRs [31] and support the idea that β_2AR-based AKHR model represents the inactive conformation of the receptor. The inactive state of rhodopsin is characterised by an 'ionic lock' involving arginine in the highly conserved DRY motif of H3 and residues in H6 [1,26,32]. MD simulations of β_2AR molecule, including the third intracellular loop, have demonstrated an on and off formation of the 'ionic lock' throughout the simulation [1]. However, the long intracellular distance between H3 and H6 in AKHR is not favourable for the formation of a salt bridge in the intracellular side. AKHR could be an example of a Class A GPCR that does not have the 'ionic lock' between H3 and H6 in its inactive state [26]. Alternatively, the length of the simulation, 100ns, could be too short for the 'ionic lock' to form, but in their simulations of β_2AR, Dror, R. O *et al* found an 'ionic lock' to form within 30-150 ns [1]. To stabilise the molecule, AKHR forms an 'on and off' salt bridge between Glu238, in H5, and Lys265, in H6. Since the distance between the oxygen atom OE1, in Glu238 and nitrogen atom NZ, in Lys265, fluctuates between 0.68 and 0.84 nm during dynamics.

4.3.1.3.2 Conformational Changes in loop domains

All loops in the extracellular region were drawn outwards and upwards creating space for ligand binding. The orientation of the second loop is similar to that of β_2AR, which is kept in an upward position above the helix bundle to allow free movement of diffusible ligands [33]. Unlike in β_2AR, the N-terminus in AKHR is ordered and maintains an upward position above the helix bundle leaving room for the ligand to access the putative binding pocket. Figure 4.6a gives the ribbon structure of the backbone atoms of the N-terminus illustrating the general conformation of the domain. Although this extracellular domain is flexible, it has a well-defined structure and a conserved backbone conformation. Similarly, in the intracellular region, the most flexible part of the molecule, ICL3, has a conserved conformation (Figure 4.6b).



(a) (b)

Figure 4.6: (a) The structural similarities of N-terminus as depicted by the superimposed backbone ribbons. The region of highest fluctuations from residue 23-41 maintains a helical turn in most of the conformations. b) Superimposition of the backbone ribbon of third intracellular loop (ICL3). The loop is the most flexible part of the molecule but has conserved conformation and maintains a helical turn in most of its conformations. Sticks indicate the change on orientation of the side chains in Arg253.

The MD simulations of β_2AR in a membrane, including a modeled ICL3, have also shown that the loop has a well-defined structure and its presence lead to a decrease in the distance between H3 and H6 resulting in the formation of an 'ionic lock' [1]. This feature is not found in the engineered β_2AR crystal structures [22,28] nor in β_2AR-based, AKHR. From visual inspection of the trajectory and the secondary structure analysis, the N-terminus and ICL3 both contain a helix for the greater part of the simulation, which is also found in the ordered N-terminus of rhodopsin and the modeled ICL3 in β_2AR [1]. The second intracellular loop was drawn away from the helix bundle in such a way that its residues interact with lipid molecules.

4.3.1.4 Secondary Structure Analysis

The secondary structure variations of each residue were determined by DSSP [34] and plotted against time for each structure (Figure 4.7). The program analyses the secondary structure of a protein by a pattern-recognition process of hydrogen-bonding patterns and geometrical features extracted from X-ray coordinates. Although AKHR is a flexible membrane protein, most structural features found in the initial structures were maintained during molecular dynamics.

4.3.1.4.1 Secondary structure Analysis of helix bundles

GPCRs are characterized by seven trans-membrane α-helices that form the core of the receptors. The flexibility and stability of these helices are crucial for ligand binding and receptor activation. During free MD simulation, the helices (blue regions in Figure 4.7), showed a high degree of stability due to the presence of the motion restricting membrane, interhelical interactions and a disulphide bond between Cys197 in ECL2 and Cys212 at the extracellular end of helix3 [35].

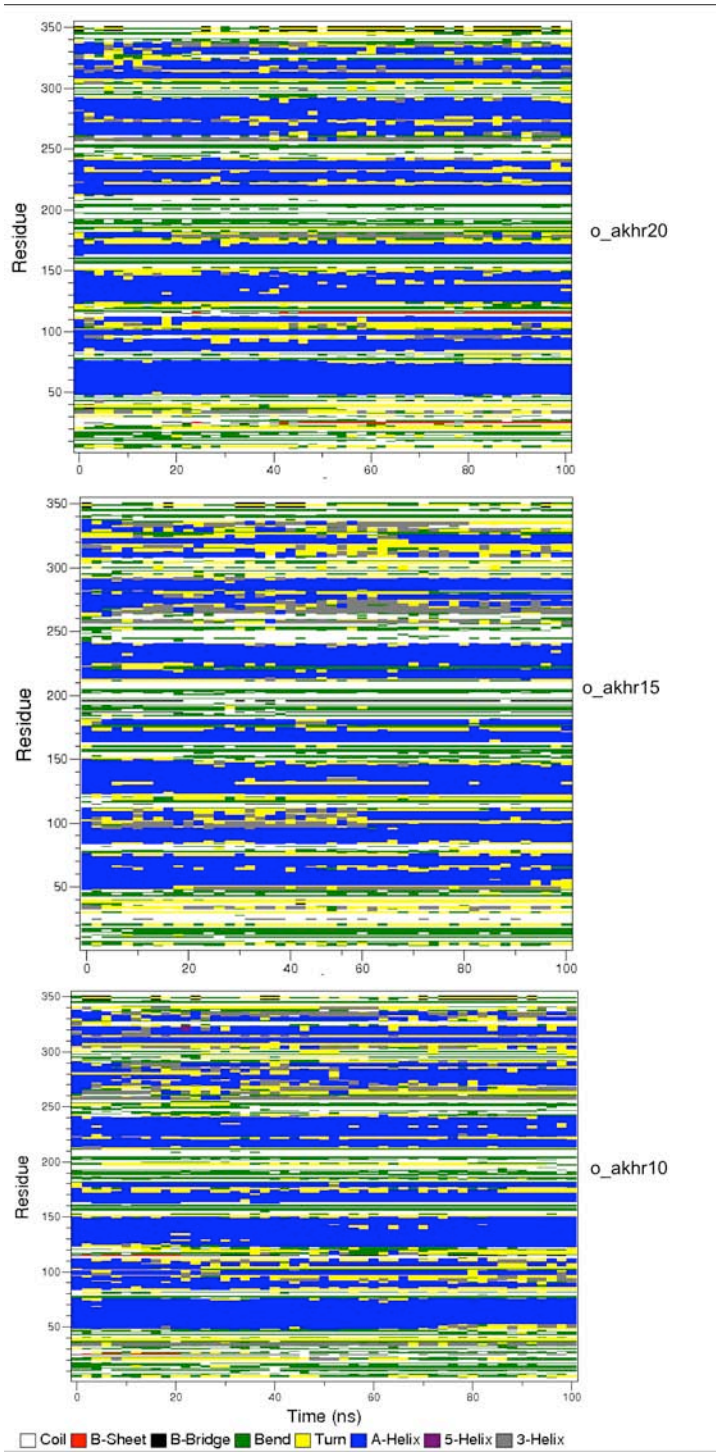


Figure 4.7: Secondary structure for each residue as a function of time for o_akhr10, o_akhr15, o_akhr20 simulations.

All three structures, o_akhr10, o_akhr15 and o_akhr20, had high percentages of α-helicity in the 7TM region. The highest percentage was in H1 (residues 47-75), of o_akhr20 with only two intracellular residues, forming turns (yellow) towards the end of the simulation. In contrast, two extracellular residues, His47 and Ile48 in o_akhr10 switched from α-helicity to turns and 3-helix from 75ns of simulation and in o_akhr15 a turn involving Ser61, Ala62 and Thr63 was on and off from 5ns until the end of the simulation.

Most helices in AKHR were characterised by deformations due to the presence of proline and serine, especially in H2, H3, H5, H6 and H7. Experimental work by Fu, D. *et al*, 1996 [36] has shown that proline-induced kinks involving the highly conserved motifs in H5, H6 and H7 are common features in transmembrane helices and cause twists in the helix. The presence of serine and threonine in the helices modulated the formation of kinks [37] and resulted in turns and bends at serine positions in H5.

Due to the presence of proline in H2, (residues 84-111), the helix had the most distortions involving mostly extracellular residues Met101 to Pro103, Gly107 and the highly conserved Asp94. Figure 4.8 shows the structural fluctuations in H2 at 20ns intervals. Pro103 induces flexibility in the helix causing the secondary structure to fluctuate between turns, bends and 3-helix turns during the simulation. The highly conserved Asp94 plays an important role in interhelical interactions that stabilise the helix bundle and is involved in the formation of a salt bridge with Ser135 in H3 (Chapter three).

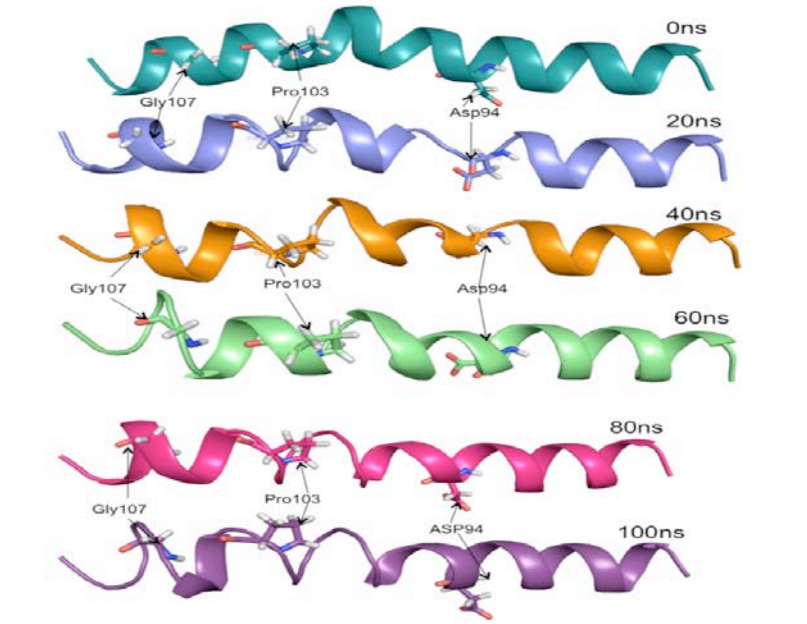


Figure 4.8: Fluctuations in structural features in helix 2 (H2) depicting alternating helical turns and coils at Asp94, Pro103 and Gly107 positions

The changes from α-helix to a turn (yellow) destroy this salt bridge and that caused the helix to drift away from H3.

Although H3 was relatively stable, in o_akhr20, Ser136 and Phe137 formed short turns between 60ns and 70ns and the last intracellular residue switching to and from α-helicity during the dynamics (Figure 4.7). In o_akhr10 and o_akhr15 the number of residues with α-helicity was reduced by 3-4 in the intracellular side of the receptor. H4 (residues 162-176) maintained α-helicity but increased in length by three residues in the extracellular side which then turned into a 3-helix at 20ns in all structures and deformation of the helix was observed at Leu174. Apart from the formation of turns in H5 (212-241) due to the presence of proline and serine, the helix was very stable in all three AKHR structures.

On the other hand, H6 (residues 262-292) lost α-helicity in most of its intracellular and extracellular residues from about 20ns to 80ns in o_akhr10 and o_akhr15 during the simulation. In o_akhr20, the α-helix was stable and turns were formed involving the highly conserved motif CWTPY [37] that contains a proline. H7 (residues 305-325) contains a proline in the highly conserved motif NPVVY, that caused the α-helix to shift to 3-helix from Ser317-Pro321 for the first 20ns and then a turn was formed for the greater part of the simulation in o_akhr10 and o_akhr20. However, in o_akhr15, this helix was very unstable and lost α-helicity in all residues between 60ns and 70ns. There was an outward drift of H7 and a change from α-helix to 3-helix and a turn from Thr315-Ser317 just above the proline induced kink. Residues in H7 participate in ligand binding in most GPCRs [6, 9], so these changes could be crucial in the activation of AKHR. This receptor also contains the eighth helix (H8) perpendicular to the membrane plane, which was stabilised by polar interactions with residues in H1 and H7 in all but o_akhr15. The secondary structures of the residues in this helix alternated between α-helix and 3-helix and in o_akhr15, it disappeared after 60ns.

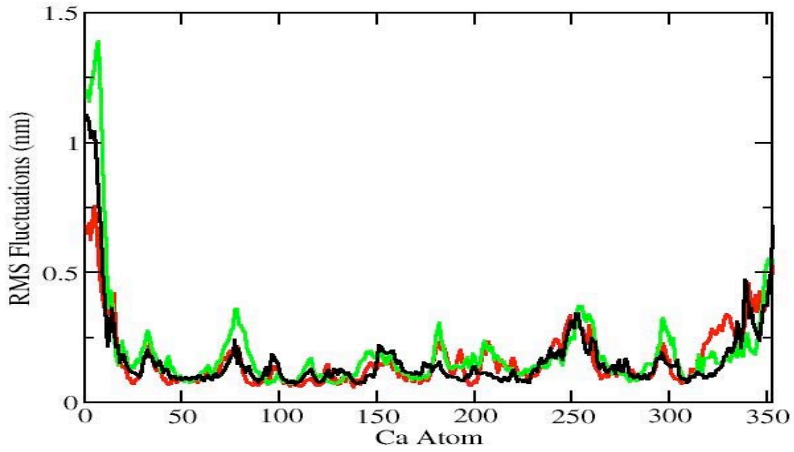
4.3.1.4.2 Secondary Structure Analysis of loop domains

The extracellular and intracellular loops, N- and C-termini were characterised by bends, coils and turns. However, distinct structural features were seen in the N-terminus, ECL1 and ICL3. The N-terminus and ICL3 contained a 3-helix from Glu32-Tyr35 and Val258-Gly260, which were 'on and off' in all simulations. These structural features are supported by the presence of a helical turn in the crystal structure of rhodopsin from Tyr29-Leu31 [39] and the modelled third intracellular loop of β_2AR as found by Dror R. O et al [1]. In o_akhr20, a β-sheet in the N-

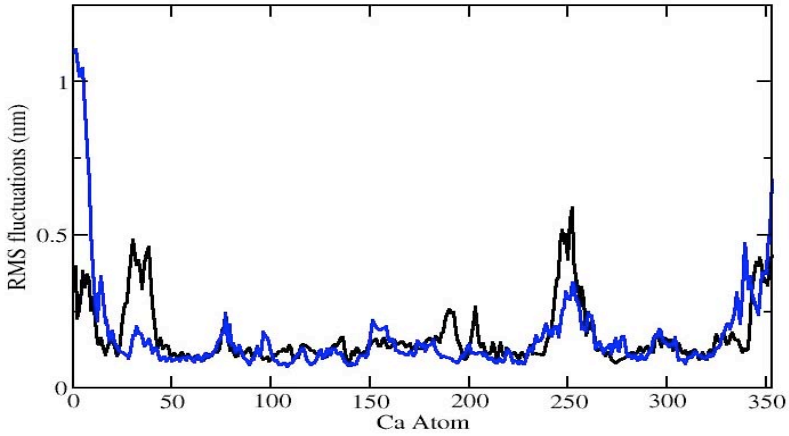
terminus and the first extracellular loop, ECL1, first appeared from 25ns and reappeared from 40ns until the end of the simulation. In o_akhr10, the β-sheet existed for the first 20ns and was absent in o_akhr15. Since these domains are not usually included in crystal structure determinations, these features could be common to all Class A GPCRs and might be crucial in receptor activation and ligand binding. In contrast to the second extracellular loop in rhodopsin and β_2AR, AKHR does not have a helical turn or β-sheets in its second extracellular loop. The loop maintained an upward position above the helix bundle to pave way for diffusible ligands.

4.3.2 Molecular Dynamics simulations of rhodopsin-based AKHR

The starting structures, c_akhr10, c_akhr15 and c_akhr20, for the three 100ns MD simulations of the rhodopsin-based AKHR, were extracted from a trajectory generated from MD simulation of the model in a membrane, with fixed helices at 10, 15 and 20ns. During MD simulation, the overall structure of the molecule and predicted helix regions were maintained. As observed in the β_2AR-based structures, the molecule is highly flexible and fluctuations of the Ca atoms depend on the structural domains as shown by the rmsf values in Figure 4.9a. All three simulations had similar fluctuations. Even though movements in H2 (82-111) were relatively high, the helix bundles were fairly rigid, especially in simulation c_akhr20 (black) where an effective hydrophobic environment around the bundles was created by the membrane. As expected, the highest movements were found in the termini as well as the extra and intracellular regions of the molecule.



(a)



(b)

Figure 4.9: The rms fluctuations of c_akhr10 (red), c_akhr15 (green) and c_akhr20 (black) as a function of time during the 100ns MD simulations. b) A comparison of fluctuations in c_akhr20 (blue) and o_akhr20 (black).

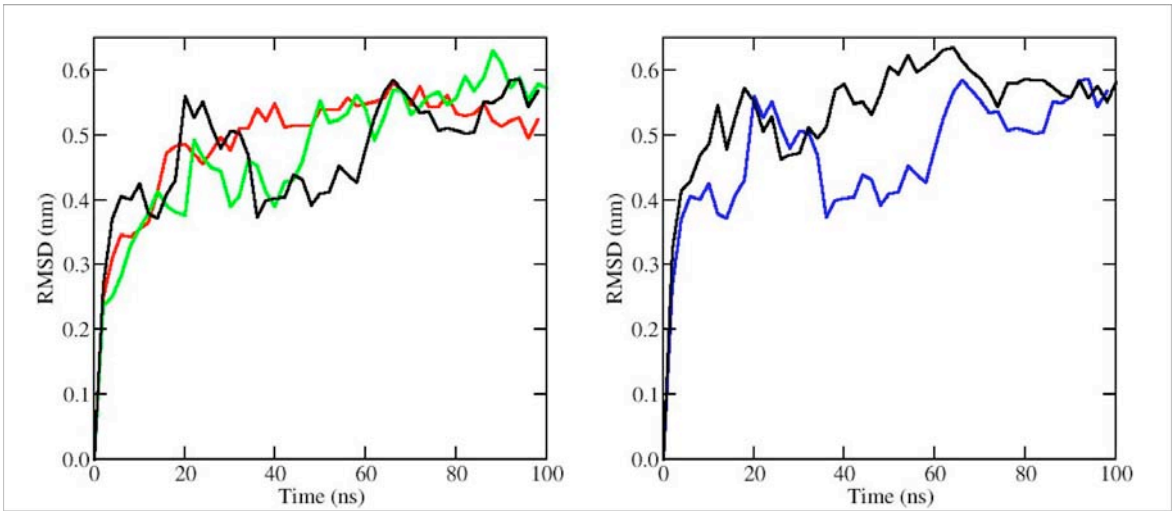
Highest fluctuations with rmsf of ~ 0.6- 1.4nm were found in the first residues in the N-terminus (1-46). This part of the structural domain is stretched away from the rest of the molecule, intermolecular interactions are minimal and have no contact with residues in neither the molecule nor the membrane.

A comparison of simulations o_akhr20 and c_akhr20 points out that fluctuations in the open and closed conformations of AKHR are very similar (Figure

4.9b). Except for the N-terminus, o_akhr20 has a higher rmsf values in most of its extracellular and intracellular regions, especially in ICL3 (243-264), ECL2 (177-211) and residues 23-41 in the N-terminus which have rmsf of ~ 0.5nm. In rhodopsin-based AKHR, the residues are closely packed, flexibility is reduced due to steric hindrance and presence of a dense network of H-bonds and hydrophobic interactions.

4.3.2.3 Deviations of molecular structures from the initial structures.

Figure 4.10 shows the root mean square deviations (rmsd) during the 100ns MD simulations. These values indicate the structural equilibration and the range of molecular deviations from the initial structures of simulations c_akhr10, c_akhr15 and c_akhr20 (Figure 4.10a). In all three simulations, there were significant deviations from the initial structures that gave rise to initial rmsd of ~0.3 to 0.4nm.



(a) (b)

Figure 4.10: a) Variation of the rmsd values during 100ns MD simulations of the closed AKHR structures. The rmsd reached steady state after 30ns but there were fluctuations until 70ns in c_akhr15 (green) and c_akhr20 (black). (b) Comparison of rmsd of o_akhr20 and c_akhr20 simulations. The systems are very similar and both approach steady state at ~ 0.59nm.

Although the three systems are similar, simulation c_akhr10 reached a steady state after 30ns and all systems attained stable conformations after 70ns and maintained rmsd of ~0.59nm. The same rmsd was attained in the β_2AR-based structures and the similarity between c_akhr20 and o_akhr20, is illustrated in Figure 4.10b. Both systems had a sudden increase in rmsd to ~0.4nm within the first 2ns of simulation and finally gained stability 70ns. Therefore, for a given molecule, the final rmsd is independent of the starting structure used.

4.3.2.4 Compactness of the protein molecules

From the variations of radius of gyration (R*g*) values of the structures of rhodopsin-based AKHR during 100ns MD simulations in Figure 4.11, the most compact molecule was in simulation c_akhr10 (red) with average R*g* of ~2.40nm.

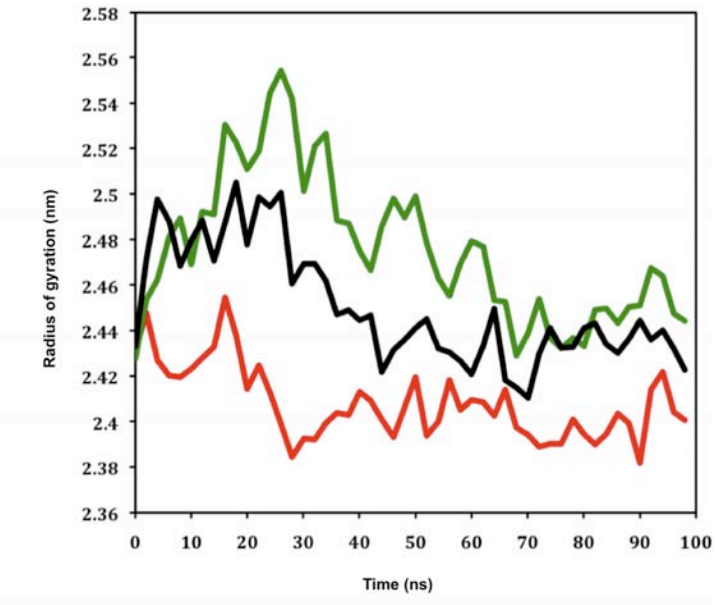


Figure 4.11: a) R*g* variations of rhodopsin-based AKHR during 100ns MD simulations of c_akhr10 (red), c_akhr15 (green) and c_akhr20 (black). The systems reached steady state after about 50ns. Structures from c_akhr10 have the lowest average Rg value ~2.40nm and are the most compact.

In simulation c_akhr15 (green), the molecule was loosely packed resulting in an average R*g* of ~2.46nm. However all systems reached steady state after 50ns and approach final R*g* of ~2.40-2.45nm. As expected, the 'closed' rhodopsin-based structures are more compact than the β_2AR-based AKHR. This is illustrated by ~5% change in R*g* of rhodopsin-based structre in contrast to ~14% in β_2AR-based AKHR.

4.3.2.1 Conformation Changes during simulations

The closed conformation of the rhodopsin-based AKHR was maintained during the 100ns MD simulations in all simulations. However, an overlay of structures from the three simulations gave an rmsd of 0.3 $\pm$ 0.02nm from the vacuum structure, showing that there are conformational and structural changes that cannot be ignored.

4.3.2.1.1 Conformational Changes in the Helix bundle

The general structure and anticlockwise arrangement of the helix bundles was maintained but there was a slight movement of helices away from each other especially in H1, H3, H6 and H7. In Figure 4.12, the arrangements and positions of the helices of the structures in simulation c_akhr20 at 20ns intervals are shown. The movements of helix 1 illustrate the shift of residues and preservation of the closed conformation in the rhodopsin-based structures (H1). H1 shifted downwards and slightly outwards by ~0.18nm from its initial position, in contrast to the marked outward divergence of ~0.6nm in the β_2AR-based structure. The small outward movement of the helix increases its distance between H2 and H7 and enlarges the binding space of the receptor.

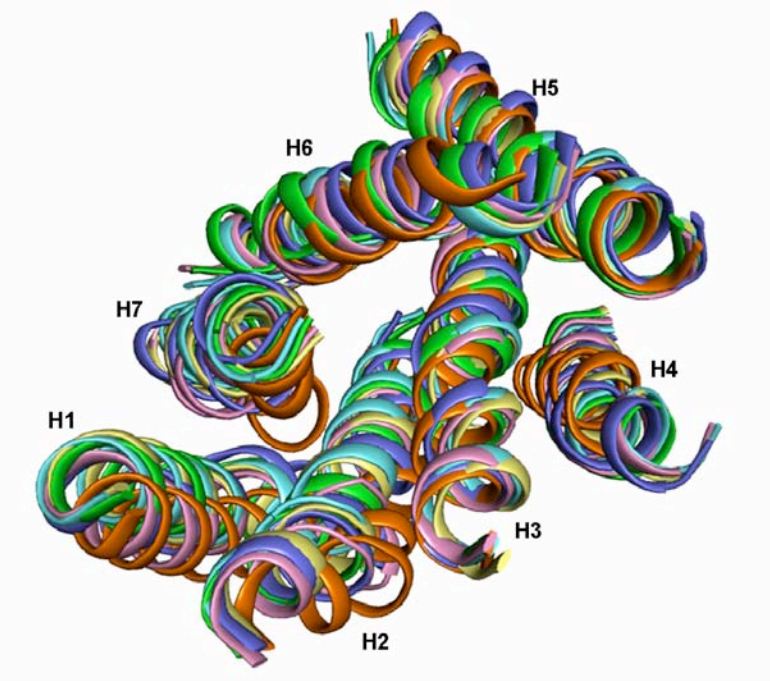


Figure 4.12: An overlay of the helices of rhodopsin-based AKHR indicating their positions at, 0ns (orange), 20ns (purple), 40ns (pink), 60ns (yellow), 80ns (cyan) and 100ns (green).

However, the bend at Ser61 and Thr63 was preserved and this kept the extracellular half of the helix drawn towards the centre of the helix bundle, upholding the open conformation. The closed conformation of a GPCR is found only in the crystal structure of rhodopsin and is essential for the protection of the covalently bound ligand, 11-cis-retinal [33]. These movements could indicate the position of the helix when AKHR is in one of its activated state.

During MD simulation of β_2AR there was an inward movement of H3 and H6, resulting in intermittently formation of an 'ionic lock' between intracellular residues [1]. Even if the same movements were observed in the β_2AR-based AKHR, an 'ionic lock' was not formed. Also ionic interaction between the highly conserved Arg146

(H3) and Lys263 (H6) in the initial structure of the rhodopsin-based AKHR was broken (Figure 4.13).

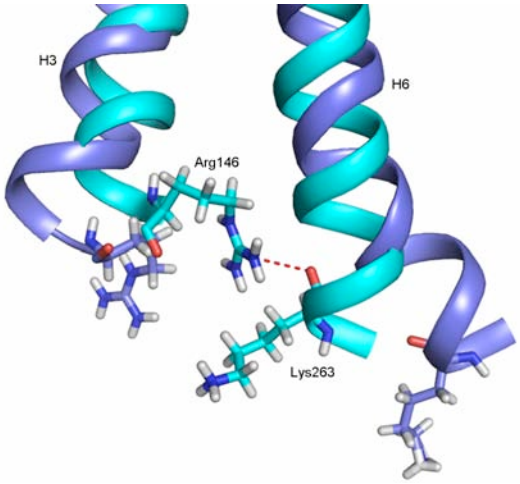


Figure 4.13: An overlay of the intracellular end of helix 3 (H3) and helix 6 (H6) in the starting structure (cyan) and c_akhr20 (purple) after 20ns of simulation. Sticks give the positions of Arg146 and Lys263 side chains in H3 and H6 respectively. The red dotted line represents the ionic bond between Arg146 and Lys263.

This is because, upon MD simulation, the helices moved away from each other, increasing the distance between Arg143 and Lys263 from 0.28nm to 1.42nm (Figure 4.13). An outward movement of these helices is typical during activation of GPCRs and in rhodopsin this leads to breaking of the 'ionic lock' [32,33,40]. The closed extracellular region and the outward movement in the intracellular region of the helix bundles of AKHR is in line with conformations in activated GPCRs and supports the idea that the rhodopsin-based AKHR could represent the bound conformation of the receptor.

4.3.2.1.2 Conformational Changes in loop domains

From the rmsf values in Figure 4.9, there was minimal movement of the loop domains in rhodopsin-based AKHR compared to β_2AR-based AKHR. In the extracellular region the loops were slightly drawn outward to preserve the positions of the loop

domains over the helix bundles and the closed conformation. Although the N-terminus had conserved conformation from residues 10-46, its amino end had undefined conformation, from Met1 to Asn10 (Figure 4.14).

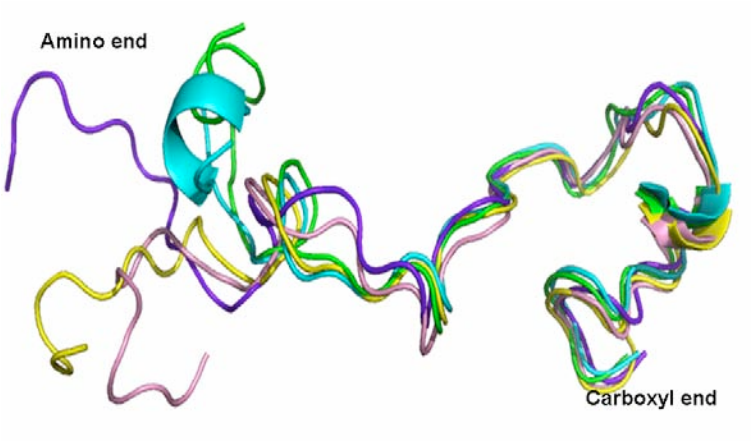


Figure 4.14: An overlay of the ribbon structure of the N-terminus of rhodopsin-based AKHR giving conformations at 20ns interval during the simulation. The amino end, Met1-Asn10, is highly flexible and has undefined structure.

Though the overall structure of the carboxyl end is similar to that of β_2AR-based AKHR, the terminus is collapsed over the helix bundles, a position similar to that of the N-terminus in the crystal structure of rhodopsin [40].

Even though the second extracellular loop (ECL2) does not have β-sheets as found in the crystal structure of rhodospin [39], it is crumpled over the helix bundle and lies below the N-terminus, ECL1 and ECL3. The loop is kept in position by the Cys197-Cys212 disulphide bond and formed a cap over the helix bundle closing possible binding sites. In contrast, in the crystal structure of β_2AR [22] and β_2AR-based AKHR, ECL2 is withdrawn away from the centre of the helix bundle, exposing it more to the solvent and is positioned above the ECL1 and ECL3 to leave space for any in-coming diffusible ligand.

Whilst ICL2 was drawn out and upwards from the helix bundle, the most flexible part of the intracellular loop domains, ICL3 (22 residues), was opened up as H6 moved away from H3 and H5. However, the loop has a preserved conformation of the backbone atoms and the structure is well defined (Figure 4.15).

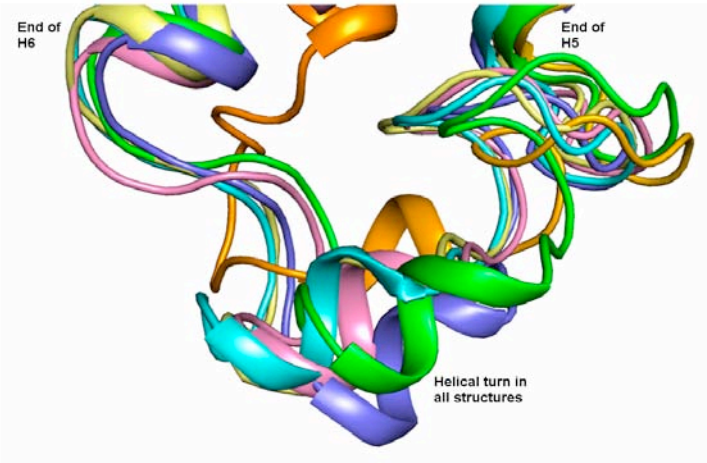


Figure 4.15: An overlay of the third intracellular loop in rhodopsin-based AKHR at 20ns intervals during the simulation.

A helical turn that was observed in the modelled β_2AR ICL3 loop [1] and in previously discussed β_2AR-based AKHR is seen from Ser250-Ser254 in the rhodopsin-based AKHR. In contrast to β_2AR-based AKHR, the secondary structure occupies the most flexible part of the loop and hence its position shifts during simulation. The recurrence of the helical turn, in ICL3 of different receptors could be an indicator of the presence and importance of this structural feature in Class A GPCRs.

4.3.2.2 Secondary Structure Analysis

The DSSP programme was used to analyse the secondary structure of rhodopsin-based AKHR molecules and the results are given in Figure 4.16. It was observed that the helical regions (blue) of the receptor were maintained during the 100ns MD simulations. Whilst some structural features in the highly flexible extracellular and intracellular loop regions were preserved, the intracellular helix (H8), which is normally parallel to the membrane surface, was surprisingly unstable in all three simulations.

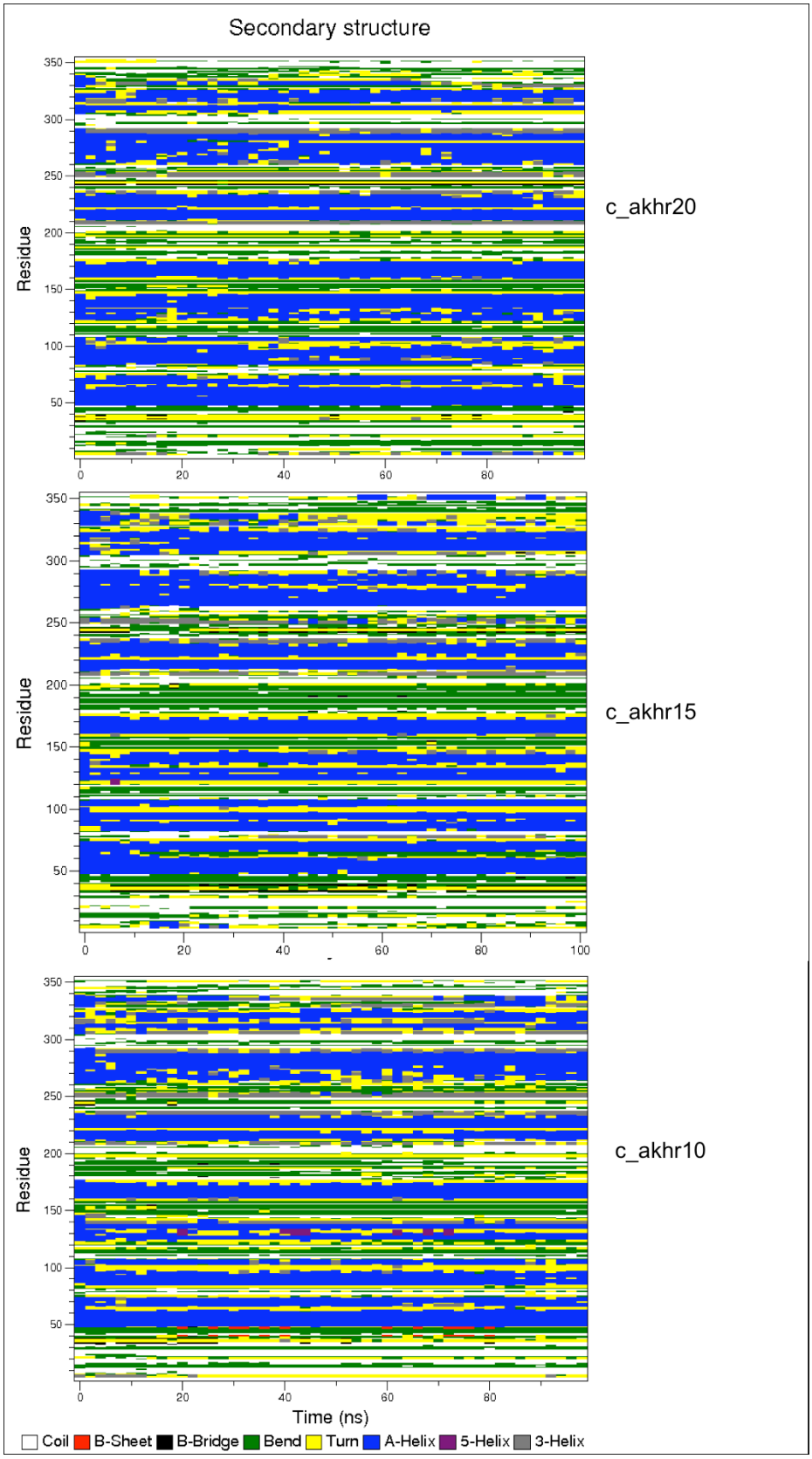


Figure 4.16: Secondary structure for each residue as a function of time for akhr10, akhr15, akhr20 simulations.

4.3.2.2.1 Secondary Structure Analysis of helix bundle

In all three simulations, the α-helices were relatively stable, even though the presence of proline, serine and threonine induced distortions in H2, H3, H4, H5, H6 and H7 [40,41]. The percentage α-helicity increased from simulation c_akhr10 to c_akhr20 and this could be attributed to the contribution of the lipid molecules to the stability of the helix bundles. The membrane creates a hydrophobic environment for the helix bundles and induces helix structures in the transmembrane regions [41]. Tight packing of lipid molecules around the helix bundles in c_akhr20 allows for efficient intercalation of the aromatic side chains with the lipid chains [42] and increases the formation of interhelical hydrogen bonds [43,44,45].

Whilst helix distortions were absent in H1 of β_2AR-based AKHR, in rhodopsin-based AKHR the helix is characterised by deformations due to the loss of α-helicity from Ala54 to Gly64 in c_akhr10 and Tyr54 to Thr56 in c_akhr20. The turns provided a flexible hinge for the helix to be drawn towards the centre of the helix bundle in the closed conformation. On the other hand, H2 displayed proline-induced kinks in all simulations. Unexpectedly 30% of residues in H3 lost α-helicity, especially in c_akhr10 where some residues formed 5-helix. In addition to the proline-induced kinks in H5, at Met220 -Thr222, and H6, at Val278-Trp281, the helices reduced in length by three residues at both the extracellular and intracellular ends in c_akhr10 and c_akhr15. Except for the loss of α-helicity by residues at the intracellular end of H6, both helices were fairly stable in c_akhr20. In simulation of c_akhr10, more than 50% of residues in H7, formed turns and the helix almost disappeared during the simulation. However, α-helicity was maintained in c_akhr15 and c_akhr20 as the lipid molecules were closely packed around the helix bundle.

Surprisingly, H8, which was very stable in β_2AR-based AKHR and is commonly found in Class A GPCR [26], disappeared after 30ns in all simulations.

Helix-helix interactions contribute more to the final stage of structure formation of membrane proteins [41] and stabilise GPCRs in their inactive states. The instability of most α-helices in the rhodopsin-based structure could be due to unfavourable contacts between the side chains of the residues and steric hindrance. This is because the closed conformation of the rhodopsin template resulted in a closed packing of residues in the helix bundle of the receptor model. In an attempt to avoid these contacts, a high percentage of the residues in the helix bundle lost α-helicity. It was observed that the largely hydrophobic side chains of an α-helix contact the hydrophobic membrane core [46] especially when the lipid molecules were closely packed to stabilise the helix bundle. This was not the case in c_akhr10 and c_akhr15 simulations where the lipid molecules were loosely packed, hence high distortions were observed.

4.3.2.2.2 Secondary Structure Analysis of loop domains

Coils, turns and bends were a common feature in the termini, extra and intracellular loop domains of the receptor in all simulations. The N-terminus also had flashes of a 3_{10}-helix from Asn3-Ala6 for the first 12ns which was 'on and off' from 20ns to 80ns. Although a second helix appeared at a similar position to that in the N-terminus of β_2AR-based AKHR and the crystal structure of rhodopsin [39], a β-sheet was formed briefly from residue 39-40 during the c_akhr10 simulation. In addition, the salt-bridge found between Tyr24 and Thr115 in β_2AR -based AKHR was absent in the rhodopsin-based structures. The N-terminus and EC1 are further apart from each other and the side chain of Thr115 is oriented away from the centred of the helical

bundle, reducing the chances of contacts with residues in the N-terminus. Whilst β-sheets found in ECL2 of rhodopsin [39], are absent in the same loop of AKHR, a transient 3-helix was formed close to the extracellular end of H5, which appeared to be an extracellular extension of the helix.

In the intracellular region, a 3-helix was formed in ICL3 throughout the simulation. This structural feature involved residues Ser250-Ser254 at the centre of the loop and is comparable to the helical turn found in the third intracellular loop of β_2AR during MD simulations in a membrane [1]. In the β_2AR-based structures, the helix appeared at a more rigid position involving residues Asp257-Gly260, which are close to the intracellular end of H6. As has been noted earlier, the presence of a conserved helical turn in ICL3 could be important in G-protein activation.

4.4 Conclusion

MD simulations to optimise the β_2AR-based and the rhodopsin-based adipokinetic hormone receptor in their inactive states have been successfully performed. MD simulations of GPCRs in vacuum are usually performed with fixed helices to maintain α-helicity but in membrane simulations, this is not necessary as the membrane stabilised the helices. In its true environment, β_2AR-based AKHR prefers a wider helix bundle whereas the rhodopsin-based structure maintained a closed conformation. The open conformation would allow for the free movement of the large AKH peptides in and out of the binding site. Although the rhodopsin-based model had an ionic bond involving the highly conserved Arg146, it was observed that this ionic bond is absent in both AKHR models during molecular dynamic simulations. The extracellular and intracellular domains have conserved conformation as shown by the highly flexible N-terminus and third intracellular loop.

These structural domains consistently form β-sheets and helical turns, especially in β_2AR-based AKHR. These structural features could be important in ligand binding and receptor activation. An important observation is that β_2AR-based AKHR has stable, seven transmembrane helices, which are crucial for ligand binding and receptor activation in GPCRs. The stability, conformational and structural changes of AKHR during MD simulations indicate that close packing of the membrane molecules around the helical bundle, as in o_akhr20, stabilised the helices and maintains α-helicity. The 3D structure of AKHR from the mosquito and the study of its molecular dynamics gives insights into the structures of insect GPCRs and pave way for docking studies that could lead to the design of non-peptide mimetics that can be used as alternative insecticides for the malaria mosquito.

4.5 References

1) Dror, R. O., *et al*. Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Nat. Acad. Sci*, **106**(12), 4689-4694 (2009).
2) van der Spoel , Vogel H. J and Berendsen H.J.C. Molecular Dynamic Simulations of N-terminal Peptides From a Nucleodite Binding Protein, *PROTEINS: structure, Function and Genetics*, **24**, 450-466 (1996).
3) Ivetac, A and Sansom, M. S. P. Molecular dynamics simulations and membrane protein structure quality. *J. Eur. Biophys*, **37**, 403-409 (2008).
4) Balázs Jójárt, Róbert Kiss, Béla Viskolcz, and György M. Kerserű. Activation Mechanism of the Human Histamine H4 Receptor – An Explicit membrane Molecular Dynamics Simulation Study. *J. Chem. Inf. Model,* **48***,* 1199-1210, (2008).
5) Rivail, L., Chipot, C., Maigret, B., Bestel, I. Sicsic, S., and M. Tarek. Large – scale molecular dynamics of a G protein coupled receptor, the human 5-HT4 serotonin receptor, in a lipid bilayer. *J. Molec. Structure (THEOCHEM)* **817**, 19-26 (2007).
6) Spijker, P., Vaidehi, N., Freddolino, P. L., Hilber, P. A. J. and Goddard III, W. A. Dynamic behaviour of fully solvated B2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. *PNAS*, **103**, (130), 4882-4887 (2006).
7) Anézo, C., de Vries, A. H., Höltje, H-D., Tieleman, D. P. and Marrink S-J. Methodology Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B* **107***,* 9424-9433, (2003)
8) Tieleman, D. P., Marrink, S.J. and Berendsen, H.J.C.. A computer perspective of membranes: Molecular dynamics studies of lipid bilayers systems**,** *Biochim. Biophys Acta* **1331**, 235-270 (1997).
9) Tieleman, D. P., van der Spoel, D., and Berendsen, H.J.C. Molecular dynamics simulations of dodecylphosphocholine micelles at three different aggregate sizes: micellar structure and lipid chain relaxation**,** *J. Phys. Chem. B* **104**, 6380-6388 (2000).
10) Delano, W., L. "The PyMol Molecular graphics system." Delano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org

11) Humphrey, W., Dalke, A. and Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics*, **14**, 33-38 (1996).

12) van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., E. and Berendsen, H., J., C. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.* **26,** 1701-1719 (2005)

13) Kaminski, G., A., Friesner, R., A., Tirado-Rives, J., and Jorgensen, W., L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, **105**, 6474-6487 (2001).

14) Darden, T. A., York, D. M., Pedersen, L. G. Particle Mesh Ewald: an *N* log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).

15) Essmann U., Perera L., Berkowitz, M. L., Darden, T, Lee, H., Pedersen L.G. A smooth particle mesh Ewald method. *J. Chem. Phys*, **103**, 8577-8593 (1995).

16) Shrivastava I. H., Capener, C. E., Forrest, L. R. & Sansom, M. S. P. Structure and dynamics of K Channel Pore-Lining Helices: A Comparative Simulation Study. *J. Biophys*. **78**, 79-92 (2000).

17) Elling, C. E., Nielsen, S. M. & Shwartz T. W. Conversion od antagonist-binding site to metal-ion site in the tachykini NK-1 receptor. *Nature*, **375,** 74-77 (1995).

18) Schlessinger, A and Rost, B., Protein Flexibility and Rigidity Predicted From Sequence. *Proteins: Structure, Functions, and Bioinformatics,* **61**, 115-126 (2005).

19) Lobanov M.Y., Bogatyreva, N. S., and Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compatness, *Molec. Bio*, **42**, 623-628 (2008).

20) Katsuaki, I., Naoto, Y., Yosihiro, U. and Takashi, I. Compact Packing of Lipocalin-type Prostaglandin D Synthase Induced by Binding of Lipophilic Ligands. *J. Biochem*, **145**(2), 169-175 (2009).

21) Elling, C. E. and Schwartz T.W., Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.* **15**, 6213-6219 (1996).

22) Cherezov, V. *et al.* High – resolution crystal structure of an engineered human β2- adrenergic G- Protein coupled Receptor. *Science* **318**, 1358-1264 (2007).

23) Okada, T., Fujiyoshi, Y., Silow, M., Navarro, J., Landau, E. M., and Shichida, Y. Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc. Natl Acad Sci, USA*, 99, 5982-5987 (2002).

24) Mirzadegan, T., Benko, G., Filipek, S. and Palczewski, K. Sequence analysis of G-protein-coupled receptors: similarities to rhodopsin. *Biochem*, **42**, 2759-67 (2003).

25) Ballesteros, J. A., Shi, L., and Javitch, J. A. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. *Mol Pharmacol*, **60,** 1-19 (2001b).

26) Kristiansen, K. Molecular mechanism of ligand binding, signalling, and regulation within the superfamily of G-protein-coupled receptors: molecular modelling and mutagenesis approaches to receptor structure and function. *Pharmac Therapeutics* **103**, 21-80 (2004).

27) Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Reviews*, **21**, 90-113 (2000).

28) Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2 adrenergic G-protein coupled receptor. *Nature* 450(7168), 383-387 (2007).

29) Hanson, M. A. et al. A specific cholesterol binding site is established by the 2.8Å structure of the human beta2-adrenergic receptor. *Structure*, **16**(6), 897-905 (2008).

30) Kaufmann, C. and Brown, M. R. Regulation of carbohydrates metabolism and flight performance by a hypertrehalosaemic hormone in the mosquito *Anopheles gambiae*. *J. Ins. Phys.* **54**, 367-377 (2008).

31) Jójárt, B. and Martinek T. A. Performance of the General Amber Force Field in Modeling Aqueous POPC Membrane Bilayers, *J. Comput. Chem.* **28**(12) 2051-2058 (2007).

32) Kobilka, B. G- protein coupled receptor structure and activation. *Biochim Biophys. Acta* **1768**(4), 794-807 (2007).

33) Kobilka, B. and Schertler, G. F. X. New G-protein coupled receptor crystal structures: Insights and limitations. *Trends Pharmacol Sci* **29**(2), 79-83 (2008).

34) Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).

35) Fu, D., Ballesteros, J. A., Weinstein, H., Chen, J. and Javitch, J. A. Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding site crevice.*Biochem,* **35**, 11278-11285 (1996).

36) Deupi, X., Olivella, M., Govaets, C., Ballesteros, J. A., Campillo, M. and Pardo, L. Ser and Thr residues modulated the conformation of Pro-kinked transmembrane α-helices. *Biophys. J.* **86,** 105-115 (2004).

37) David, T. J. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics,* **23**(5), 538-544 (2007).

38) Palcezewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science* **289** (5480), 739 – 45 (2000).

39) Fanelli, F. and De Benedetti P. G. Computational Modelling Approaches to Structure-Function Analysis of G Protein-Coupled Receptors. *Chem Rev*, **105**, 3297-3351 (2005).

40) Hironori Kokubo, and Yuko Okamoto, Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange Monte Carlo simulation, *Chemical Physics Letters* , **392**, (1-3), 168-175 (2004).

41) Johansson, A. C. V., and Lindahl**,** E. Amino-Acid Solvation Structure in Transmembrane Helices from Molecular Dynamics Simulations, *Biophysical Journal*, **91** (12), 4450-4463, (2006).

42) Curran, A. R., and Engelman, D. M. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr Opin Struct Biol* **13**, 412–417 (2003).

43) Choma, C, Gratkowski, H., Lear, J. D., and DeGrado, W. F. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.*, **7,** 161–166 (2000).

44) Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger A. T., and Engelman, D. M. Interhelical hydrogen bonding drives strong interactions in membrane proteins, *Nat. Struc.t Biol.*, **7**, 154 - 160 (2000)

45) Meindl-Beinker, N. M., Lundin, C., Nilsson, I., White, S. H., and von Heijne , G. Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly, *EMBO Rep*. **7**(11), 1111–1116 (2006).

46) Lemmon, M. A., and Engelman, D. M. Specificity and promiscuity in membrane helix interactions, *FEB Letters,* **346** (1), 17-20 (1994)

CHAPTER FIVE

Docking calculations: Identification of binding pocket in AKHR and MD simulations of bound receptor.

5.1 Summary

The binding site and accessibility of the binding pocket in AKHR were determined by perfoming blind docking (BD) calculations of the adipokinetic hormone, Del-CC (pGlu-Lys-Asn-Phe-Ser-Pro-Asn-Trp-Gly-Asn-NH_2) using Autodock4 and Autodock Tools [1,2,3]. It was observed that the helix bundle and the extracellular domains defined the binding pocket of AKHR. The open β_2AR-based AKHR structure facilitated hormone binding, whilst the closed rhodopsin-based model had limited access of the binding pocket. To determine the effect of ligand binding on the conformation of the receptor, MD simulations of AKHR bound to an active insect AKH were performed. Upon MD simulation of β_2AR-based AKHR-ligand complex, there was a sudden shift from an open to a closed conformation of the helix bundle in the extracellular side of the receptor within the first nanoseconds of simulation. Simultaneously, in the intracellular side helices H3, H5 and H6 moved away from each other. The protein-ligand complex was stabilised by an intense network of H-bonds and formation of salt bridges between residues in the helices and the extracellular loop domains, and the ligand. Tyr285 in H6 was bound to the ligand throughout the simulation and could play an important role in ligand binding and receptor activation.

5.2 Experimental Methods

5.2.1 Docking Calculations

5.2.1.1 Identifying the binding pocket of AKHR

Blind docking (BD) calculations [4,5,6] of Del_CC were perfomed using Autodock4 and Autodock Tools [1] to identify the binding site and accessibility of the binding pocket of the β_2AR- and rhodopsin-based 3D structures of AKHR. BD is a method that involves scanning the entire extracellular surface of the receptor by the ligand to identify suitable binding sites in the receptor. Autodock has been found to be useful in blind docking where the location of the binding site is not known [4,5,7,8,9,10]. The highly active adipokinetic hormone, Del-CC (pGlu-Lys-Asn-Phe-Ser-Pro-Asn-Trp-Gly-Asn-NH_2), [2,3] which regulates mobilisation of carbohydrates in the blister beetles, was used as the ligand since its structure was readily available in our laboratory.

Docking parameters for macromolecules and the ligand (Appendices D and E) were set based on those used by Hetenyi *et al* [4]. The β_2AR-based receptor model was used in the determination of the binding site and the two models were used in determining the accessibility of the pocket from the extracellular side. Based on the positions of the binding pockets in rhodopsin and β_2AR [11], a variety of grid maps were generated in the extracellular region of β_2AR-based AKHR, using the AutoGrid4 program [1,4,5]. The most suitable grid map with grid points of 64 x 64 x 50 in xyz, and a spacing of 0.05nm, was used for the for blind docking calculations. The grid box was centred at the Lsy306 HZ1 atom, to include all residues with side chains oriented towards the centre of the helix bundle. The ligand was flexible with 30 active torsions and the Lamarckian genetic algorithm (LGA) was employed to

generate conformations of the ligand within the binding pocket [1,6,7,8]. For a population size of 150, the maximum number of generations and energy evaluations were set at 100 x 10^6 and all other parameters were set to default AutoDock4 values. The molecule was subjected to 50 trials of BD for 100 runs, to search for the binding site in AKHR. Cluster analysis was executed at a cut-off of 2Å. For the first four trials, the conformation of lowest energy was obtained from the most populated cluster and the results were used to carry out more focused BD calculations. Results from these calculations confirmed the position of the binding pocket of AKHR. From the protein-ligand complex with lowest estimated free energy of binding (ΔG_b) [6] the binding pocket of AKHR was identified. To determine the accessibility of the binding site, in the rhodopsin-based model, the procedure was repeated using grid points 64 x 50 x 66 in xyz.

5.2.1.2 Identifying the binding mode of Del_CC

To determine the binding mode of Del_CC and refine the docking, further calculations were performed for 100 runs using a rigid receptor molecule and a flexible ligand with 32 released torsion angles. Both ligand and receptor parameter files were prepared using Autodock Tools (Appendix). Grid points (64 x 64 x 50 in xyz) which focused on the preferred binding pocket, were applied to generate grid maps with a spacing of 0.05nm using the program AutoGrid4. A total of 500 x 10^6 generations and energy evaluations, and LGA were applied for a population of 300. From the results, the conformation of lowest ΔG_b was used to further optimise the protein-ligand complex.

5.2.2 Molecular dynamics (MD) Simulations of bound AKHR in a Membrane

To further optimise the docking, MD simulations were carried out with both protein and ligand molecules flexible. The protein-ligand complex of lowest ΔG_b from section 5.2.1.2, was inserted into a previously prepared membrane (Chapter 4) to represent its true environment. This gave the receptor/membrane complex shown in Figure 5.1 in which water molecules are included.

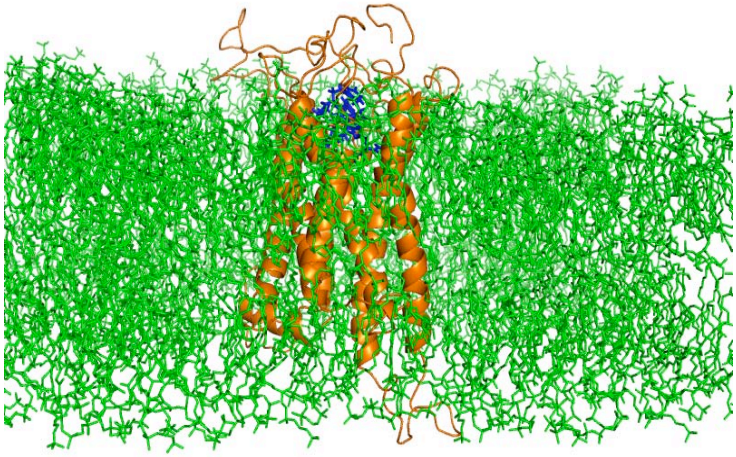


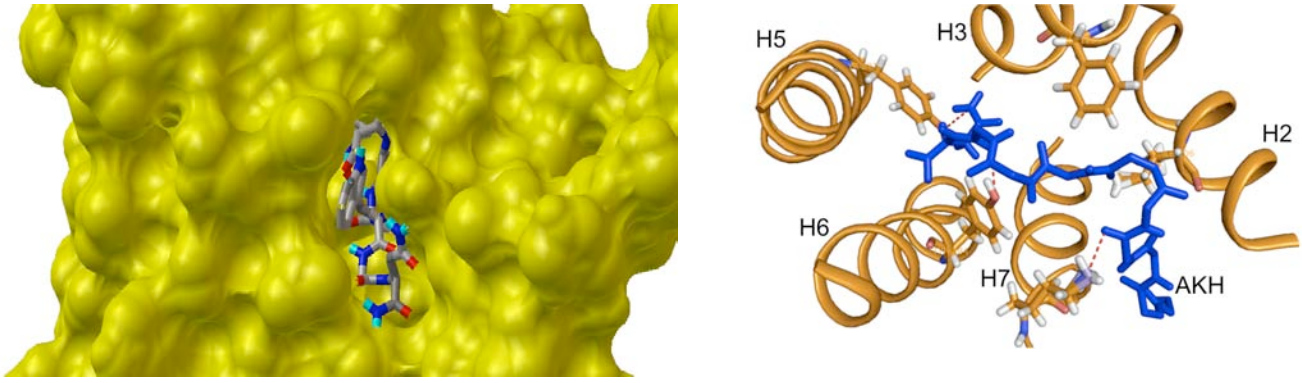
Figure 5.1: A complex of docked AKH/Del_CC (blue) and the receptor, AKHR (orange) inserted into a POPC membrane (green). Water molecules are not shown.

Based on the MD simulation results in Chapter four, the system was energy minimised for 3000 steps of L-bfgs and equilibrated by subjecting it to position restrained MD simulations for 20ns to effectively soak the protein by the lipid and water molecules. The starting structure extracted after 20ns was subjected to 100ns MD simulations by applying the protocol and parameters described in Chapter four. Structures extracted from the MD trajectory at 10ns intervals were rescored to obtain the final values of the estimated free energy of binding. The modified scoring function of Autodock 3.0 described by Hetenyi, C et al [12] for large and flexible peptidic molecules was applied.

5.3 Results and Discussion

5.3.1 Identifying the binding pocket of AKHR

From the 50 trials of blind docking, the ligand, Del_CC, occupied a position defined by the helix bundle, the N-terminal and the extracellular loop domains. In this putative binding site the ligand was in close contact with residues in helices 2, 3, 5, 6, and 7, the N-terminus, ECL2 and ECL3. Figure 5.2a gives the surface of the extracellular side of the receptor indicating the position of the agonist in the binding pocket and the interactions between Del_CC and helix residues are given in Figure 5.2b.



(a) **(b)**

Figure 5.2: a) Surface of AKHR (yellow) indicating the ligand, Del_CC (coloured by atom) occupying the binding pocket in the receptor. b) Helices 2, 3, 5, 6, and 7 defining the binding pocket of adipokinetic hormone receptor (orange). The ligand, (blue) forms hydrogen bonds with Gln209 (second extracellular loop, not shown), Lys307 (third extracellular loop), Tyr285 (H6) and Tyr221 (H5).

The position of the putative binding pocket and involvement of the N-terminus, extracellular loop domain and the extracellular end of the helices, in AKHR is in good agreement with experimental data from mutational mapping of binding sites for peptide GPCRs from both vertebrates and invertebrates [13,14]. From the study of

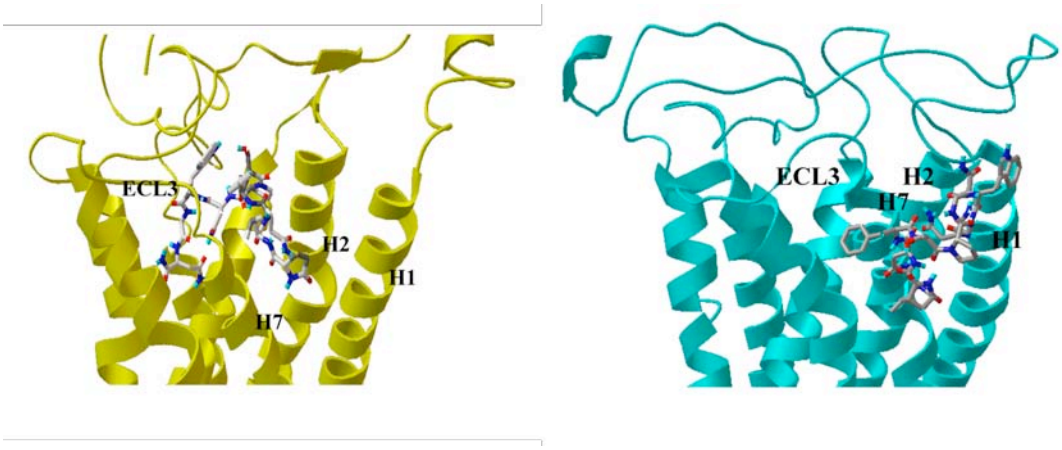
the tachykinin system, the exchange of the extracellular loop segments between neurokinin-1 and neurokinin-3 receptors have shown that the extracellular domain and the helix bundle are both involved in binding [15,16]. Chimera studies of insect peptide GPCR, Pheromone-biosynthesis-activating-neuropeptide receptor (PBAN-R), also demonstrated the importance of the extracellular domain and the disulphide bridge in ligand binding [17,18]. This position was also similar to the location of the binding pocket in the human gonadotropin releasing hormone receptor [19] but differs from that of rhodopsin and β_2AR, which are buried inside the helices [14,20].

Having obtained the position of the preferred binding site of AKHR, more focused docking was carried out to determine the binding mode of Del_CC. It was found out that in the conformers of lowest estimated free energy of binding, the ligand formed H-bonds and salt bridges with Tyr221, Tyr285, Gly209 and Lys307 as shown in Figure5.2b. ΔG_b for ligand/receptor binding was -38.5 kJ/mol with an estimated inhibition constant of 168 nM. This indicates that the receptor has high affinity for Del_CC. The protein-ligand complex is stabilised by an intense network of H-bonds and ionic interactions between the receptor and ligand residues.

5.3.1.1 Accessibility of the binding pocket in β_2AR- and rhodopsin-based AKHR

AKHs occur as octa-, nona-, or decapeptides that are released from the brain into the circulatory system of the insect body [21,22]. At the fat body, the ligands enter the binding pocket of their GPCRs from the extracellular side [22,23,24]. Accessibility of the binding site by these diffusible ligands is therefore crucial for ligand binding. When the hormone, Del_CC, diffused into the binding pockets of both AKHR models, it easily accessed the binding pocket of the open β_2AR-based structure

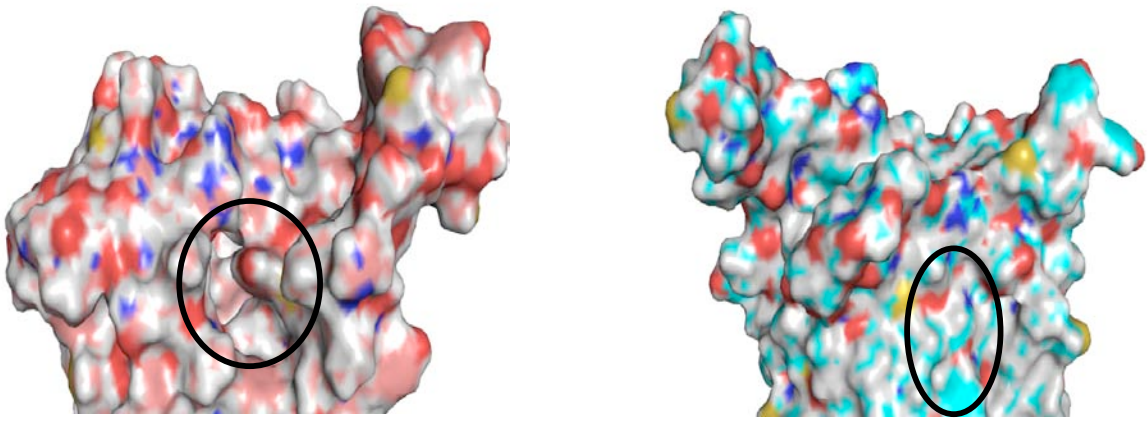
(Figure 5.3a) to occupy the putative binding site. In contrast, in the closed, rhodopsin-based conformation of AKHR, while it was possible to force the ligand into the identified binding pocket, the space between the helices, loops and the N-terminus was too small to allow the ligand free access to the pocket from the extracellular side. Instead it was found to bind to the periphery of the binding pocket between helices 1 and 7 (Figure 5.3b).



(a) **(b)**

Figure 5.3: Binding pocket accessibility of (a) the β_2AR-based and (b) the rhodopsin-based AKHR. The ligand is right inside the binding pocket in the β_2AR based AKHR but only occupies a peripheral position in the rhodopsin based AKHR.

As observed earlier, the use of rhodopsin and β_2AR as templates for the helix bundles results in models of AKHR with different conformations. β_2AR-based AKHR has an open conformation, that gives room for free movement of the ligands in and out of the binding pocket due to open cavities on the extracellular surface (Figure 5.4a). On the other hand, the rhodopsin-based structure has a closed conformation that does not leave accessible cavities on the extracellular surface of the receptor (Figure 5.4b).



(a) (b)

Figure 5.4: Surface structure of a) the β_2AR-based 3D structure of AKHR. The black ring shows the deep cavity between H1, H7 and N-terminus in the extracellular region of the receptor. b) The rhodopsin-based 3D structure of AKHR. There was no opening between H1, H7 and N-terminus as indicated by the black ring.

The loop domains and the helices of closed AKHR are drawn close to each other, therefore, the accessibility of the binding pocket is reduced and hormone access is inhibited. It is important to note that, when the grid map was focused on the putative binding pocket during docking with rhodopsin-based AKHR, the ligand was forced into the binding site and the lowest ΔG_b was higher than +100 kJ/mol. It is postulated then, that the β_2AR-based structure could represent the conformation of the bound/inactive AKHR, while the rhodopsin-based structure could give information on the active, bound, state of the receptor.

5.3.2 Molecular dynamics Simulations of bound AKHR in a Membrane

Since docking calculations were executed with a rigid protein molecule and Autodock4 performed structural optimisation of the ligand, MD simulations were performed to further refine and optimise the β_2AR-based AKHR-ligand complex of lowest ΔG_b [6] in a typical bilayer membrane environment. The MD simulations were aimed at checking the changes in conformation and structural features of AKHR

after binding to an agonist. In addition, the stability and mode of binding of the adipokinetic hormone, Del_CC, in the putative binding site were investigated.

5.3.2.1 Conformational Changes in bound AKHR

There were drastic conformational changes in the whole receptor molecule within the first 1.5ns of MD simulations of bound AKHR. These were maintained throughout the simulation as the molecule converted from an open, inactive, to a closed, active, state. This drastic receptor response to activation by the AKH is in contrast to the slow transition of β_2AR to its active conformation during dynamics of the adrenaline (epinephrine) bound receptor [25]. However, the fast attainment of the active conformation by AKHR is consistent with the high activation rates observed when noradrenaline induced changes in fluorescently labelled β_2AR in a detergent, with half-lives of 2.8 and 70s for the fast and slow phases and within ~40ms [26,27,28,29].

Changes in conformation of the protein molecule in the presence of the ligand were induced by the mobility of the atoms. A comparison of the mobility of the atoms in unbound (o_akhr20) and bound AKHR is given by the rmsf of the Ca atoms for each residue during a 100ns simulation (Figure 5.5). The results reflected similar mobility in most parts of the molecules in both simulations. As expected large fluctuations, rmsf of ~ 0.4-0.5nm, were seen in the loop domains and the termini. In contrast to rmsf of ~0.1nm for the helices of o_akhr20 (black), the helix bundles of bound AKHR were highly mobile as shown by high rmsf of ~0.2nm for H1 (47-75), H2 (84-111), H3 (122-149) and H4 (162-176).

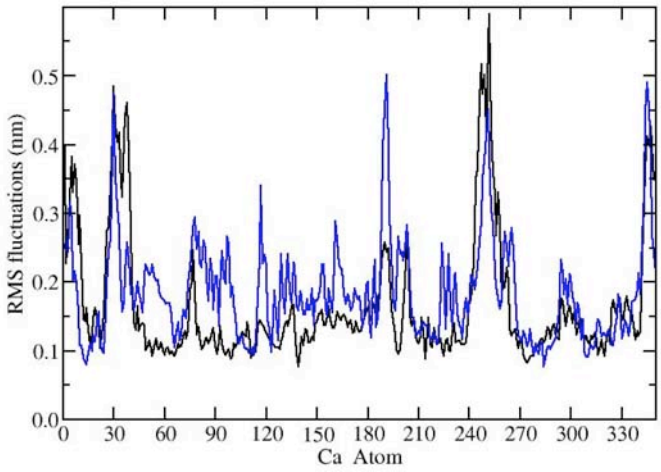
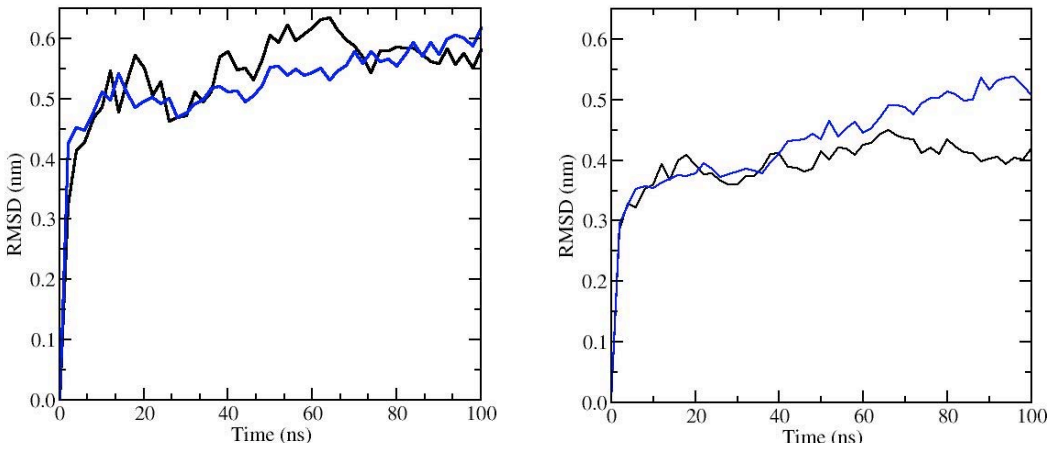


Figure 5.5: The rms fluctuations of o_akhr20 (black) and bound AKHR (blue) as a function of time, during 100ns MD simulations.

This could be attributed to the changes from α-helicity to turns and bends especially in H1 and H7 (308-325) as the helices interacted with the ligand. Mobility of helices was also observed in the photoactivated form of rhodopsin as illustrated by the movements of the middle part of H3 [30]. In o_akhr20, the helices had low flexibility and dynamics as indicated by low B-factor values since the tightly packed membrane molecules restricted motion. However flexibility and mobility of helices in the presence of an agonist are essential for receptor activation [31,32] and accounts for the drastic change in conformation of the receptor during ligand binding.

The high molecular movements and the structural changes account for the high deviations of the bound AKHR molecule from its starting structure as shown by the rmsd in Figure 5.5 for both unbound (black) and bound (blue) AKHR. Both simulations deviated from their starting structures but the bound receptor deviated the most (Figure 5.5a).



(a) (b)

Figure 5.5: Variation of the rmsd values during 100ns MD simulations of unbound (black) and bound (blue) AKHR. (a) Rmsd for both unbound AKHR structures and the bound structure initially increased to 0.30 – 0.40nm. b) Helices in bound AKHR attained equilibration after 40ns and there was an increase until the end of the simulation.

Within the first 1.5ns of simulation, there was a steep rise in rmsd to ~0.45nm as helices were drawn inwards in the extracellular region and outwards in the intracellular region of bound AKHR. From about 10-40ns, the system reached steady state with an rmsd of ~0.53nm. High fluctuations from 2-18ns were due mainly to the collapse of the highly flexible extracellular and intracellular domains. A closer look at the rmsd of the helices indicated that conformational changes in helices due to ligand binding contributed more to the deviations than did the loop regions (Figure 5.5b). The helices reached steady state from 2 – 40ns with an rmsd of ~0.35nm. However, there was a steady rise in rmsd from 40 – 100ns due mainly to changes from α-helical to 3-helical, turns and bends of some residues in the helix bundle. This indicates that the molecule had passed its steady state period and given more simulation time, the rmsd would continue to increase.

Irrespective of the high deviations and fluctuations in bound AKHR molecules, the radius of gyration (Rg) values [33,34] shown in Figure 5.6, indicated that the molecule was more compact than unbound AKHR in both systems. The initial Rg was ~2.43nm and an average Rg of ~2.47nm was attained during the simulation.

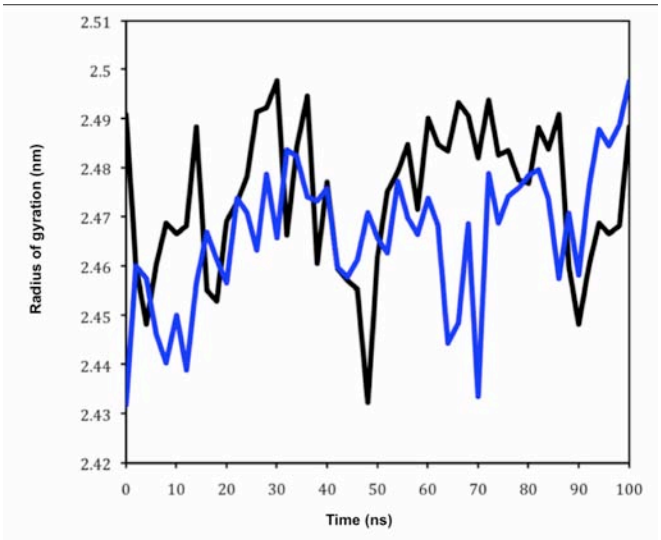


Figure 5.6: Radius of gyration as a function of time for o_akhr20 (black) and bound AKHR (blue) during a 100ns MD simulation.

Fluctuations in Rg of bound AKHR could be due to the changes in secondary structure within the helix bundle as shown from the analysis results by DSSP. On the other, hand, the initial Rg for the unbound receptor was higher by ~6% and its average by ~1%. From these results, it can be concluded that ligand binding increases compactness of the receptor molecule.

5.3.2.1.1 Conformational Changes in helix bundle

Although the overall structure of the AKHR molecule was maintained, ligand binding induced a variety of conformational changes in the helices of the β_2AR-based AKHR model. There was an inward movement of the helices in the extracellular region and an outward shift in the intracellular region within the first 1.5ns of simulation. The change in conformation is clearly shown by an overlay of unbound and bound AKHR molecules after simulation (rmsd) shown in Figure 5.7 (movie on CD).

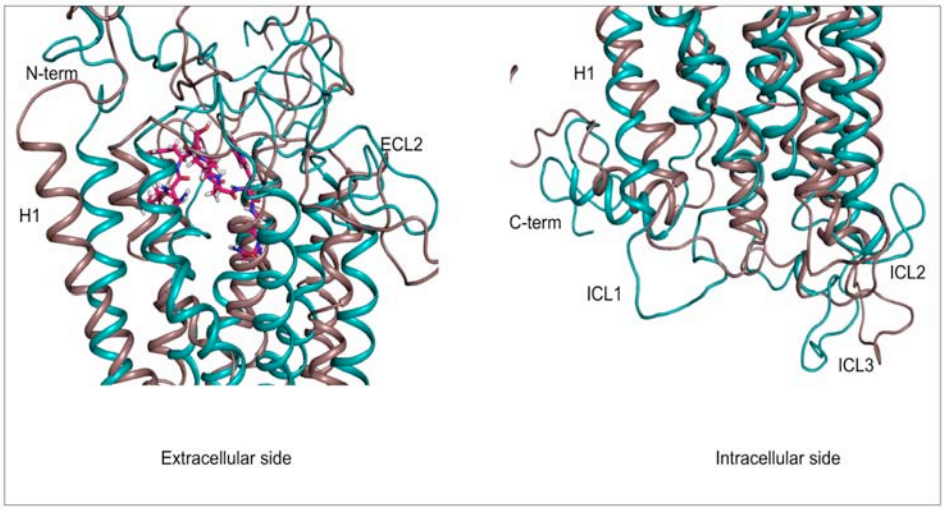


Figure 5.7a: An overlay of unbound (brown), and bound (green) AKHR helices. In both structures, the helix bundles are very stable and most residues maintained α-helicity. An inward movement of the extracellular regions of the helices results in a closed conformation, similar to that of rhodopsin-based AKHR. Consequently, in the intracellular region the helices moved away from each other and the loops, ICL1, ICL2 and ICL3 were pushed outwards.

In the extracellular side three helices H1, H5 and H7 had a significant tilt towards the centre of the helix bundle. To illustrate shifts of H1 in unbound and bound AKHR respectively, Figure 5.7b shows the positions of this helix relative to its position in the starting structure. The orientation of the extracellular terminal residue, His47, in the helix is also shown.

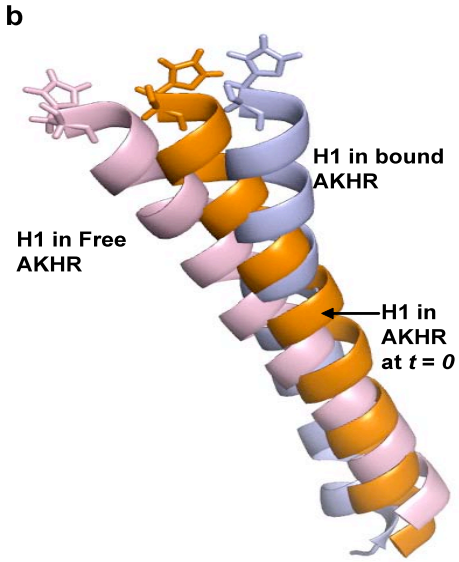


Figure 5.7b: Positions of helix 1 in unbound/free AKHR (pink), and bound AKHR (light blue) relative to the initial structure of AKHR (orange) after MD simulation for 40ns. In free AKHR H1 is deflected outward by ~0.6 nm. In contrast, there is an inward deflection by ~0.5nm and a downward shift of H1 in bound AKHR due to ligand binding. Sticks show the orientation of His47 at the extracellular end of the helix.

The Ca atom in His47 at the extracellular end of H1 was shifted outwards by 0.6 $\pm$ 0.02nm in free/unbound AKHR. In contrast, the helix was drawn towards the centre of the helix bundle by ~0.4nm from its initial position in bound AKHR, to give a total distance of 1.0 $\pm$ 0.2nm between His47 in unbound and bound AKHR. This distance is equal to the distance between the ends of H1 in the β_2AR-based and rhodopsin-based models as shown in Figure 3.7.

Upon activation by an agonist, conformation changes took place to accommodate the bound ligand. During activation of GPCRs, such conformational changes are important for the transmission of signal from the extracellular region to the intracellular region of the receptor, activation of the heterotrimeric G-protein and production of a cascade of signal [31,32]. Similar changes have been observed in β_2AR, when the presence of an agonist, epinephrine, caused some of the helices to tilt and become closer to each other [35], resulting in a closed conformation in the extracellular region. These changes are crucial in that they reflect the shift from inactive to active state of the receptor.

In the intracellular side of AKHR H1, H3, H5 and H6 are drawn away from the centre of the bundle, giving it an open conformation in that region. Experimental data support the shift of H6 away from H3 in the intracellular region of activated GPCRs [25,31]. In rhodopsin the movement of the cytoplasmic end of H6 was demonstrated by, for example, restraints due to engineered metal ion binding [36]. In β_2AR the most important movements were found to involve the outward divergence of H6 in the intracellular region or its clockwise rotation leading to the disruption of the 'ionic lock' [24]. Similarly, in bound AKHR, H6 was heaved away from both H3 and H5. This led to the destruction of the ionic interactions between H5 and H6, which could lead to increased basal activity, and activation of the G-protein.

Rhodopsin has a covalently bound 11-cis-retinal molecule within its binding crevice [37,38]. This ligand is coupled to Lys296 in H7 through formation of a Schiff base. The receptor therefore maintains a closed conformation in the extracellular region to protect the ligand from hydrolysis [11]. Likewise, the presence of an agonist in AKHR, initiated a tilt and an inward movement of the helix

bundles, and the loops collapsed over the helix bundles to gain a closed conformation that protected the ligand. The closed conformation of AKHR and the outward movement of H6, in the intracellular region indicate that AKHR/Del_CC complex gives an activated state of the receptor. This conformation can be used to provide more information on the binding pocket, binding mechanism in insect GPCRs and improve the calculated binding energies for agonists [39]. In addition to the results in Chapter four, these observations supported the idea that the open β_2AR-based AKHR model represented the inactive state of the receptor that shifted to a closed conformation upon receptor activation.

5.3.2.2 Secondary Structure Analysis of bound AKHR

The changes in secondary structure of each residue in bound AKHR molecule, as determined by the DSSP programme [40], are presented in Figure 5.8.

5.3.2.2.1 Secondary structure Analysis of helix bundle

As shown earlier, the blue regions indicated the α-helices, which form the core of the receptor and play an important role in ligand binding. Even though ligand binding stabilised most of the helices, it was seen that the percentage of α-helicity fluctuated during simulation and drastic structural changes were induced in most helices. Changes were clearly seen in H1 where α-helicity in the extracellular side was stable until 40ns, and then His47-Leu49 shifted to a turn. On the other hand, the intracellular residues shifted from α -helicity to turns, B-bridges and 3-helicity throughout the simulation. The percentage of α -helicity in most residues was reduced by more than 30% for the greater part of the simulation.

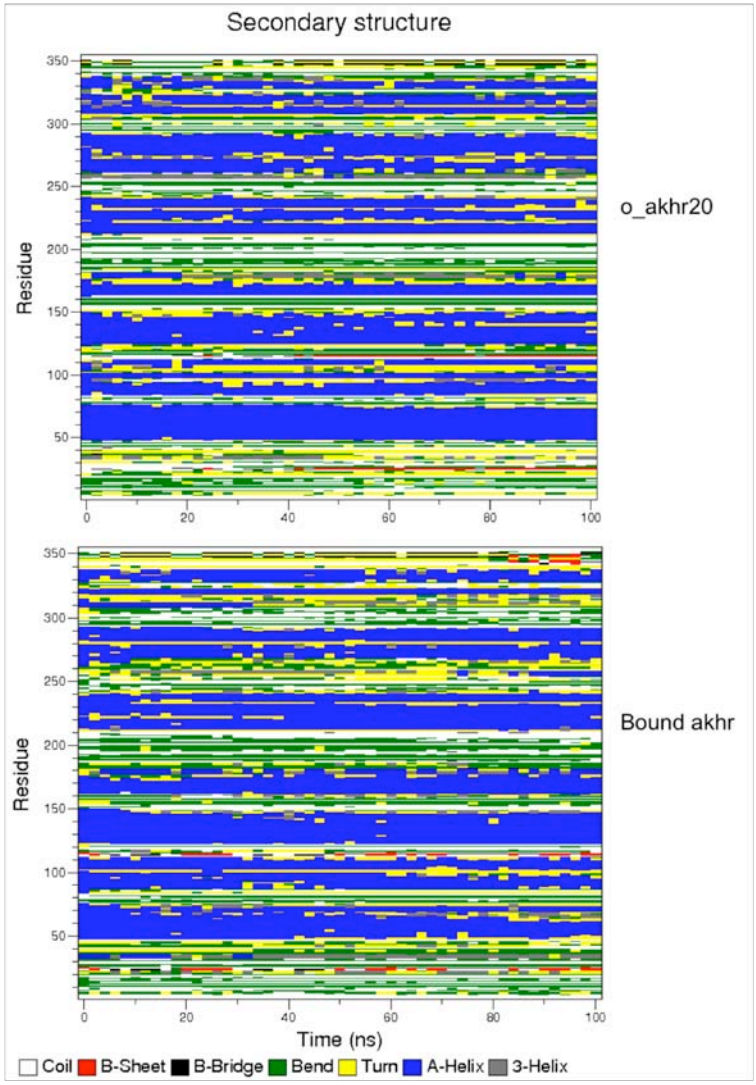


Figure 5.8: Change in secondary structure of residues in bound AKHR during 100ns MD simulation of o_akhr20 (top) and bound AKHR (bottom).

The change in structural features in this helix was due to the inward movement in the extracellular side and the outward movements in the intracellular region that accounts for the change in conformation and the high deviations from the starting structure discussed earlier.

As shown by the o_akhr20 results, in unbound AKHR, H2 (residues 84-111) had the most structural distortions involving mostly extracellular residues Met101 to

Pro103, Gly107 and the highly conserved Asp94. In bound AKHR, the interactions between H2 residues and the ligand stabilised the helix. However, in addition to the turn at Pro103 position, two to three residues in the extracellular side lost α-helicity. Also in H3 the turn that was formed in o_akhr20, was absent in the bound structure. Although turns were formed in H4 during simulations, there was a higher percentage of α -helicity compared to o_akhr20 especially in the extracellular side.

Proline induced kinks, enhanced by deformations due the presence of serine and threonine, were a common feature in most of the helices especially in H5, H6 and H7 [41,42,43] and resulted in turns and bends at serine positions in H5. These kinks and distortions acted as hinges for the free outward movements of the helices in the intracellular side, during ligand binding. As a result, 2-3 intracellular residues in H5 and H6 lost α-helicity. On the other hand, there was an inward drift of H7 and a change from α-helicity to 3-helicity and a turn from Thr315-Ser317 just above the proline induced kink. The high extent of distortions and kinks in this helix induced flexibility, which could be important in ligand binding and receptor activation.

The eighth helix, a common attribute in Class A GPCRs [13,38,44], was stabilised by an intrinsic network of H-bonds with H1 residues. However, turns where intermittently formed between 10 and 20 ns and after 50ns of simulation. The main structural features in most residues alternated between α-helicity and 3-helicity. Since the eighth helix is believed to play an important role in G-protein activation [32], these changes could be important for that purpose in AKHR.

5.3.2.2.2 Secondary structure Analysis of loop domains

Due to high flexibility, the extra and intracellular loops, and the termini were characterised by bends, coils and turns similar to those observed in the unbound

receptor. However, the 3-helix formed in the N-terminus (Glu32-Tyr35) and the third intracellular loop (Val258-Gly260) in o_akhr20 was preserved throughout the simulation of the bound receptor. This confirmed the presence of these structural features in AKHR, which are also found in rhodopsin [38] and the modelled third intracellular loop of β_2AR [44].

On the other hand the β-sheet (red) that only appeared after 40ns in the N-terminus (Tyr24-Tyr25) and ECL1 (Trp114-Gly117) of o_akhr20 was 'on and off' in the bound structure throughout the simulation due to participation of Tyr110 in ligand binding. There was closing and opening of side chains of Thr115 in ECL1 and consequently, the salt bridge between Tyr24 and Thr115 was intermittent resulting in the 'on and off' β-sheets. As in rhodopsin [45], these β-sheets formed a shield that aided in stabilising the ligand inside the binding pocket. The protective effect was also enhanced by the inward collapse of the second extracellular loop. This loop was kept in position by the Cys121-Cys197 disulphide bond. As highlighted in Chapter four, the recurring helical turn and β-sheets in the N-terminus and the extracellular loop domains indicate that they could be present in most Class A GPCRs and are important in ligand binding and receptor activation.

5.3.2.3 Mode of binding and stability of Del_CC

Docking calculations involving a flexible Del_CC molecule and a rigid AKHR molecule generated the receptor/ligand complex. The complex was inserted into a POPC bilayer membrane to more closely represent its true environment. This was then subjected to 100ns MD simulations with both ligand and receptor molecules flexible. During the simulation, the ligand was stable and remained in the putative

binding pocket of AKHR. There was a slight change in conformation of the ligand, which involved an upward shift of the middle part consisting of Asn3-Asn7. This part of the molecule was exposed to the extracellular loop domains and the N-terminus (Figure 5.9).

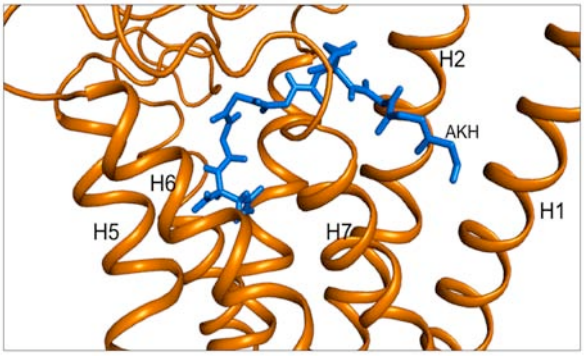


Figure 5.9: Position and conformation of AKH (Del_CC) inside the binding pocket of AKHR.

The N- and C-termini of the ligand were drawn closer by approximately 0.4 nm and downwards to interact with residues in the helices. This conformation was stabilised by a dense network of intramolecular ionic bonds involving Asn3, Phe4, Pro6 and Asn7 (Figure 5.10). Due to the closed conformation of the helix bundle and an inward collapse of the extracellular domain of AKHR, the binding pocket reduced in size to protect the ligand. In addition, the adipokinetic hormone, Del_CC, was stabilised by H-bonds, salt bridges and hydrophobic interactions between ligand residues and Tyr110 (H2), Ile106 (H2), Phe126 (H3), Tyr221 (H5), Tyr285 (H6), Trp292 (H6), Phe312 (H7), Ile211 (ECL2), Gln209 (ECL2) and Lys307 (ECL3) of the receptor.

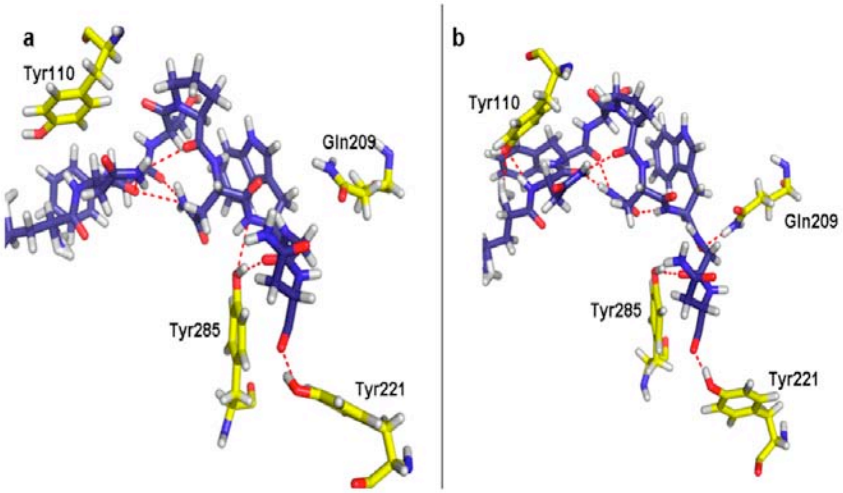


Figure 5.10: Interactions between the ligand, Del_CC (blue) and some of the residues in the binding pocket of AKHR after a) 20 ns and b) 40 ns of simulation. On and off hydrogen bonds and salt bridges are formed between Tyr110 in H2 with ligand residues Leu2 and Asn3, Gln209 in ECL2 with Trp8 and Tyr221 in H5 with Asn10. Tyr285 formed salt bridges with either Trp8 or Gly9 or both.

The participation of both the helix bundle and the extracellular domains is in agreement with results from mutational mapping of binding sites for peptide GPCRs by Fong T. H. et al, 1992 [46], and chimera studies of insect peptide GPCRs by Rafaeli, A, 2009 [19].

The ligand forms polar contacts with three tyrosine residues in H2, H5 and H6, and glutamine in the second extracellular loop. In Figure 5.10, it is shown that hydrogen bonding between Tyr110 (H2) and Gln209 (ECL2) is 'on and off' throughout the simulation. After 20ns (Figure 5.10a), both Tyr110 and Gln209 were drawn upwards increasing the distance between their side chains and the residues

Leu2, Asn3 and Trp8 in the agonist. This gives the residues an open orientation, which switches back to a closed position, as shown in Figure 5.10b.

Tyr285 and Tyr221 in H6 and H5 respectively played important roles in ligand binding by continually forming salt bridges with ligand residues. Tyr285 formed ionic bonds with either or both Trp8 and Gly9 throughout the simulation as shown in Figure 5.10. The clockwise and anticlockwise movements of H6 and the upward orientation of the side chain of Tyr285 at the centre of the helix bundle, makes it easy for the residue to change position of the OH atom and interact with more than one residue of the ligand. The flexibility of H5 due to proline-induced kinks gives Tyr221 different orientations so that it has access to the C-terminus of the ligand. The participation of H6 and H5 side chains in agonist binding was established by mutation analysis of β_2AR with agonists like epinephrine, isoproterenol and related catecholamine compounds [26,35]. Hence, with more docking calculations the participation of these helices in activation of AKHR could be confirmed.

5.3.2.3 Binding affinities

Since Del_CC is a long and flexible ligand, docking calculations were rescored using Autodock3 [12]. Structures were extracted from the 100ns MD trajectory at 10ns intervals and the results from rescoring are given in table 5.1.

Table 5.1: Estimated free energy of binding, inhibition constant and final binding energy of Del_CC.

Conformations of Del_CC	ΔG_b (kJ/mol)	KI (nM)	E_b (kJ/mol)
1_0ns	-38.50	168	---
2_10ns	13.73	----	-5.52
3_20ns	-30.23	5800	-69.38
4_30ns	25.12	----	-17.58
5_40ns	-4.10	19	-44.67
6_50ns	-18.97	47600	-64.64
7_60ns	-36.68	363	-77.37
8_70ns	-33.95	114	-76.45
9_80ns	-30.90	3920	-74.40

❖ (2_10ns denotes the second conformer extracted after 10ns. Other conformers are similarly labelled).

From the results it is shown that binding affinity for Del_CC fluctuates and most of the studied conformations had high affinities. The average estimated free energy of binding was -33.92kJ/mol, which indicates a generally stable ligand/receptor complex.

5.4 Conclusion

The binding pocket in the adipokinetic hormone receptor was succefully identified by performing docking calculations of an insect hormone, Del_CC and MD simulations of the receptor-ligand complex. As in most peptide GPCRs, the helix bundles, the extracellular loop domains and the N-terminus defined its active site. Although both open β_2AR-based and closed rhodopsin-based AKHR could accommodate the ligand, the β_2AR-based/ligand complex had higher binding affinity.

In adition, the open conformation facilitated ligand binding by providing easy access to the binding pocket.

During MD simulation, there was fast activation of AKHR that resulted in a change from an open to a closed conformation of the molecule in the extracellular region. Simultaneously, the intracellular side of the receptor opened up, activating the receptor. The agonist Del_CC, was stabilised by polar contacts with residues in Helices 2, 5, 6 and the second extracellular loop. Throughout the simulation, Tyr285 constantly interacted with the agonist and its importance in ligand binding and receptor activation could be verified by performing more docking calculations of AKH from *Anopheles gambiae*. The stability of the AKHR molecule and the identification of its binding pocket should go a long way in the design of non-peptide mimitics that can be used as alternative insecticides.

5.5 References

1. Morris, G., M. *et al*. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **19,** 14, 1639-1662 (1998).
2. Nair, M., Jackson, G. E., and Gade, G. Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics. *J. Comp. – aided Mol. Des.***15**, 259 – 270 (2001)
3. Gade, G. and Auerswald, L. Flight substrates in blister beetles (Coleoptera: Meloidae) and their regulation by neuropeptides of the AKH/RPCH family. *Eur. J. Entomol.* **96,** 331-335 (1999).
4. Hetényi, C., and van der Spoel, D. Efficient docking of peptides without prior knowledge of the binding site. *Prot. Sci.***11**, 1729-1737 (2002).
5. Hetenyi, C. and van der Spoel, D. (2006) Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Letters*, **580**(5): 1447-1450.
6. Manetti, F., Tintori, C., Armand-Ugón, M., Clotet-Codina, I., Massa, S., Ragno, R., Esté, J. A., and Botta, M. A Combination of Molecular Dynamics and Docking Calculations to explore the Binding Mode of ADS-J1, a Polyatomic Compound Endowed with Anti-HIV-1 Activity. *J. Chem. Inf. Model.* **46**, 1344-1351 (2006).
7. Kovacs, M., Toth, J., Hetenyi, C., Malnasi-Csizmadia, A., and Sellers, J.R. Mechanism of blebbistatin inhibition of myosin II. *J. Biol. Chem*, **279**(34), 35557-35563 (2004).
8. Bikadi, Z., Hazai, E., Zsila, F., and Lockwood, S.F. Molecular modeling of non-covalent binding of homochiral (3S, 3'S)-astaxanthin to matrix metalloproteinase-13 (MMP-13). *Bioorg. Med. Chem.*, **14**(16), 5451-5458 (2006).
9. Hazai, E., Bikadi, Z., Zsila, F., and Lockwood, S.F. Molecular modeling of the non-covalent binding of the dietary tomato carotenoids lycopene and lycophyll, and selected oxidative metabolites with 5-lipoxygenase. *Bioorg. Med. Chem.*, **14**(20), 6859-6867 (2006).

10. Iorga, B., Herlem, D., Barre, E., and Guillou, C. Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the "blind docking" approach. *J. Molec. Model.*, **12**(3), 366-372 (2006) .
11. Kobilka, B. and Schertler, G. F. X. New G-protein coupled receptor crystal structures: Insights and limitations. *Trends Pharmacol Sci* **29**(2), 79-83 (2008).
12. Hetényi, C., Paragi, G., Maran, U., Timár, Z., Karelson, M., and Penke, B. Combination of a Modified Scoring Function with Two-Dimensional Descriptors for Calculation of Binding Affinities of Bulky, Flexible Ligands to Proteins. *J. Am. Chem. Soc.* **128**(4), 1232-1239 (2006).
13. Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Reviews*, **21**, 90-113 (2000).
14. Berthold, M., and Bartfai, T. Modes of Peptide Binding in G Protein-Coupled Receptors. *Neurochem. Res.* **22**(8), 1023-1031 (1997).
15. Elling, C. E. and Schwartz T.W., Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering. *EMBO J.* **15**, 6213-6219 (1996).
16. Elling, C. E., Nielsen, S. M. and Shwartz T. W. Conversion of antagonist-binding site to metal-ion site in the tachykini NK-1 receptor. *Nature*, **375,** 74-77 (1995).
17. Rafaeli, A. Pheromone biosynthesis activating neuropeptide (PBAN): Regulatory role and mode of action. *Gen. Comp. Endocri.,* **162**, 69-78 (2009).
18. Choi, M-Y., Fuerst, E-J., Rafaeli, A. and Jurenka, R. Role of extracellular domains in PBAN/pyrokinin GPCRs from insects using chimer receptors. *Insec. Biochem and Mol. Bio.* **37,** 296-306 (2007).
19. Sealfon, S., C., Weinstein, H. and Millar, R., P. Molecular mechanisms of ligand interaction with the gonadotropin releasing hormone receptor. *Endocr. Rev.* **18,** 2, 180-205 (1997).
20. Bhattacharya, S., Hall, S. E., Li, H., and Vaidehi, N. Ligand stabilization states of human β2 - adrenergic receptor: Insight into G- protein coupled receptor activation. *Biophys.* **94**, 2027-2042 (2008).
21. Riehle, M. A., Garczynski, S. F., Crim, J., W., Hill, C. A. & Brown, M. R. Neuropeptides and peptide hormones in *Anopheles gambiae*. *Science* **298**, 172-175 (2002).

22. Gade, G. and Auerswold, L. Flight substrates and their regulation by a member of the AKH/RPCH family of neuropeptides in cerambycidae. *J. Ins. Phys.* **46**, 1575 – 1584 (2000).
23. van der Horst, D. J. Insect adipokinetic hormones: release and integration of flight energy metabolisation. *Comp. Biochem. and Phys. Part B* **136**, 217 – 226 (2003)**.**
24. Kaufmann, C. and Brown, M. R. Regulation of carbohydrates metabolism and flight performance by a hypertrehalosaemic hormone in the mosquito *Anopheles gambiae*. *J. Ins. Phys.* **54**, 367-377 (2008).
25. Huber, T., Menon, S., and Sakmar, T. P. Structural Basis for Ligand Binding and Specificity in Adrenergic Receptors: Implications for GPCR-Targeted Drug Discovery. *Biochem.* **47**,(42), 11013-11023(2008)
26. Swaminath, G., Deupi, X., Lee, T. W., Zhu, W., Thian, F. S., Kobilka, T. S., and Kobilka, B. Probing the β2-adrenoceptor binding site with catechol reveals differences in binding and activation by agonist and partial agonists. *J. Biol. Chem.* **280**, 22165-22171 (2005).
27. Yao. X., Prnot, C., Deupi, X., Ratnala, V. R., Swaminath, G., Farrens, D and Kobilka, B. Couplng ligand structure to specific conformational switches in the β2-adrenoceptor. *Nat. Chem. Biol.* **2**, 417-422 (2006).
28. Vilardaga, J. P., Bunemann, M., Krasel, C., Castro, M., and Lohse, M. J. Measurement of the millisecond activation awitch of G protein-coupled receptorsin living cells. *Nat. Biotechnol.* **21**, 807-812 (2003).
29. Lohse, M. J., Hoffmann, C., Nikolaev, V. O., Vilardaga, J. P. and Bunemann, M. Kinetic analysis of G protein-coupled receptor signaling using flourescence resonance energy transfer in living cells. *Adv. Protein. Chem.* **74**, 167-188 (2007).
30. Salom D., et al. Crystal structure of a photoactivated deprotinated intermediate of rhodopsin. *Prot. Natl. Acad. Sci*., **103**(44), 16123-16128 (2006).
31. Kristiansen, K. Molecular mechanism of ligand binding, signalling, and regulation within the superfamily of G-protein-coupled receptors: molecular modelling and mutagenesis approaches to receptor structure and function. *Pharmac Therapeutics* **103**, 21-80 (2004).

32. Fanelli, F. and De Benedetti P. G. Computational Modelling Approaches to Structure-Function Analysis of G Protein-Coupled Receptors. *Chem Rev*, **105**, 3297-3351 (2005).
33. Lobanov M.Y., Bogatyreva, N. S., and Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compatness, *Molec. Bio*, **42**, 623-628 (2008).
34. Katsuaki, I., Naoto, Y., Yosihiro, U. and Takashi, I. Compact Packing of Lipocalin-type Prostaglandin D Synthase Induced by Binding of Lipophilic Ligands. *J. Biochem*, **145**(2), 169-175 (2009).
35. Katritch, V., Reynolds, K. A., Cherezov, V., Hanson, M. A., Roth, C. B., Yeager, M., and Abagyan, R. Analysis of full and partial agonists binding to B2-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit*. **22**, 307-318 (2009).
36. Sheikh, S. P., Zvyaga. T. A., Lichtarge, O., Sakmar, T. P., and Bourne, H. R. Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature,* **383**, 347-350 (1996).
37. Stenkemp, R. E., Teller, D. C. and Palczewski, P. Crystal structure of Rhodopsin: A G-protein coupled receptor, *Chem. Biochem*. **3**, 963 – 967 (2002).
38. Palcezewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science* **289** (5480), 739 – 45 (2000).
39. Curran, A. R., and Engelman, D. M. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr Opin Struct Biol* **13**, 412–417 (2003).
40. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
41. Perlman, J. H. *et al*. Interactions between Conserved Residues in Transmembrane Helices 1, 2, and 7 of the Thyrotropin-releasing Hormone Receptor. *J. Bio. Chem,* **272**(18), 11937-11942 (1997).
42. Fu, D., Ballesteros, J. A., Weinstein, H., Chen, J. and Javitch, J. A. Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding site crevice. *Biochem,* **35**, 11278-11285 (1996).

43. Deupi, X., Olivella, M., Govaets, C., Ballesteros, J. A., Campillo, M. and Pardo, L. Ser and Thr residues modulated the conformation of Pro-kinked transmembrane α-helices. *Biophys. J.* **86,** 105-115 (2004).
44. Cherezov, V. *et al.* High – resolution crystal structure of an engineered human β2- adrenergic G- Protein coupled Receptor. *Science* **318**, 1358-1264 (2007).
45. Kobilka, B. G- protein coupled receptor structure and activation. *Biochim Biophys. Acta* **1768**(4), 794-807 (2007).
46. Fong, T. M., Yu, H., Huang, R.R.C. and Strader, C. D. The extracellular domain of the neurokinin-1 receptor is required for high-affinity binding of peptides. *Biochem*, **31**, 11806-11811(1992).

CHAPTER SIX

AKH from *Anopheles gambiae* : NMR Experiments, Conformational Search and Docking experiments.

6.1: Summary

The primary sequence of the most important adipokinetic hormone from the malaria mosquito, Anoga_akh (pGlu-Leu-Thr-Phe-Thr-Pro-Ala-Trp-NH_2) was recently published [1]. Since no crystal structure of this neuropeptide is available, in this Chapter focus was on using NMR results to build the 3D structure of Anoga_akh and perform docking calculations of the ligand with AKHR. Although the binding site of AKHR was identified in Chapter Five, it was important to perform docking calculations with AKH from the mosquito so that any artifects arising from determining the binding pocket and activation of the receptor using AKH from the blister beetle, Del_CC, were eliminated.

Conformational search was done using interproton distances from Nuclear Magnetic Resonance (NMR) results during distance restrained molecular dynamics simulations of Anoga_akh. The results have shown that Anoga_akh prefers an almost cyclic conformation. This conformation is stabilised by a H-bond between the C-terminus and Leu2. Docking calculations have shown that the binding pocket of AKHR is described by both the helix bundle and the extracellular domains. Binding affinity for the ligand was higher than for Del_CC. MD simulations of Anoga_akh/AKHR complex in a membrane have indicated that the ligand induced conformational changes in the receptor molecule. The extracellular half of the molecule attained a closed conformation whilst the intracellular was opened up. Both

polar and hydrophobic interactions stabilised the ligand and Ile106 proved to be crucial for binding of Anoga_akh.

6.2: Nuclear Magnetic Resonance (NMR)

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the most powerful and widely used techniques in the study of structures of macromolecular biological molecules, including peptides and proteins in membrane bilayers [2,3,4,5]. Information obtained from both one and two-dimensional spectra have been used to determine their secondary structures and orientations in solution and membrane [6]. From the two dimensional Total Correlation Spectroscopy (TOCSY) [7] and the primary sequence of a peptide, proton chemical shifts are assigned. It has been shown that there is correlation between the chemical shifts of Hα protons and the secondary structure of a peptide. Thus from the assignments the secondary structure of the peptide can be predicted.

Sequential assignments of the amino acid can be obtained using the nuclear Overhauser effect (nOe) [8]. NOe is a through space, dipolar interaction, which is distance dependent. These interactions can be obtained using either the NOESY [9] or ROESY [10] pulse sequence. By superimposing the NOESY spectrum onto the assigned TOCSY spectrum, the sequential pattern of d_{NN} (i, i+1) and $d_{\alpha N}$ (i,i+1) crosspeaks [8] can be established. The assignment of the crosspeaks can be facilitated by floating the chirality in preliminary distance geometry calculations [11]. Additional structural information, distance measurements between pairs of different nuclei and between pairs of like nuclei have been obtained on peptides in bilayers using magic angle spinning methods [6]. Interproton distances from the 2D nOe

crosspeak intensities are estimated by the isolated spin pair approximation (ISPA)[12]:

$$\boldsymbol{r}_{ij} = \boldsymbol{r}_{ref}\,(\boldsymbol{a}_{ref}/\boldsymbol{a}_{ij})^{1/6},$$

Where $\boldsymbol{r}_{ij}$ is the interproton distances to be estimated and $\boldsymbol{a}_{ij}$ is the corresponding 2D nOe crosspeak intensity. NMR interproton distances restraints have been successfully applied during molecular dynamics simulations for conformational search of peptides in water and in a membrane mimicking solvent like dodecylphosphocholine (DPC) and dimethyl sulphoxide (DMSO) [13,14,15]. Therefore, in this work, the chemical shifts and the interproton distances were used to determine the flexibility and model free order parameters of AKH from the malaria mosquito (Anoga_akh). The interproton distances were used as restraints during the molecular dynamics simulation of the peptide in water.

6.3: Experimental

6.3.1: NMR

Sample Preparation

Using the recently published primary sequence of the active adipokinetic hormone from *Anopheles gambiae* (Anoga_akh, (pGlu-Leu-Thr-Phe-Thr-Pro-Ala-Trp-NH_2)) [1], the hormone was synthesised at GL Biochem (Shanghai) Ltd, China. The 98% pure white powder was used in NMR experiments with no further purification. The solvents used were water/D_2O and a membrane mimicking solution of deuterated dodecylphosphocholine (DPC-d_{38}).

For the aqueous phase, 1.6mg of anoga_akh was dissolved in 0.5ml of 10:1 (v/v) H_2O:D_2O solution buffered at pH 4 by 50mM sodium phosphate buffer solution.

The same mass of anoga_akh was dissolved in approximately 150mM DPC-d_{38} micelle also buffered at pH 4 by 50mM sodium phosphate buffer. In both cases trimethylsilylpropionate (TSP) was used as the chemical shift reference.

^{1}H NMR Experiments

^{1}H NMR experiments of anoga_akh were performed on a Bruker Avance 400Hz (Department of Chemistry, University of Cape Town) and a DRX-500MHz (Analytical Chemistry, University of Debrecen). The two dimensional NMR spectra of the peptide in water and DPC were collected in the phase sensitive mode from TOCSY (mixing time, 50ms) [7,16], NOESY (mixing time, 100ms) [9] and ROESY (mixing time, 150ms) [10] using the WATERGATE pulse to suppress the water resonance [17,18]. Spectral assignments were based on the method of Wüthrich [9] and ISPA [12] was used to calculate the interproton distances.

6.3.2: Restrained MD simulations

The starting structure for conformational search of Anoga_akh was built using InsightII 2005 and energy minimised by using 3000 steps of steepest descent algorithm. The molecule was subjected to 10ns molecular dynamic simulations in vacuum and water using GROMACS version 4.0.4 [19]. All simulations were performed with a time step of 2 fs, using the OPLS-AA/L all – atom [20] force field and constant temperature, volume and number of particles (NVT). A cut-off 1.0nm was used for van der Waals interactions and electrostatic interactions for real space calculations.

In vacuum

Anoga_akh was inserted into a simulation box and energy minimised using 500steps of l-BFGS minimiser. Distance restraints between specific protons were applied

according to calculated NMR interproton distances (Appendix G). The listed distances were used as the lower limits and two upper limits were obtained by increasing the values by ~ 10%. The molecule was subjected to 10ns distance restrained MD simulations at 600K and 100 structures collected. To allow the molecule to overcome an energy barriers and since only one starting structure was used, conformational search was performed at high temperature [15]. Simulated annealing for 0.1ns was performed to gradually decrease the temperature of each collected structure to 300K. The conformations were energy minimised, cluster analysis was performed with a cut-off distance of <0.2 nm. The conformation of lowest energy was identified from the largest cluster.

In water

The minimum energy conformation from the simulations in vacuum was used as the starting structure of Anoga_akh for 10ns MD simulations in water. The water atoms were based on the SPC water model [21]. To allow the water to soak the peptide, position-restrained MD simulations for 0.1 ns were performed. All simulations were performed with periodic boundary conditions [22]. The peptide and solvent were coupled separately using the Berendsen coupling method to a temperature of 300K using a coupling constant, tau (τ)_t, of 1 x 10^{-4} ns.

Distance restrained MD simulation, based on the NMR results was performed for 10ns and 100 structures were collected for analysis. Cluster analysis of the structures was performed based on a cut-off value of <0.2 nm for superimposing backbone atoms. The minimum energy conformation from the largest cluster gave the structure of Anoga_akh.

6.3.3: Docking Calculations

Based on blind docking results obtained in Chapter Five, docking calculations of Anoga_akh were performed to determine the estimated free energy of binding and the binding mode of the ligand to AKHR using the β_2AR-based receptor model. The AutoGrid4 program [23] was used to prepare a grid map with grid points of 64 x 64 x 50 in xyz, and a spacing of 0.05nm, for use during docking calculations. The grid box was placed in such a way that it centred at the midpoint of the putative binding pocket. Docking calculations were performed with a rigid receptor but flexible ligand (32 active torsions angles). Both ligand and receptor parameter files were prepared using Autodock Tools (Appendix). For 30 docking trials, the Lamarckian genetic algorithm (LGA) was employed to generate conformations of the ligand within the binding pocket [23,24,25,26]. A total of 5 x 10^8 generations and energy evaluations were applied for a population of 300. Cluster analysis was executed at a cut-off of 2Å.

From the results, the conformation of lowest estimated free energy of binding, ΔG_b , was used to further refine the docking calculations using MD simulations with both protein and ligand molecules flexible. The protein-ligand complex of lowest ΔG_b was inserted into a previously prepared membrane (Chapter Four) to represent its true environment. The system was energy minimised using 3000 steps of L-bfgs and equilibrated by subjecting it to position restrained MD simulations for 20ns to allow the lipid and water molecules to effectively soak the protein. The structure extracted after 20ns was subjected to 100ns MD simulations whilst applying the protocol and parameters described in Chapter Four. Structures extracted from the MD trajectory at 10ns intervals were rescored to obtain the final values of the estimated free energy of

binding using the modified scoring function of Autodock 3.0 described by Hetenyi, 2006 [27] for large and flexible peptidic molecules.

6.4: Results and discussion

6.4.1 NMR

Spectral assignment and interproton distances

The ^{1}H chemical shifts asignments for Anoga_akh in water and DPC were successfully performed and the results are given in Tables 6.1 and 6.2. The assignments were done by referring to chemical shift tables, the primary sequence of anoga_akh, and by identifying the individual amino acid spin system from TOCSY spectra.

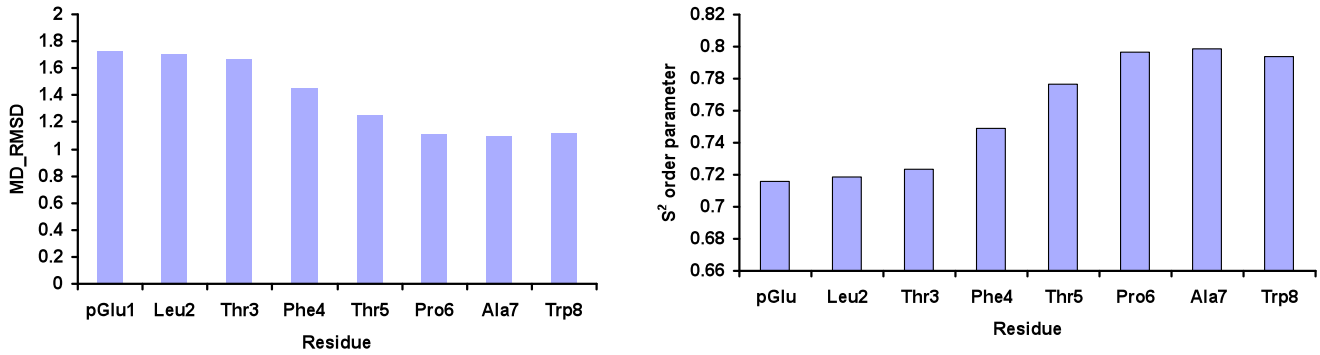
Table 6.1: ^{1}H chemical shift (ppm) assignments for Anoga_akh in 9:1 H_2O/D_2O solution

Residue	NH	Hα	β	β'	γ	γ'	δ	δ'	Other
pGlu1	7.752	4.183	2.228		1.868	2.355			
Leu2	8.275	4,236	1.445		1.355		0.750	0.702	
Thr3	7.946	4.233	3.949		0.958				
Phe4	8.154	4.494	2.853						7.08–7.16
Thr5	7.961	4.314	3.797		0.968				
Pro6	--	4.0100	1.219	1.856	1.656		3.407		
Ala7	8.200	4.004	1.145						
Trp8 – NH_2	7.592	4.457	3.121						7.031(2), 7.508(4), 6.97-7.03

Table 6.2: ^{1}H chemical shift (ppm) assignments for Anoga_akh in DPC solution

Residue	NH	Hα	β	β'	γ	γ'	δ	δ'	Other
pGlu1	7.914	4.266	1.880; 2.245; 2.395						
Leu2	8.408	4.300	1.391; 1.513				0.754	0.777	
Thr3	8.092	4.226	4.052	0.995					
Phe4	8.013	4.564	2.933						7.11 92,60; 7.04 (3,4,5)
Thr5	7.933	4.338	3.914	0.984					
Pro6	--	4.118	1.294; 1.793		1.637; 1.562		3.391		
Ala7	8.166	4.152	1.198						
Trp8 – NH_2	7.687	4.490	3.10; 3.16						7.154; 7.452, 6.962; 9.886; 7.333; $CONH_2$ 7.32, 7.15

The results indicate that there was a close relationship between chemical shifts for each residue peptide in water and in DPC. Using the RCI tool [28] available on http://wishart.biology.ualberta.ca, the chemical shifts were used to determine the flexibilty and the order parameters, S^2, of Anoga_akh (Figure 6.1).



(a) (b)

Figure 6.1: a) Molecular dynamics backbone root mean square deviation and b) order parameter, S2, of Anoga_akh calculated using RCI server [28].

The results show that the N-terminal end from pGlu1 to Phe 4 has higher backbone deviations and fluctuations (MD_RMSD ranging from 1.4 -1.7) than the C-terminal end (Thr5-Trp8) which has much lower MD_RMSD of 1.1 – 1.2 (Figure 6.1a). The molecule however is generally ordered with order parameters ranging from ~0.72 – 0.80 (Figure 6.1b). As expected the highly flexible N-ternimus end is less ordered than the C-terminus.

Temperature coefficient and Hydrogen bond formation

The temperature coefficients of the amide chemical shifts were determined from a plot of the chemical shifts against temperature (Appendices H and I). The temperature coefficient is used as an indicator for hydrogen bond formation by amide protons (NH) [15]. The chemical shifts, ^{1}H-^{1}H coupling constancts (Hz) and the corresponding calculated phi (ϕ) angles of some residues in Anoga_akh are given in Table 6.3.

Table 6.3: H^N temperature coefficients (ppb/K), ^{1}H-^{1}H coupling (Hz) and the calculated ϕ angles.

Residue	H^N temperature coefficient (ppb/K)	$^{3}J_{HNH\alpha}$ coupling constant (Hz)	Calculated ϕ angles (^{o})
pGlu1	-	-	-
Leu2	3.4	6.8	-68
Thr3	4.2	8.0	-90
Phe4	2.8	7.5	-86
Thr5	3.6	8.0	-90
Pro6	-	-	-
Ala7	1.6	5.5	-70
Trp8	-	-	-

The temperature coefficient results indicate that the NH of most residues in Anoga_akh participate in formation of intramolecular hydrogen bond. For Ala7, the temperature coefficient of 1.6 was the lowest and reveals that its proton is shielded from the solvent by the aromatic Trp8 aromatic side chains and therefore should form pronounced intramolecular hydrogen bonds.

Interproton distances

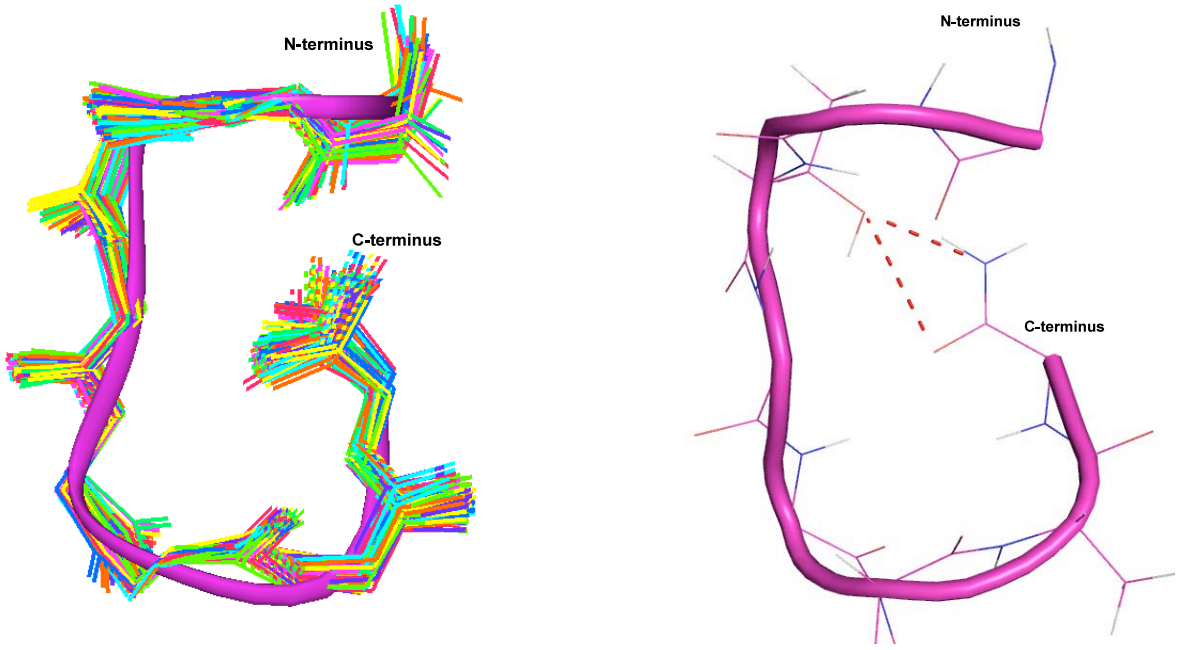
Interproton distances from the nOe crosspeaks were calculated by the Isolated Spin Pair Approximation (ISPA) method [12], and pseudoatom corrections were done according to Wüthrich, 1986 [8] (Appendices F and G). ISPA was a suitable approximation since short mixing times of 100ms and 150ms were used in NOESY and ROESY respectively, and effects of spin diffusion were small [14]. The first and second upper limits upper were obtained by adding ~10% and 20% of each of the measured distances since nOe measurements are biased towards short internuclear distances [15]. Pro6 Hb – Hb1 internuclear distance, 0.18 nm, was used as the reference. Although, in water, 17 distances ranging from ~0.26 – 0.43 nm were obtained and 32 distances in DPC, from ~0.27 (Leu2 NH – pGlu1 NH) – 0.74 (Leu2 Hd2 - Phe4 H2,6) nm most distances were similar, for example, distance between Leu NH and pGlu1 Hb2 was 0.43 nm and 0.42 nm, respectively. These internuclear distances were used in distance restrained MD simulation of Anoga_akh.

6.4.2: MD simulations of Anoga_akh

6.4.2.1 Conformational Search

Interproton distances and restraints obtained from NMR experiments in water were used during 10ns MD simulations at 600K in vacuum to search for conformation of the mosquito peptide. High temperatures were used to allow the molecule explore

high energy conformers [13,14,15]. 100 structures were collected separated and subjected to 0.1ns simulated annealing and energy minimised. Cluster analysis of the structures at 0.2 nm cut-off distance gave one cluster. The conformation of lowest energy was further subjected to 10ns MD in water at 300K. Another set of 100 structures was collected and each structure was energy minimised. Cluster analysis gave only one cluster indicating that the confomation of the peptide is highly conserved (Figure 6.2a).



(a) (b)

Figure 6.2: a) An overlay of 100 structures from MD simulations of Anoga_akh in water at 300K. b) The minimum energy conformation of Anoga_akh. The red dotted lines indicate a salt bridge between the oxygen atom in Trp8, OH in Leu2 and nitrogen in NH_2.

The identified minimum energy structure indicates that the peptide prefers a cyclic conformation. The salt bridges between Leu2 and Trp8 atoms and intramolecular hydrogen bonds complete the circle (Figure 6.2b). This conformation

was similar to the conformation of Gonadotropic releasing hormones (GnRH) [13] and the Crustacean cardioactive peptide (CCAP) [15]. To complete the circle in these peptides intramolecular interactions like H-bonds [13] or disulphides bonds as in CCAP [15] are used. Similarly, in Anoga_akh a salt-bridge was used to complete the circle.

Figure 6.3 shows a comparison between phi angles calculated from 10ns trajectory in water, using GROMACS tools [19] and those predicted from the NMR chemical shifts by the programme PREDITOR [29].

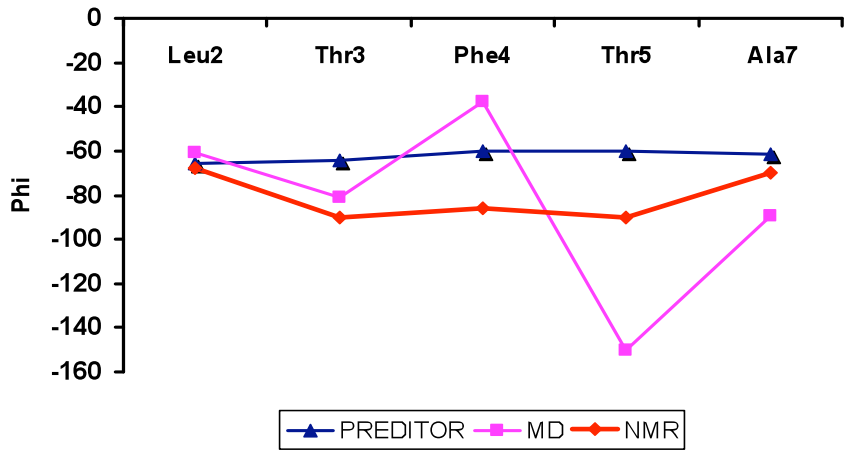


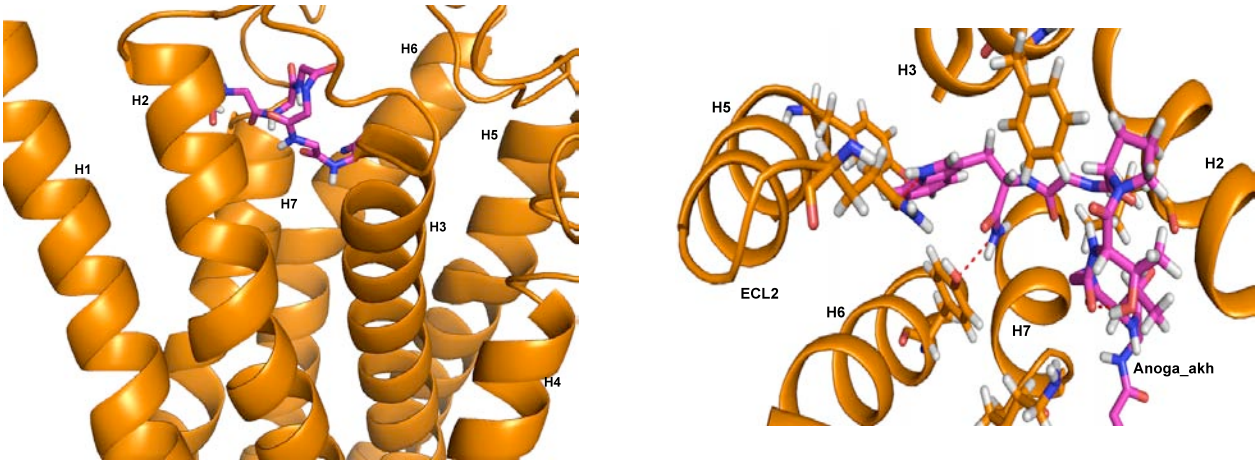
Figure 6.3: Comparison of phi torsion angles calculated from NMR chemical shifts by PREDITOR [29], estimated from MD simulations in water and from calculations using $^{3}J_{HNH\alpha}$ coupling constants.

All 100 conformations obtained from the MD simulations, were used in the calculation of the average dihedral angle, phi. For an average standard deviation of ~ $\pm$ 15 for each residue, it was observed that phi angles from MD calculations were comparable to the PREDITOR and NMR results. With the exception of Thr5, MD simulations reproduced experimental results. Therefore the conformation of Anoga_akh form MD simulation should be close to the real structure.

6.4.3 Docking Calculations

6.4.3.1 AKHR Binding pocket and binding mode of Anoga_akh

Docking calculations of Anoga_akh to its receptor using Autodock4 programme [23] confirmed the position of the binding pocket of adipokinetic hormone receptor (AKHR) and the involvement of the extracellular domains in ligand binding as shown in Figure 6.4a. These observations are supported by experimental results from mutational mapping of peptide receptor binding site [30,31]. The ligand successfully occupied the putative binding pocket and was in contact with residues in helices 2, 3, 5, 6 and 7, ECL2 and the N-terminus (Figure 6.4b).



(a) (b)

Figure 6.4: a) Side view of bound AKHR (orange) shows the position of Anoga_akh (pink) and b) Extracellular view of Anoga_akh/AKHR complex indicating the involvement of H2, H3, H5, H6, H7 and ECL2 in ligand binding.

From 30 docking trials, lowest estimated free energy of binding (ΔG_b) of -47.3kJ/mol) and an inhibition constant of 5.22nM were obtained. This shows that the receptor had higher binding affinity for AKH from *Anopheles gambiae* than for Del_CC from

the blister beetle (ΔG_b of -38.5kJ/mol). However, in this conformation Tyr285 and Gln206 participated in stabilising the complex through formation of salt bridges with NH_2 and Trp8 at the C-terminus of Anoga_akh. The conformation was used as the starting structure in MD simulations of the receptor-ligand complex in its true hydrophobic environment.

6.4.4 Membrane simulations of Anoga_akh/AKHR complex

To refine docking experiments and determine the effect of Anoga_akh on the conformation and structural features of AKHR, 100ns MD simulations were performed with a flexible receptor-ligand complex inserted in a pre-equilibrated POPC membrane model. Refinement and optimisation of the complex was important since docking calculations were performed with a rigid receptor molecule and mobile ligand molecule. MD simulations with both molecules flexibile provided a way of determining the stability and binding modes of Anoga_akh in the complex during molecular motion.

6.4.4.1 Conformational changes in bound AKHR

The agonist, Anoga_akh, induced conformational and structural changes in the receptor molecule during MD simulations as the molecule changed from an 'open' to a 'closed' conformation in the extracellular region, whilst the intracellur region opened up. The differences in conformation between bound and unbound AKHR were supported by an rmsd of ~ 0.32 nm obtained from an overlay of Ca atoms of the two receptor molecules. Initial changes in conformation of the receptor in the presence of Anoga_akh occurred within the first 2 ns of simulation. This indicates slow attainment of the active state by the receptor in the presence of Anoga_akh.

However, the slow transition from inactive to active state of AKHR is similar to the rate of conversion of β_2AR when activated by its agonist, adrenaline [33]. In comparison, drastic changes occurred in ~1.5 ns in the presence of Del_CC (Chapter Five), indicating fast change from inactive to active state by the receptor.

Figure 6.5 gives a comparison of the flexibilty of the receptor Ca atoms in unbound (o_akhr20) and bound AKHR (Anoga_akh/AKHR complex). The rmsf results indicate that there was similar mobility of the atoms in the two receptor molecules and high fluctuations were observed in the extracellular and the intracellular domains, with rmsf as high as ~0.65 nm. Although helices average rmsf was ~0.1nms, there was increased mobilty of atoms in the extracellular end of H1 (47-75), and H2 (84-98) which experienced great conformational changes due to ligand binding.

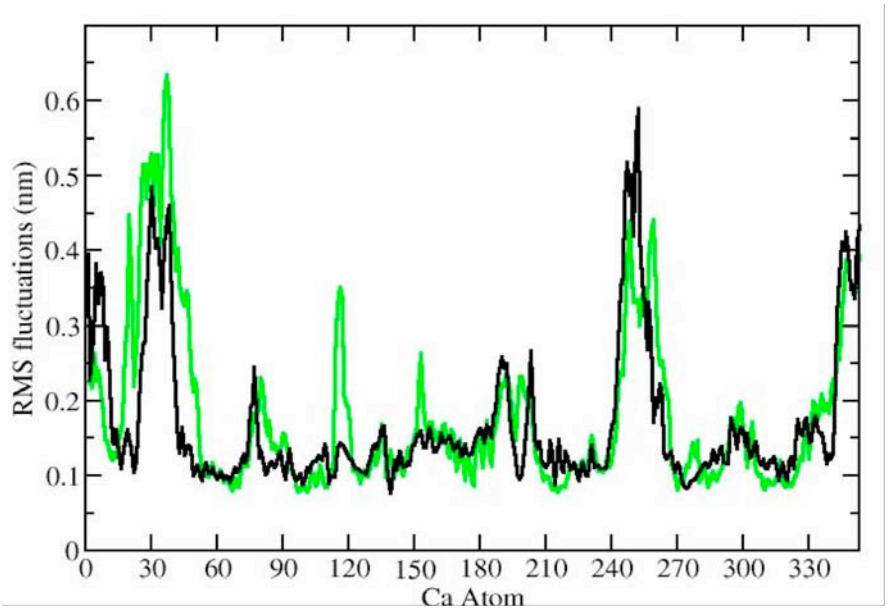


Figure 6.5: Variations of Root mean square fluctuations of bound (green) and unbound (black) AKHR during 100ns MD simualtions.

During 100ns MD simulations, movements of the receptor resulted in deviations of bound AKHR molecule from its starting structure by an initial rmsd of ~ 0.35 nm within the fisrt 2ns (Figure 6.6).

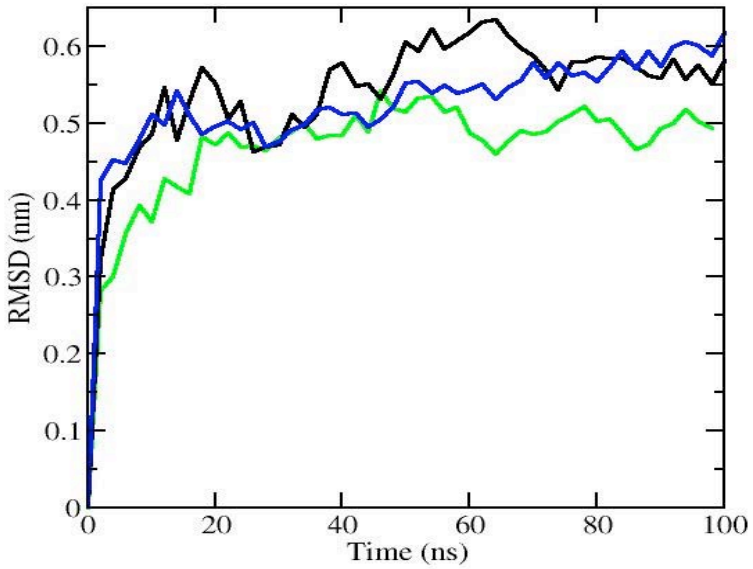


Figure 6.6: Comparison between molecuar deviations of unbound (black) and AKHR bound to Anoga_akh (green) and AKHR bound to Del_CC (blue).

Deviations were due to the inward movement of helices in the extracellular region and an outward movemnet in the intracellular side. The system however attained a steady state after ~10ns and this was maintained until the end of simulation. In contrast, AKHR bound to Del_CC had much higher deviations and the drastic change in conformation is indicated by an initial rmsd of ~ 0.42 nm, whilst the outward movement of helices in unbound AKHR an initial rise to ~0.38 nm. Of the three simulations, AKHR bound to Anoga_akh deviated the least from its starting structure.

Comparison of the compactness of the receptor molecule during MD simulation showed that unbound and bound molecules had similar compactness for the greater part of the simulations, as illustrated in Figure 6.7.

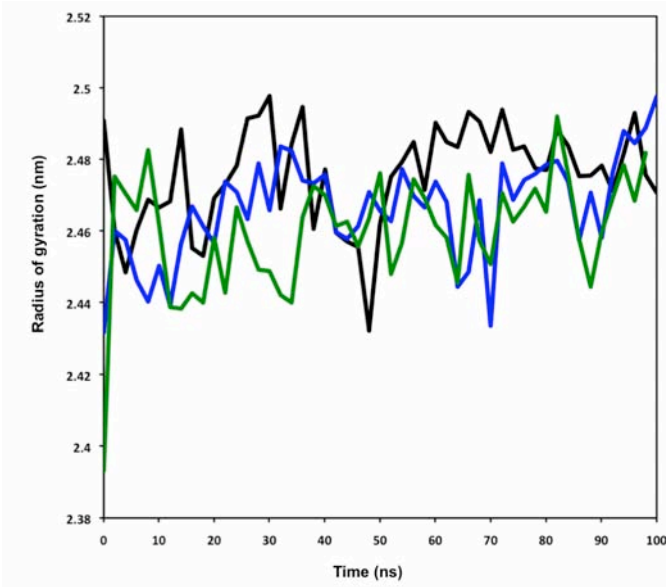


Figure 6.7: Comparison of changes in radius of gyration of AKHR bound to Anoga_akh (green) and bound to Del_CC (blue) and unbound (back) AKHR during 100ns MD simulations.

However, in bound AKHR, there was increase in Rg from ~2.41 to ~2.50 nm from 0 - 10ns, as helices and the loop domains shifted to accommodate and protect the ligand. This sudden rise in Rg, also observed with AKHR bound to Del_CC, corresponds to the initial rise of the rmsd shown in Figure 6.6, indicating major movements in the receptor. The molecule remained in the activated state until the end of the simulation period maintaining an Rg of ~2.48 nm. In contrast, unbound AKHR had a sudden decrease in Rg from ~ 2.49 to ~ 2.45nm, reflecting an increase in compactness in the hydrophobic environment. The different changes in compactness at the beginning of simulation clearly demonstrate that when bound to an agonist, the confomation and structural features of AKHR change to accommodate the ligand.

6.4.4.1.1 Conformational changes in the helix bundle

From the beginning of the MD simulations, AKHR experienced conformational changes, most important of which involved an inward movement of the extracellular side of the helix bundle and an outward withdrawal of the intracellular side due to activation of the receptor, Figure 6.8. Such changes in conformation of the helical bundle are crucial for transmission of signals from the extracellular to the intracellular region of the cell and lead to activation of the G-protein [30,33]. In extracellular regions H1, H2, H5 and H6 were drawn closer to each other. The position of H1 in bound AKHR (gray) compared to its location in the unbound molecule reveals an inward and upward movement by ~ 0.68 nm. Similar movements were witnessed by helices 2,5 and 6.

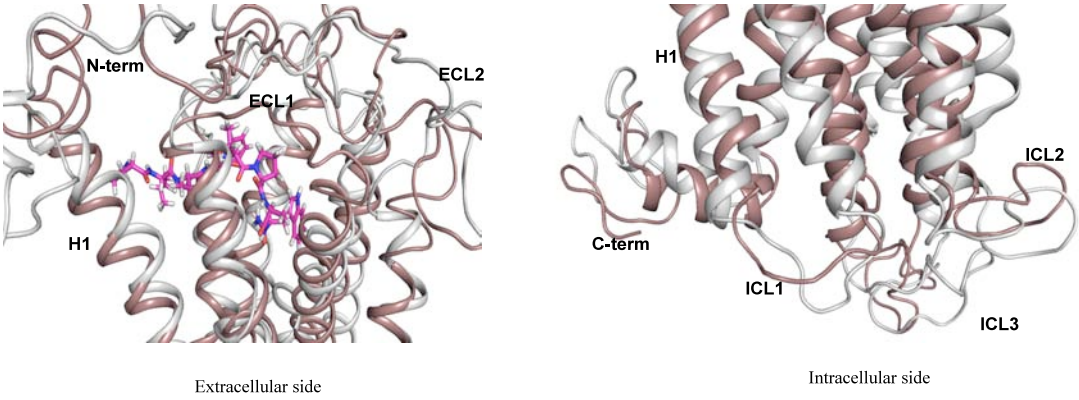


Figure 6.8: An overlay of unbound (brown) and bound (gray) AKHR. There was an inward movement of the helices in the extracellular region, and an outward movement in the intracellular region.

On the other hand, in the intracellular region there was an outward movement of helices 1, 2, 3 and 5 (Figure 6.8). Experimental results have demonstrated the importance of the drift of H6 and H3 away from each [33,34]. The same movement was observed in AKHR as it attains an activated state. Similar movements were also observed during activation of rhodopsin [35] and β_2AR [32], where an outward divergence of H6 was crucial in disrupting the 'ionic lock' and induced activation of

the heterotrimeric G-protein. In AKHR the outward movement resulted in breaking of the ionic interactions between H5 and H6, which could increase the basal activity of the receptor and initiate activation of G-protein.

The inward and outward movements of helices caused the extracellular loop domains to collapse over the helix bundle, especially ECL1 and ECL2. Consequently, a closed conformation that protected the agonist was produced. On the contrary, the intracellular loop domains were drawn outwards. Such changes are usually found during activation of Class A GPCRs [30,33] and they illustrate that Anoga_akh activated the receptor.

Conformational changes in Anoga-akh

Binding of Anoga_akh to AKHR induced conformation changes from an almost cyclic molecule to an L-shaped one (Figure 6.9). The N-terminus of the ligand was very flexible because of few intramolecular interactions. An overlay of the ligand molecules before and after docking gave an rmsd of ~0.28 nm.

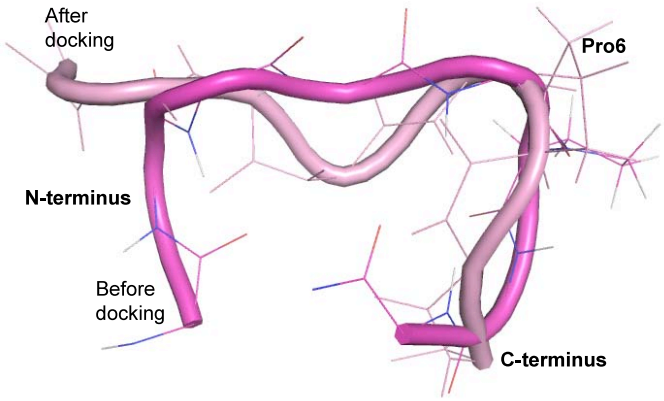


Figure 6.9: Comparison of conformations of Anoga_akh before (deep pink) and after (light pink) docking.

In the ligand/receptor complex the conformer is stabilised by H-bonds, salt bridges and hydrophobic interaction with the receptor molecule as discussed in the next section.

6.4.4.2 Binding mode and binding affinity of Anoga_akh

Experimental results from mutational mapping and chimera studies of peptide GPCRs have shown that both the helix bundle and the extracellular loop domains participate in ligand binding (36,37). The inward movement of the extracellular side of the helix bundles and the collapse of the loop domains brought extracellular residues closer to the ligand and binding pocket of AKHR was defined by H2, H3, H5, H6, H7, ECL1 and ECL2. Anoga_akh was therefore stabilised by H-bonds, salt-bridges and hydrophobic interactions involving residues like Ile106, Tyr110, Phe126, Tyr221, and Tyr285 (Figure 6.10) in these domains.

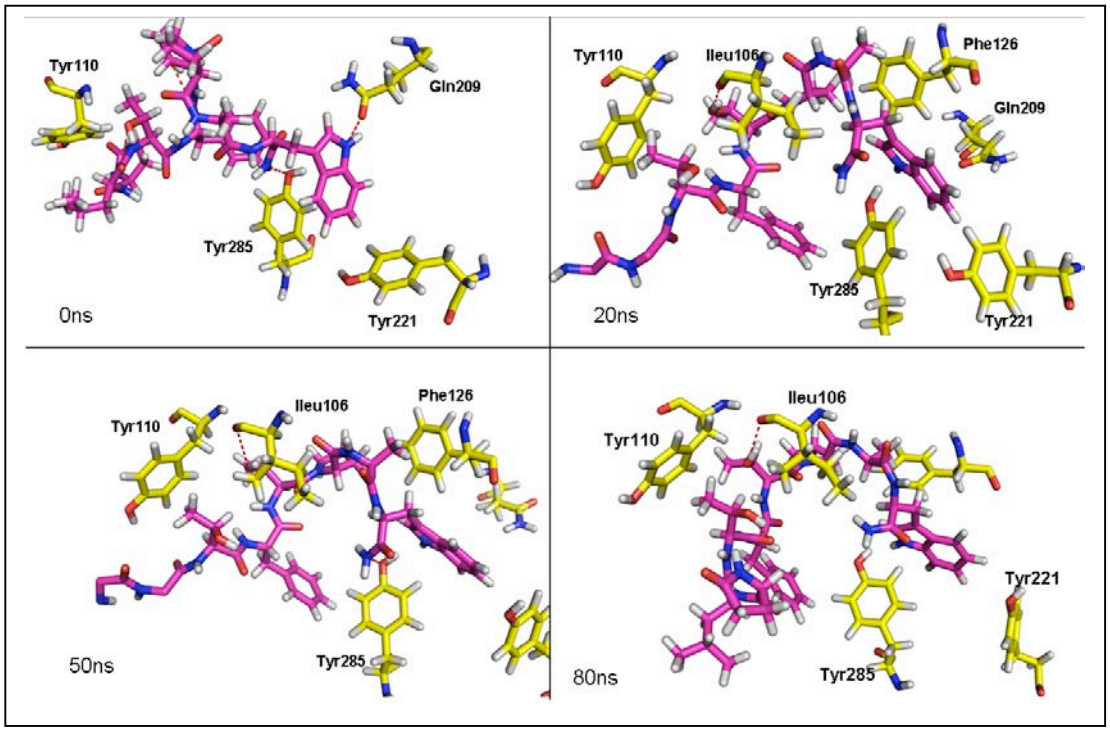


Figure 6.10: Interactions between Anoga-akh (pink) and some of the residues in AKHR after 0ns, 20ns, 50ns and 80ns of MD simulations.

At instances when no polar contacts were present between the ligand and the receptor, the ligand was kept in position by hydrophobic interactions with Tyr110, Tyr221, Tyr285, Phe126 (Figure 6.10) and non-polar residues in the N-terminus and ECL3. In addition to the hydrophobic interactions, Tyr285 occassionally formed a salt bridge with the nitrogen atom of NH_2 at the C-terminus of the ligand, for example at 0 and 50ns (Figure 6.10). The most importatnt residue in binding Anoga_akh was Ile106 (H2) which constantly formed H-bonds or a salt-brdge with Thr5 in the ligand. Participation of this residue, which was not seen in the initial structure (at 0ns) accounts for the inward movement of H2 during dynamics.

From the 100ns trajectory of the Anoga_akh/AKHR complex, structures were extracted at 10ns interval and rescored to obtain final values of the estimated free energy of binding. Calculations were performed by the modified scoring function of Autudock3.0 designed by Hetenyi, 2006 [27]. Table 6.4 gives $\Delta G_{b,}$ inhibiton constant (KI) and final binding energy (E_b), obtained from the calculation.

Table 6.4: Estimated free energy of binding, inhibition constant and final binding energy of Anoga_akh

Anoga_akh conformations	**ΔG_b(kJ/mol)**	**KI (nM)**	**E_b (kJ/mol)**
_0ns	-47.30	5.22	
2_10ns	-35.71	560	-73.39
3_20ns	-43.04	29.40	-83.99
4_30ns	-15.99	-----	-56.94
5_40ns	-32.66	1900	-73.31
6_50ns	-29.56	6720	-66.28
7_60ns	-27.97	12600	-67.87
8_70ns	-36.23	457	-74.40
9_80ns	-50.20	1.63	-88.89

- (2_10ns denotes the second conformer extracted after 10ns. Other conformers are similarly labelled).

The highest binding affinity was obtained after 80ns, ΔG_b of -50.20 kJ/mol and the ligand experienced least resistance to binding giving KI of 1.63nM. The conformation had higher affinity than in the initial docking calculations results (ΔG_b of -47.3kJ/mol). As shown in Figure 6.10, H-bond between the oxygen atom in Ile106 and HG1 atom in Thr5, and hydrophobic interactions stabilise the conformation. This highlighted the importance of Ile106 in binding of Anoga_akh to AKHR. It is important to note that, although binding affinity for Anoga_akh varied throughout molecular dynamics, it was much higher (average ΔG_b = -33.92kJ/mol), than that for Del_CC (average ΔG_b = -19.02kJ/mol), from blister beetle.

6.5: Conclusion

The most stable conformation of adipokinetic hormone from the malaria mosquito was determined by using NMR data and distance restrained molecular dynamics simulations. The molecule preferred a cyclic conformation and a salt-bridge completed the circle. Docking calculations of Anoga_akh gave a complex in which the ligand occupied a binding pocket described by the helix bundle and the extracellular domains. The Anoga_akh/AKHR complex was stabilised by both polar and hydrophobic interactions involving residues in helices and loops. Ile106 played an important role in binding the agonist and the receptor had a very high binding affinity for Anoga_akh. Upon binding, the conformation of the agonist changed from cyclic to L-shaped.

MD simulations revealed that there was slow conformational change of the receptor molecule from an open to a closed conformation in the extracellular region

within the first few nanoseconds. Subsequently, the intracellular region changed from a closed to an open conformation, a change that is crucial in activation of G-proteins. Identification of the binding pocket of the mosquito AKHR, binding affinity and binding mode of Anoga_akh should go a long way in the design of non-peptide mimetics that can be used as alternative insecticides.

6.7: References

1. Kaufmann, C.and Brown, M. R. Regulation of carbohydrates metabolism and flight performance by a hypertrehalosaemic hormone in the mosquito *Anopheles gambiae*. *J. Ins. Phys.* **54**, 367-377 (2008).
2. Hunter, H. N., Fulton, D. B., Ganz, T., and Vogel, H. J. The Solution Structure of Human Hepcidin, a Peptide Hormone with Antimicrobial Activity That Is Invoolved in Iron Uptake and Hereditary Hemochromatosis. *J. Biol. Chem.*, **277**(40), 37597-37603 (2002).
3. Walse, B., Kihlberg, J., Drakenberg, T. Conformation of desmopressin, an analogue of the peptide hormone vasopressin, in aqueous solution as determined by NMR spectroscopy. *Eur. J. Biochem.,* **252**, 428-440 (1998).
4. Kim, H. J., Howell, S. C., Van Horn, W. D., Jeon , Y. H., and Sanders, C. R. Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **55**(4), 335-360,(2009).
5. Brünger, A. T., Robert L. Campbell, G. Marius Clore, Angela M. Gronenborn, Martin Karplus, Gregory A. Petsko, and Martha M. Teeter. Solution of a Protein Crystal Structure with a Model Obtained from NMR Interproton Distance Restraint. S*cience*, **235**(4792), 1049 – 1053 (1987)
6. Cross, T. A., and Opella, S. J., Solid-state NMR structural studies of peptides and proteins in membranes, *Current Opinion in Structural Biology,* **4**(4), 574-581(1994).
7. Braunschweiler, L., Ernst, R. R. Coherence Transfer By Isotropic Mixing - Application To Proton Correlation Spectroscopy. *J Magn. Reson.,* **53**(3), 521-528 (1983).
8. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R.R. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546-4553 (1979).
9. Wüthrich, K. *NMR of Proteins and Nucleic Acids*, Wiley, 178 (1986).
10. Hull, W. E. Experimental Aspects of Two-Dimensional NMR. In: W.R Croasmun and R.M.K Carlson, Editors, *The Two-Dimensional NMR*

Spectroscopy—Applications for Chemists and Biochemists (2nd ed.), VCH, New York, 337–343 (1994).

11. Weber, P. L., Morrison, R., and Hare, D. Determining Stereo-specific 1H Nuclear Magnetic Resonance Assignments from Distance Geometry calculations. *J. Molec. Biol.,* **204**, 483-487 (1988).
12. Thomas, P. D., Basus, V. J., and James, T. L. Protein solution structure determination from two-dimensional nuclear Overhauser effect experiments: Effect of approximations on the accuracy of derived structures. *Pro. Natl. Acad. Sci. USA.* **88,** 1237-1241 (1991).
13. Maliekal, J. C. Conformational and Docking Studies of Gonadotropin Releasing Hormone and Its Analogs by NMR Spectroscopy and Molecular Modeling. *PhD Thesis*, University of Cape Town (1999).
14. Stone, S. R. Structural Characterisation of the Solution and Membrane-associated Conformations of Human Little Gastrin and Its Bioactive Fragments by NMR Spectroscopy and Molecular Modeling, *PhD Thesis*, University of Cape Town (2006).
15. Jackson, G. E., Solution conformations of an insect neuropeptide: Crustacean cardioactive peptide (CCAP), *Peptides,* **30**, 557-564 (2009).
16. Bax, A/ Freeman, R., and Morris, G. Correlation of Proton Chemical Shifts by Two-Dimensional Fourier Transform NMR. *J. Magn. Reson.,* 42, 164-168 (1981).
17. Piotto, M., Saudek, V., and Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J.Biomolec. NMR*., **2**, 661-665 (1992).
18. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. Gradient-tailored water suppression for proton-nitrogen-15 HSQC experiments optimized to retain full sensitivity. *J. Magnt. Reson, Series A,* **102**, 241-245 (1993).
19. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., E. and Berendsen, H., J., C. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.* **26,** 1701-1719 (2005)
20. Kaminski, G., A., Friesner, R., A., Tirado-Rives, J., and Jorgensen, W., L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins

via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, **105**, 6474-6487 (2001)

21. Beresden, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. Interaction models for water in relation to protein hydration, in *Intemolecular forces* (Pullman, B., Ed.), 331-342, D. Reidel, Dordrecht (1981).
22. Dzwinel, W., Kitowski, J., and Moscinski, J. 'Checker Board' Periodic Boundary Conditions in Molecular Dynamics Codes. *Molec Simul*, **7**, 171-179 (1991).
23. Morris, G., M. *et al*. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* **19,** 14, 1639-1662 (1998).
24. Manetti, F., Tintori, C., Armand-Ugón, M., Clotet-Codina, I., Massa, S., Ragno, R., Esté, J. A., and Botta, M. A Combination of Molecular Dynamics and Docking Calculations to explore the Binding Mode of ADS-J1, a Polyatomic Compound Endowed with Anti-HIV-1 Activity. *J. Chem. Inf. Model.* **46**, 1344-1351 (2006).
25. Kovacs, M., Toth, J., Hetenyi, C., Malnasi-Csizmadia, A., and Sellers, J.R. Mechanism of blebbistatin inhibition of myosin II. *Journal of Biological Chemistry*, **279**(34), 35557-35563 (2004).
26. Bikadi, Z., Hazai, E., Zsila, F., and Lockwood, S.F. Molecular modeling of non-covalent binding of homochiral (3S,3 ' S)-astaxanthin to matrix metalloproteinase-13 (MMP-13). *Bioorganic & Medicinal Chemistry*, **14**(16), 5451-5458 (2006).
27. Hetényi, C., Paragi, G., Maran, U., Timár, Z., Karelson, M., and Penke, B. Combination of a Modified Scoring Function with Two-Dimensional Descriptors for Calculation of Binding Affinities of Bulky, Flexible Ligands to Proteins. *J. Am. Chem. Soc.* **128**(4), 1232-1239 (2006).
28. Berjanskii, M. V., David S. Wishart, D. S. A Simple Method To Predict Protein Flexibility Using Secondary Chemical Shifts. *J. Am. Chem. Soc.*, **127** (43), 14970 -14971 (2005).
29. Berjanskii, M. V, Neal, S., Wishart, D. S. PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.* 1(34), (Web Server issue), W63-9 (2006).

30. Gether, U. Uncovering Molecular Mechanisms Involved in Activation of G Protein-Coupled Receptors. *Endocr. Reviews*, **21**, 90-113 (2000).
31. Berthold, M., and Bartfai, T. Modes of Peptide Binding in G Protein-Coupled Receptors. *Neurochem. Res* .**22**(8), 1023-1031 (1997)
32. Huber, T., Menon, S., and Sakmar, T. P. Structural Basis for Ligand Binding and Specificity in Adrenergic Receptors: Implications for GPCR-Targeted Drug Discovery. *Biochem.* **47**,(42), 11013-11023(2008)
33. Kristiansen, K. Molecular mechanism of ligand binding, signalling, and regulation within the superfamily of G-protein-coupled receptors: molecular modelling and mutagenesis approaches to receptor structure and function. *Pharmac Therapeutics* **103**, 21-80 (2004).
34. Fanelli, F. and De Benedetti P. G. Computational Modelling Approaches to Structure-Function Analysis of G Protein-Coupled Receptors. *Chem Rev*, **105**, 3297-3351 (2005).
35. Sheikh, S. P., Zvyaga. T. A., Lichtarge, O., Sakmar, T. P., and Bourne, H. R. Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature,* **383**, 347-350 (1996).
36. Fong, T. M., Yu, H., Huang, R.R.C. and Strader, C. D. The extracellular domain of the neurokinin-1 receptor is required for high-affinity binding of peptides. *Biochem*, **31**, 11806-11811(1992).
37. Choi, M-Y., Fuerst, E-J., Rafaeli, A. and Jurenka, R. Role of extracellular domains in PBAN/pyrokinin GPCRs from insects using chimer receptors. *Insec. Biochem and Mol. Bio.* **37,** 296-306 (2007).

CHAPTER SEVEN

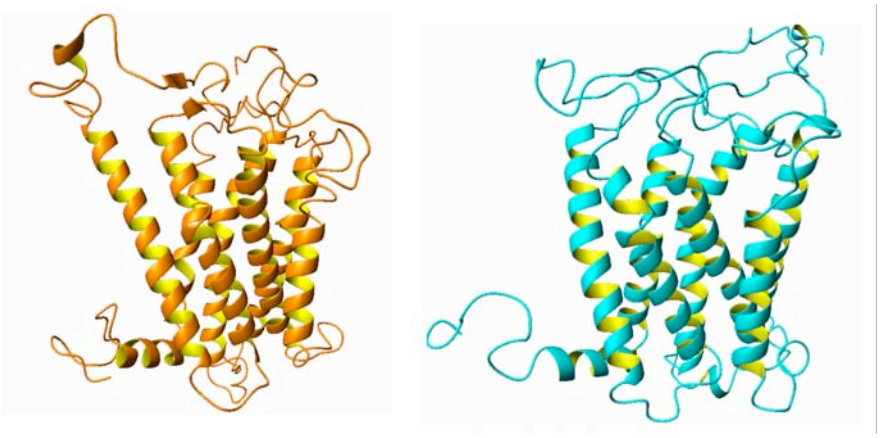
Summary and Conclusions

7.0: Summary

G-protein coupled receptors occupy a central position in the physiology of insects and are excellent drug targets in humans [1]. The 3D structures of the receptors are crucial for ligand binding but to date no 3D structure of an insect G-protein coupled receptor (GPCR) is available. Knowledge of the 3D structure of the GPCR for adipokinetic hormone (AKHR) from the malaria mosquito is crucial for the development of insect specific insecticides. Since the determination of structures of GPCRs using experimental methods has proved to be very difficult, computational molecular modelling was used to build 3D structures of AKHR.

7.0.1 Three-dimensional structures of AKHR

From the genome of *Anopheles gambiae,* the primary sequence of a putative GPCR of AKH was identified and recently published [2,3]. From the PSIPRED and MEMSAT3 predictions, it was found that the receptor has seven transmembrane helices consisting of 218 residues. The primary sequence was used to construct two models of the receptor using homology modelling to build the 7 transmembrane helix bundles based on the crystal structures of beta2-adrenergic receptor (β_2AR) [4] and rhodopsin [5]. Conformational search of the flexible loop regions was performed using high temperature simulated annealing and restrained molecular dynamic simulations. The rhodopsin-based and β_2AR-based AKHR molecules (Figure 7.1) were obtained by joining the loops to their respective helices and optimising the structures using MD simulations.



(a) (b)

Figure 7.1: a) The beta2 based structure and b) the rhodopsin-based structure.

The molecular models have similar overall structures and both have an anticlockwise arrangement of seven α -helices. Both contain the cystein-cystein disulphide bond and the eighth helix that is found in most Class A GPCRs. The 7TMs are stabilised by interhelical interactions involving the highly conserved residues [6], for example Asp94 (H2) and Asn319 (H7). As with most GPCRs, the helices are characterised by kinks that provide flexible hinges for conformational change during activation of the receptor [7] especially Met, Thr or Tyr precedes proline [8]

An overlay of the models reflected that, there are marked differences (rmsd of 0.79 nm) in the two molecular models. Most importantly, the rhodopsin-based model has a 'closed' extracellular region while the β_2AR-based structure is 'open'. Since rhodopsin has a closed structure that protects its covalently bound ligand [5], it was postulated that the β_2AR-based structure symbolizes the unbound form of AKHR whilst the rhodopsin-based represents the bound form. In addition, helices in the β_2AR-based molecule are longer and there are β -strands in the ECL1 and the N-

terminus, and a helix in the N-terminus. These features are not found in the rhodopsin-based structure.

The rhodopsin-based and β_2AR-based AKHR molecules were refined by performing extensive MD simulations in a POPC membrane to mimic the receptor's true environment. Both molecules were very stable during simulations and the overall structure was maintained. The membrane played an important role in stabilising the helical bundle. Upon optimisation of the β_2AR-based AKHR molecule, in the extracellular region, helices moved apart (Figure 4.5). This resulted in a wider extracellular region than in the template, leaving a cavity for ligand binding. Also the N-terminus was ordered and consistently formed a helical turn from Glu32-Tyr35 (Figure 4.6) and β -strands involving Tyr24-Tyr25. Similar features were also found in the third intracellular loop (Figure 4.6). On the other hand, the rhodopsin-based molecule maintained the closed conformation with a slight outward shift of the helices (Figure 4.12). The ionic bond found between H3 and H6 was broken during simulations (Figure 4.13). A helical turn was also found in the N-terminus and the third intracellular loop (Figure 4.14 and 4.15). The opening up of the β_2AR-based AKHR supports the idea that the molecule represents the unbound form of the receptor.

7.0.2 Binding pocket and Activity of AKHR

Separate docking calculations with adipokinetic hormones from *Anopheles gambiae* (Anoga_akh (pQLTFTPAWa)), modelled based on interproton distances from NMR experiments, and the blister beetle (Del_CC (pQLNFSPNWGNa)) [9], show that helix bundles and the extracellular domains of the receptor participate in hormone binding.

The binding pocket is defined by helices 2, 3, 5, 6 and 7, the second and third extracellular loops and the N-terminus (Figure 7.2).

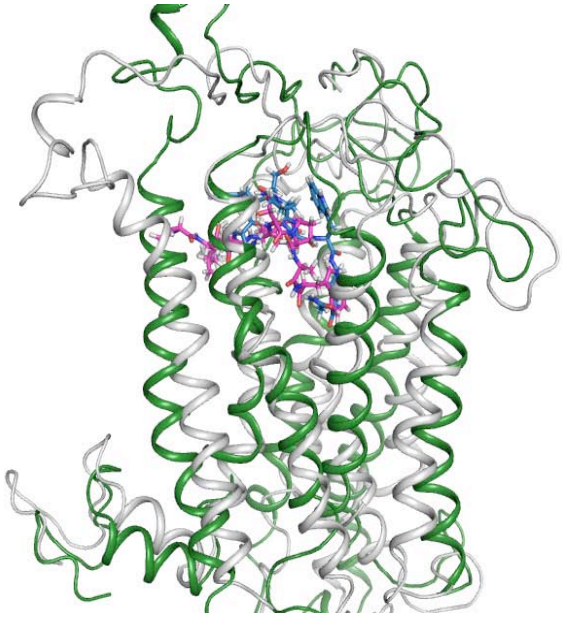


Figure 7.2 Comparison of the position of Anoga_akh (pink) and Del_CC (blue) in the receptor, AKHR. Conformations of the receptor after activation are shown by green (Del_CC activated) and gray (Anoga_akh activated) ribbons.

The binding pocket in the 'open' β_2AR-based model was easily accessed by the ligands, thus hormone binding was facilitated.

Molecular dynamics of the 'open' β_2AR-based AKHR/ligand complex have revealed that the receptor 'closes' up in the extracellular region upon activation. In the intracellular region, helices 'open' up, moving H3 and H6 apart from each other. The moving apart of these helices is crucial for receptor activation [1]. In the presence of Anoga_akh, there was a slow change from inactive to active receptor, a behaviour also observed in the activation of β_2AR by adrenaline [10]. On the other hand, Del_CC induced fast conformational changes in the receptor and the helices were drawn much closer to each other. Such a response to ligand binding was also observed when β_2AR was activated by noradrenaline [11,12]. The changes in

conformation due to ligand binding illustrate that inactive AKHR has an 'open' conformation whilst the activated form is 'closed'.

Binding of the ligands involved ionic as well as hydrophobic interactions with residues in the helix bundle and the second extracellular loop. Del_CC formed ionic bonds with mostly Tyr110, Gln209, Tyr221 and Tyr285. It was observed that Tyr285 was crucial for binding Del_CC. This ligand is more polar than Anoga_akh and contains two Asn residues that were pivotal in binding. On the other hand Ile106 was important for binding of Anoga_akh through Thr5. Due to low polarity of Anoga_akh, the aromatic side chains of Phe126 and the three Tyr residues helped to stabilise the ligand through hydrophobic contacts.

Although Anoga_akh slowly induced conformational changes, binding affinity for this ligand was higher than for Del_CC. This is evidenced by the initial estimated free energy of binding (ΔG_b) of -47.3kJ/mol for Anoga_akh compared to -38.9 kJ/mol for Del_CC. During molecular dynamics of the ligand/receptor complexes, average ΔG_b for Anoga_akh was ~ -33.92 kJ/mol (Table 6.3), whilst that for Del_CC was ~ -19.02 kJ/mol (Table 5.1). Differences in binding affinity, binding mode and the response of the receptor when activated by different ligands could imply receptor selectivity.

7.1 Conclusions

In this thesis, the primary sequence of an adipokinetic hormone receptor from the malaria mosquito, *Anopheles gambiae*, was successfully used to construct two 3D structures of the receptor, the rhodopsin-based and the beta2-adrenergic-based molecular models. The structural features and the general conformations of the

models are in agreement with most mutational data and could give an insight into the 3D structures of insect GPCRs. However, the beta2-adrenergic receptor has an 'open' conformation that represents the unbound form of the receptor but does not have the 'ionic-lock'. The rhodpsin-based structure is 'closed' and symbolises the bound form of the adipokinetic hormone receptor.

During membrane MD simulations to optimise the models, the β_2AR-based molecule attained a wider helix bundle that maintained the 'open' conformation. The change in conformation indicates the differences between the model and its template. On the other hand, there was a slight outward shift of helices in the rhodopsin-based structure and the 'closed' conformation was also preserved. The ionic bond involving the highly conserved Arg146 was broken as helices separated. The highly flexible N-terminus and third intracellular loop are structured and they consistently formed β-sheets and helical turns, especially in β_2AR-based AKHR. These structural features could be important in ligand binding and receptor activation.

Docking calculations of insect adipokinetic hormones, Del_CC, from the blister beetle and Anoga_akh, from *Anopheles gambiae*, revealed that the helix bundles, the extracellular loop domains and the N-terminus describe the binding site of the adipokinetic hormone receptor. This observation is in agreement with experimental results from mapping of the binding sites in peptide GPCRs. On the other hand, rhodopsin and beta2-adrenergic receptors have their binding sites inside their helix bundles. The 'open' conformation facilitated ligand binding by providing easy access to the binding pocket, whilst the binding pocket in the 'closed' molecule could not be accessed by the ligands from outside the receptor.

MD simulations of the adipokinetic hormone receptor/ligand complex have shown that ligand binding causes the receptor to change from an 'open' to a 'closed'

conformation in the extracellular side. Simultaneously, the intracellular side of the receptor 'opened' up as the ligands activated the receptor. The mosquito adipokinetic hormone induced a slow conversion of the receptor from inactive to active form and had high binding affinity. In contrast, Del_CC caused fast and drastic changes in conformation but had low binding affinity. In the adipokinetic hormone receptor Tyr285 was importance in binding Del_CC whilst Ile106 played a crucial role in binding of Anoga_akh. From these observations it can be concluded that the beta2-adrenergic receptor represents the unbound form of the receptor whilst the rhodpsin-based molecule represents one of its activated states.

Although at atomic-level, the molecular model of AKHR and AKHR/AKH might differ from the structures determined by x-ray crystallography or by NMR spectroscopic methods, they can be reliably used for further studies of insect GPCRs. Since adipokinetic hormone receptor is involved in the generation of energy during insect flight, knowledge of its 3D structure, binding site and activity could lead to the design of non-peptide mimetics that can block the binding pocket and lead to the development of species specific insecticides. Hence, by using genomic information to build the structure of the receptor and determine its activity, the fight against malaria could be enhanced.

7.2 Future work

Molecular dynamic simulations of G-protein coupled receptors in a membrane are presently carried out for more than 500 ns, therefore it would necessary to perform longer simulations using bigger computer clusters and/or the Coarse grained molecular dynamic simulations [13,14].

The information obtained in this work could be used in designing agonists and antagonists that block the binding site and give insect specific insectides.

Once adipokinetic hormone receptor is bound to a G-protein inside the cell, docking of a G-protein to the receptor to determine how the receptor would function can be performed.

7.3 References

1. Kobilka, B. G- protein coupled receptor structure and activation. *Biochim Biophys. Acta* **1768**(4), 794-807 (2007).
2. Kaufmann, C. & Brown, M. R. Adipokinetic hormones in the malaria mosquito, *Anopheles Gambiae*: Identification and expression of genes for two peptides and a putative receptor. *Ins. Biochem. and Mol. Biol.* **36**, 466 – 481 (2006).
3. Belmont, M., Cazzamali, G., Williamson, M., Hauser, F. & Grimmelikhuijzen, C. J. P. Identification of four evolutionary related G – protein coupled receptors from the malaria mosquito *Anopheles gambiae*. *Biochem. and Res. Comm.* **344**, 160 – 165 (2006).
4. Rasmussen, S. G. F. *et al.* Crystal structure of the human beta2 adrenergic G-protein coupled receptor. *Nature* **450**(7168), 383-387 (2007).
5. Palczewski, K. *et al.* Crystal structure of rhodopsin: A G- protein coupled receptor. *Science* **289** (5480), 739 – 45 (2000).
6. Chelikani, P. *et al.* Role of group conserved residues in the helical core of β2-adrenergic receptor. *PNAS* **104**(17), 7027 – 7032 (2007).
7. Fu, D., Ballesteros, J. A., Weinstein, H., Chen, J. and Javitch, J. A. Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding site crevice.*Biochem,* **35**, 11278-11285 (1996).
8. Deupi, X., Olivella, M., Govaets, C., Ballesteros, J. A., Campillo, M. and Pardo, L. Ser and Thr residues modulated the conformation of Pro-kinked transmembrane α-helices. *Biophys. J.* **86,** 105-115 (2004).
9. Nair, M., Jackson, G. E. & Gade, G. Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics. *J. Comp. – aided Mol. Des.***15**, 259 – 270 (2001).
10. Huber, T., Menon, S., and Sakmar, T. P. Structural Basis for Ligand Binding and Specificity in Adrenergic Receptors: Implications for GPCR-Targeted Drug Discovery. *Biochem.* **47**,(42), 11013-11023(2008)
11. Swaminath, G., Deupi, X., Lee, T. W., Zhu, W., Thian, F. S., Kobilka, T. S., and Kobilka, B. Probing the B2 adrenoceptor binding site with catechol reveals differences in binding and activation by agonist and partial agonists. *J. Biol. Chem.* **280**, 22165-22171 (2005).

12. Yao. X., Prnot, C., Deupi, X., Ratnala, V. R., Swaminath, G., Farrens, D and Kobilka, B. Couplng ligand structure to specific conformational switches in the β2-adrenoceptor. *Nat. Chem. Biol.* **2**, 417-422 (2006).
13. Rudd, R. E., and Broughton, J. Q. Coarse-grained molecular dynamics and the atomic limit of finite elements. *Phys. Rev. B* **58***,* R5893–R5896 (1998).
14. Scott, K., Bond, P., Ivetac, A., Chetwynd, A., Khalid, S., Sansom, M. Coarse-Grained MD Simulations of Membrane Protein-Bilayer Self-Assembly. *Structure*, **16**(4), 621-630 (2008).

APPENDICES

Appendix A

Predicted Secondary structure of AKHR.

```
Conf: 9962666558874478998875787789999866778997867781899899999999999
Pred: CCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHH
 AA:  MPNTMAAHINQRIEDHRNLLADWSYYANETAGEEYYEMPIDMRFNSGHILSIMVYTTLMV
              10        20        30        40        50        60
Conf: 9999899965788855887797789999999999999999986799999996196551 37
Pred: HHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCEECHH
 AA:  FSATGNLTVLSILAQRKVRASSRINIMLAHLAIAADLLVTFLMMPLEIGWAYTVRWTAGD
              70        80        90       100       110       120
Conf: 9999999999999999999999998810689703767611477998999999999857
Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHEEEEECCCCCCCCHHHHHHHHHHHHHHHHHHH
 AA:  LMCRVMAFFRTFGLYLSSFILICISVDRYFAVLKPLKVHEHRAVLMIAAAWIMSGLCSLP
             130       140       150       160       170       180
Conf: 7644678885068996599981899679999999999999999999999999999999
Pred: HHHEEEEEEECCCCCCEEEEEECCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
 AA:  QAFIFHLEGHPNITGYQQCVTYHYFEEEIYQIIYNVLVMCLMYTFPLIVILCYGSIYYEI
             190       200       210       220       230       240
Conf: 9876024442100110346788899999999999999999767999999999646556 5
Pred: HHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCC
 AA:  FSRTNPRNLESFRRSSIDVLGRAKRKTLRMTIMIVIVFVVCWTPYYVMSLWYWLDKESAK
             250       260       270       280       290       300
Conf: 567999999999999987879999814899999999856588888288588999
Pred: CHHHHHHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHCCCCCCCCCCCCCCCC
 AA:  NVDQRIQKGLFLFASTNSCMNPVVYGVFNVRKKHTKKLLKTTHEKSCGSHLTMRA
```

Conf: Confidence (0=low, 9=high), Pred: Predicted secondary structure (H=helix, E=strand, C=coil), AA: Target sequence. Blue shows a predicted helix in the n-terminus, red indicates residues in the seven transmembrane helices, cyan shows the eighth helix in the intracellular region and black gives residues in loops, N- and C-termini.

Appendix B

Sequence alignment of AKHR and Beta2 helices.

```
akhrTm1             --HILSIMVYTTLMVFSATGNLTVLSILAQRK        30
2RH1a               DEVWVVGMGIVMSLIVLAIVFGNVLVITAIAK        32
                        :  *  .  ::. *. . .** * *  *
akhrTM2             -INIMLAHLAIADLLVTFLMMPLEIGWAYT          29
2RH1A               VTNYFITSLACADLVMGLAVVPFGAAHILM          30
                     .  :     .***:: : ::*:  .
akhrTM3             --LMCRVMAFFRTFGLYLSSFILICISVDRYF---     30
2RH1a               GNFWCEFWTSIDVLCVTASIETLCVIAVDRYFAIT     35
                      : *.. : : .: :  *   *  *:*****
akhrTM4             EHRAVLMIAAAWIMSGLCSL----                20
2RH1a               KNKARVIILMVWIVSGLTSFLPIQ                24
                    :::* ::*  .**:*** *:
akhrTM5             --YNVLVMCLMYTFPLIVILCYGSIYYEIFSR---     30
2RH1a               QAYAIASSIVSFYVPLVIMVFVYSRVFQEAKRQLK     35
                      * :      : .**::::  :*: :: ..*
akhrTM6             -KRKTLRMTIMIVIVFVVCWTPYYVMSLWYWL        31
2RH1a               KEHKALKTLGIIMGTFTLCWLPFFIVNIVHVI        32
                     ::*.*:   :*  .*.:** *::::.: : :
akhrTM7             ------LFLFASTNSCMNPVVYG                  17
2RH1a               KEVYILLNWIGYVNSGFNPLIYC                 23
                      :   *  :. .** :**::*
akhrTM8             NVRKKHTKKLLK-                           12
2RH1a               RDFRIAFQELLCL                           13
                       :   ::**
```

AkhrTM1-8 are residues in AKHR helices, 2RH1a indicate residues in beta2 helices. Asterisk (*) indicates identical residues aligned, and colon (:) and dots (.) indicate similar residues. Number of residues used for each helix is given in the last column.

Appendix C

Sequence alignment of AKHR and Rhodopsin helices

```
akhrTM1         ---HILSIMVYTTLMVFSATGNLTVLSILAQ            28
1f88A           PWQFSMLAAYMFLLIMLGFPINFLTLYVTVQ            31
                   . :       *:::  . *: .*.: .*
akhrTM2         -INIMLAHLAIADLLVTFLMMPLE------             24
1f88A           PLNYILLNLAVADLFMVFGGFTTTLYTSLH             30
                 :* :* :**:***::.*  :
akhrTM3         GDLMCRVMAFFRTFGLYLSSFILICISVDRY---         31
1f88A           GPTGCNLEGFFATLGGEIALWSLVVLAIERYVVV         34
                *   *.: .** *:*  :: : *: ::::**
akhrTM4         EHRAVLMIAAAWIMSGLCSLP---                   21
1f88A           ENHAIMGVAFTWVMALACAAPPLV                   24
                *::*:: :* :*:*:  *: *
akhrTM5         EIYQIIYNVLVMCLMYTFPLIVILCYGS-              26
1f88A           ETNNESFVIYMFVVHFIIPLIVIFFCYGQ              27
                *  :  : : :: : : :*****:  :.
akhrTM6         RAKRKTLRMTIMIVIVFVVCWTPYYVMSLWYW-          32
1f88A           KAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFT          33
                :*::::. **.*::**.*::** **  :::: :
akhrTM7         --IQKGLFLFASTNSCMNPVVYG                    21
1f88A           PIFMTIPAFFAKTSAVYNPVIYI                    23
                  : .  .:**.*.:  ***:*
akhrTM8         NVRKKHTKKLLKTT                             14
1F88a           MMNKQFRNCMVTTL                             14
                 :.*:. : ::.*
```

1F88a indicate residues in rhodopsin helices. Asterisk (*) indicates identical residues aligned, and colon (:) and dots (.) indicate similar residues. Number of residues used for each helix is given in the last column.

Appendix D

Blind docking parameters of Del_CC to AKHR.

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 akhr_11_1.maps.fld	# grid_data_file
map akhr_11_1.A.map	# atom-specific affinity map
map akhr_11_1.C.map	# atom-specific affinity map
map akhr_11_1.HD.map	# atom-specific affinity map
map akhr_11_1.OA.map	# atom-specific affinity map
map akhr_11_1.N.map	# atom-specific affinity map
elecmap akhr_11_1.e.map	# electrostatics map
desolvmap akhr_11_1.d.map	# desolvation map
move akh_11_1.pdbqt	# small molecule
about 13.6675 28.2697 20.9725	# small molecule center
tran0 13.668 28.270 20.973	# initial coordinates/A or random
quat0 random	# initial quaternion
ndihe 30	# number of active torsions
dihe0 random	# initial dihedrals (relative) or random
tstep 2.0	# translation step/A
qstep 50.0	# quaternion step/deg
dstep 50.0	# torsion step/deg
torsdof 32 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 150	# number of individuals in population

ga_num_evals 100000000 # maximum number of energy evaluations

ga_num_generations 100000000 # 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

These parameters were used to identify the binding pocket in both the rhodopsin –based and the beta2 –based models.

Appendix E

Refined docking parameters using beta2 – based model and Del_CC.

Parameter and quantity	Description
move akh_12.pdbqt	# small molecule
ndihe 32	# number of active torsions
torsdof 32 0.274000	# torsional degrees of freedom and coefficient
ga_pop_size 300	# number of individuals in population
ga_num_evals 500000000	# maximum number of energy evaluations
ga_num_generations 500000000	# maximum number of generations
ga_run 100	# do this many hybrid GA-LS runs
analysis	# perform a ranked cluster analysis

These parameters were used in addition to the Autodock4 default parameters as shown in Table 1a.

Appendix F

Interproton distances from NMR (nm) used during Molecular dynamic simulations. pGlu-LTFTPAW-NH_2 peptide in H_2O-D_2O 10:1, pH ca. 4 (phosphate buffer) T=290K, ROESY 150 ms (*in brackets: pseudoatom corrections according to: Wüthrich, K., NMR of Proteins & Nucleic Acids, p. 178, Wiley 1986)

Distances represent *lower limits* between protons/pseudoatoms involved.

Proton Number	Assignment	Distance (Á) corrected*
1	LNH-pEbb'	4.3(m)
2	ANH-Pbb'	4.7(m)
3	FNH-T3g	4.3(m)
4	T3NH-Lbb' (or g)	4.2(m)
6	T5NH-Fbb'	4.0(m)
7	WNH-Aa	3.3(m)
8	WNH-Ab	3.1
9	CONH2a-Wbb'	4.2(m)
10	CONH2a-Wa	2.6
11	CONH2b-Wa	3.7
12	LNH-T3NH	2.8
13	FNH-T3NH	2.8
14	LNH-pENH	3.2
15	ANH-WNH	3.0
16	T5a-Pd.d'	3.6(m)
17	T5b-Pd,d'	3.6(m)
23	Pb-Pb'	1.8 REF!

Appendix G

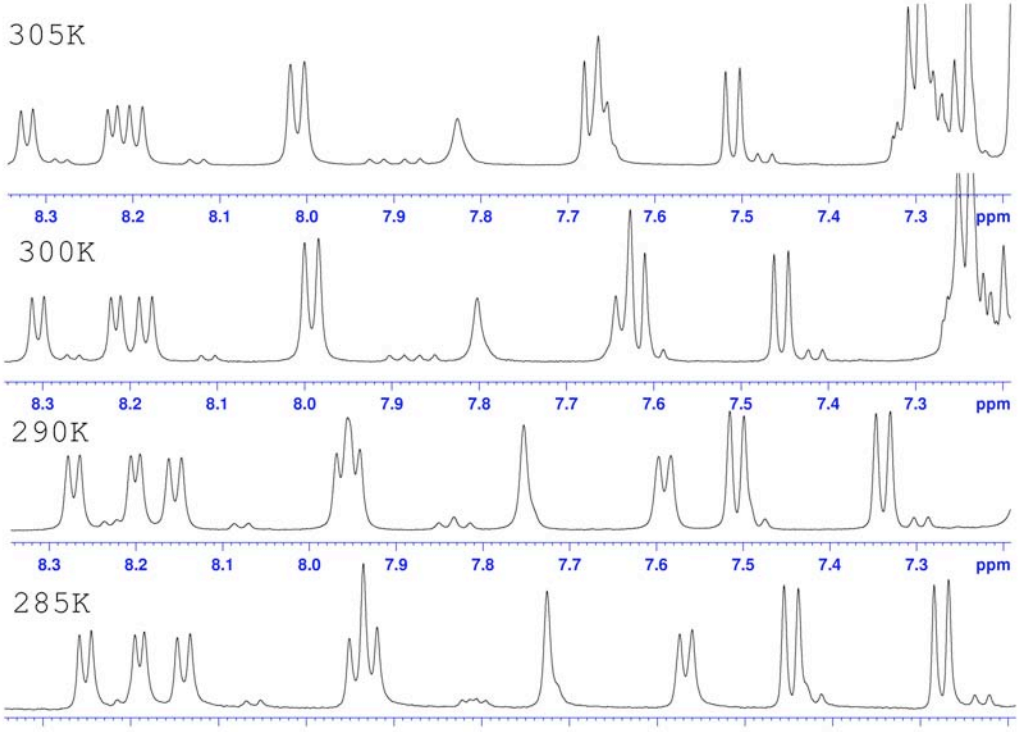
pGlu-LTFTPAW-NH_2 peptide H_2O-D_2O 9:1, DPCC micelles in pH ca. 4 phosphate buffer, T=290K, NOESY 100 ms
(*in brackets: pseudoatom corrections according to:
Wüthrich, K., NMR of Proteins & Nucleic Acids, p. 178, Wiley 1986)
Long-range contacts in red
Distances represent *lower limits* between protons/pseudoatoms involved.

#	Assignment	Distance (Å) corrected*
1	LNH-pEbb'	4.2(m)
3	ANH-Pb	3.9(m)
4	T3NH-Lbb' (or g)	3.9(m)
6	T3NH-Ldd'	5.2(q)
7	FNH-T3g	3.8(m)
8	FNH-Lbb' (or g)	4.4(m)
10	T3NH-La	3.0
12	FNH-T3b	3.6(m)
12a	FNH-T3a	3.0
13	T5NH-Fa	3.0
14	T5NH-Fbb'	3.5(m)
15	WNH-Aa	3.0
16	WNH-Ab	3.5(m)
17	Ldd'-F2,6	7.4(qr)
18	Ldd'-F3,4,5	7.2(qr)
19	T3g-F3,4,5	6.5(qr)
20	T3g-F2,6	7.0(mr)
21	T5g-W2	5.4(m)
22	Ab-W2	4.4(m)
23	Lbb' (or g)-F3,4,5	6.3(mr)
25	Pbb' (or gg') -F3,4,5	6.4(mr)

26	Lbb’ (or g)- F2,6	6.5(mr)
27	Pgg’ (or bb’)-F2.6	6.0(mr)
29	LNH-T3NH	2.9
30	LNH-pENH	2.7
31	ANH-WNH	2.7
32	T5NH-F3,4,5	5.0(r)
36	T5a-Pdd'	3.7(m)
37	T5b-Pdd'	3.9(m)
38	Wb-Ab	5.6(mm)
40	T5a-Pbb’	4.3(m)
41	T5a-Pgg’	4.0(m)

Appendix H

NMR spectra of Anoga_akh in water at varying temperatures



Chemical shifts (ppm)

Appendix I

A plot of H^N Chemical shifts in water against time