A single AKH neuropeptide activating three different fly AKH-receptors: an insecticide study via computational methods.

by

Ibrahim A. Abdulganiyyu

BSc (Hon) Chemistry

Thesis presented for the degree of Doctor of Philosophy
in the Department of Chemistry
University of Cape Town

Supervisor**: Prof. Graham E Jackson**

Co-Supervisor**: Prof. Heather G. Marco**

NOVEMBER 30, 2020

Table of Contents

DECLARATION

By submitting this thesis electronically, I at this moment declare that this thesis leading to the award of *Doctor of Chemistry* in the faculty of science, University of Cape Town, is a product of my work, conducted under the supervision of Professor G. E. Jackson and Professor Heather Marco. That I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that it has not been submitted before for a degree or examination at this or any other university. All materials consulted, and sources of information are duly and adequately acknowledged.

Signed by candidate 30/11/2020

---------- -- ----------

Ibrahim A. Abdulganiyyu Date

DEDICATION

This thesis is dedicated to Almighty ALLAH, who made it possible for me to go through this training and also my lovely parent and my families who supported me in every way possible.

ACKNOWLEDGEMENT

In the name of Allah (SWT), the Most Gracious and Most Merciful. All praise is due to Allah (SWT), the Lord of the world, who honoured humanity and favoured them markedly above a significant part of His creation. May His peace and blessing be upon His Noble **prophet** Muhammad, his entire household, companion and those who follow him till the end of time (Amin).

Nuclear magnetic measurements (NMR) were carried out at School of Chemistry, Bio21 Institute, University of Melbourne, Melbourne, Vic. 3010, Australia. Special thanks to Frances Separovic and Marc-Antoine Sani.

I would like to thank Prof. Graham E. Jackson and Prof. Neil for introducing me to the field of NMR.

I gratefully acknowledge the encouragement, tolerance and assistance of my Supervisor.

Prof. Graham E. Jackson, you have been very patient with me and always shared your knowledge, even when I had a hard time of it! You have made me believe in myself.

My immense gratitude goes Heather G. Marco, for her contributions during the entire period of my studies, also thank you for the care, encouragement, and support.

My profound gratitude goes to my dear one Sumayya Ahmad Tijjani for her love, support, patience, and tolerance displayed during the entire program, to you, I say thank you and will forever cherish you.

Also, I wish to extend a warm thank-you to Dr Stephen Fineberg. Your help and guidance have been irreplaceable, and you are a constant source of knowledge and great ideas!

To my fellow students and members of the Jackson group Ahmad, Fatin Khadija tolu Daud, Amouna and Hanna. You had made my days fun, even when every experiment went down the drain.

I will not forget Prof Garba Uba Goje for the fatherly support and guidance, for, without it, this study would not have anyway been possible. Also, special thanks to Ismail Adefeso (Amir), Shakirudeen Lawal and the entire Dua Group for making my stay in Cape Town a worthy one.

Most of all, I gratefully acknowledge the encouragement, tolerance, and assistance of all the students and staff of the department for their immense contribution to my academic build-up and contributions to my total development.

My sincere appreciation also goes to my kids Fatima Ibrahim and Khadija Ibrahim, Hanifa Sunusi and Nasir Sanusi. My siblings Late Yusuf, Aminu, Hassan, Halira and Hassana. To you, I say a big thank you.

Specially thanks to TETFUND via Jigawa state polytechnic Dutse, ASSUP and its leadership, especially the chair Basiru Yusuf, I say a big thank you for your assistance. Without your support, this study would not have materialised.

Finally, I thank all Jigawa state polytechnic staffs, especially the rector Ali Abdu and the registrar Ilyas Jahun, the Gumel emirate, my cosine Abdulaziz Chiroma, Bulama Abdulaziz

and my best friend Abdulkadri Mahmud Ibrahim, Kazim Olamilakan, Adamu Haruna, Abdulrazak Abubakar, Abdulmajid Aminu, Sani Y. Adam, Muhammad Yusuf Gumel, Mamiya, Yawale babawarwade, Kabiru Roni, Ibrahim Tahir, the entire MCSF U07CH and all those whom their names were too numerous to mention. For their assistance during the entire course of my study. May Allah reward you all abundantly (Amin).

Publications and conference contributions

Publications

1. Ibrahim A. Abdulganiyyu, Krzysztof Kaczmarek, Janusz Zabrocki, Ronald J. Nachman, Elisabeth Marchal, Sam Schellens, Heleen Verlinden, Jozef Vanden Broeck, Heather Marco and Graham E. Jackson, Conformational analysis of a cyclic AKH neuropeptide analogue that elicits selective activity on locust versus honeybee receptor; *Insect Biochemistry and Molecular Biology*, https://doi.org/10.1016/j.ibmb.2020.103362**.** ***Accepted*** 15 March 2020 ***Published online***: 28 July 2020.

2. Ibrahim A. Abdulganiyyu, Marc-Antoine Sani, Frances Separovic Heather Marco and Graham E. Jackson, The adipokinetic hormone, Phote-HrTH, from the fruit fly: solution structure and receptor binding model; *Australian Journal of Chemistry*, https://doi.org/10.1071/CH19461. ***Accepted***: 11 November 2019 ***Published online***: 17 January 2020.

Conference contributions

1. Poster presentation at the 43rd national convention SACI National Convention, 2019, Pretoria, South Africa.

2. Poster presentation at the Postgraduate Symposium 2019, University of Cape Town, South Africa.

3. Poster presentation at the 18th Asian Chemical Congress and the 20th General Assembly of the Federation of Asian Chemical Societies 2019, Taipei, Taiwan.

ABSTRACT

Flies are a widely distributed pest insect that poses a significant threat to food security. Flight is essential for the dispersal of the adult flies to find new food sources and ideal breeding spots. The supply of metabolic fuel to power the flight muscles of insects is regulated by adipokinetic hormones (AKHs). The fruit fly, *Drosophila melanogaster*, the flesh fly, *Sarcophaga crassipalpis*, and the oriental fruit fly, *Bactrocera dorsalis* all have the same AKH that is present in the blowfly, *Phormia terraenovae*; this AKH has the code-name Phote-HrTH. Binding of the AKH to the extracellular binding site of a G protein-coupled receptor causes its activation. In this thesis, the structure of Phote-HrTH in SDS micelle solution was determined using NMR restrained molecular dynamics. The peptide was found to bind to the micelle and be reasonably rigid, with an S^2 order parameter of 0.96. The translated protein sequence of the AKH receptor from the fruit fly, *Drosophila melanogaster*, the flesh fly, *Sarcophaga crassipalpis*, and the oriental fruit fly, *Bactrocera dorsalis* were used to construct two models for each receptor: *Drome*-AKHR, Sarcr-AKHR, and Bacdo-AKHR. It is proposed that these two models represent the active and inactive state of the receptor. The models based on the crystal structure of the β-2 adrenergic receptor were found to bind Phote-HrTH with a predicted binding free energy of –107 kJ mol^{-1} for *Drome*-AKHR, –102 kJ mol^{-1} for Sarcr-AKHR and –102 kJ mol^{-1} for Bacdo-AKHR. Under molecular dynamics simulation, in a POPC membrane, the β-2AR receptor-like complexes transformed to rhodopsin-like. The identification and characterisation of the ligand-binding site of each receptor provide novel information on ligand-receptor interactions, which could lead to the development of species-specific control substances to use discriminately against these pest flies.

Keywords

Adipokinetic hormones, Adipokinetic hormone-receptor, Beta2-adrenergic receptors, Docking, G-protein, coupled receptors, GROMACS, Homology modelling, Molecular dynamics simulation.

1

Introduction

1.0 Background

Insects are the most abundant and most diverse group of animals on this planet. It is not surprising, therefore, that insects account for many pests as seen from a human perspective. They compete for the same food (agricultural and stored food pests)[1], and they act as vectors of numerous pathogens to both man and domesticated animals with severe consequences (medical pests and transmitters of disease)[2,3]. However, not all insects are pests, and this can be seen in the large insect order Diptera, which comprises about 152 000 described fly species [4]. A few examples of useful fly species are the pollinators of flowers to result in fruit and seed production, like hoverflies (Family: Syrphidae)[5–7]; flies that are used in wound care, like larvae of the family Calliphoridae that feed on necrotic tissues and are also a source of pharmacologically active substances such as antimicrobial, antiviral and antitumor compounds [8–10] and fly larvae of the black soldier fly *Hermetia illucens* (Family: Stratiomyidae) that are used at an industrial level as organic waste removers and simultaneously as a source of economically valuable products, such as oils and dried protein feed for chicken and fish [11]. Other fly species may have both beneficial and pest status and all insects play a role in the ecosystem [6,7]. Adult flies, like many other insects, have wings that they use for aerial

locomotion (flight) to find food sources, for dispersal or for finding a mate [12]. Insect flight is always based on aerobic metabolism [13] and is dependent on efficient oxidation of carbohydrates and lipids in the flight muscles [14–17]. Carbohydrates comprise the primary supply of fuel for short flights and are stored as glycogen in the fat body and trehalose in the haemolymph [15,18]. At the start of a flight, trehalose is transported to the flight muscles where it is hydrolysed to glucose for use [15]. Lipid reserves are critical for more prolonged and demanding flights and are stored in the fat body as triacylglycerol (TAG) [19,20]. TAG needs to be transformed into stereospecific 1,2 diaclyglycerol (DG) having an SN-1,2-configuration [21–23], before being released into the haemolymph and transported to the flight muscles. Membrane-bound lipase enzymes catalyse the hydrolysis of DG to produce fatty acids, and the successive oxidation leads to energy release [24]. Mobilisation of stored metabolites from the fat body into the haemolymph is regulated by members of a neuropeptide family, generically known as the adipokinetic hormone family (see Section **1.3.0** for further understanding**)** [15–17]. Adipokinetic hormones (AKHs) are synthesised and stored in intrinsic neurons of the corpus cardiacum (CC) in the head of insects, released into the hemolymph as required, and act as extracellular messengers by binding to their cognate G-protein coupled receptor (GPCR) on fat body cells. [14–17] Binding of the AKH to the extracellular domain of its cognate AKH-receptor (the so-called AKHR) elicits an intracellular signal transduction cascade that ends in the activation of critical enzymes for substrate mobilisation, like glycogen phosphorylase and lipases [15,25–28]. Given the importance of the AKH signalling system in insect metabolism and the fact that similar endocrine pathways are selected by pharmacology as targets of drug treatment of human diseases, it is no wonder that insect hormonal systems are being looked at with great interest as potential targets for specific pest control [29–33]. Thus, the AKH-AKHR system of insects is being considered as a potential target for novel pest control agents (Marco and Gäde, 2020).

There is an ever-louder call for such alternatives to the current chemical pesticides that are being used globally. Apart from adverse effects on the environment and human health [34–36], the unspecific action of commercial chemical pesticides has resulted in "superbugs", i.e. pest insect populations that are resistant to the pesticide compounds [37,38], and are also implicated in the observed decline in insect biomass and biodiversity [35,39,40]. It has become necessary, therefore, to consider all the ecological services offered by different insect species, including being part of the natural food chain that sustains other animals, like birds and fish, and to steer away from the indiscriminate conventional way of combatting insects towards bio-rational pest management. This has spurred on the search for "green insecticides", i.e. species- (or group-) specific control agents (peptide mimetics that are based on the insect's hormones) to target and control pest insect numbers with little or no adverse effect on beneficial insects, other organisms and the environment. One idea for developing a so-called green insecticide is to target the GPCRs of pest insects[29–33,41,42] in the hope that these are species-specific. For this to be successful and applicable, research into ligand-receptor interactions is needed for both pest and beneficial insects to ascertain significant differences/similarities in the signalling components and to identify those components that are specific to pest insects which may be further investigated for drug development.

In this context, the current thesis was developed to focus on the AKH-AKHR interactions in three different fly species: *Drosophila melanogaster* (vinegar/fruit fly), *Bactrocera dorsalis* (oriental fruit fly) and *Sarcophaga crassipalpis* (flesh fly). All three fly species produce the same AKH in the corpora cardiaca, viz. an octapeptide (pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2) that was discovered in the blowfly, *Phormia terraenovae* and is thus, code-named Phote-HrTH, and the respective AKHR is known from each species [43–45]. This provides us with an ideal experimental system in which we can investigate how an endogenous ligand (Phote-HrTH) may interact with its endogenous AKH receptor in beneficial, and pest fly species. The

resulting models will shed light on whether the same amino acids are involved with ligand binding and whether the same conformation is maintained in all species. Hence, insight can ultimately be gained into whether a species-specific effect can be predicted between the three fly species and possibly others who also produce Phote-HrTH as a potential lead for peptide mimetics. Computational techniques were employed here to study AKH-AKHR interactions. Via nuclear magnetic resonance (NMR) restrained molecular dynamic (MD), the solution structure of the ligand (Phote-HrTH) was elucidated. Homology modelling techniques were used to construct two 3D models for each of the fly species, using β2AR and rhodopsin crystal structures as templates. The two receptor models built for each species differ in a few critical features, the most important of which is that the β2AR-based models have an open conformation, while the rhodopsin-based models have a closed structure[46]. To enable us to imitate the receptor's precise nature, long MD simulations (100 ns) were conducted in a model membrane to determine stability and conformational changes during agonist binding.

In order to validate the model built of the fruit fly AKH-receptor (Drome-AKHR), *in-silico* structure-activity studies were conducted and compared with biological activity data.

1.2.0 Biology of Diptera (true flies)

"True flies" (order Diptera) are distinctly different from other insects that are colloquially known as a type of "fly." However, belonging to another order, such as butterflies (order Lepidoptera), scorpionflies (order Mecoptera), dragonflies (order Odonata), stoneflies (order Plecoptera) and mayflies (Ephemeroptera). Whereas other winged insects have four wings (2 pairs of wings), dipterans are distinct in that they have only two wings (1 pair of wings) – an ancestral pair of hindwings have evolved into small structures called halteres, which serve as sensory organs that detect rotation of the body during flight and provides positional feedback to the flight muscles via the nervous system) [47]. Diptera's diverse order includes mosquitoes,

midges, crane flies, marsh flies, flesh flies, fruit flies, house flies, robber flies, and tsetse flies, to mention but a few. Dipterans have a holometabolous lifecycle (in other words, a complete metamorphosis that includes four life stages, i.e., egg, larva, pupa, and imago or adult). The larval and adult forms of a fly species may be classified as pests or beneficial according to their diet (blood, pollen/nectar, rotting organic matter, fruits and vegetables, and other insects). Beneficial services of flies include being part of the ecological food web as a food source for other animals, such as reptiles, fish, birds, and predatory insects; ecosystem engineering, by conditioning and enriching soil; accelerated removal and decomposition of biological matter, quality indicators of water bodies, pollinators of many plants, medical intervention in wound healing and source of antibiotics [48,49]. Additionally, certain fly species are used in forensic entomology as markers for estimating corpse death, used as model organisms for insect physiology, biochemistry, and genetics. [50–52]

The insect pests can be divided into medical and agricultural pests. Some fly species, especially blood-sucking flies, pose a significant threat to man's health. In feeding, they can transmit severe diseases such as trypanosomiasis (sleeping sleekness), malaria, cholera, and yellow fever [2,47]. The agricultural pests are adult and larvae feeding on fruits and vegetables even before they are fully ripe [1]. Adult flies transmit disease vectors to other animals, e.g., equine infectious anaemia anthrax in cattle and sheep [3]. The line between good and wrong concerning the status (beneficial or pest) of an insect is not always as clearly and finally delineated as expressed above. Therefore, to avoid future ecological problems arising from the drastic alteration of species composition, we should not apply broad-spectrum pesticides that kill invertebrates indiscriminately.

This thesis focuses on three different fly species with different economic importance attached to each but with the same endogenous mature AKH peptide and a structurally similar (but not identical) AKHR for their metabolic processes. The selected dipteran species are shown in

Figure 1.1: the vinegar/fruit fly, *Drosophila melanogaster*, has both beneficial and pest insect status (agricultural pest and genetic model for developmental studies) [53]; the oriental fruit fly, *Bactrocera dorsalis*, is a significant and invasive pest of fruit orchards [54], and the flesh fly, *Sarcophaga crassipalpis*, is a common laboratory animal used in the study of gene expression and the study of diapause in insects [55], thus, a beneficial insect.

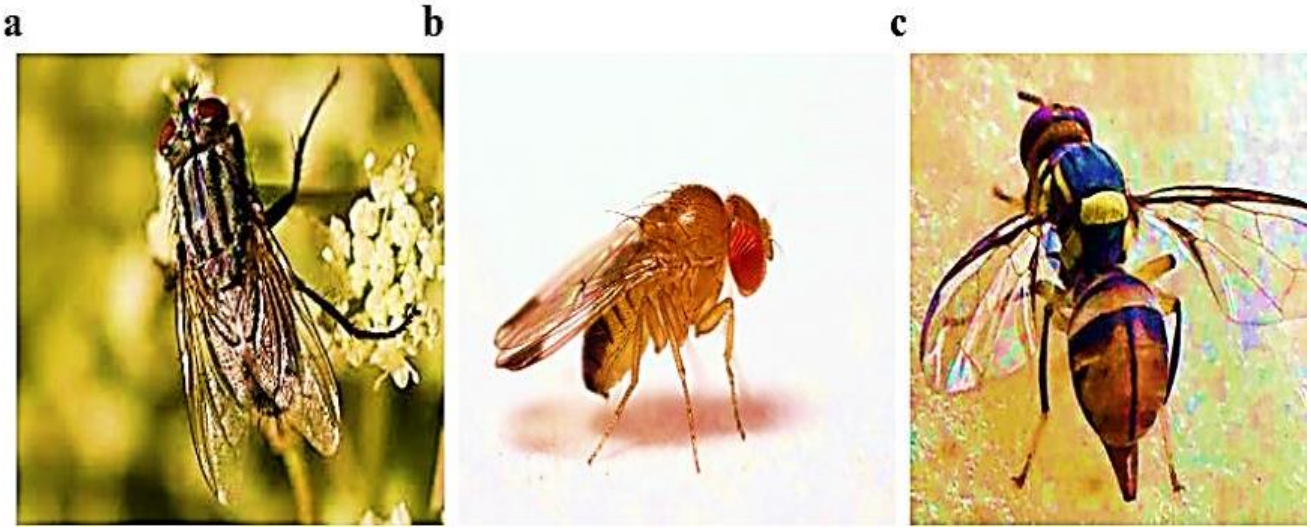


Figure 1. 1: a) Dorsal view of adult female *Sarcophaga crassipalpis* taken from Bänziger and Pape, 2004 [56]. *b*) side view of adult *D. melanogaster, taken from* Mansourian et al., 2018 [57]. c) Dorsal view of adult *Bactrocera dorsalis* taken from Schutze et al., 2015 [58].

1.3.0 Adipokinetic Hormones (AKHs) in Diptera

Foraging flies are active fliers, and for this, they require large amounts of energy. The supply of fuel metabolites for oxidation in insects' flight muscles is regulated by a neuropeptide family generically referred to as the adipokinetic hormone (AKH) family. AKHs are synthesised in intrinsic neurons of the *corpora cardiaca* (CC, a pair of retro-cerebral glands). In dipterans, the CC form part of the ring gland which is a complex structure formed of the larval prothoracic gland fused with the CC and the corpora allata during development and located in the anterior thorax in the adult stage[59].

The first AKH was isolated from the locust, *Locusta migratoria* [60], while the first dipteran AKHs were sequenced from the horse fly, *Tabanus atratus* in 1989 [61]. Since the first AKH sequence, over 80 AKH sequences are now known from insects [62]. The distinctive features of insect AKHs are that they are either octa-, nona- or decapeptides with blocked termini (a pyroglutamate at the N-terminus) and an amidated C-terminus; a Phe or Tyr in position 4, Trp in position 8 and Gly in position 9 [63]. In 1990, an octapeptide AKH with the primary sequence pGlu–Leu–Thr–Phe–Ser–Pro–Asp–Trp–NH_2 was isolated and sequenced from the blowfly *Phormia terraenovae.* It was shown to have an aspartic acid in position 7, which resulted in the peptide having an overall negative charge [64]. This is the first identified charged member of the AKH family, and because it demonstrated hypertrehalosaemic activity *in vivo* in *P. terraenovae*, it was, thus, named Phote-HrTH [64,65]. The same peptide is produced in fruit flies, e.g., *D. melanogaster*[45,66] and *B. dorsalis*[44], the flesh fly (*S. crassipalpis*) [41,43], as well as in a variety of other fly pest species[67,68]. Phote-HrTH has been called Drome-AKH in *D. melanogaster* [45] and Bacdo-AKH in *B. dorsalis* [44].

Whereas some insects produce more than one AKH, for example, two AKHs are synthesised in the malaria mosquito *Anopheles gambiae* [69] and tsetse flies[70], and the crane fly *Tipula paludosa* makes three AKHs, most other Diptera have only one AKH [71]. In Diptera, AKHs may function as hypertrehalosaemic hormones (mobilising carbohydrates), hyperlipidaemia hormones (mobilising lipids), or as hyperprolinaemic hormones (mobilising proline as fuel metabolite) (Gäde and Marco, 2013). AKHs work by binding to a GPCR [16,28,72–74] situated in the plasma membrane of adipocytes. The binding of the peptide ligand leads to the activation of intracellular pathways that trigger the mobilisation of carbohydrate or lipid reserve, and proline in some insects [25,26,28,75–77].

1.4.0 G-protein coupled receptors (GPCRs)

1.4.1 Secondary Structure and Function

GPCRs are known as seven transmembrane receptors (7TM receptors), the most diverse and significant group of membrane receptors [78] and considered the major superfamily of membrane-bound proteins [76,79–83]. GPCRs share a similar overall tertiary structure: 7TM α - helices, with alternating intracellular and extracellular loops connecting the helices, and a G protein that associates with the intracellular region of the C-terminus of the receptor protein [27,43,45,69,83–87]. **Figure 1.2** is a schematic diagram of the β-2-adrenergic receptor (β2AR) showing the structural domains commonly found in most GPCRs.

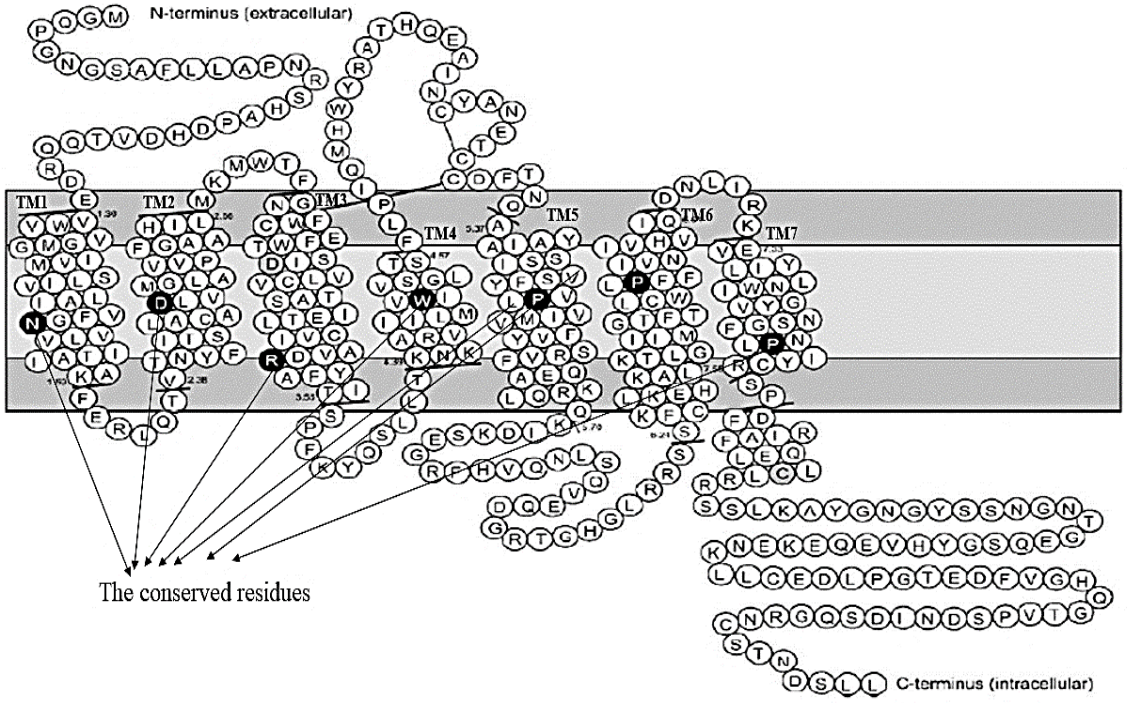


Figure 1. 2: Schematic 2D diagram showing the overall transmembrane topology of the β-2-adrenergic receptor (β2AR). TM1-TM7 represents transmembrane helices 1-7. The core region is represented in light grey and the two water-lipid interface regions of the lipid in dark grey. The conserved residues in each of the 7 transmembranes are in black[88].

Each of the 7 TM helices possesses the conserved residue motifs patterns found in most class A GPCRs, namely Asn in TM1, Asp in TM2, Arg in TM3, Trp in TM4, Pro in TM5, Pro in TM6, and Pro in TM7 (see **Figure 1.2**). [89–92]. GPCRs are also characterised by the presence of disulphide bonds linking the first two extracellular cell loops (ECL I and -II); these loops help in the stabilisation of the transmembrane helices and also take part in ligand-receptor binding [46,93–97]. The Rhodopsin receptor has one disulphide bond between Cys110 and Cys187, while the β2AR receptor has two disulphide bonds between Cys106/Cys191 and Cys184/Cys190. Both receptor molecules contain a palmitoylation site in the intracellular region. For the Rhodopsin molecule, the site is between Cys322 and Cys323 while for the β2AR molecule the palmitoylation site is Cys341 [96,98–101].

GPCR ligands comprise neurotransmitters which include, Ca^{2+} ions, proteins, amines, and hormones [102–106]. The binding of a ligand to the extracellular region of a GPCR causes a conformational change in the 7 TM regions, which activates the receptor [106–111]. Senses, such as taste, smell, vision, pain, along with secretion, metabolism and neurotransmission are biological processes where GPCRs play intermediary roles [112]. Because of the physiological importance of GPCRs, receptor-ligand interactions have been targeted successfully with therapeutic drugs in medicine[33,46,113,114], and are being considered as suitable targets for insect control agents [29–33,41,114–116].

Most ligands bind as antagonists (inhibitors of receptor activity, antagonists mimic ligands that bind to a receptor and prevent receptor activation by a natural ligand). Moreover, some ligands bind as agonists (are mimetic of the natural ligand and produces a similar biological effect as the natural ligand when it binds to the receptor) to the GPCR [117–120]. At the same time, some ligands are allosteric modulators (change the receptor's response to stimulus by increasing or decreasing the strength of the signal a receptor sends into the cell) for GCPRs [88,121–123].

1.4.2 Classification of GPCRs

GPCRs are categorised into six prominent families based on sequence homology and functional similarity [117,124–126]. These are Class1 or A (Rhodopsin-like family), Class 2 or B (Secretin receptor family), Class 3 or C (Pheromone /Metabotropic glutamate), Class 4 or D (Fungal mating Pheromone receptors), Class 5 or E (Cyclic AMP or Taste-like receptors), and Class 6 or F (Frizzled / smoothened) GPCRs. Amongst these GPCR families, class 1 or A has the most significant number of identified/known structures. The first GCPR structure determined from this family was rhodopsin [127,128]. Most type A GPCR structures were elucidated based on homology with the rhodopsin structure. Amongst the GPCRs built from rhodopsin is beta-2 Adrenergic receptor (β2AR) in the year 2007, after which other GCPRs have been elucidated using either rhodopsin or β2AR as a template [125,129,130]. As of May 2019, the ChEMBL database had documented 772 GPCRs, amongst which only 30 crystal structures have been recorded. ChEMBL_20 contains 688 Class A, 44 Class B, 35 Class C, and 1 of each Class D, E and F sequences[131–134].

1.4.3 Adipokinetic hormone Receptor (AKHR) of the dipteran species studied in this thesis.

The AKHR is a fundamental part of the AKH signal transduction pathway. The AKHR was first sequenced in *D. melanogaster* and *Bombyx mori* (silkworm) and identified as belonging to the class A GPCR superfamily [45,135]. The AKHRs were discovered to be evolutionarily and structurally related to the gonadotropin-releasing hormone receptor from vertebrates [45]. Other dipteran AKHRs were cloned and characterised from malaria and yellow fever mosquitoes, *A. gambiae* and *Aedes aegypti* [69,136] from *G. morsitans morsitans* [137], *S. crassipalpis* [43] and *B. dorsalis* [44]. With the increase in whole-genome sequencing of animals, many more AKH and AKHR genes are annotated, and sequences can be predicted for Diptera and other insects (see

for example Flybase – an online bioinformatics database: https://flybase.org and Vectorbase – bioinformatics resources for invertebrate vectors of human pathogens: https://vectorbase.org).

An alignment of the primary amino acid sequence of the *S. crassipalpis* AKHR [43], *D. melanogaster* AKHR [43], and the *B. dorsalis* AKHR [44] are presented in **Figure 1.3**.

Figure 1. 3: Multiple primary amino acid sequence alignment of adipokinetic hormone receptors (AKHRs). The flies are represented with Genbank database accession numbers. The flesh fly *S. crassipalpis* (AOA1B2RVC93), the fruit fly *D. melanogaster* (AAN10047.1), and the oriental fruit fly *B. dorsalis* AQX83416 [43–45].

Figure 1.3, the primary amino acid sequence length of the flesh AKHRs, is between 424 and 455 amino acids with high identity (over 70%) to each other [43,44]. These fly AKH receptors have the seven-transmembrane topology that is characteristic of the GPCR superfamily [43,44,76]. Furthermore, the amino acid residues, patterns, and motifs generally found in the 7 TM helices

of other Class A GPCRs are also found in all the AKHRs; see **Table 1.1**, where X signifies any amino acid residue.

Table 1. 1. The highly conserved residues, patterns, and motifs found in AKHRs and Class A GPCRs and those identified in three fly AKHRs. X represents amino acid residues common to other Class A GPCRs.

Helix	Residue/Motif/Patternsin GCPRs	Residue/Motif/Patterns In the flesh fly	Residue/Motif/Patterns In the fruit fly	Residue/Motif/Patterns In the oriental fruit fly
1	N	N	N	N
2	LAxxD	LAIAD	LAIAD	LAIAD
3	(D/E)R(Y/H)	DRY	DRY	DRY
4	W	W	W	W
5	(F/Y)XXPXXXXXXXY	YACPLITFIYCY	YAFPLITFIYCY	YAFPLITFIYCY
6	(F/Y)XXXWXPYY	FIICWTPYY	FIICWTPYY	FVICWTPYY
7	(N/D)PXXY	NPIVY	NPLVY	NPLVY

These conserved residues play a vital role via helical interactions to stabilise the helices. Also, they are essential for receptor activation and ligand binding.

AKHR was shown to be expressed in the fat body and specific nervous tissues of cockroach and tobacco hornworm, *Manduca sexta* [138,139]. Also, the AKHR expression is found in the abdomen and thorax of *A. aegypti* and *A. gambiae*– presumably in the fat body [69,136]. They were thus confirming the role of AKHs in mobilising stored metabolites from the fat body. During receptor activation (see **Figure 1.4**), insects which fully oxidise carbohydrates for flight (like cockroaches), or insects which oxidise carbohydrates in conjunction with lipids for flight (like the locusts) or insects which oxidise proline for flight (some beetles). Essential for this regard is that they possess the endogenous AKHs as a ligand which bind to the extracellular part of Gq-protein-coupled receptor, to activate a phospholipase C (PLC). The resulting inositol trisphosphate (IP_3) releases Ca^{2+} from internal stores. In addition, the influx of extracellular

Ca^{2+} is amplified and, via a kinase cascade, glycogen phosphorylase is activated, which lead to the production of glucose-1-phosphate. glucose-1-phosphate is transformed into trehalose, which is discharged into the haemolymph. In insects which use lipids for prolonged flight (like the locust) or proline for flight (certain beetles), adenylate cyclase (AC) is activated after the AKHs bind to their respective Gs-protein-coupled receptor. Also, in locusts, when AC is activated, cyclic AMP is synthesised. The resulting cyclic AMP, together with the messengers both intracellular and extracellular Ca^{2+}, activate a triacylglycerol lipase, which results in the production of 1,2 diacylglycerols (in locusts) or free fatty acids (in fruit beetle) (see **Figure 1.4**) [16,140].

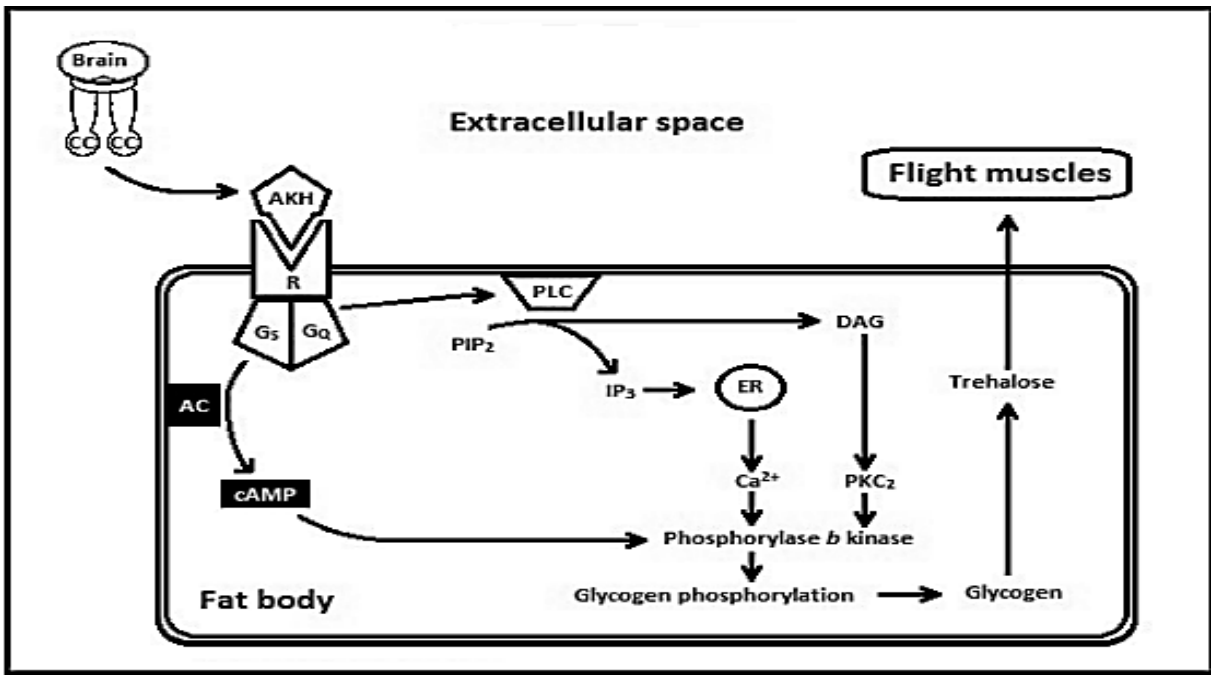


Figure 1. 4: AKH controlled mobilisation of trehalose during high energy demanding flight in the fat body of insects (documented in *L. migratoria*, *P. sinuata*, and *P. americana*). Abbreviations - CC: corpus cardiacum; AKH: adipokinetic hormone; R: AKH receptor; G_s/G_q: G-protein; PLC: phospholipase C; PIP_2: phosphatidylinositol bisphosphate; DAG: diacylglycerol; IP_3: inositol triphosphate; ER: endoplasmic reticulum; AC: adenylate cyclase; cAMP: cyclic AMP; PKC_2: protein kinase C and black-boxed AC and cAMP represent a specific pathway in *L. migratoria*. It was adapted from[16,140].

GPCRs are the largest family of cell surface receptors. They are essential targets for drug pharmacology in humans, as bioactive peptides and proteins signal via GPCRs. If one interferes with this ligand-receptor binding, one could potentially influence or disrupt critical physiological processes in the body. This was effectively achieved with the development of a group of drugs, generically known as β-blockers, and used to treat a variety of cardiac ailments, hypertension, migraine and anxiety [63]. In the same vein, it is thought that the GPCRs of pest invertebrates, including the AKHRs of insects, maybe suitable targets for new insect control agents with a more significant deal of species specificity than the conventional chemical pesticides. One of the steps towards achieving this goal is to know the AKHR binding site, the 3D structure of both ligand and receptor. Unfortunately only very few 3D structures of most GPCR/AKHRs are available. [63].

1.4.4 GPCRs class A activation

Ligand binding stimulates the activation of class A GPCRs from the extracellular side of the receptor through conformational changes apparent in the α-helices [29,90,139,141,142]. Opsins (colour pigments and rhodopsin) are photosensitive GPCRs that differ from other Class A GPCRs/AKHRs as photons of light activate them [143]. The rhodopsin ligand (chromophore, 11-cis-retinal), which is protonated by a Schiff base, is covalently bonded to a Lys residue in Helix7 [89,101,144–147]. Within femtoseconds on exposure to light, the receptor-ligand changes its conformation from 11-cis to all-trans [29,89,100,101,148]. The activation of the rhodopsin GCPR receptor involves neutralisation of the Schiff's base counterion, which in turn leads to a conformal change of the molecule, particularly from the intracellular region that is related to the protonation of two cytoplasmic acidic residues [90]. Experimental studies back the photo-activation leading to the movements of TM3 and TM6, causing a break in the DRY ionic lock [90].

Molecular studies have shown that the binding cavity of some small agonists, viz; adenosine, 3-peptide thyrotropin-releasing hormone, biogenic amines, and nucleotides are between TM3, TM4, TM5, TM6, and TM7[149–152]. Also, site-directed mutagenesis experiments indicated that some residues in ECL2 are essential for the binding of antagonists and agonists [153]. The existence of a disulfide bond between ECL2 and ECL3 is known to covalently link the loop close to the binding site [153], making the ECL2 form a lid over the bonded ligand to prevent it from being released. Also, from mutagenesis of serotonin, it was suggested that ECL2 and ECL1 play an essential role in receptor conformal change [154] or even create a channel for ligand penetration into the binding cavity[90]. The same hypothesis of ligand channelling has also been postulated for rhodopsin [155,156]. The participation of the extracellular domains in the binding of large peptide ligands has been documented in mutational studies of GPCRs, and it has been shown that the residues in the extracellular loop and N-terminus interrelate directly [157–161]. Thyrotropin-releasing hormone (TRH), angiotensin II, neuropeptide Y (NPY), gonadotropin-releasing hormone (GnRH) and chemokines are amongst the receptors reported with such findings [162–166]. Experimental data have shown that residues found in the upper parts of TM2, TM3, TM5, TM6, and TM7 interact directly with particular neuropeptides, which include δ opioid, ETB endothelin and Y1NPY in humans [167–169]. Only recently, the importance of the extracellular domains during ligand/receptor binding in molecular docking calculations relating to the insect neuropeptide binding to the pheromone biosynthesis and activating AKH-receptor was reported [170–172]. In biogenic amine receptors, identified binding cavities are different from the crucial positions assumed to interact with ligands [158].

AKHRs / GPCRs are generally recognised to exist in equilibrium between the active and inactive states. Upon ligand binding, there exists a molecular inter-switch between both states. When bound to an agonist, the activated state is favoured, leading to the formation of the agonist-activated receptor-AKH/ G-protein complex [86,98,173]. Similarly, antagonists bind to the

receptor with a superior affinity for the inactive state when compared to the active state [33,174,175]. However, neutral antagonists bind with an equivalent affinity for both inactive and activated states and do not have any effect on the constitutive activity.

1.4.5 Signal transduction and G-proteins

AKHR/GPCRs act on heterotrimeric G-proteins in the form of guanosine diphosphate (GDP) and guanosine triphosphate (GTP). The heterotrimeric subunits are denoted the α subunit, the β subunit, and the γ subunit [90,158]. The α- and β subunits are attached to the plasma membrane. For a receptor to be activated, the ligand first binds to the identified cavity with the best pose, leading to a change in conformation of the GCPR. The conformal changes allow the GPCR to act as a guanine nucleotide exchange factor (GEF), and the G-α subunit then binds to either GDP (to become inactive) or GTP (to become active). When becoming inactive, the two G protein subunits disconnect into two parts: the GTP-bound G-α subunit and a β-γ subunit dimer [90,158,176–178]. The separate units at this stage are no longer active but remain fixed to the plasma membrane. This will enable them to diffuse and interact with other G-proteins [179–181]. To reactivate the system again, GTP is first hydrolysed back to GDP, which causes a revival of α -GDP and βγ-subunits once again, causing the receptor to assume the active heterotrimer state [90].

1.4.6 GPCRs structure

GPCRs proteins with seven-transmembrane (7TM) **α**-helices in an anticlockwise arrangement, when observed from the N-terminus (extracellular region) (see **Figure 1.5).** [182–187]. Three loops connect the helices on the extracellular region, making up the N-terminus and another three loops connect the helices on the intracellular region, making up the C-terminus[182]. From one GPCR to another, the loops differ in length. GCPRs contain conserved residues that influence

the discrimination of ligand from the extracellular side, whereas the conserved residues in the cytoplasmic region affect G-protein coupling selectivity[187–189].

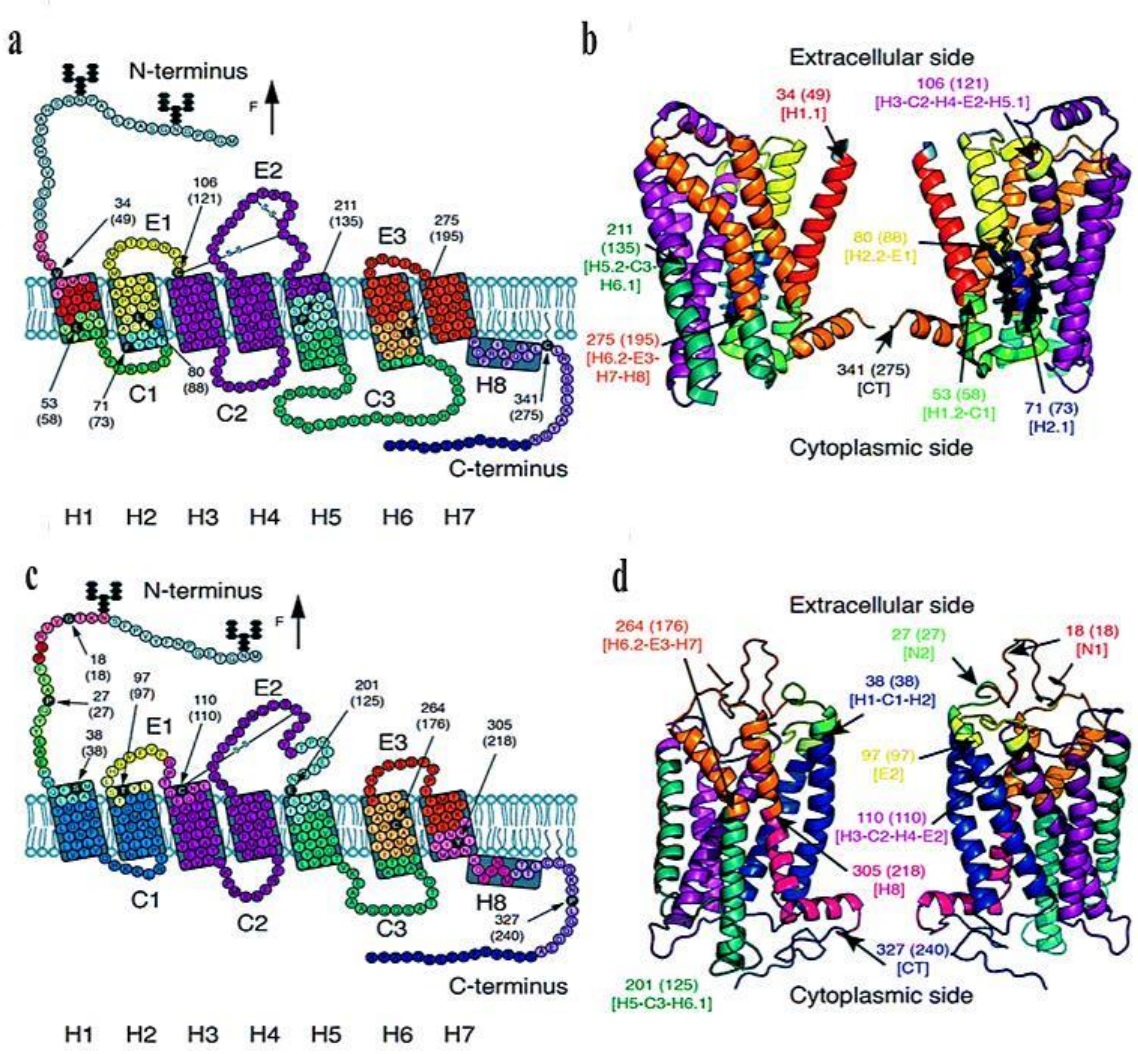


Figure 1. 5: (a) Schematic diagram showing amino acid sequence and (b) 3D structure of human β2 adrenergic receptor (PDB ID 3D4S). (c) Schematic diagram showing amino acid sequence and (d) 3D structure of bovine Rhodopsin (PDB ID 1U19). The black amino acid residues signify the end of the stable structural segment and the beginning of the next stable structural layer in the sequence. The seven α-helices of both GCPR representations are labelled H1-H7. The extracellular loops are denoted E1, E2, E3, while the cytoplasmic (intracellular loops) are C1, C2, C3. A small tail-like helix usually found in most GCPRs at the C terminal is denoted H8 adapted from Gruber 2010 [123].

The putative binding site for glycoprotein hormones and the largest polypeptides are located at the N-terminus [190–192]. The expected ligand binding sites of GPCRs are the N-terminus or the

extracellular loops, predominantly ECL2 [106,123,190,193–196]. Recently, over 109 GPCRs have been used as targets for developing new drug [92,197]. In all the 109 targeted GPCRs, the extracellular part of the receptor has been identified as the most preferred ligand-binding site. To enable sufficient elucidation of the signal transduction pathway, efficient selectivity of ligands by GPCRs is of supreme importance for both side effects and correction of drug-drug interaction [198,199].

Binding affinity and molecular docking studies have proposed that related sequence motifs could identify ligands with similar chemical features and same biological space [186,187,200,201]. However, in the binding site and the residues, physicochemical properties can be compared. Thus, in-class A GPCRs, the presence of conserved residues in the transmembrane regions, is vital in receptor activations. Also, ECL2 is essential to most GPCRs as it provides stability to the system, [120,195,202,203]. The stability of the system is possible because of the presence of cysteine residue in the ECL2 and conserved CWxP motif in transmembrane six (TM6) [193,204,205]. The cysteine forms an ionic lock [202]. The presence of this ionic lock helps to preserve the position of helices three and six together to maintain the protein structure [129,195].

In the subfamily of opioid and chemokine receptors, the second extracellular loop forms a β-hairpin conformation[111]. It is an excellent example because they make a common peptidic-binding motif found in GPCRs. Sequence identity across the five families of GPCRs is low (below 25%), even though they both share the same 7 TM topology in humans, and there exists variation in the N terminal domain [111,206]. Though, among subfamilies, the structural sequence resemblance is better between subtypes[186,207]. Hence, within subtypes, templates that permit predictions can be used by comparative modelling, which is precise enough for some applications.

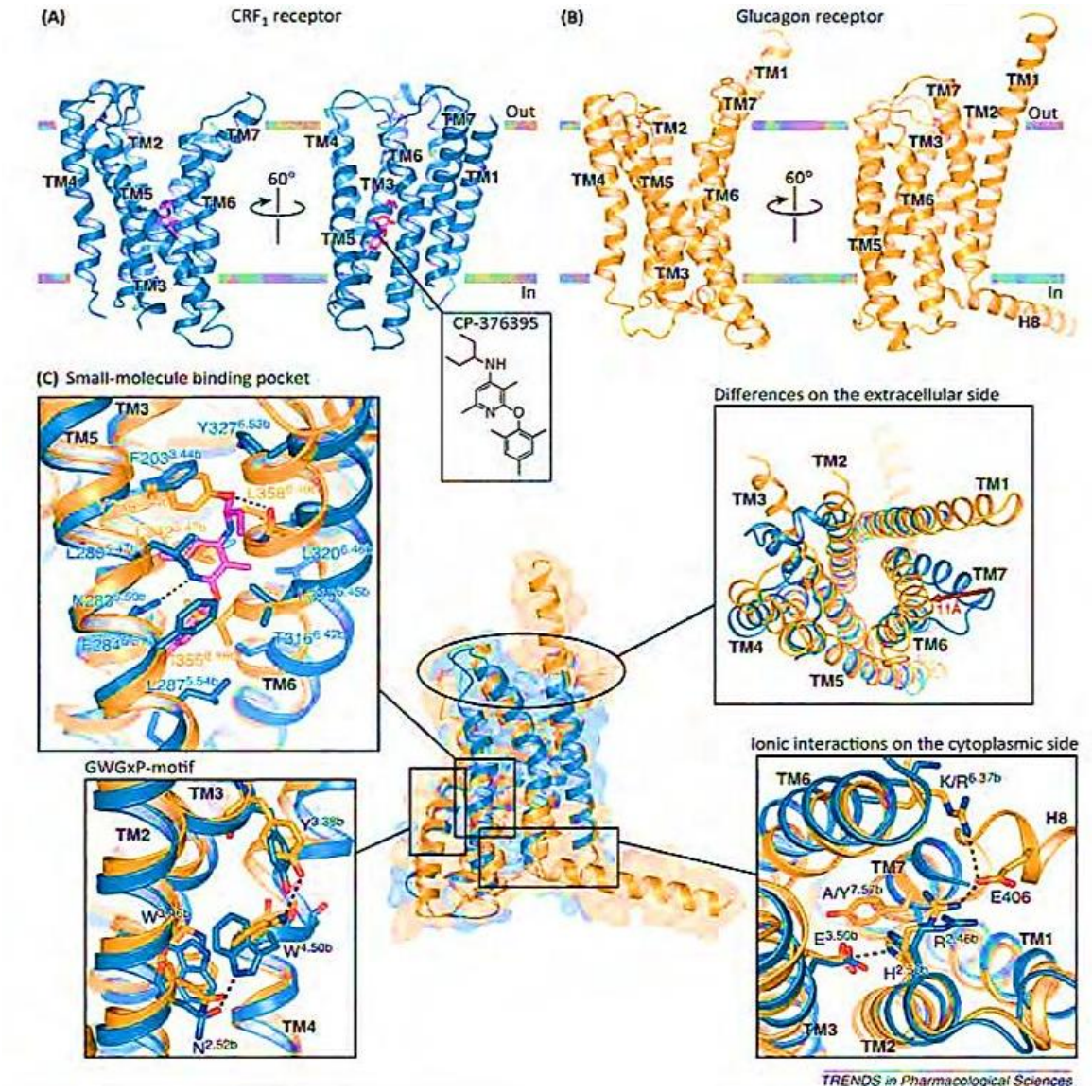


Figure 1. 6: (a) Crystal structure of corticotropin-releasing factor receptor 1 (CRF1) (PDB ID 4K5Y) within a membrane. (b) Crystal structure of glucagon receptor (PDB ID 4L6R) within a membrane. (C) An overlay of the glucagon and CRF1 receptors [114]. In purple sticks, the disulfide bond between TM3 and ECL2 is represented. The red arrow with a distance of 11Å between the Ca-atoms of glucagon receptor and that of the CRF1 receptor emphasise the structural difference between them. The alignment of both receptors indicated the binding site and showed the antagonist binding configuration around CP-376395, and a structural re-conformation around TM6 for both receptors adapted from Hollenstein 2014 [124].

1.4.7 GPCRs Three Dimensional (3D) Structures

To get a proper understanding regarding the molecular functions of protein, and to enable the practical design of experiments especially molecular understanding of disease-related mutation, site-directed mutagenesis and structure-based design of specific inhibitors three dimensional (3D) structures of receptors are required [208,209].

3D structures of GPCRs are limited despite their significance in physiological processes. The reason for this is most likely that the determination of 3D structures is very complicated. Also, experimental methods like X-ray crystallography and NMR cannot easily be applied for 3D structure determination, due to the flexibility of membrane-bound proteins outside a membrane and the difficulty in crystallising a membrane-bound molecule [29,128,141,190,210–212]. The low solubility and relatively large molecular size of the receptors make it impractical to apply experimental techniques like NMR[29,128,141,190,211–213]. As a result, to date, the mammalian rhodopsin and the human Beta2-adrenergic receptor (β2AR)[96,98,141] are the only resolved crystal structures of a GPCR [89,201,205]. Even with these crystal structures, the highly mobile ICL3 has been removed. Due to the above-listed complexity of the experimental techniques and to reduce time and cost, the pharmaceutical industry has resorted to computational methods to obtain their molecular models of GPCRs.

1.4.8 Crystal structures of Beta2- adrenergic receptors (β2AR) and Rhodopsin

β2AR and rhodopsin possess comparable overall molecular structures [154] [29,86,128,148,154]. They both belong to Class A GPCRs [130,214,215], and rhodopsin belongs to a subfamily known as opsin, [29,88,96,98,130] while β2AR belongs to the amine family [29,46,96,113,216]. Both rhodopsin and β2AR experience similar conformational changes upon activation [46].

1.4.9 Rhodopsin

Rhodopsin is located in the retina of the eye and is stabilised structurally by a covalently bounded 11–cis retinal ligand [29,46,113,128,141,217–221]. The first crystal structure [89], available with PDB ID 1F88A (see **Figure 1.7** a &c) had a resolution of 2.8Å. The resolved molecule had 348 residues and a total of 194 TM residues.

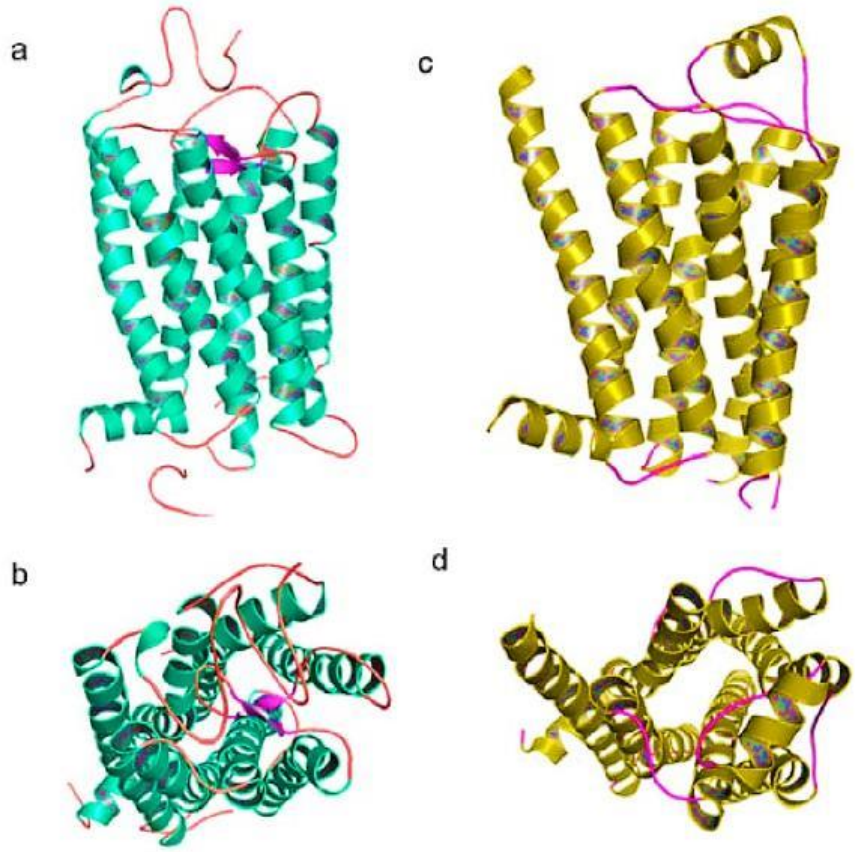


Figure 1. 7: (a) Side view of the rhodopsin (PDB, 1F88A) crystal structure showing the α-helix (ribbon in green), and loops (purple). b) A view of the rhodopsin crystal **structure from the extracellular region, demonstrating the closed conformation of the molecule. c) Side view of the β-2 adrenergic receptor (PDB, 2rh1A) crystal structure showing the α-helix (ribbon in gold), and the loops (pink). d) A view of the β-2 adrenergic receptor from the extracellular region, demonstrating the open conformation of the helix bundle [156].**

The rhodopsin crystal structure, like any GPCR class A structure involves 7 TM helices with a tail like small helix parallel to the membrane inside the cell, it possesses 3 ICL and 3 ECL regions with a β sheet in ECL2. The molecule has an N-terminus which forms a cap-like

structure over the molecule [29,98,222,223]. The protein has the disulfide bond found in most GPCRs, between Cys110 of TM 3 and Cys187 in the second extracellular loop.

Until 2007, the rhodopsin molecule (PDB, 2rh1A) was the only known GPCR crystal structure, and so was used as a template for building 3D molecules of most GPCR structures [156]. Now, another crystal structure is available, which is (PDB code, 2X72) [224].

1.4.10 Beta2-adrenergic receptor

The β2AR crystal structure was published in 2007. The structure was achieved by replacing part of ECL3 with T4 lysozyme to reduce its flexibility. The resolution was 2.4Å [96,98,113] (**Figure 1.7 c & d**), and is available in the PDB database (ID 2rh1A). The β2AR molecule has a broader range of signalling behaviour and can react to a spectrum of several diffusible ligands or peptides, coupled to more than one G-protein. Like the rhodopsin molecule, the binding site is the same [29,46,223]. When compared to rhodopsin, β2AR has a more open structure when viewed from the extracellular region, and two disulfide bonds constrain its ECL2 to enable free access to the ligand-binding cavity [29,46].

Modelling of large molecules, such as AKHR is dependant on available template structures. Template 3D structures constructed using computational techniques may slightly differ with 3D structures constructed using an experimental technique (by NMR or x-ray crystallography). Nevertheless, the crystal structures of β2AR and rhodopsin provide valuable information regarding the structure and activity of Class A GPCRs, and has, consequently, simplified structure-based drug design.

1.5.0 Research Motivation and Statement of the Problem

Insect control via commercially available non-specific toxic chemicals has resulted, over the years, in some of the harmful insects gaining resistance to the widely used insecticides (Wang et al., 2018; Gould et al., 2018). Furthermore, a potential risk to food production may arise in

the near future due to indiscriminate killing of pollinator insects (collateral damage) as a result of continued use of conventional broad-spectrum toxic chemicals[225,226]. Some of these chemicals are carcinogenic, and somehow, these chemicals find their way into the stomach leading to severe health problems [227,228]. This has spurred on the search for "green insecticides" (a species-specific pesticide that is selective in its action. Negatively affecting the pest insect and causing no harm to the beneficial insect, other animals and the environment) that can serve as a viable and responsible alternative to current practices. One idea for developing a green insecticide is to target the GPCRs that are involved in endocrine signalling in insects. Because the regulation of metabolism is of such fundamental importance in insects, the adipokinetic hormone signalling is investigated in this thesis. For the AKH system to be seriously evaluated as a potential application in green pesticides, research into structure-based ligand-AKHR interactions is needed for pest and beneficial insects to find those that are specific to pest insects for further drug development. Essential in this regard, is the availability of 3D AKHR models. Even though the relationship between physiological process and 3D structures are vital, very limited 3D structures of insects AKHRs are available. The first models of an insect AKHR were published in 2011 for the receptor in the malaria mosquito *A. gambiae* [229]; aside from this, there are no 3D models available for AKHRs of other dipterans.

Literature has shown that about 50% of pharmaceutical industries target GPCRs, and the molecules are elucidated using homology models [128,170,172,214,221,230–238]. From an accurate AKHR model, the binding cavity can be determined. With the aid of the insect AKH neuropeptide, structure-based design of non-peptide ligands (mimetics) that can act with high potency at the receptor as antagonists or agonists could be developed [239]. Doing so could help in the production of insect-specific insecticides that could reduce the vector capacity of these insects by decreasing the amount of energy released during flight. Moreover, it could lead to the discovery of neuropeptide analogues that are selective in their activity and negatively

influencing the targeted pest species without harm to beneficial insect species. Thus, enhancing the eradication of some of the damages caused by some of these insects. For this reason, this study focuses on the determination of the 3D structure and locating and comparing how one AKH, which is common to three fly species, binds to its cognate AKHR in each of the fly species. The AKH receptors of the three fly species are similar but not identical (see Chapter 3, 4 and 5).

1.6.0 Aim and Objectives

1.6.1 Aim

To explain via modelling how a single adipokinetic hormone (Phote-HrTH) can activate receptors from three different insects, namely, the flesh fly *Sarcophaga crassipalpis*, the vinegar fly (also called fruit fly) *Drosophila melanogaster* and the oriental fruit fly *Bactrocera dorsalis*.

1.6.2 Specific objectives and expected significance of the research

- Determine the solution structure of Phote-HrTH using NMR constrained molecular dynamics.
    - Record NMR spectra of Phote-HrTH in DPC micelle solution.
    - Use the measured nOe's to determine average internuclear distances.
    - Conduct NMR restrained molecular dynamics in water and DPC micelle solution.
- Use each primary sequence of the adipokinetic hormone receptor for all three insects to build three-dimensional structures of each, using the crystal structures of rhodopsin and β2AR as templates.
- Refine the structures using molecular dynamics in a membrane.

- Identify the binding pockets of the adipokinetic hormone receptors using site finder program and confirming the result with site map program.
- Validate the *Drosophila melanogaster* β2AR model using experimental ligand mutation data.
- Perform molecular dynamics of the receptor/ligand complex in a membrane.
- Compare and contrast the results for the different insect species at a molecular level.

1.7.0 Questions to be addressed in this work

1. Are the three constructed 3D structures of the AKHR, based on two different homology model targets, i.e. beta2-adrenergic receptor and rhodopsin molecules, the same? If not, is there any correlation between the structures?
2. What structural and conformational changes occur during MD simulations of the constructed models when in a membrane?
3. How do these results compare to experimental data of other GPCRs?
4. How does the binding cavity location in these constructed AKHRs compare to those in Beta2-adrenergic receptor and rhodopsin? How does each adipokinetic hormone bind to its cognate receptor differ from one another? How does this position agree with experimental results from other peptide GPCRs?
5. What are the binding energies of the AKH to each receptor?
6. What structural and conformational changes occur during the MD simulation of the beta2-adrenergic-based model?
7. What is the correlation between the experimental data and molecular data on structure-activity?

2

Construction of the Three-dimensional Structure of Phote-HrTH via NMR Restrained Molecular Dynamics

Summary

In this chapter, we present the 3D structure of Phote-HrTH (pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2). The structure was elucidated using NMR restrained molecular dynamics simulations in both water and micelle solution.

Elucidation of the Phote-HrTH peptide structure was conducted with the aid of Ccp-NMR. Both one and two-dimensional NMR spectra were used to resolve the secondary structure and orientation in solution and DPC micelle solution [107,240].

The three-dimensional structure of Phote-HrTH has a β-type I turn H-bonding stabilises that. The β-turn is due to Pro6 and is stabilised by a moderately strong, transient H-bond between Trp8(NH) and Thr3(CO) as well as many other H-bonds between the side chains and backbone.

2.1.1 Introduction

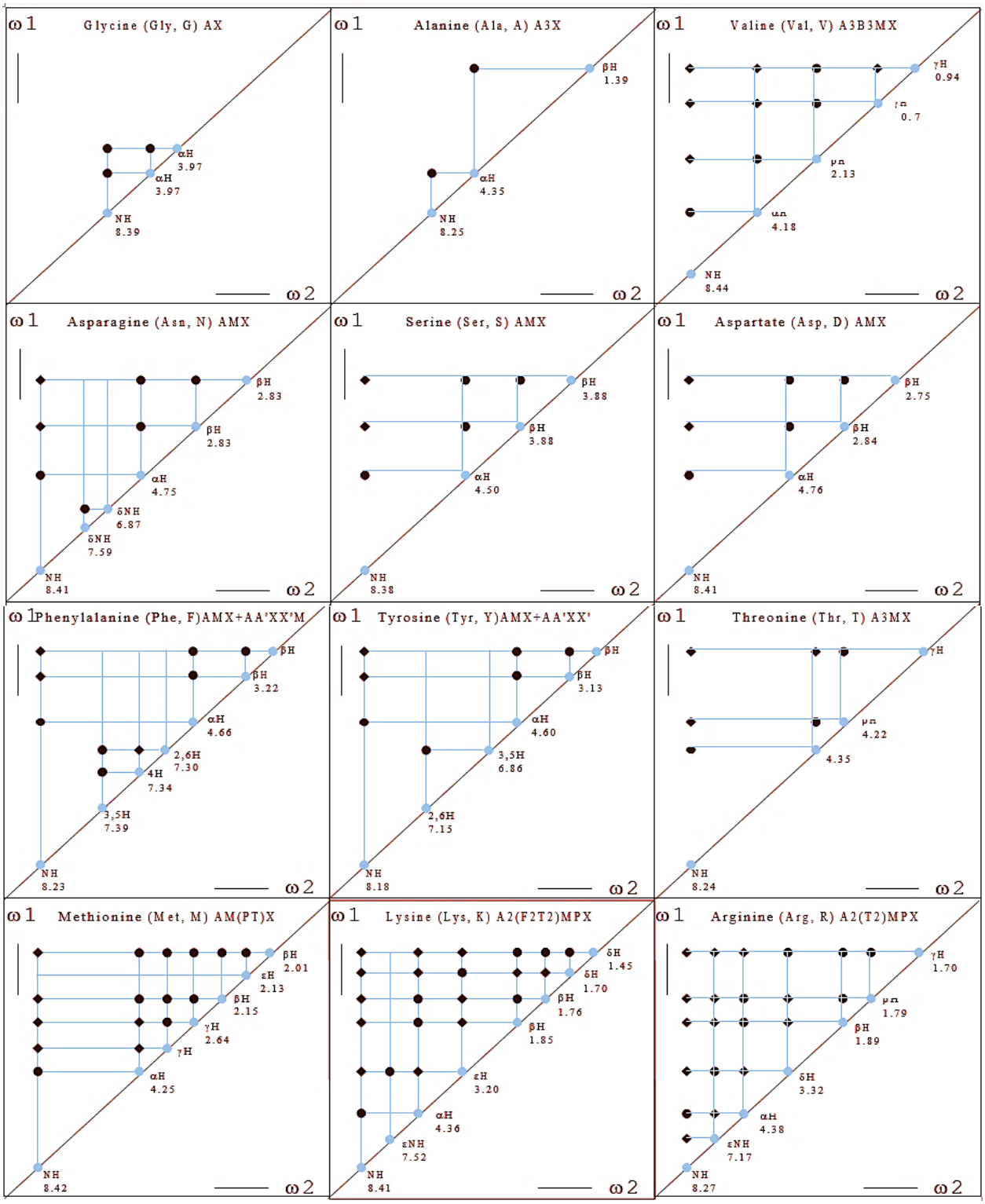
2.1.2 Overview of some NMR structure elucidation programs.

Experimental determination of the solution structure of Phote-HrTH was accomplished using NMR. NMR is a valuable tool for the elucidation of solution structures of biological molecules [241–250]. The solution structure elucidation was accomplished using the collaborative computing project for NMR (CcpNmr Analysis), a graphics-based data analysis, NMR resonance assignment and spectrum visualisation program [240]. The CcpNmr Analysis program helps in the assignment of NMR spectra and the generation of structural constraints. Distance constraints are derived by integration of the NOESY cross-peaks and relating these to internuclear distances. DANGLE[251] is used to calculate torsion angles from chemical shifts. These restraints can then be fed into ARIA (ambiguous restraints for iterative assignment) and CYANA (combined assignment and dynamics algorithm for NMR applications) to calculate possible structures of the molecule. The success of such a strategy for structure calculation depends on a complete list of chemical shift assignments [252]. An experienced spectroscopist can calculate structures manually for small molecules, but it is cumbersome and time-consuming. ARIA or CYANA function through several refining cycles of conformational computations to reach a structural model with the least deviation from the experimental data [240,253]. The calculated structure with minimum energy has maximum compatibility with the experimental data. This iterative process utilises preliminary structures from earlier cycles to reduce the ambiguities of cross-peak assignments. Additional selection criteria for cross-peaks and NOE- assignments are added after each cycle, and it yields better structures in the process. Cycle one is thus the most crucial one since it is the first for which a previous structure is not available [240,253]. The key conformational data-input are upper distance limits resulting from NOEs [240,254].

2.1.3 Brief information relating to chemical shifts

All nuclei resonate at a particular frequency, but the precise value of this frequency can vary for even the same type of nuclei, it all depends on the local environment. This phenomenon is called the chemical shift. Chemical shifts arise from small changes in electron density around a specific nucleus, and the electronic environment influences the difference in shifts seen for different nuclei. That is to say; the electrons induce a limited additional electronic field either opposing to or adding to the external magnetic field. This causes shielded and de-shielded nuclei, respectively [255]. Hence, a nucleus resonates at slightly different frequencies, depending on the surrounding electron densities. Chemical shifts are reported relative to a reference standard to establish that the values are independent of the applied field. Thus, a standard table of values could be used by all spectroscopists [256].

Standard chemical shift values for nuclei bonded to different neighbours are tabulated and are used to identify the signal arising from different functional groups in a spectrum. For example, hydrogen atoms of an amino acid resonate at a specific frequency depending on which atom it is bonded to. As such, there is a specific pattern of signals from each amino acid, and this is used to identify each residue type in a given protein spectrum. The signals expected for the 20 standard amino acids in correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) ^{1}H-^{1}H 2D correlation spectra are shown in **Figure 2.1**. [257]. The chemical shift values correspond to amino acids in random coil structures.


ω1
Glycine (Gly, G) AX
αH 3.97
αH 3.97
NH 8.39
ω2
ω1
Alanine (Ala, A) A3X
βH 1.39
αH 4.35
NH 8.25
ω2
ω1
Valine (Val, V) A3B3MX
γH 0.94
γH 0.7
βH 2.13
αH 4.18
NH 8.44
ω2
ω1
Asparagine (Asn, N) AMX
βH 2.83
βH 2.83
αH 4.75
δNH 6.87
δNH 7.59
NH 8.41
ω2
ω1
Serine (Ser, S) AMX
βH 3.88
βH 3.88
αH 4.50
NH 8.38
ω2
ω1
Aspartate (Asp, D) AMX
βH 2.75
βH 2.84
αH 4.76
NH 8.41
ω2
ω1
Phenylalanine (Phe, F)AMX+AA'XX'M
βH
βH 3.22
αH 4.66
2,6H 7.30
4H 7.34
3,5H 7.39
NH 8.23
ω2
ω1
Tyrosine (Tyr, Y)AMX+AA'XX'
βH
βH 3.13
αH 4.60
3,5H 6.86
2,6H 7.15
NH 8.18
ω2
ω1
Threonine (Thr, T) A3MX
γH
γH 4.22
4.35
NH 8.24
ω2
ω1
Methionine (Met, M) AM(PT)X
βH 2.01
εH 2.13
βH 2.15
γH 2.64
γH
αH 4.25
NH 8.42
ω2
ω1
Lysine (Lys, K) A2(F2T2)MPX
δH 1.45
δH 1.70
βH 1.76
βH 1.85
εH 3.20
αH 4.36
εNH 7.52
NH 8.41
ω2
ω1
Arginine (Arg, R) A2(T2)MPX
γH 1.70
γH 1.79
βH 1.89
δH 3.32
αH 4.38
εNH 7.17
NH 8.27
ω2

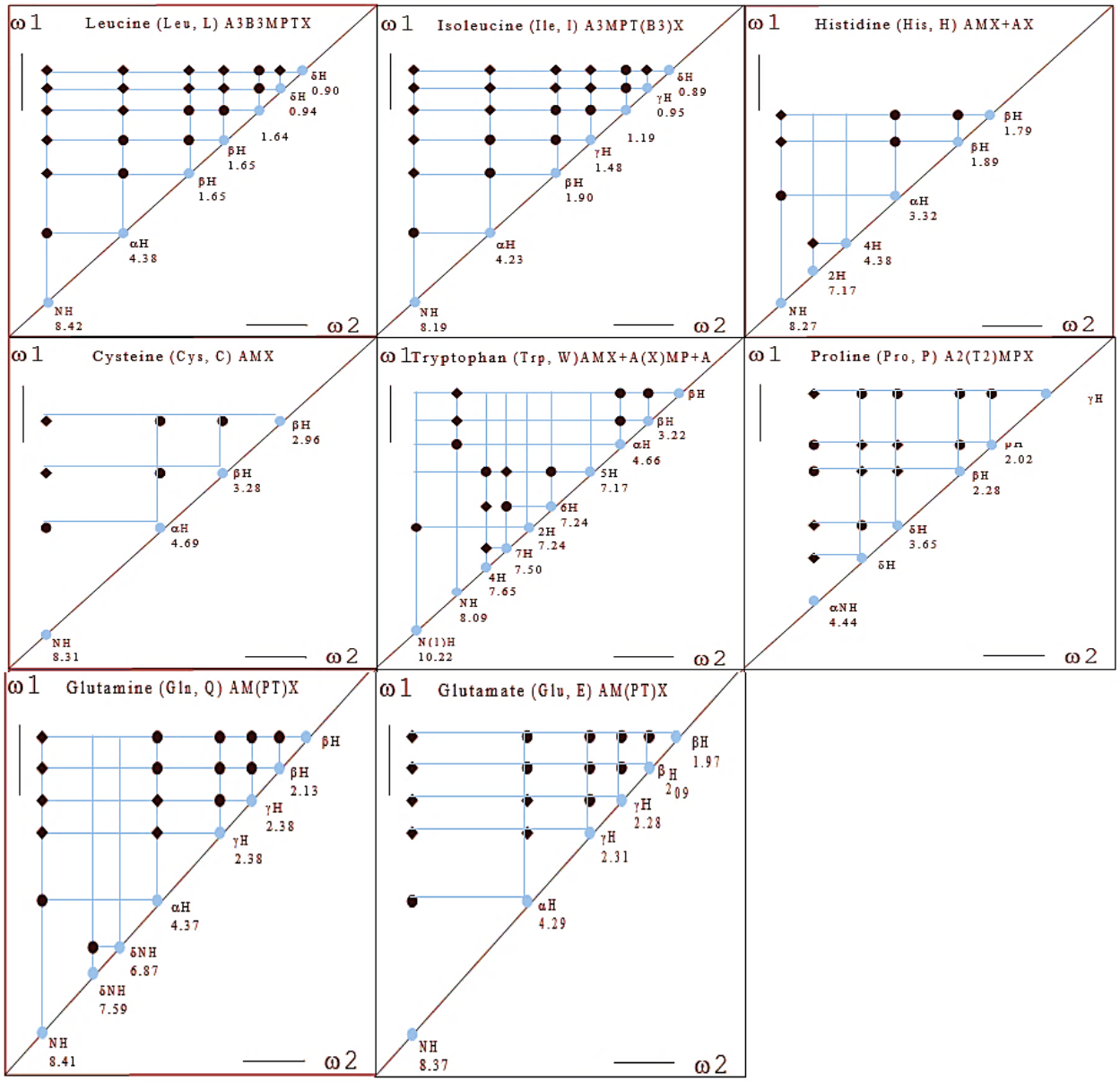


Figure 2. 1. Schematic diagram showing the spin systems of the 20 standard amino acid residues in a COSY/TOCSY ^{1}H-2D NMR experiment. Only peaks on the left-hand side of the diagonal are shown. In an actual spectrum, identical peaks reflected about the diagonal can be visible. [258] A characteristic pattern of signals results from each amino acid from which the amino acid can be recognised. All peaks may be shown in a TOCSY spectrum. Filled grey circles show diagonal peaks; filled black circles illustrate cross-peaks which are observable in a COSY spectrum, filled black squares show new cross-peaks observable in a TOCSY spectrum [257]. Random coil ^{1}H chemical shift standards are included [257].

The spin systems have complicated names, letters and numbers. The letters refer to the different

coupled spins. The numbers signify the number of magnetically equivalent nuclei. For a better explanation, the reader is referred to the book "NMR of proteins and nucleic acids" by Wüthrich chapter five [259].

In this thesis, only the specific patterns of the spin systems and the labelling of the atoms are of importance. **Figure 2.2** exemplifies how the atoms in an amino acid are numbered. The first carbon in the side chain (attached to the carbonyl carbon) is called the α-carbon. The attached proton(s) is(are) the α-proton(s). The successive side-chain atoms are named by following the Greek alphabet. Thus, a proton is labelled α, β, γ, δ or ε depending on which carbon in the side chain it is attached. The proton attached to the backbone amide is labelled HN the protons on the aromatic side chains are labelled according to the standard numbering

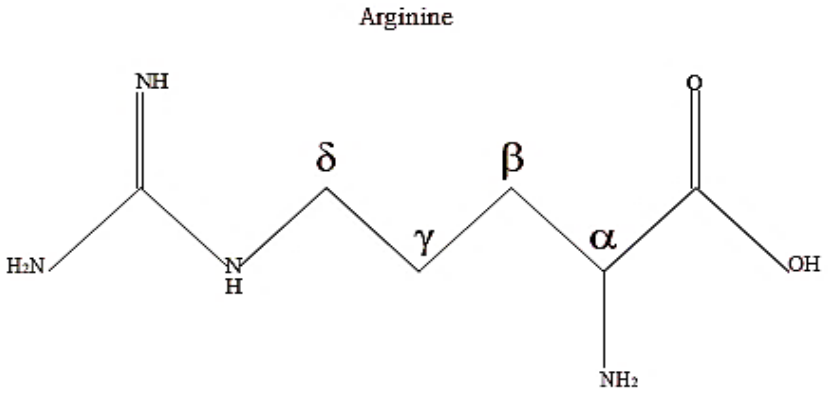


Figure 2. 2. Schematic diagram showing the amino acid arginine with side-chain carbons labelled as is standard amino acid numbering system, taken from [260].

Chemical shifts are among the most detailed data collected from an NMR-spectrum of a peptide. The values are sensitive to the local environment, which is dependent on the secondary structure of the peptide. [261–263].^{15}N, Hα- and amid-protons experience a net up-field shift when in an α-helical conformation and a net downfield shift when in a β-sheet conformation.^{13}C-shifts display the reverse trend, with downfield shifts for α-helical conformations and up-field shifts for β-sheet

conformations. From the above trends, it is possible to measure a specific part of a structure's flexibility. If chemical shift values are close to random coil values, then the peptide is likely to be flexible, whereas, chemical shifts far from random coil value point at a somewhat rigid structure [264].

2.1.4 Chemical shift index (CSI)

The chemical shift index is a widely used graphical method in protein NMR. It is used to identify and display the type of secondary structures (beta strands, random coil and helices) in polypeptides utilising only the backbone chemical shift information. CSI was first designed for analysing $^{1}H\alpha$ chemical shifts, and later the ^{13}C backbone shifts. CSI makes use of the already known principle that is $^{1}H\alpha$, the chemical shift of amino acid residues in helices tend to move towards the right side of an NMR spectrum (shifted upfield) relative to their random coil values and moves left side of an NMR spectrum (downfield) in β-strands. A similar movement of the $^{1}H\alpha$ is also found in the ^{13}C chemical shift [262,265,266].

2.1.5 Backbone angles (torsion angle)

Backbone angles are angles found between two intersecting planes. By comparing chemical shifts which were derived from a new compound with those of an already known structure, phi ($\varphi-$) and psi ($\psi-$) angles can be predicted [81,267]. These angles can then be used as an additional parameter, alongside NOEs and J-couplings when calculating the structure, see **Figure 2.3.**

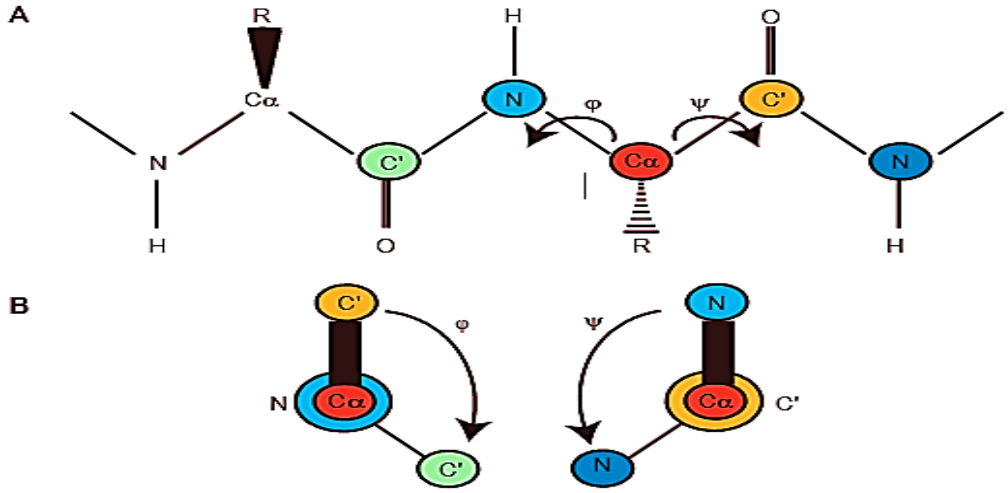


Figure 2. 3. Amino acid torsion angles. φ Is the torsion angle around the N-Cα, while ψ is the torsion angle around Cα –C' bond. (A) Plane of a backbone peptide with the side chains oriented above of the plane (filled wedge) or below the plane (striped wedge). The arrows designate the rotational bonds. (B) Projection of the peptide in part (A), observed along with the bond of rotation from the atom drawn as a small circle to the atom drawn as a big circle. The bold black line represents the bond between the atom drawn as a small circle and its neighbouring atom. The thin black line represents the bond between the atom drawn as a big circle and its neighbouring atom. The atoms are coloured as in (A).

2.1.6 Scalar (spin-spin) couplings

Interactions between nuclear spins that occur during electron bonding are known as scalar or spin-spin interactions. The interaction is measured by the scalar coupling constant $^{n}J_{ab}$, where n represents the number of covalent bonds separating the two nuclei a and b [268]. With scalar coupling (J-coupling), magnetisation is conveyed through bonds: All the spins predictably have an effect on each other, and this can aid in the determination and labelling of functional groups [255]. In a J-coupled correlation spectrum, all protons within an amino acid have the tendency to couple with each other. Correlation with protons on adjacent amino acids is so small that it is not visible. The basis for this is described under the TOCSY section. The essential couplings from a spectroscopist's point of view are n=1-4 [269]. Couplings between nuclei cause splitting

of the NMR-signal. Weakly and strongly coupled two-spin systems are referred to as AX- and AB-spin systems, respectively[269]. J-values for n=1-3 are reliant not only on internuclear distance but also molecular structure. As such, one can obtain possible parameters for structure calculations from scalar couplings[270]. Particularly ^{3}J-couplings (vicinal couplings) are useful in this respect. The value is dependent on substituents, the distance between the two carbon atoms involved, H-C-C-H bond angle, along with torsional and dihedral angles, with the latter angle being of most interest [270]. ^{3}J-couplings vary in magnitude in a bond-angle dependent manner. The relationship between the coupling constant J and the dihedral angle φ (see Figure 2.3.) is given by the Karplus equation [256]. Thus, J-couplings can be used to measure dihedral angles and thus conformational arrangements around bonds.

2.1.7 Linewidth and signal: noise ratio

As the molecular mass increases, so does the rotational correlation time. This has a profound effect on the linewidth of the peaks; the longer the correlation time, the broader the lines [269]. Thus, the peaks become less well defined, and they overlap. The decreased sensitivity can be understood by considering the small ^{3}J scalar couplings that coherence transfer relies on upon as the linewidths broaden, they become similar or even more significant than the actual J-values we try to measure [269]. Linewidth thus affects the resolution of a spectrum, but it also affects the signal-to-noise ratio [269]. Signal-to-noise ratio, or sensitivity, is proportional to the square root of the number of scans in the experiment [271,272].

2.1.8 Two-dimensional (2D) NMR- approaches TOCSY, NOESY, COSY and HSQC.

Two-dimensional NMR was first presented during the summer school lecture by Jean Jenner

at Ampére in former Yugoslavia in 1971 [273], and the principles he summarised still govern current techniques. 2D-experiment can be summarised into four stages. They include; sample preparation, evolution, mixing and acquisition (see **Figure 2.4**) [269,274–276]

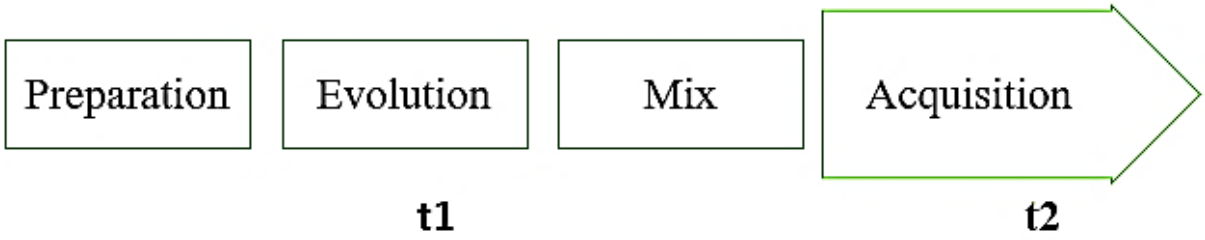


Figure 2. 4. The framework of a 2D experiment, each box represents a stage in the 2D NMR pulse sequence. The two-time domains t1 and t2 are acquired during the evolution period and the acquisition period, respectively taken from [260].

The initial stage, which is the preparation phase, involves one or more pulse, perturbing the system from the original equilibrium state to a preparatory state. The evolution period (2nd stage) involves a sequential increase in t_1 to achieve the F_1 dimension in a broad spectrum. In the third stage, the mixing coherence is moved from one spin to another. This stage of the experiment describes the kind of correlation to be observed between the two dimensions (the information present in the spectrum) hence the vital information of a 2D spectrum. The last stage involves the acquisition and recording of the FID, which give rise to the F2 (t2) dimension [277]. At each stage, the evolutions of spins are recorded, and a possible image of events is gained. This means that one can trail the movement of the magnetisation from one nucleus to another either through space, bond or even both.

In general, the correlation grows with mixing time, the longer the mixing time the more correlation. On the other hand, because of nuclear relaxation, the signal decays with time (T_1

relaxation time). Thus the two effects counter each other, and a compromise time is used. [241].

2.1.9 Correlation spectroscopy (COSY)

One of the first 2D techniques was COSY technique [274]. Cross peaks result from protons divided by three or two bonds [269]. A limitation of this technique is that the diagonal and cross-peaks in the spectrum are out of phase by 90° in both dimensions and as such, require substantial weighting to improve cross-peak line shapes [245]. Secondly, correlations between spins more than three bonds apart cannot quickly be assigned. Also, the elements of a multiplet are in antiphase; this means that if the coupling is less than the peak linewidth, it performs self-cancellation to reduced peak intensities [245,278].

2.1.10 Total correlation spectroscopy (TOCSY)

TOCSY is also known as homonuclear Hartmann-Hahn (HOHAHA). TOCSY utilises magnetism, which can be transported via some couplings throughout the mixing time [245]. When the Hartmann-Hahn prerequisite is satisfied, pulses are then applied [271,279,280]. Under these conditions, there is an efficient transferred of magnetisation amongst coupled nuclei. Only intra-residual couplings are observable in TOCSY. [269].

In a TOCSY experiment, it is possible to obtain data regarding spins that are coupled to one another as a spin system. With the effective use of long mixing time, it is possible to transfer magnetism via a spin system and the whole set of cross-peaks for spin-coupled systems can be seen [271]. Regrettably, amino acids having short chains loose intensity during long mixing times. Thus, when recording spectra, it is best to use and record different mixing times and to compare the results [279]. TOCSY is commonly used for detecting each amino acid residue in a protein concerning the type of residue, but not to a specific residue. For example, cysteine, histidine, phenylalanine, tryptophan and tyrosine residues all give rise to the same array of connected

cross-peaks (**see Figure 2.5**). It is possible from a TOCSY spectrum to only identify a spin system as belonging to this group, but not which amino acid in particular. Alanine, however, has a different signal and a specific assignment can be made from TOCSY. Even so, if there is more than one alanine residue in the sequence, one can only state that a signal comes from an alanine but not which one[245].

2.1.11 Recognition of spin-systems (TOCSY)

The initial step in the assignment procedure is to decide which peaks belong to the same spin-system; that is to the same amino acid. A TOCSY-spectrum provides this kind of information. As mentioned in the description of TOCSY-spectra above, only the type of spin-system can be determined in this step [260]. **Figure 2.5.** demonstrates how the peaks in a spin-system are connected in different regions of a TOCSY-spectrum.

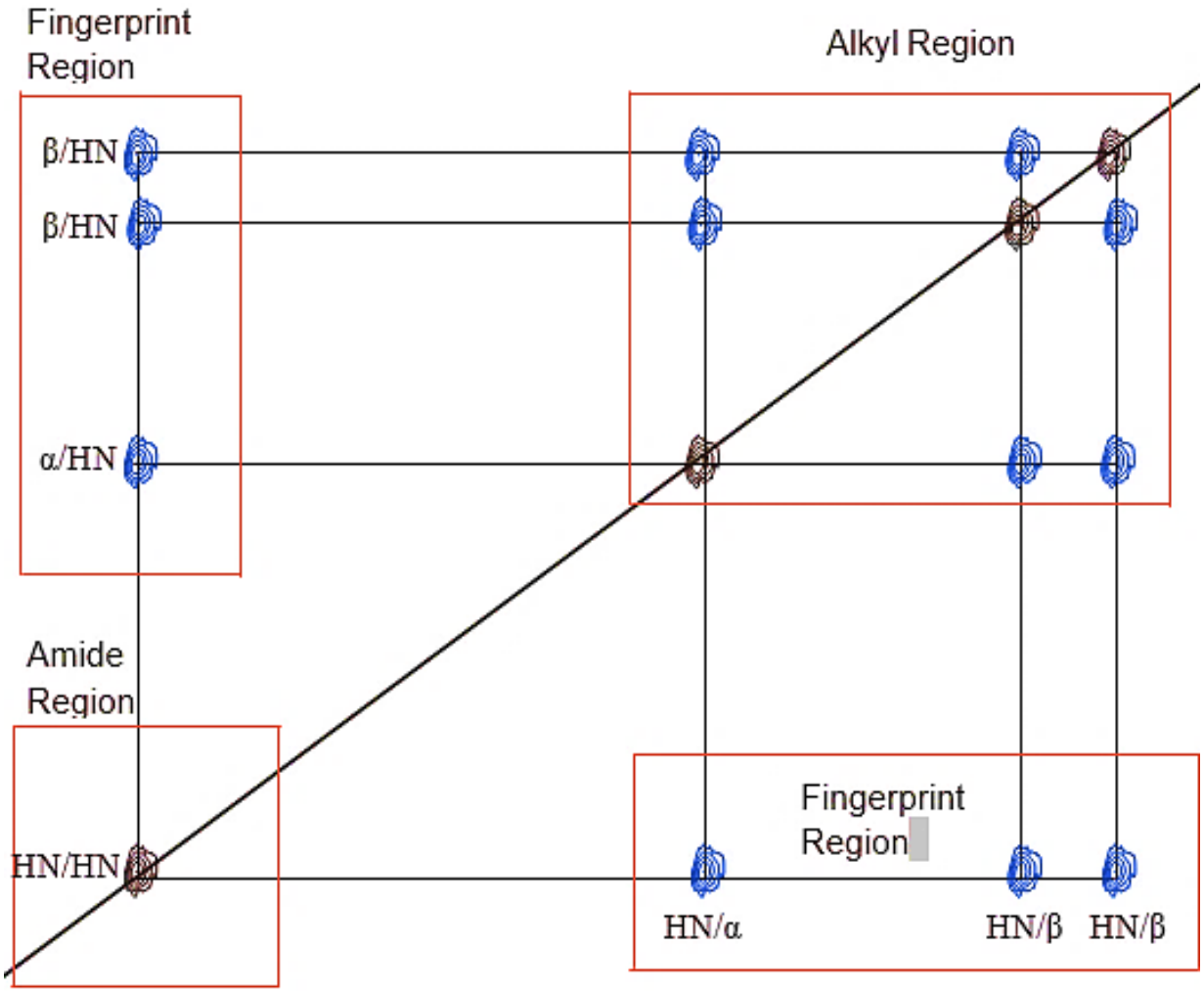


Figure 2. 5. Example of TOCSY correlation diagram. Chemical shift scale in ppm taken from Haugen et al.,[260].

Figure 2.5 exemplified spin system identification in a TOCSY spectrum. HN refers to the proton connected to the backbone amide nitrogen, α to the proton connected to the Cα in the side chain and β to the Cβ. Similarly, protons connected to carbons in longer side chains are labelled according to the carbon it is associated with. Thus, the peaks are labelled according to the protons that contribute to a particular peak in the same order as the magnetisation travels(Haugen et al., 2005)

By tracing all the peaks aligned through a line vertically from an amide-shift on the left-hand side of the diagonal (the amide region), the potentially scalar coupled peaks are identified. To

verify that the peaks are coupled and not the result of another spin system, the corresponding peaks originating from the α-proton and β-protons in the alkyl region are also traced. An array of peaks like the one in **Figure 2.5** shows all possible connections in a TOCSY spectrum for an AMX spin-system. Names of the different atoms are also given in **Figure 2.5** exemplified by the peaks in the fingerprint region. In short, a peak is named by stating the origin of the magnetisation (first) and the destination of the magnetisation (last). Thus, the peak identified as α/HN is a result of magnetisation travelling from the α-proton to the amide proton, whereas the HN/α peak results from magnetisation travelling in the opposite direction; from the amide-proton to the α-proton.

2.1.12 NOESY (Nuclear Over Hauser Effect Spectroscopy) and Sequential assignment of the spin systems

NOESY depend on the nuclear Over Hauser effect, which links nuclei that are correlated through space [271]. Having recognised the spin systems, the next step is to sequentially assign the spin systems by identifying the intraresidual NOE couplings (see **Figure 2.6**.) (Haugen & Sophie, 2005). The sequential assignment resulting from the outstanding dipolar NOE-peaks and distance restraints are obtained from the remaining dipolar NOE-peaks [282].

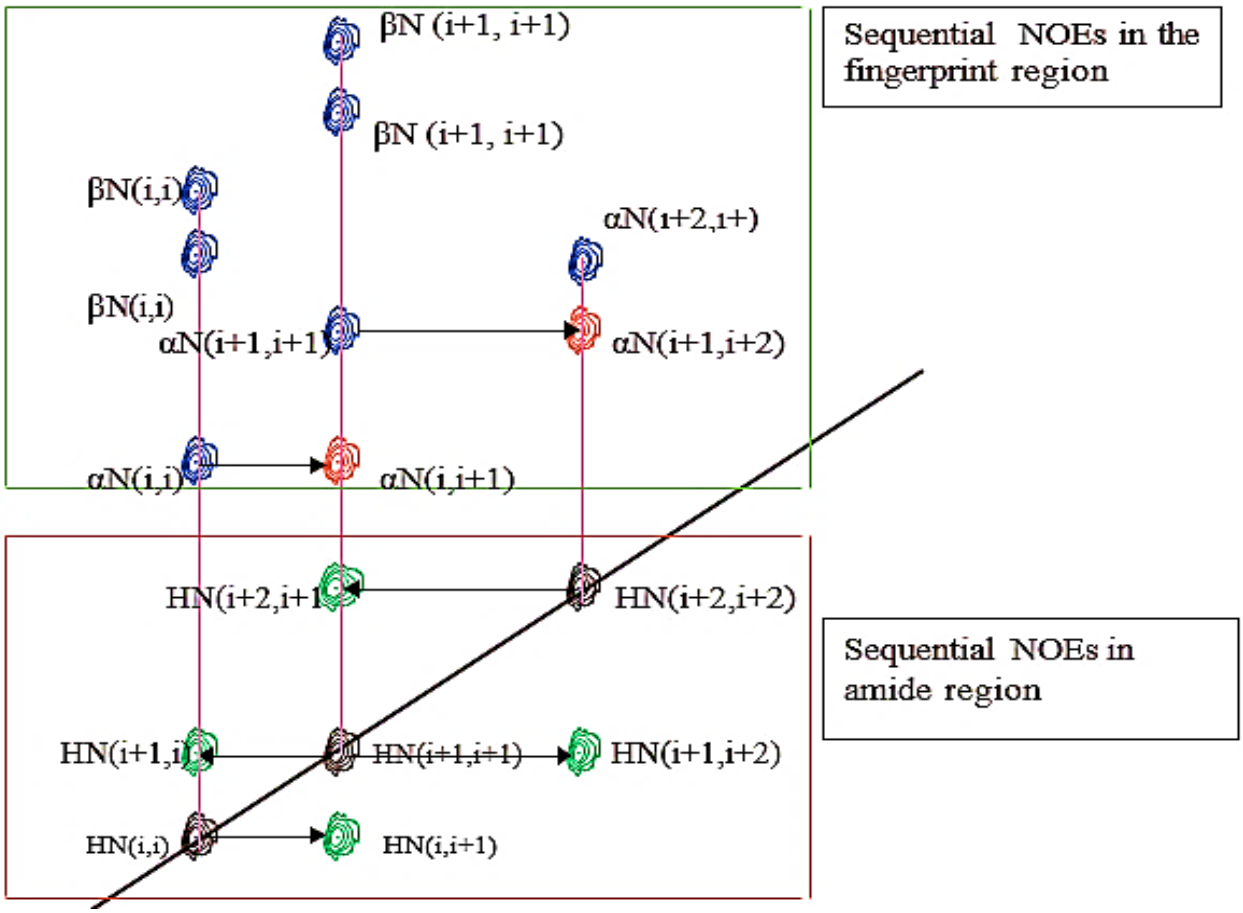


Figure 2. 6. Example of NOESY correlation diagram in ppm taken from Haugen et al., [260]. Blue and black peaks are intraresidue peaks already identified in the TOCSY spectrum. The red and green peaks are interresidue α-to- amide and amide-to-amide peaks, respectively. The horizontal arrows indicate how the cross-peaks are perused.

In the fingerprint region, sequential couplings between side-chain protons are identified. Starting at the α-proton of residue i and moving horizontally, it is possible to find a strong peak along with the amide shift of residue i+1. This is the interresidual NOE peak between the α-proton of residue i and the amide proton of residue i+1. Following the same approach, starting at the α-proton of residue i+1 and moving horizontally, one should find a strong interresidual NOE peak along with the amide shift of residue i+2. In the same approach, cross-peaks from β-protons to amide protons and other side-chain protons supports the connection of the spin systems established by tracing cross-peaks between α-protons and amide protons (not illustrated in **Figure 2.6**.): Starting at the β/γ-proton of residue i and moving horizontally along the ω2-axis, one should find a peak along with the amide shift of residue i+1. This is

the inter-residual NOE peak between the β/γ-proton of residue i and the amide proton of residue i+1. Inter-residual cross-peaks between the amide proton of a residue and side-chain protons other than the α-proton of a neighbouring residue are generally weaker than the cross peak between α-proton and amide proton on neighbouring residues. One could also alternatively search for sequential connections in the opposite direction: Moving vertically up or down from the α-proton of residue i, one should find a strong inter-residual NOE peak. By looking horizontally along the ω_2-shift of this inter-residual peak, one should find the α-proton of residue i-1. This can be visualised by following the arrows from residue i+1 to i in **Figure 2.6** in the opposite direction as to what is drawn. Usually, a string of three consecutive spin systems is enough to determine the location of these spin systems. However, sometimes more than three consecutive spin systems have to be identified to pick out a unique sequence. The aim is to locate a string of consecutive spin systems that can only be positioned one place in the sequence. If there is more than one possible position for the string of spin systems identified (irrespectively of the number of consecutive spin systems), neighbouring spin systems must be located. The number of residues in the molecule, type of spin systems represented in the sequence and their distribution along the sequence determines how many consecutive spin systems are needed to assign that particular sequence with confidence.

2.1.13 Heteronuclear single quantum coherence spectroscopy (HSQC)

HSQC is a pulse sequence, which allows magnetisation to transfer from a nucleus of a sensitive proton to another insensitive nucleus (^{13}C- or ^{15}N-) via direct ^{1}J-couplings. This is thus a ^{1}J-coupled correlation technique and can be used in the assignment of ^{13}C- or ^{15}N peaks based on their proton assignments. For carbon, one can expect correlation peaks throughout the entire carbon-bonded side chain of amino acid. For nitrogen, mainly the amide backbone cross-peaks are of interest, but for some of those amino acids with amides in their side chains, additional peaks are expected. The ^{13}C- and ^{15}N chemical shifts are used for structural calculations.

2.1.14 Solvent suppression

Most NMR experiments involving peptides and proteins are performed in aqueous solution-90%H20:10%D20 [283,284]. The significant challenges with this approach are the generation of a dominant spectral peak resulting from the high proton concentration of the solvent. This high-intensity solvent peak could lead to a problem of dynamic range. The primary aim of solvent suppression is to decrease the size of the NMR signal entering the receiver coil. Solvent suppression has two primary known methods, namely:

a. Pre-saturation.
b. Solvent non-excitation methods.

Solvent non-excitation methods are only suitable for solvent suppression if the chemical exchange between the solvent and sample protons, either do not occur or occur at a rate slower than their longitudinal relaxation rate [245]. The rate of exchange can, to a certain extent, be controlled by pH and temperature [284].

2.1.15 Structure calculation, the CcpNmr analysis-program

A significant difficulty with NMR is the reliability with which all information can be accurately assigned and labelled. Having identified all spin systems and labelled them according to the amino acid sequence, additional information yielding distance constraints such as NOE- cross-peaks, volumes of these peaks, three-bond J-couplings and chemical shifts are used to calculate a plausible 3D-structure [285]. The success of such a strategy for structure calculation depends on an as complete a list of chemical shift assignments as possible [252]. An experienced spectroscopist can calculate structures manually for small molecules, but it is cumbersome and time-consuming. Preferably, one of several computerised model-simulating programmes is used. Here the CcpNmr Analysis program has embedded in it servers like CYANA (Combined assignment and dynamics Algorithm for NMR Applications), and ARIA (Ambiguous

Restraints for Iterative Assignment) for automated NMR structure calculation [240,253,286,287]. What ARIA and CYANA do is perform several refining cycles of conformational computations to reach a structural model with the least deviation from the experimental data (Herrmann *et al*., 2002; Linge *et al*., 2003). The calculated structure with minimum energy is the structure that has maximum compatibility with the experimental data. This iterative process utilises preliminary structures from earlier cycles to reduce the ambiguities of cross-peak assignments. Current selection criteria for cross-peaks and NOE assignments are added after each cycle, yielding better structures in the process. Cycle one is thus the most crucial as a previous structure is not available [240,253,286,287]. The important conformational data-input are upper distance limits derived from nuclear Over Hauser effects (NOEs) [286]. A difficulty with NMR is the unreliability with which all information can be accurately assigned and labelled. As mentioned above, distances constraints from NOEs are essential for structure calculation. Many of the NOE cross-peaks cannot be unambiguously assigned, i.e. there is more than one possible assignment for the peaks. CcpNmr overcomes this obstacle in several ways. Firstly, all ambiguous cross-peaks are treated as the superposition of the signals from each of its multiple possible assignments and weighing them inversely proportionally to their interatomic distance. If all the correct assignments are among the assignments that are selected, the distance constraint is never falsified, even though incorrect assignments are also included [252,286,289]. Constraint combination takes the above a step further; assignment from different, often unrelated, cross-peaks are combined to generate distance restraints.

Furthermore, the concept of network anchoring is a powerful tool for eliminating false conclusions. The basis of network anchoring is that each NOE assignment must be connected to all other assignments in a network without contradictions. This ensures that constraints from unstructured coil-regions are not falsely forced into a defined structure[252,286,289].

The actual calculation performed by ARIA or CYANA is carried out by defining a target function that measures the agreement between a structure and a given set of constraints [240,253,286,290]. A low target function value reflects a confirmation that satisfies the constraints more closely than another with the higher function value. Thus, the best agreement between constraints and a calculated 3D structure is defined as the set giving the lowest target function value. Simulations done with different chemical shift assignments left out show that at least 90% of the shifts must be assigned for a satisfactory result. If some "essential" assignments that contribute actively to the structure are left out, a higher percentage of assignment might be necessary for a correct structure calculation[240,252].

2.1.16 Computational methods

2.1.17 Molecular mechanics (MM)

MM is one of the first computational tools used for modelling large molecules (i.e., nucleic acids and proteins). It ignores the quantum properties of electrons and takes into account a straightforward explanation of nuclei and electrons as point charges based on the Born-Oppenheimer approximation [291]. The point charges are connected by springs (bonds) and other possible functions that describe electrostatic interactions and van der Waals. The potential energy functions of the system, bond angles, bond distances and other parameters are first determined empirically based on specific sets of atoms. The potential functions and the experimentally determined parameters used to define the potential energy function of a system is known as a forcefield.

2.1.18 Force field

Forcefields are parameters and a collection of functions that are utilised to define the potential energy of a system. The terms used in resolving the total energy of a system are divided into non-bonded interactions and bonded interactions. In a general force field equation, the non-

bonded interactions are further divided into van der Waals and electrostatic interactions. Also, the bonded interactions are subdivided into angle bending, bond stretching, improper dihedral angles and proper dihedral angles

2.1.19 Bonded Interactions

These interactions are based on fixed atoms, namely, bond angle, torsional angle interactions and bond stretching. Bond stretching is defined as a harmonic potential that considers the compression and stretching of bonds which make atoms behave as if they are connected by springs. The energy of the bond varies with the square of its displacement from the natural bond length. The stretching of the bond can be described by the addition of quartic terms to the harmonic potential and the Morse potential.

$$v(l) = \frac{k}{2}(l - l_0)^2 \qquad 2.3$$

Where the potential energy stored in oscillator is given by **υ, k** is the spring constant, the final bond length **l** and the natural bond length $\mathbf{l_0}$. The energy of each bond in the system is defined by a reference value and force constant, and the harmonic potential can also be described as angle bending denoted by **θ**.

$$v(\theta) = \frac{k}{2}\left(\theta_i - \theta_{i,0}\right)^2 \qquad 2.4$$

Torsional or dihedral potentials describe the energy contributions of improper and proper dihedral angles. Though some MM force field does not make use torsional potentials, instead to achieve the desired energy profile, they make use of non-bonded interactions between 1-4 atoms in a dihedral angle. Torsional potentials are often established using a cosine series expansion where potential energy stored in oscillator is given by **υ, w** is the angular frequency

.

$$\nu(w) = \sum_{n=0}^{\mathbb{N}} \frac{\nu_n}{2}[1 + cos(\cap w - \gamma)] \quad 2.5$$

Equation 2.6 below shows the improper torsion, and the improper torsion angle is maintained at 0^0 or 180^0

$$\nu(w) = k(1 - cos\, 2w) \quad 2.6$$

2.1.20 Non-bonded interactions

Electrostatic interactions describe the forces experienced by atoms as a result of their charges. They are treated using the classical pair-wise additive Coulombic term.

$$\nu(r_{ij}) = \frac{q_i q_j}{4\varepsilon_0 r_i j} \quad 2.7$$

Where the potential energy is given by **υ, r_{ij}** is the distance between ***i*** and ***j***, $q_i q_j$ are the atomic charge and ε_o is the permeability of free space. London forces are generally known as Van der Waals interactions. They mostly described as either repulsive or attractive forces between instantaneous dipoles. The van der Waals interactions are usually treated by Lennard Jones potential and are often shortened at a cut-off distance. To smooth truncated forces and thermal noise around the cut-off radius, different functions can also be applied.

$$\nu(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad 2.8$$

The general class 1 forcefield is described by equation 2.9.

$$V(x) = \sum_{bond} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angle} \frac{k_i}{2}\left(\theta - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{v_n}{2}[1 + cos(n\omega - \gamma)\,] +$$

$$\sum_{i=j}^{N} \sum_{j=i+1} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{i_j}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \qquad 2.9$$

2.1.21 Restraints and constraints

During simulations, constraints are applied to help fix angles or bonds. Constraints are different from restraints. Constraints are used to position the restrained entities as close as possible to value. Restraints are applied by improving an additional term to the potential energy function of a system, and a penalty is applied to the energy of the system when the bond or angle deviates from the reference value. Distance restraints are used to refine structures using NMR data. Targeted restraint is mostly generated from NOESY data [292]. Targeted distance restraints can be applied using the following relationships (see equation 2.10) below a specified lower bound and between two specified upper bounds and linear beyond the largest bound[292].

$$v_{d_r}(r_{ij}) = \begin{cases} \frac{1}{2} k_d r \left(r_{i_j} - r_0\right)^2 & for \quad r_{ij} < r_0 \\ \frac{1}{2} kdr\left(r_{ij} - r_1\right)^2 & for \quad r_0 \le r_{i_j} < r_1 \\ \frac{1}{2} k_{d_r(r_2 - r)(2r_{ij} - r_2 - r)} & for \quad r_1 \le r_{ij} < r_2 \\ & for \quad r_2 \le r_{ij} \end{cases} \qquad 2.10$$

2.1.22 Minimization Energy (EM)

EM is executed to remove steric clashes that may lead to instabilities of the system during molecular dynamics simulation. EM is achieved by searching for the global minimum energy potential of a macromolecule which is zero. Sometimes, the configurational space dimensionality of a macromolecule and the number of local minima the molecule have can be very high, making it impossible to sample the whole conformational space. To enable the

determination of the global minimum within a short period, it is much easier to determine the nearest local minimum by systematically moving down the steepest local gradient of the potential energy function. Some EM methods usually applied include the L-BFGS and steepest descents.

2.1.23 L-BFGS

The L-BFGS is achieved by successively producing improved approximations of the Hessian matrix from a previous step, and a fixed number of corrections is used[293].

2.1.24 Steepest descents

The Steepest descents EM is categorised under first-order minimisation algorithms. It does not consider any history built up in previous steps, and instead, it takes a step in the direction parallel to the force of the system. The motion is usually downhill, and the size of the step can be adjusted to achieve a quick result[294].

2.1.25 Temperature and pressure coupling

In terms of energy potential, MD simulations can be executed under different conditions like NVT and NPT where the N = signifies the number of molecules in the system, V = volume, P = pressure, T = temperature. Usually, N and T are kept constant, and either P or V allowed to vary. For example, while studying the flexibility and stability of a GCPR molecule in a mimetic membrane all placed in a water box, the NPT ensemble is mostly preferred[294,295]. Likewise, in our systems, which are also highly flexible, for a proper understanding of how peptides activate a receptor leading to a conformational change, we performed MD simulations under NPT condition [296,297].

2.1.26 Simulated annealing (SA)

SA is a method used to search for the minimum energy of a molecular system. During SA, the temperature of the system is increased such that the system occupies high-energy states of its

conformational space and can cross energy barriers[217,298–300]. The temperature is then dropped slowly to allow the system cool without becoming trapped in high energy conformations. At this stage, the lowest energy conformer could be attained. At absolute zero, it is expected that the system should occupy the lowest energy state that corresponds to the global minimum energy conformation[217,298–300].

2.1.27 Molecular dynamics simulations

Molecular dynamics simulation technique encompasses the incorporation of Newtonian equations of motion for numerous molecules interacting over the control of a force field (F) t [298,301] see equation 2.3 and 2.4.

$$\boldsymbol{F = Ma} \qquad \textbf{2.3}$$

Where **F** is the force field, **M** is mass, and **a** is the acceleration

$$\frac{d^2xi}{dt^2} - \frac{F_{x_i}}{m_i} \qquad \textbf{2.4}$$

Where, $\mathbf{m_i}$ = mass of an atom along one coordinate $\mathbf{x_i}$, $\mathbf{F_{xi}}$ = force on the atom in that direction.

Integrating the equations of motion help define how the accelerations, velocities and positions of a molecule in a system can vary with time [298,301].

2.1.28 Algorithms

The potential energy function in molecular dynamics is made up of 3N atomic positions, which makes it extremely difficult to get an analytical solution to the equations of motion. To proffer a better solution that regard integration algorithms have been developed from the Taylor series expansion equation. The algorithms are the leap-frog algorithm, the Verlet algorithm, velocity Verlet algorithm and the Beeman's algorithm[302–304].

2.1.29 Varlet algorithm (VA)

The VA [305] is the most commonly used algorithm for integration in MD simulations[298]. See the following equations.

$$\mathbf{r}(\mathbf{t} + \boldsymbol{\delta}\mathbf{t}) = \mathbf{2r}(\mathbf{t}) - \mathbf{r}(\mathbf{t} - \boldsymbol{\delta}\mathbf{t}) + \boldsymbol{\delta}\mathbf{t2a}(\mathbf{t}) \quad \mathbf{2.5}$$

$$\mathbf{v}(\mathbf{t}) = \frac{[\mathbf{r}(\mathbf{t} + \boldsymbol{\delta}\mathbf{t}) - \mathbf{r}(\mathbf{t})]}{\mathbf{2}}\boldsymbol{\delta}\mathbf{t} \quad \mathbf{2.6}$$

$$\boldsymbol{v}(\boldsymbol{t} + .\boldsymbol{\delta t}) = \frac{[\boldsymbol{r}(\boldsymbol{t} + \boldsymbol{\delta t}) - \boldsymbol{r}(\boldsymbol{t})]}{\boldsymbol{\delta t}} \quad \mathbf{2.7}$$

Where r = positions, **a** = accelerations, **t =** time and **r (t - δ t)** is the positions from the previous step while **t + δt, r (t + δt) =** is used to calculate the new positions at time **t. δt =** is the time step.

2.1.30 Leap-frog algorithms (LFA)

The LFA [303], is a deviation of the VA with the following equation **2.8** and **2.9** [298].

$$\boldsymbol{r}(\boldsymbol{t}\boldsymbol{\delta t}) = \boldsymbol{r}(\boldsymbol{t}) + \boldsymbol{v}\left(\boldsymbol{t} + \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\delta t}\right)\boldsymbol{\delta t} \quad \mathbf{2.8}$$

$$\boldsymbol{V}\left(\boldsymbol{t} + \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\delta t}\right) = \boldsymbol{v}\left(\boldsymbol{t} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\delta t}\right) + \boldsymbol{a}(\boldsymbol{t})\boldsymbol{\delta t} \quad \mathbf{2.9}$$

Where **r =** positions, **a** = accelerations, $\left(\boldsymbol{t} + \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\delta t}\right)$ = time and **t + δt =** leap-frog velocity at a particular position of time [298,301]. The LFA is cheaper and easy to apply when dealing with large molecules during MD. Choice of time step during the simulation

The time step to be used when conducting MD simulations is crucial because if for any reason, a wrong time step is used it could lead to severe artefacts (i.e. thermostat artefacts, the origin of layer structure artefacts, collective motion artefacts and nonlinear resonance artefacts) [298,306].

It is advisable, during simulation, to use as small a time step as possible. However, shorter time steps mean more computational time. In practice, a compromise is used. [298,307]. Conducting MD simulation with Groningen machine for chemical simulation (GROMACS), a time step of 1 fs or 2 fs is normally used[307,308].

2.1.31 Periodic boundary conditions (PBC)

PBC assist in the minimisation of the edge effects in a finite system [298,307,309–311]. PBC is usually used for approximation of an infinite (extensive) system, by using a small part of a unit cell (**Figure 2.7**). Theoretically, any cell shape not limited to, hexagonal prism, cube, hexagonal prism, elongated dodecahedron, rhombic dodecahedron and truncated octahedron, can be used as long as it covers all space by translation operations of the inner box in a 3D format. Typically, a minimum image convention is used where a particle in the central box does not see more than one image of another particle within the box.

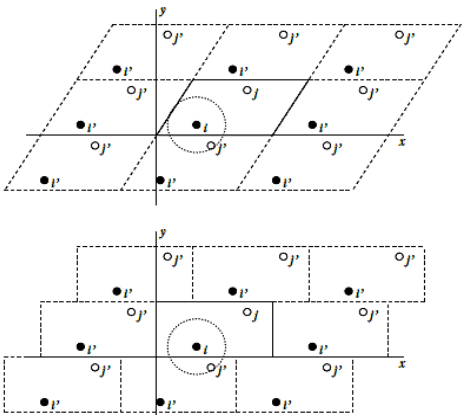


Figure 2. 7. Schematic diagram showing periodic boundary condition in two dimensions. Where x is the length of the box in one direction, and dx is the distance direction of the vector from one object i to another to object j. The horizontal x and the vertical y represent the two Cartesian coordinates [307,311].

2.1.31.1 Elucidating the solution structure of Phote-HrTH

2.1.31.2 Preparation of sample

Phote-HrTH (pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2) was bought from Peptic Co Ltd, China. Its purity was tested with HPLC–MS and establish to be > 98% pure. NMR samples were prepared by dissolving the dry peptide in a solution consisting of 150 mM sodium dodecyl-d_{25} sulfate (SDS), 20 mM phosphate pH 4.5, 0.05 mM TSP (trimethylsilyl propionate) and 10%v/v D_2O to reach a final peptide concentration of 2 mM. Another sample was prepared by dissolving ~1 mg of Phote-HrTH in 0.5 ml of 30% deuteron-DMSO, which was 20 mM in phosphate buffer.

2.1.32 NMR experiment

All NMR spectra were obtained at 310 K on an 800 MHz Bruker Advance II spectrometer equipped with a 5mm TCI cryoprobe. ^{1}H homonuclear TOCSY (mixing time τ_{mix} = 80 ms) and NOESY (τ_{mix} = 150 and 300 ms) were acquired with 512 points in the F1 dimension and 4k points in the F2 dimension. Between 16 and 32 transients were accumulated with a 1.5 s recycle delay. The data were multiplied with a squared sine bell function shifted by 90°. The ^{1}H spectral window was set to 9600 Hz. ^{13}C-^{1}H HSQC experiments were performed with 256 points in the F1 dimension and 4k points in the F2 dimensions. 64 Transients were acquired with a 2s recycle delay. The ^{13}C spectral window was set to 33200 Hz. Non-uniform sampling ^{15}N-^{1}H HSQC experiments were performed with 25% of 128 points in the F1 dimension and 4k points in the F2 dimensions. 1024 Transients were acquired with a 1.5 s recycle delay. The ^{15}N spectral window was set to 3240 Hz.

All data dimensions were zero-filled to twice the respective FID size. ^{1}H chemical shifts were referenced to TSP at 0 ppm. Data were processed in Topspin (Bruker) then analysed using the

CCPNmr Analysis program[240,312–314]. Backbone and side chains were assigned using all experiments.

2.1.33 Structure calculations

The structure determination was achieved through the assignments of NOESY cross-peaks, after which NMR distance restraints where generated. The generated NOE distance restraints were supplemented with dihedral angle restraints predicted with DANGLE from H_α, H_N, N_H, C_α, C_β chemical shifts[251]. A standard CNS 1.1-based protocol was employed using the ARIA 2.2 interface[315]. Ten lowest energy structures were refined in a water shell and evaluated with MolProbity[316].

The lowest energy structure from ARIA 2.2 was used as a starting structure for restrained molecular dynamics using GROMACS 2018.6 [301,308,317].

2.1.34 MD simulation

2.1.35 Restrained molecular dynamics of Phote-HrTH

The starting structure of Phote-HrTH was obtained using Ccp-Nmr software [240]. Using the steepest descent algorithm, the structure was energy minimised for 50,000 steps. Using GROMACS version 2018.6, NMR distance restrained molecular dynamics simulations were conducted both in water and dodecyl phosphocholine (DPC) [301,308,317]. All molecular dynamics simulations were carried out using time-averaged NMR restraints with a disre-tau of 10 ps, a time step of 2 fs, and the OPLS-AA/L, all-atom force field. The inter-nuclear distances determined from the NOE spectra were modified by adding 10 or 20% to set the first and second upper limits of a square well potential [318]. The potential was set to zero between the lower and upper bounds and increased quadratically (Force constant 1000 kJ $mol^{-1}nm^{-1}$) beyond that. The LINCS algorithm was applied to constrain all bonds. For van der Waals and electrostatic interactions, a cut-off of 1.0 nm was used.

2.1.36 Simulation in water

Using the single point charge (SPCE) water model, a box containing the peptide, and 7000 water molecules with a cut-off of 1.0 nm was constructed. After equilibration, molecular dynamics was conducted for 10 ns at 300 K under NPT conditions. In total, 200 structures were collected at 50 ps intervals. Cluster analysis was then done using the linkage algorithm of GROMACS and a cut-off value of < 0.1 nm for superimposing backbone atoms. The minimum energy conformer from the most significant cluster gave the structure of Phote-HrTH.

2.1.37 Simulation in water/DPC mixture

The lowest energy structure from the most significant cluster from the water simulation was used as the starting conformer of Phote-HrTH in micelle solution. The peptide was placed in the centre of a 7 nm cubic box filled with a 50 DPC molecule micelle and 10 000 water molecules. [319] The micelle was translated so that the centre of the micelle was at the bottom edge of the box. This meant that using periodic boundary conditions, half the micelle was at the bottom of the box and the other half was at the top. The peptide was then placed in the centre of the box. Energy minimisation was carried out, using the steepest descent method, for 10,000 steps or the tolerance was < 10 kJ mol^{-1}or to machine precision. Two stages of system equilibration were performed to solvate the peptide and to accomplish a steady starting state of temperature, pressure, and density. The first stage of equilibration involved performing MD for 100 ps under NVT conditions at 300K followed in the second stage by a further 1 ns MD under NPT conditions. The final MD simulation was for 10 ns, during which 200 structures were. Cluster analysis was performed in the same manner as before.

2.1.38 Results and Discussion

2.1.39 Spectral assignment

Phote-HrTH is hydrophobic but readily soluble in SDS micelle solution; this signifies that the peptide interacts with the micelle. This interaction was confirmed by measuring the diffusion coefficient of the peptides, which was found to be approximately $8 \times 10{-}11\ m^2s^{-1}$, which gives a micelle size of roughly 56 DPC molecules. Given the insensitivity of the diffusion coefficient to molecular mass, this is in good agreement with our assumed size of 50 DPC molecules per micelle but, more importantly, justifies our use of the NOESY pulse sequence for such a small peptide. An expanded assignment of the NH -Hα region of both the TOCSY and NOESY spectra is given in **Figure 2.8.**

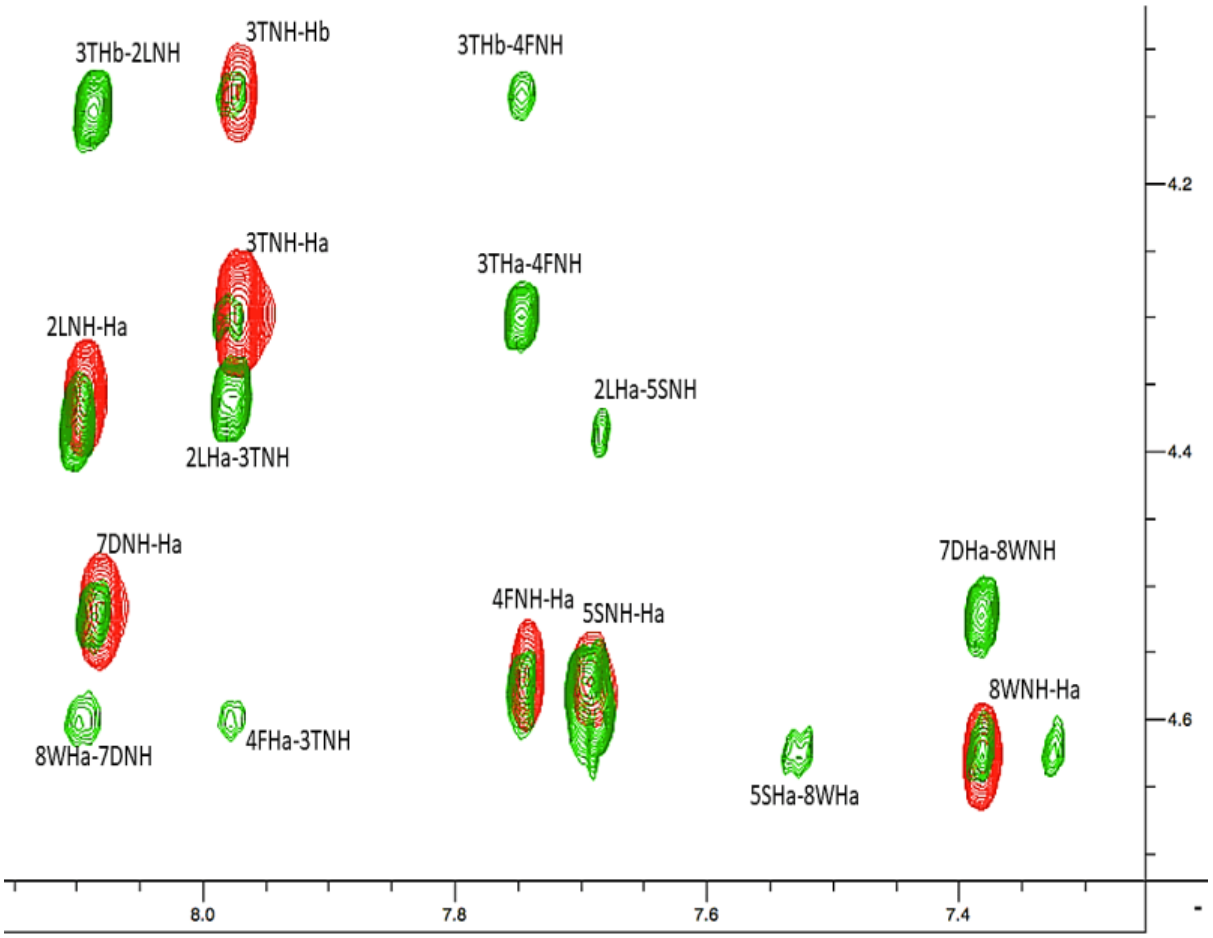


Figure 2. 8 Phote-HrTH spectral assignments of TOCSY (RED) and NOESY (green) overlay showing an expansion of the NH -Ha region with chemical shift (ppm).

The chemical shifts and assignments of Phote-HrTH, in an SDS micelle solution, are given in **Table 2.1**. These data were used to compute structure-induced chemical shift changes using the CSDb algorithm [320,321].

Table 2.1. Chemical shift (ppm) assignments for Phote-HrTH 150 mM SDS micelles, pH 4.5, 2mM peptide, Temp = 310 K referenced to TSP

Residue	N	H	HA.	HB	HG	CA	CB	CG.	Others
1 Glu	-	-	4.39	2.36*	2.53 *	59.6	28.2	32.2	
2 Leu	121.8	8.10	4.38	1.63, 1.49	1.64, 1.98	55.3	42.6	27.4	HD1 0.93, HD2 0.89, CD1 25.5, CD2 24.2
3 Thr	115.1	7.98	4.30	4.14	1.09	61.7	70.1	21.4	
4 Phe	122.0	7.75	4.58	2.98*		-	40.3		HD 7.15, HE 7.21, HZ 7.15, CD 131.9, CE 131.3, CZ 129.7
5 Ser	117.4	7.70	4.58	3.59, 3.53		-	63.8		
6 Pro	-	-	4.15	1.68*	1.54, 1.46	55.9	31.6	27.0	HD1 3.10, HD2 3.42, CD 50.3
7 Asp	120.4	8.09	4.52	2.69, 2.46		54.4	41.1	-	
8 Trp	120.8	7.39	4.63	3.24*		-	30.1		NE 128.9, HD1 7.20, HE1 9.90, HE3 7.53, HZ2 7.40, HZ3 7.02, HH2 7.09, CD1 127.5, CE3 121.1, CZ2 114.4, CZ3 121.7, CH2 124.2,

2.1.40 Peptide flexibility

The NMR chemical shift assignments are given in **Table 2.1** and were used to probe the solution structure and flexibility of Phote-HrTH in SDS. Berjanskii and Tremblay et al. [265,322] have demonstrated that by contrasting the measured chemical shifts to literature values for a random coil structure, some idea as to the structure and flexibility of the peptide can be obtained [265,322]. Structure-induced chemical shift changes (observed shifts minus random coil reference values) were analysed using the CSDb algorithm available at andersenlab.chem.washington.edu/CSDb/about.php [320,321]. **Figure 2.9** shows a plot of random coil chemical shift deviations for Phote-HrTH. Apart from Pro6, all the H_α deviations are small and downfield, while the H^N shifts are both upfield and downfield. Polypeptides with a helical structure have, H^N and H_α chemical shifts which are, on average, −0.30 ppm less than their random coil values, while β-sheet structures have shifts of ca. 0.4 ppm[323]. Thus, the results for Phote-HrTH indicate that it has some turn-like structure. Similar results were found for Declu-CC from the meloid beetle, *Decapotoma lunata*, Melme-CC from the Fruit beetle *Pachnoda sinuata* [324] and Dappu-RPCH from *Daphnia pulex*[325].

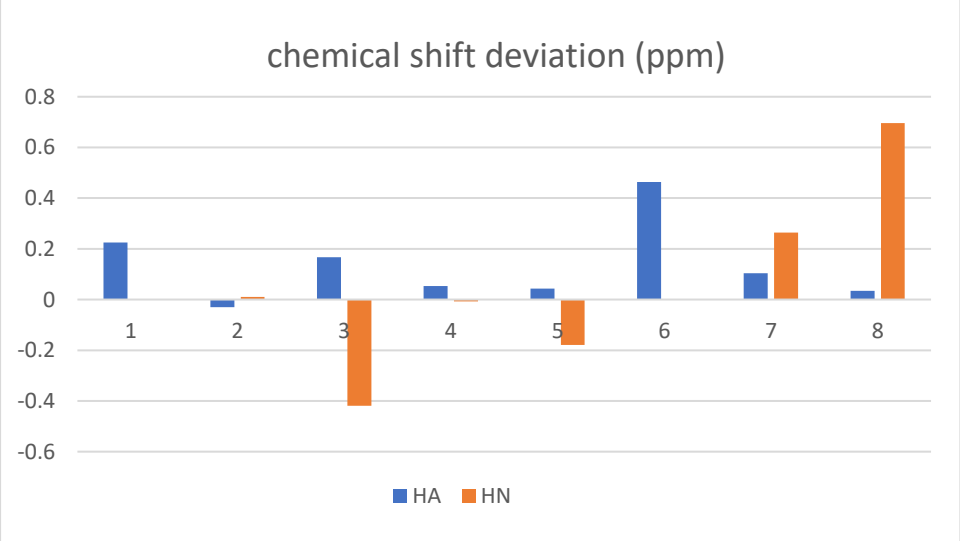


Figure 2. 9. Phote-HrTH random coil chemical shift deviation in SDS micelle solution.

Applying the Random Coil Index tool [322], the chemical shifts can also be used to estimate the model-free order parameter, S^2, of Phote-HrTH (see **Figure 2.10**). An order parameter of 1 signifies the peptide is rigid, while an order parameter of 0 signifies the peptide has no structure. **Figure 2.10** shows that Phote-HrTH is inflexible (rigid), with a maximum order parameter of 0.97 around serine, whereas the C-terminal has less ordering (S^2 = 0.96).

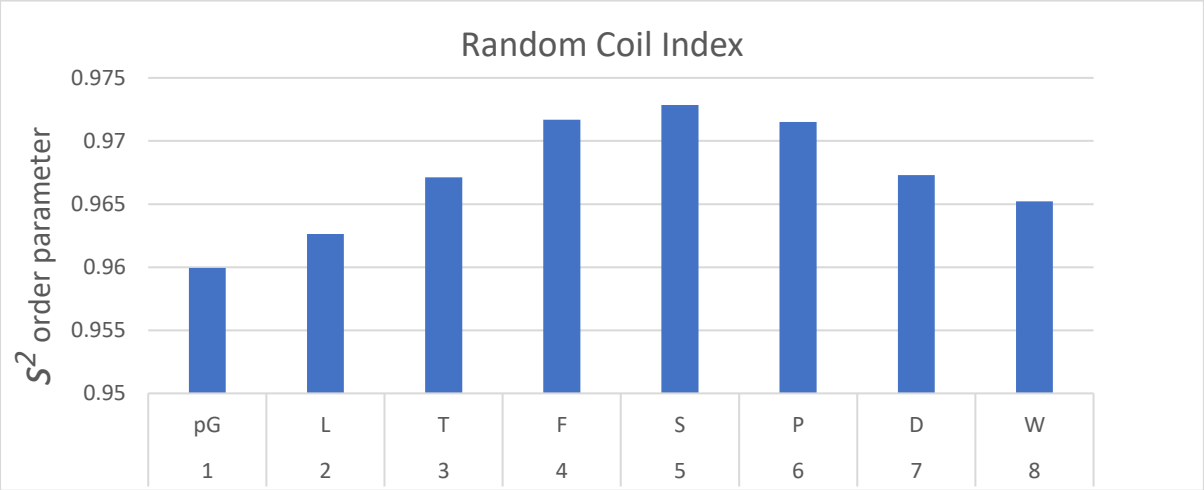


Figure 2. 10. Phote-HrTH S^2 order parameter calculated using RCI. [326]

Similarly, one can calculate the dynamic molecular root means square deviation (RMSD) and NMR RMSD. The result obtained (**Figure 2.11**) shows that, as expected, Phote-HrTH is more flexible at both the N- and C-termini and more rigid between Phe 4 – Pro 6; overall, the peptide is inflexible.

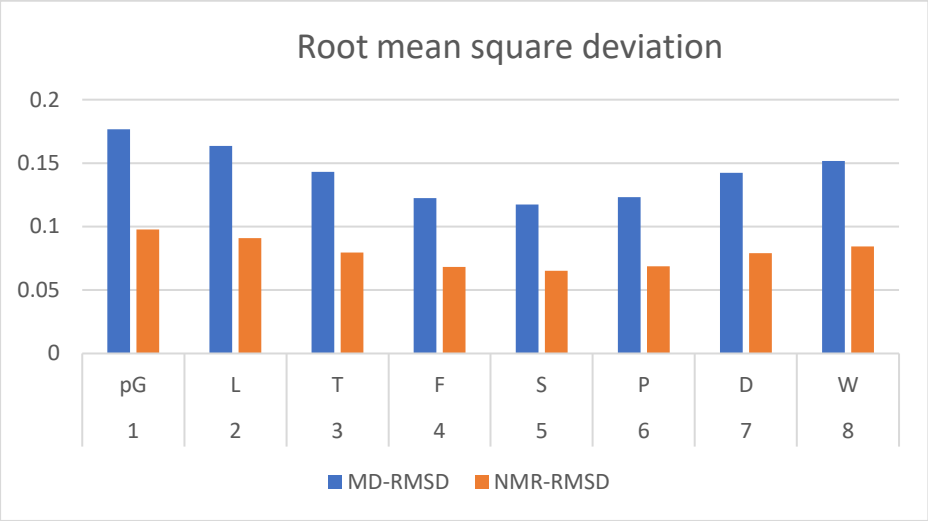


Figure 2. 11. MD and NMR root-mean-square deviations (RMSD)

2.1.41 MD Analysis

2.1.42 Simulation in water

94 NMR derived NOE distance restraints, together with dihedral angle restraints predicted with DANGLE from H_α chemical shifts, were used to refine the structure of Phote-HrTH in water, using ARIA 2.2. The results shown in **Figure 2.12b** is an overlay of the 10 lowest energy structures. All the structures had the same basic shape but differed in the orientation of some of the turns, particularly at the N-terminus. MolProbity analysis of the structures showed them to have no violations, no clashes and all torsion angles to be in the allowed or favourable Ramachandran space[316].

The lowest energy structure from the ARIA analysis was used as the starting structure for restrained molecular dynamics in water. After a 10 ns simulation, cluster analysis gave one (**Figure 2.12** a) dominant cluster, indicating that the conformation of the peptide is highly conserved. All the dihedral angles were in the allowed region of dihedral space.

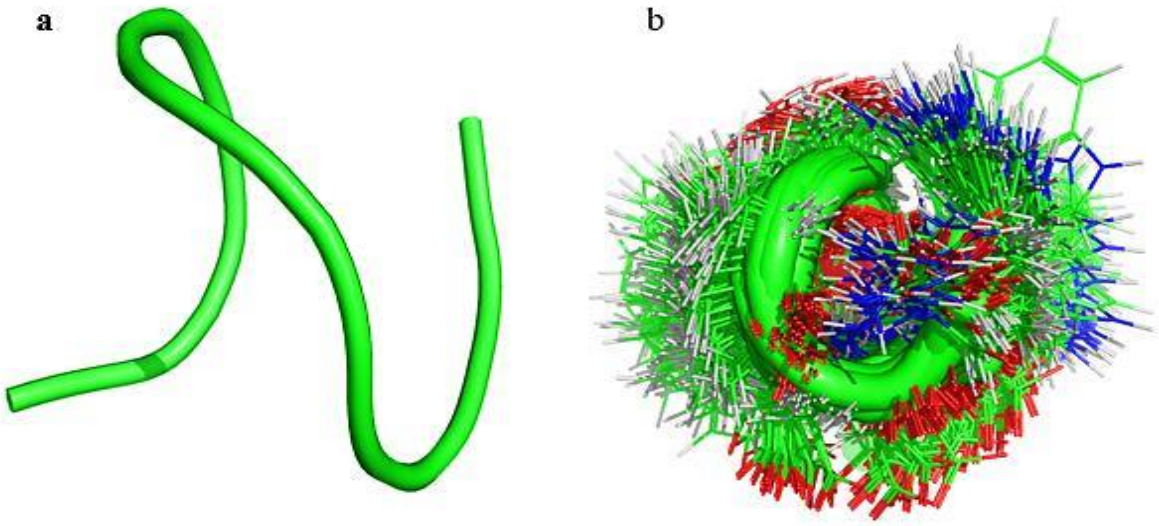


Figure 2. 12. (a) Lowest energy conformation of Phote-HrTH calculated using NMR restraints in water. (b) overlay of 10 lowest energy conformation of Phote-HrTH clusters. ARIA was used to generate these results.

The three-dimensional structure of Phote-HrTH possesses a β-type I turn and is stabilised by H-bonding. The β-turn is due to Pro6 and is stabilised by a moderately strong, transient H-bond between Trp8(NH) and Thr3(CO) as well as many other H-bonds between the side chains and backbone. Further analysis was conducted to back up the β-type I turn postulate using the COUDES server [327] available at mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#jobs::COUDES.M24372186968088. The prediction of the server was similar to our prediction in that a β-turn was found in the central residues of the motif, Ser-Pro-Asp-Trp. These observations agree with the S^2 parameters and chemical shift deviations given before (**Figure 2.9**) and the random coil parameters (see **Figure 2.10**).

2.1.42.1 Simulation in DPC.

The root conformation from the water simulation was used as the starting structure for simulation in a DPC micelle solution. The simulations were started with the peptide in the

middle of the solvent box, but it quickly diffused to interact with the micelle.

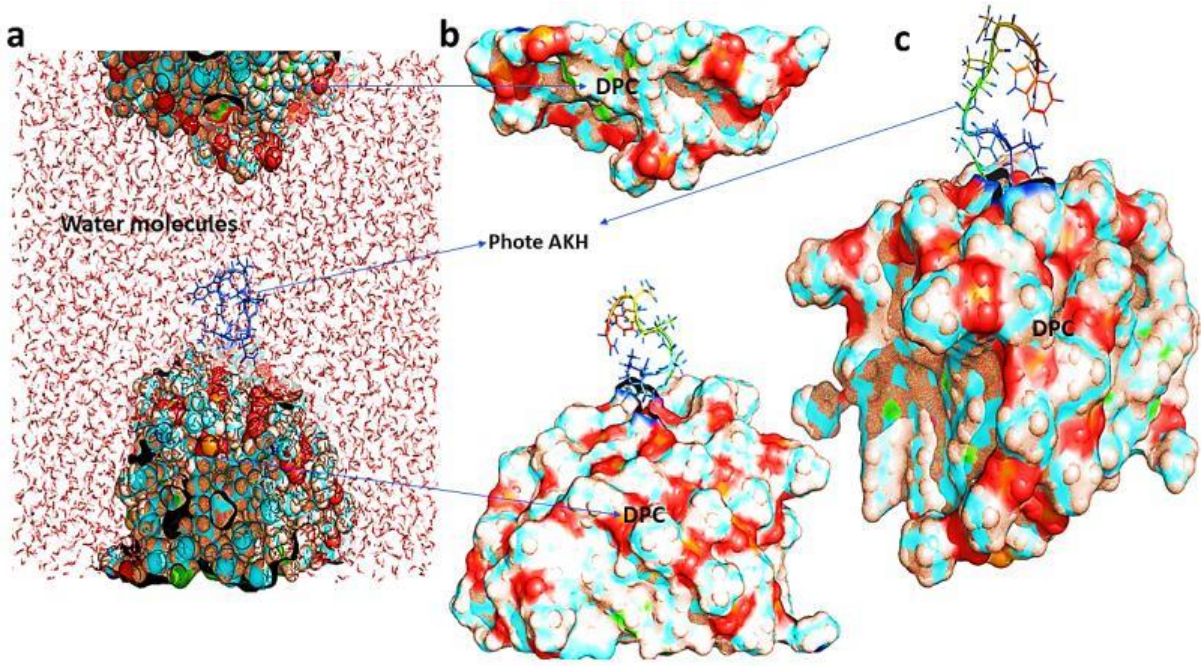


Figure 2. 13. Interaction of Phote-HrTH with DPC micelle solution, between 1ns – 2ns of MD. b) Interaction of Phote-HrTH with DPC micelle solution, between 2.5 ns – 10ns of MD. c) Interaction of Phote-HrTH with DPC micelle solution after 10ns of MD.

Depending on the starting orientation of the peptide relative to the micelle, the peptide would interact with the phospholipid and move away (**see Figure 2.13**) until a stable orientation was established. To elaborate more on this, see **Figure 2.14**, where the peptide/DPC contact area is plotted as a function of time. For Phote-HrTH, contact between the DPC and micelle is established during the first 10 ns. After that, it is stable for 20 ns before the peptide moves along the surface of the micelle. It is interesting to note that the contact area between Phote-HrTH is much higher than those reported for Declu-CC, Melme-CC and Dappu-RPCH but similar to that of Locmi-AKH-I [325]. The interaction between the peptides and the lipid surface, as shown by the contact area, is interesting as it has been postulated according to 'Membrane Compartments Theory", a flexible neuropeptide is postulated to first binds to its membrane. The membrane induces conformal changes before the peptide recognises and bind to its receptor [328,329]. Hence, surface binding is an essential step in receptor activation.

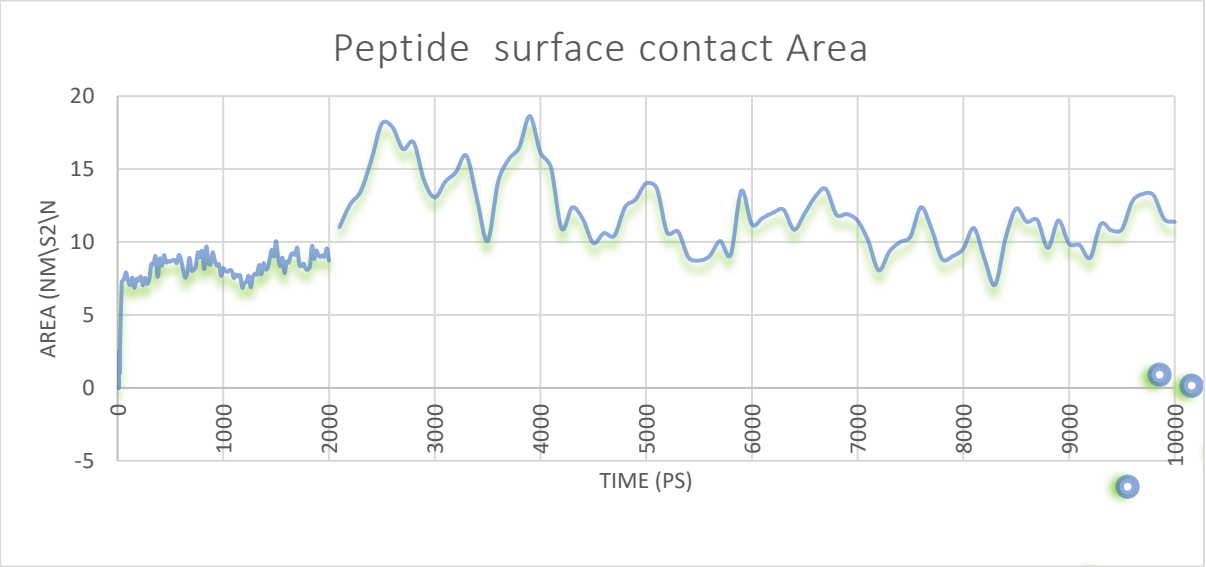


Figure 2. 14. Contact surface area (nm^2) between Phote-HrTH and DPC micelle as a function of time.

Cluster analysis of the trajectory gave a single large cluster for Phote-HrTH (**Figure 2.15**). The predominant conformation of Phote-HrTH does have a turn feature but a more open structure compared with those obtained for Declu-CC, Melme-CC, and Dappu-RPCH [325].

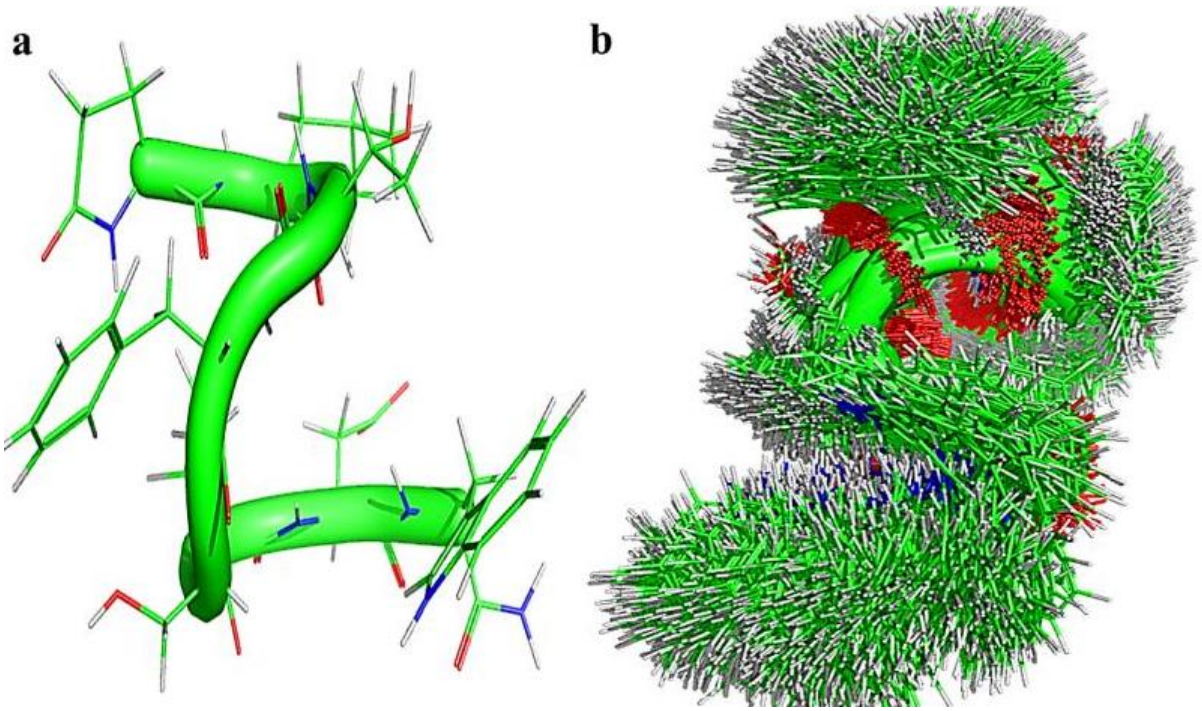


Figure 2. 15. a) The lowest energy conformation of Phote-HrTH calculated using NMR restraints in DPC micelle. b) Overlay of 50 lowest energy structures of Phote-HrTH calculated using NMR restraints in DPC micelle.

Figure 2.15 (a and b) illustrated a clear β-turn around its proline residue. A tightly coiled conformer in the DPC solution was observed when compared to the starting structure (see **Figure 2.12**).

Conclusion

Phote-HrTH has been studied in both watter and micelle solution. The results have shown that Phote-HrTH has a natural β-turn and is not very flexible, having a model-free order parameter of 0.97 in the micelle-bound state. H-bonds between Trp8(NH) and Thr3(CO) as well as H-bonds between the side chains and backbone stabilise this conformation. In general, the molecule has an opened conformation.

3

Constructing a 3D Structure of the Flesh Fly Receptor, *Sarcr*-AKHR

Summary

To predict structure-based drug-ability of insect AKHRs, the newly available genome of the flesh fly AKHR was used to construct a three dimensional (3D) model of the receptor Sarcr-AKHR [120,190,330]. *Homology modelling techniques were employed using two target crystal structures, namely the beta2-adrenergic receptor (β2AR) and rhodopsin. The β2AR structure has an open conformation meaning the receptor is inactive, while the rhodopsin structure possesses a closed, active, conformation. To adequately explain the active and inactive conformation of the constructed models, these two target models were necessary. This is because literature has shown that active rhodopsin does not allow the diffusion of a ligand into the binding site* [224,331].

3.1.0 Introduction.

The flesh fly (*S. crassipalpis)* is a carrion feeder and used in forensic entomology for estimating corpse death, as well as being used in labs as models for insect physiology and biochemistry. [50–52]. However, flesh flies are both beneficial and pest insects [6,7]. The flesh flies are associated with some deadly diseases to both humans (they carry leprosy and transmit intestinal pseudo

myiasis) and animals (causes myiasis and blood poisoning) [50–52]. Since the flesh flies have both beneficial and pest characteristics, care needs to be taking while managing them. The conventional method of managing them is by the use of chemical pesticides of which somehow, they have developed resistance[37,38]. These chemical pesticides do not just kill them but kill another beneficial insect, and they could also affect humans negatively. As such, there is a need for better and efficient insecticides (green insecticide) that will be selective in their activity, negatively affecting the targeted pest species without causing damage to valuable insect species. One way of achieving the green insecticide is to target the GPCRs of pest insects[29–33,41,42].

Neuropeptides and their G-protein-coupled receptors are considered suitable targets for new insect control agents, similar to the bases used in the enhancement of drugs for the treatment of human disease [29–33,41]. 3D structures of GPCR/AKHR and ligand are crucial for molecular explanation and understanding of ligand binding and activations of receptors. However, to date, very few 3D structures of insect GPCR/ AKHR are available. Thus, the absence of adequate literature and elucidated 3D models hinders template target virtual screening[332,333]. Also, the lack of 3D models and inadequate ligand information has hinders ligand-based approaches to drug discovery [92,334]. Adequate information on 3D structures of adipokinetic hormones and their cognate receptors from different insect pests could help in the development of mimetic neuropeptide analogues that are selective in their activity, negatively affecting the targeted pest species without causing harm to beneficial insect species.

Here we used the primary amino acid sequence from the flesh fly AKHR for the construction of 3D models of Sarcr-AKHR via homology modelling. To conduct homology modelling, there is a need for a template structure; these templates must have a sequence closely related to the protein being built. The amino acid sequence uniquely determines the structure of a protein.

Knowing the sequence, it should, at least in theory, suffice to obtain the structure [209,330,335–338]. Since we are interested in the activation of the receptor, models of both the active and inactive structures were constructed. The receptors were built based on the crystal structures of rhodopsin and beta2-adrenergic receptor (β2AR). Thus, the β2AR model is described as the open or inactive models, while the rhodopsin based model is termed the closed or active model. [239] The closed conformation of the rhodopsin-based model is similar to the rhodopsin crystal structure where the close nature protects *cis*-retinal from hydrolysis.

In this chapter, the constructed Sarcr-AKHR 3D model from the rhodopsin template will be referred to as the Rhodflf-model, while the Sarcr-AKHR 3D model constructed from the β2AR template will be referred to as the β2flf-model.

3.2.0 Experimental Methods

3.2.1 Constructing the α-helices

3.2.2 Secondary structure prediction

The primary sequence of Sarcr-AKHR was obtained from Genbank database accession number **AOC38019.1.**[43] PSIPRED 4.0[339–341]and MEMSAT-SVM[342,343] programmes, available on http://bioinf.cs.ucl.ac.uk/web_servers/, were used to predict the secondary structure and the transmembrane regions of the receptor. PSIPRED 4.0 and MEMSAT-SVM makes use of two feed-forward neural networks, which perform analysis on output obtained from PSI_BLAST [341,344].

3.2.3 Alignment of the Primary Sequence and Homology Model

To adequately explain the inactive and active state of the flesh fly AKHR, two models were proposed, one to represent the active (closed state) and the other the inactive (open state) state.

Because of the presence of the conserved amino acid residues generally found in most class A GPCRs, one could confidently say AKHRs also belong to the same class A GPCRs family. To select the best template structures that serve as our target templates, the PSI_BLAST [341,344] search tool was used. This gave the two crystal structures, β2AR (5D5A.1A) and rhodopsin (2X72.1A), available in the Protein Data Bank [345,346], as the best templates. These selected templates were used as target templates for sequence alignment of Sarcr-AKHR, using Clustalw2 available at https://www.ebi.ac.uk/Tools/msa/clustalw2[347,348].GMQE (Global Model Quality Estimation) was used for quality estimation. This combines properties from the target–template alignment and the template search method. Upon achieving an appropriate alignment, homology modelling was then conducted using the SWISS-MODEL server. This is a web-based service strictly dedicated to protein structure homology modelling. SWISS-MODEL server executes virtual protein structure modelling by satisfying spatial restraint[349].

3.2.4 **Constructing the Model**

Both models were built based on the target-template alignment using ProMod3 installed in the SWISS-MODEL server. Coordinates that are conserved between the target and the template were copied from the template to the model. Insertions and deletions were remodelled using (SWISS-MODEL Template Library (SMTL) version 28/2/2019, PDB release 22/2/2019) the fragment library. Sidechains were then rebuilt. Finally, the geometry of the resulting model was regularised by using an opls force field. Loop modelling was done with ProMod3 alongside PROMOD-II as an alternative model all available on the SWISS-MODEL server. The models were viewed and analysed using the programs PYMOL[350].

3.3.0 Results and Discussion

3.3.1 Results

3.3.2 Analysis of primary sequence

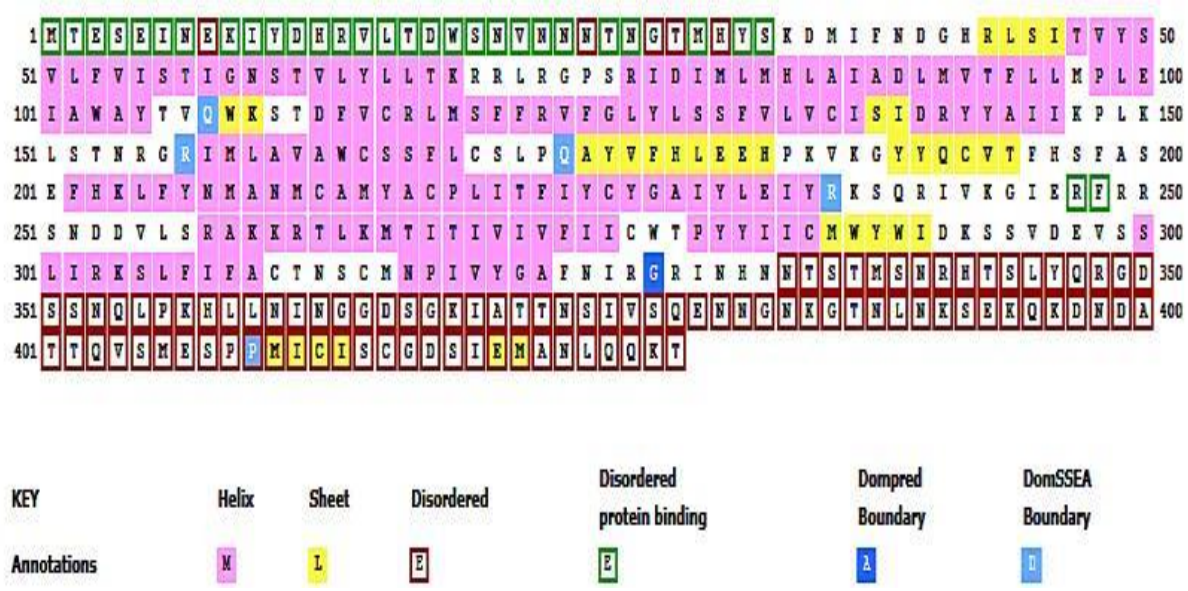


Figure 3.1: Primary amino acid sequence of Sarcr-AKHR submitted for secondary structure prediction to PSIPRED

Primary structure analysis of the 429 amino acid residues (see **Figure 3.1**) indicated that not all the amino acid sequence residue could be used for the construction of the 3D model. To understand how the transmembrane topology and the interactive receptor features might look, a 2D Protter diagram was used (**Figure 3.2)**[351].

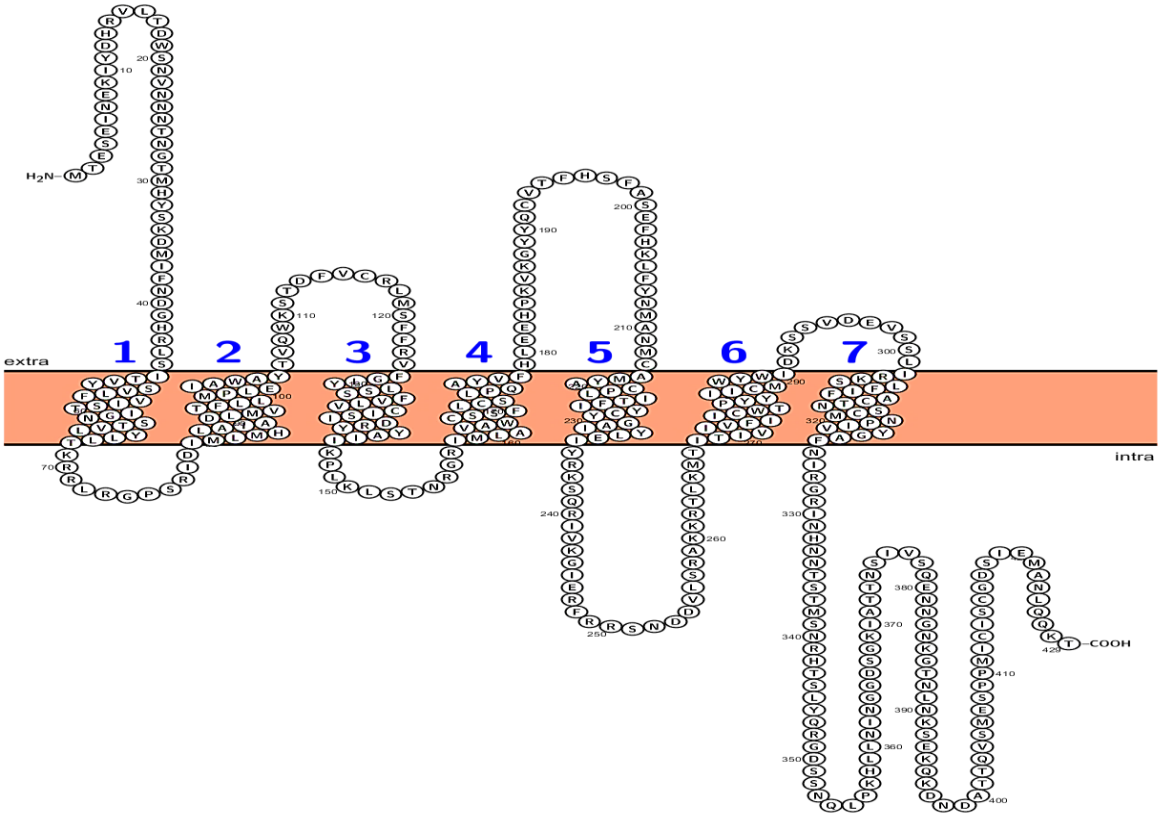


Figure 3. 2: A 2D interactive receptor diagram showing the overall transmembrane topology of Sarcr-AKHR. The numbering 1-7 indicates the transmembrane helices. In dirty brown, is the core and water-lipid interface regions of the lipid membrane.

The 2D interactive receptor diagram for Sarcr-AKHR has been carefully studied to understand and make sure the presence of the disordered part of the protein does not affect any of the conserved amino acid residues. The presence of these conserved, amino acid residues is vital when any 3D model is constructed.

Having understood the primary structure analysis and the 2D interactive receptor diagram, three programs were employed to predict the secondary structure and transmembrane region of Sarcr-AKHR accurately. The three programs are PSIPRED 4.0[352] (**Figure 3.3**), MEMSAT-SVM [343,352] (**Figure 3.4**) and MEMSAT-SVM topology predictions (**Figure 3.5**)[342,353]. The result indicated that only 340 residues make up the seven transmembrane domain (7TMs). MEMSAT-SVM is a highly accurate predictor of transmembrane helix topology because it has

the ability to discriminate between signal peptides and identify the cytosolic and extracellular loops.

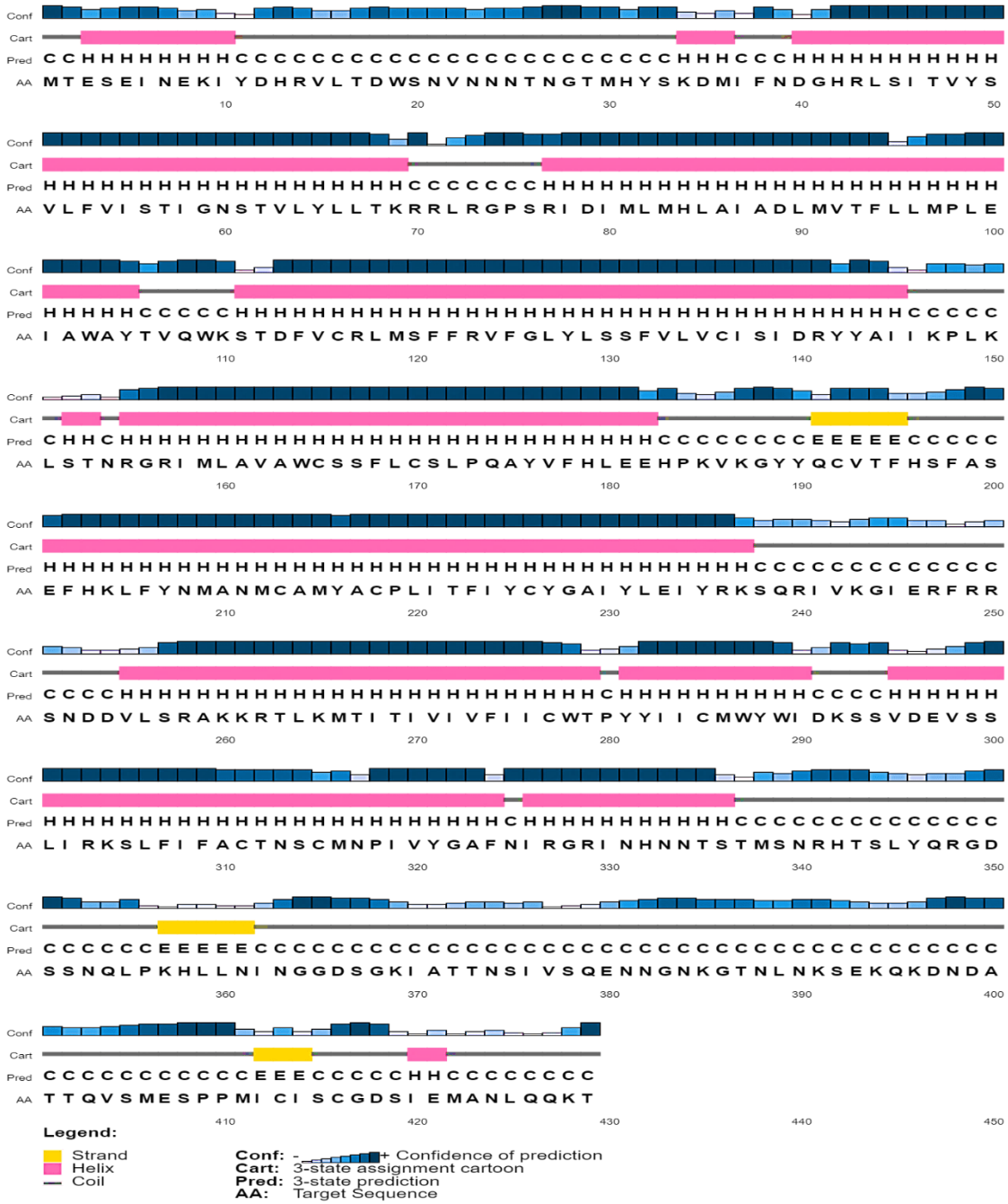


Figure 3.3: PSIPRED predicted secondary structure of Sarcr-AKHR H= helix (in yellow), S = strand (in pink) and C =coil (ash)

The predicted result shows that the primary amino acid of Sarcr-AKHR when fed into the PSIPRED (**Figure 3.3**) and MEMSAT-SVM (**Figure 3.4**) programs, further established that the amino acid sequences are distributed into seven transmembrane (TMs) helical bundles. Also, they possess three loops connected to the helices from the extracellular region, making up the N-terminus and another three loops connected to the helices from the intracellular region making up the C-terminus. To further confirm this prediction, topology analysis was conducted see **Figure 3.4.**

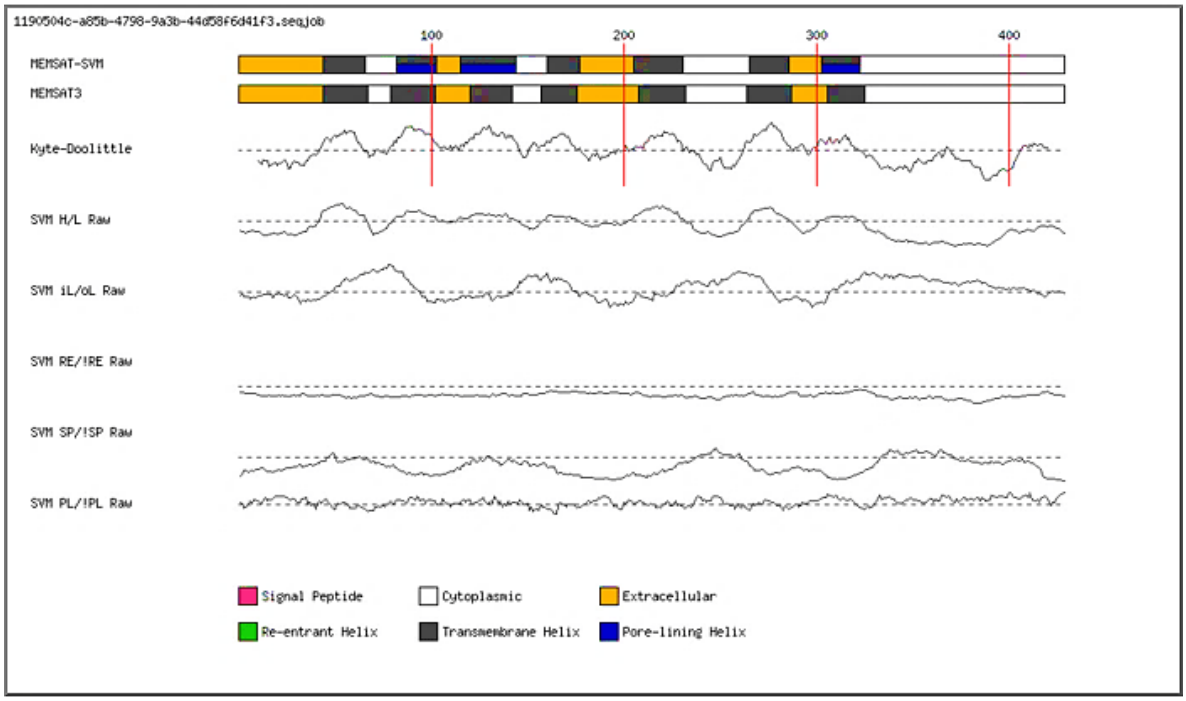


Figure 3.4: Schematic diagram of the MEMSAT3 and MEMSATSVM predictions for the query sequence of Sarcr-AKHR. Traces indicate the RAW outputs for the prediction SVMs. Dashed lines indicate the prediction threshold. PL: Pore lining residue SP: Signal peptide residue RE: Re-entrant helix residue iL/oL & H/L: Helix prediction.

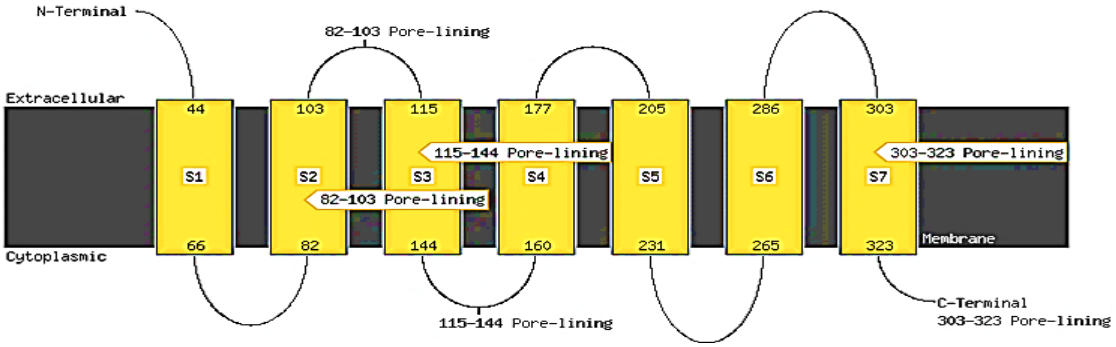


Figure 3.5: Prediction of transmembrane helices and the topology analysis of Sarcr-AKHR. The helices are represented in yellow and labelled S1-S7, the membrane (black) and the loops (thin black line) starting from the N-terminus (the extracellular region) and terminating at the C-terminus (the intracellular region)

The topology analysis predicted the helices as follows, H1(44-66), H2 (82-103), H3 (115- 144), H4 (160-177), H5 (205-231), H6 (265-286) and H7 (303-323) see **Figure 3.5**. The pore-lining was found to be at residues 82-103, 115-144, and 303 -323. The extracellular region comprises the primary amino acid residue at 44, 103, 115, 177, 205, 206, and 303 while the cytoplasmic region comprises residues 66, 82, 144, 160, 231, 265 and 323.

3.3.3 Homology modelling

Two templates were selected for homology modelling, the crystal structure of beta 2 adrenergic receptor (5D5A.1A) and rhodopsin (2X72.1A). These two templates were selected because the rhodopsin crystal structure has a covalently bound ligand and, as such, represents the active state of the receptor. The beta-2-adrenergic receptor (β2AR), on the other hand, does not have a bound ligand, and so represents the inactive state of the receptor. Clustalw2 was used to align the Sarcr-AKHR sequence to the templates as shown in **Figure 3.6**

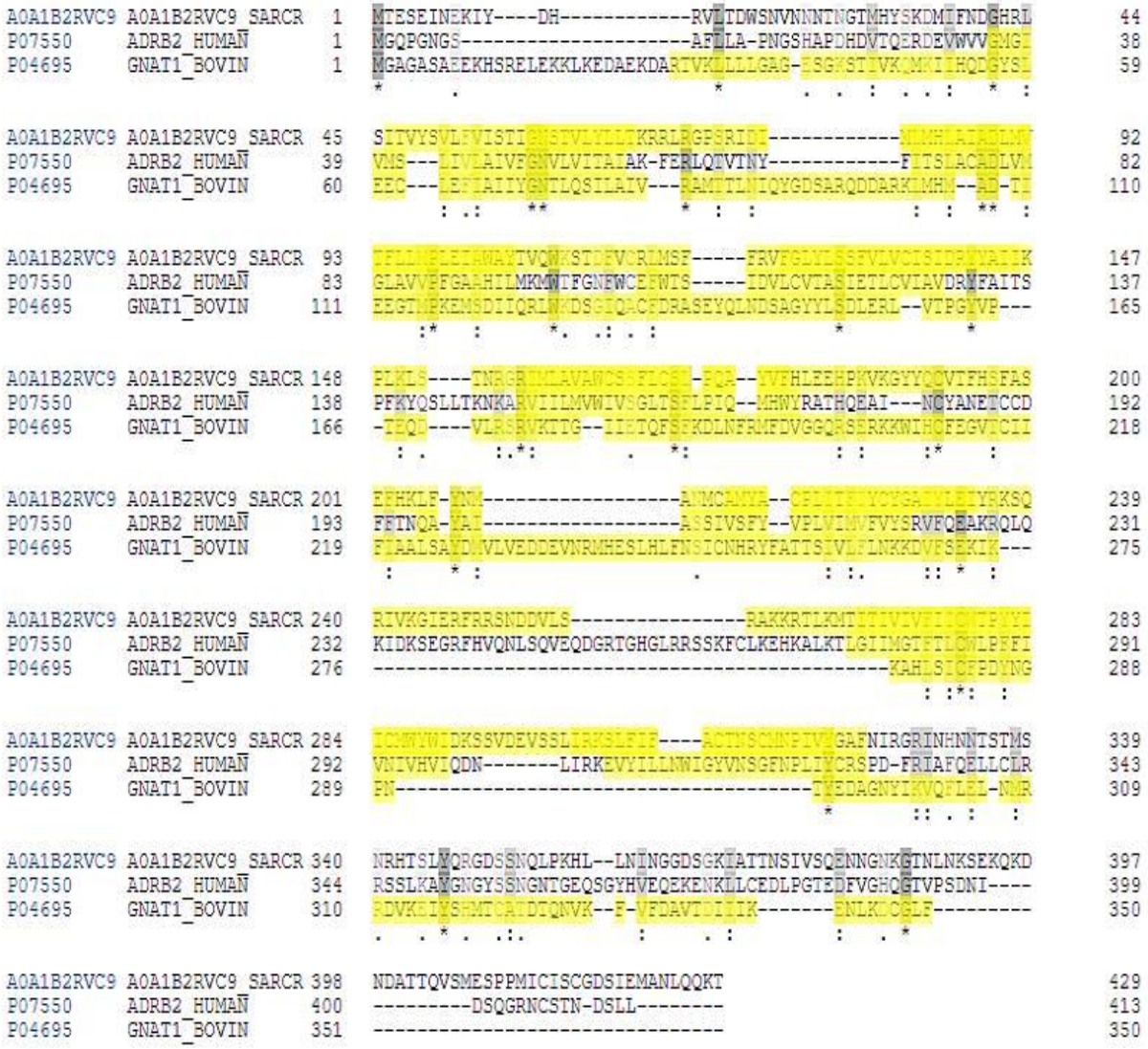


Figure 3.6: Clustalw2 sequence alignment of Sarcr-AKHR and the two selected target templates β2AR, (5D5A.1A / P07550), and rhodopsin (2X72.1A / P04695). The sequence similarity is highlighted in a black and white colour while the transmembrane domain highlighted in yellow.

The alignment (Table 3.1) showed a sequence identity ranging from 16% to 45% and sequence similarity between 52% and 88% for the different helices (helices 1-7). Sequence alignments of the transmembrane residues of Sarcr-AKHR, with those of the target templates β2AR, (5D5A.1A) and rhodopsin (2X72.1A) (see **Figure 3.7**), displayed relatively low sequence percentage identity but possessed high sequence similarity (see Table 3.1). High sequence similarity and low sequence percentage identity are not new to class A GPCRs superfamily,

because, with high percentage similarity, GPCRs / AKHRs do possess similar structural and functional features[298].

Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	SCOP Codes	Structure	CATH Entry
CERT	96.782	9e-09	-639.6	-3.1	372.0	284	304	429	5d5a		CATH Summary
CERT	93.886	2e-08	-469.0	13.5	461.0	321	348	429	Zx72		CATH Summary
CERT	90.390	4e-08	-460.8	11.1	428.0	325	350	429	4bu0		CATH Summary
CERT	90.259	4e-08	-553.5	12.6	410.0	320	363	429	1u19		CATH Summary
CERT	88.522	6e-08	-544.4	13.3	412.0	265	278	429	2z73		CATH Summary
CERT	82.355	2e-07	-544.8	24.8	412.0	299	454	429	4ea3		CATH Summary
CERT	78.793	6e-07	-515.1	4.7	314.0	277	429	429	4grv		CATH Summary
CERT	77.213	8e-07	-550.8	14.7	337.0	269	289	429	4ej4		CATH Summary
CERT	76.940	9e-07	-458.0	4.3	313.0	286	390	429	2rh1		CATH Summary
CERT	76.642	9e-07	-619.1	7.4	284.0	271	303	429	4ib4		CATH Summary

Figure 3.7: The statistical result from Psi-blast of Sarcr-AKHR illustrating the common ancestry. The best two templates with net score 96.782 and 93.886 were selected for the construction of the active and inactive model of Sarcr AKHR.

Table 3.2: Comparison of sequences of AKHR, Rhodopsin, and β2AR helices.

Helix	% Sequence identity with		% Sequence similarity with	
	β_2AR (β_{2flf})	rhodopsin (Rhodflf)	β_2AR(β_{2flf})	rhodopsin (Rhodflf)
1	18	23	61	62
2	25	16	58	52
3	29	30	68	60
4	28	40	76	80
5	23	17	65	60
6	34	23	83	74
7	43	45	88	88

The TM helix segments of both β2AR and rhodopsin show high sequence similarity; this could be the reason for their structural similarity[354,355].

The highly conserved amino acid residues that are common to Class A GPCRs [190,223,356] are also present in the constructed models, see **Figure** 3.8.

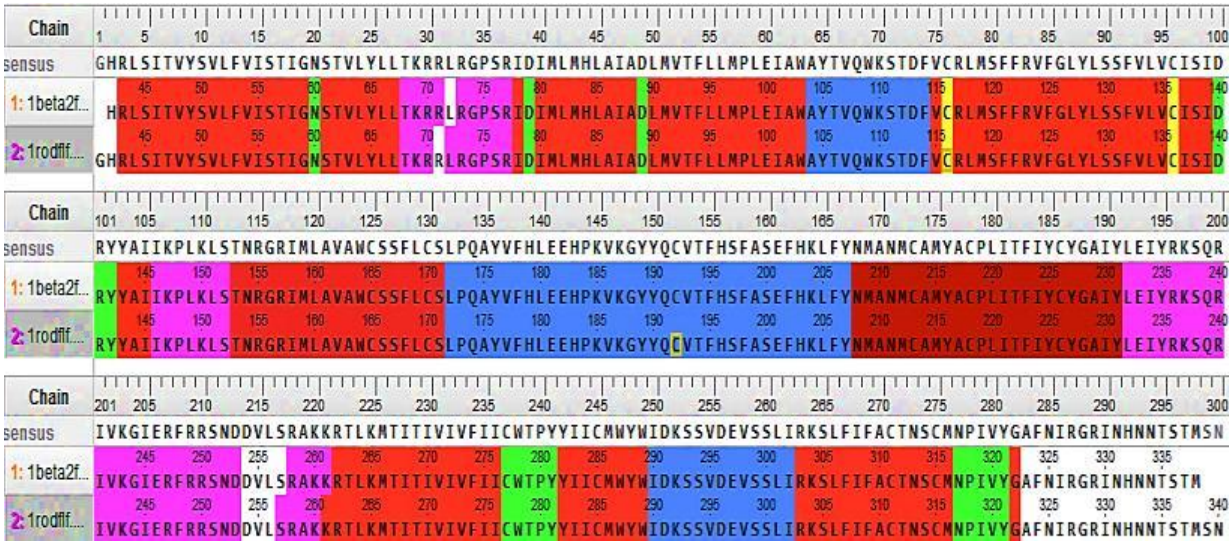
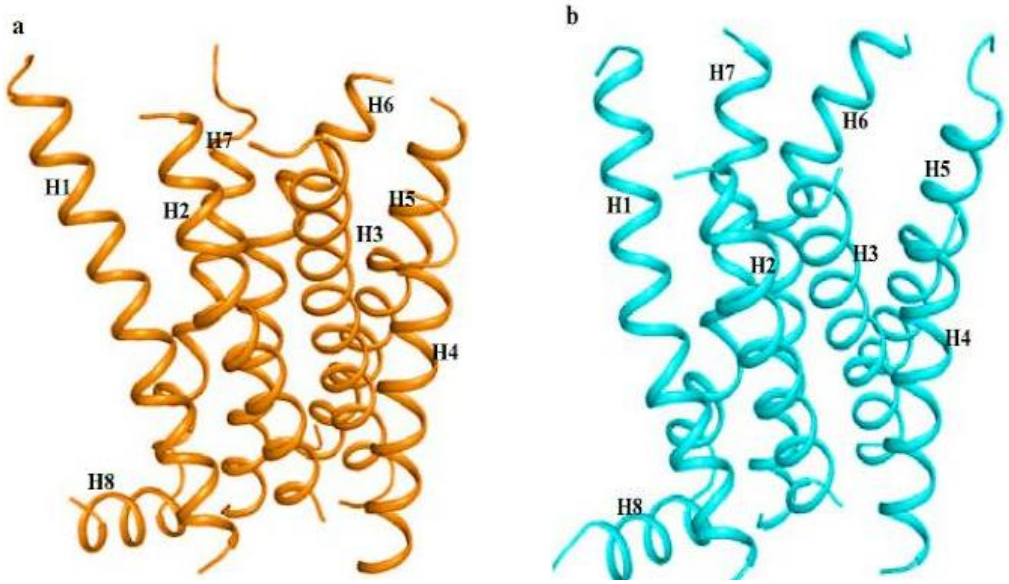


Figure 3.8: Schematic diagram representing an alignment of both models constructed from both rhodopsin and beta2 androgenic receptor of Sarcr-AKHR. Red indicates residues in the seven transmembrane helices, blue shows the extracellular regions. In contrast, the intracellular regions are represented in purple, yellow denotes region cystine ionic lock, and green shows the highly conserved residues.

Having established the similarities and alignments, the Swiss-Model server [330] was used to build two homology models of the 7 TM bundles of Sarcr-AKHR. These models were labelled β2flf-model (**Figure 3.9 a)** and Rhodflf-model (**Figure 3.9 b)**

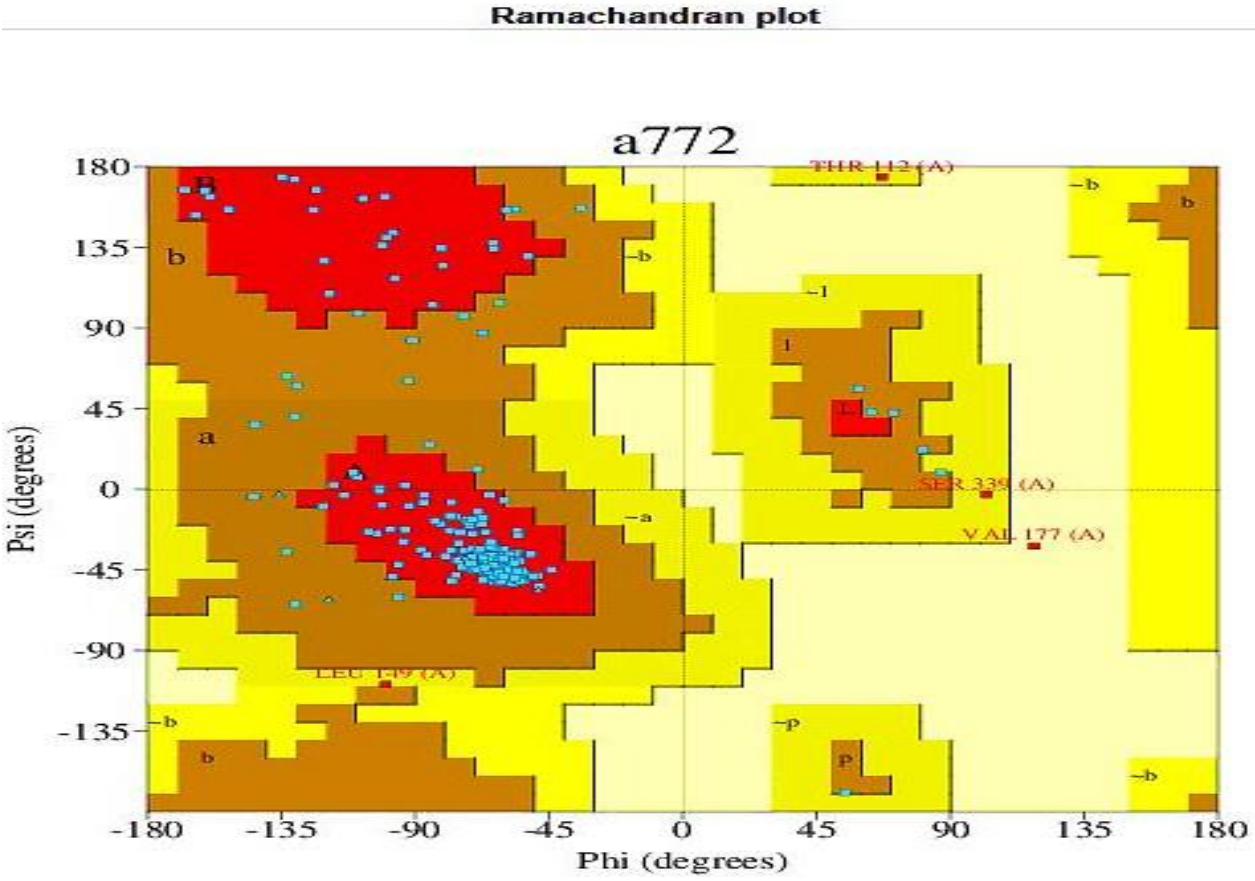


(a) **(b)**

Figure 3.9: The constructed seven transmembrane helices of the Sarcr adipokinetic hormone receptor. a) β2flf-model seven transmembrane helices of Sarcr -AKHR and b) Rhodflf-model seven transmembrane helices of Sarcr-AKHR.

The overall structure of the constructed models (**Figure 3.9**) are alike, especially the anticlockwise arrangement of the transmembrane helices. They both possess an eighth intracellular helix that is parallel to the cell membrane. Also, both structures have a tilted TM3, and the shortest transmembrane in both models is TM4.

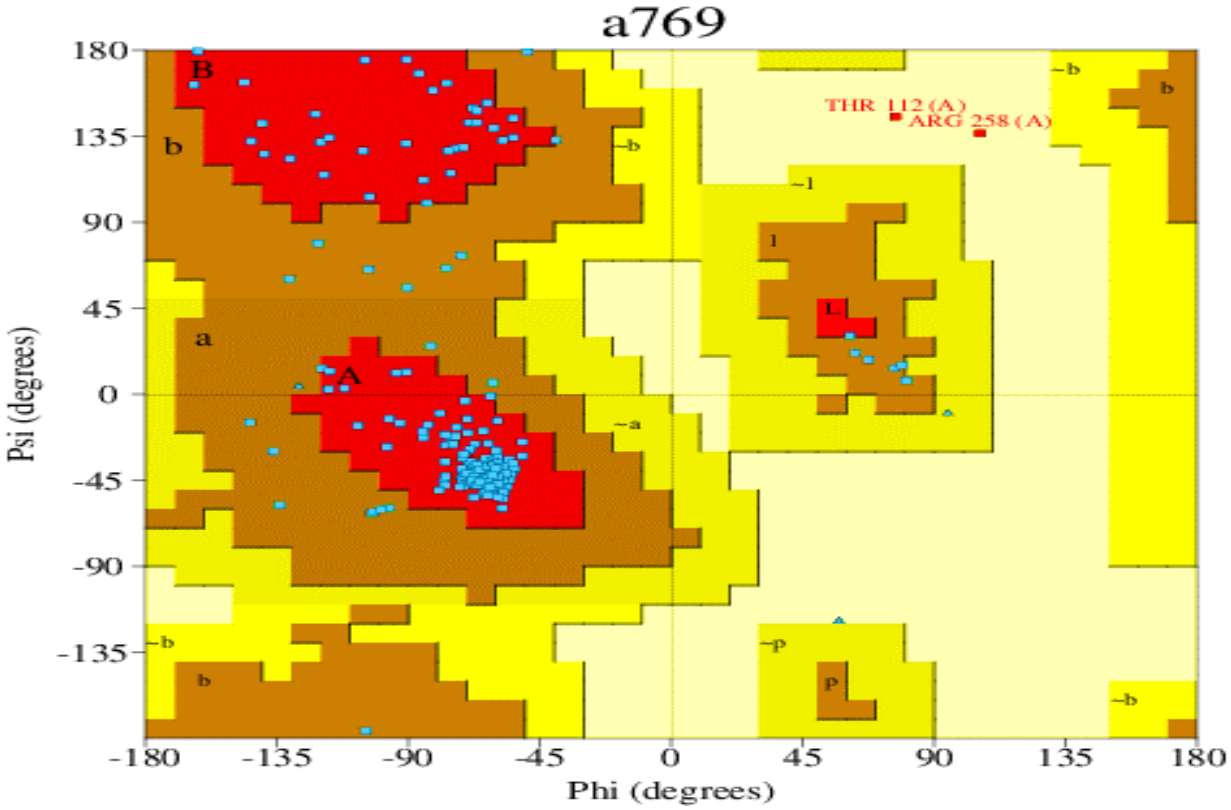
(a)



1. Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	252	89.7%*
Additional allowed regions	[a,b,l,p]	25	8.9%
Generously allowed regions	[~a,~b,~l,~p]	3	1.1%
Disallowed regions	[XX]	1	0.4%*
Non-glycine and non-proline residues		281	100.0%
End-residues (excl. Gly and Pro)		1	
Glycine residues		10	
Proline residues		8	
Total number of residues		300	

(b)



		No. of residues	%-tage
Most favoured regions	[A,B,L]	254	91.4%
Additional allowed regions	[a,b,l,p]	22	7.9%
Generously allowed regions	[~a,~b,~l,~p]	0	0.0%
Disallowed regions	[XX]	2	0.7%*
Non-glycine and non-proline residues		278	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		9	
Proline residues		8	
Total number of residues		297	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

Figure 3.10: Ramachandran plots of a) β2flf-model and b) Rhodflf-model.

The quality of the constructed models was assessed with evaluation programs, ERRAT[357] and PROCHECK[358]. Ramachandran plots of the two models are shown in **Figure 3.10**. The result shows that most of the primary amino acid residue of the constructed models are either in allowed or favoured regions. For the β2flf-model, Val177 had a φ angle 10^0 outside the

generously allowed region, while the Rhodflf-model had two residues, Thr112 andArg258, had ψ torsion angles 20^0 outside the allowed region. These problems were corrected during the molecular dynamics of the models (see appendix for the Ramachandran plots) [358]. The ERRAT program was used to check the quality of constructed models. A score of 95.73 and 89.73 for the β2flf-model and Rhodflf-models, respectively as the acceptable score range for a high-quality model is > 50 [357]. As such, we conclude that these two models are of the high-quality model and are acceptable.

3.4.0 Structural comparison of models

The constructed models (β2flf-model and Rhodflf- model) can be compared to their template GPCR structures. An overlay of the β2flf -model with its template (see **Figure 3.11 a)** gave an RMSD of 3.1Å for superimposition of Cα atoms of the whole molecule and 2.04Å for the helix bundles. Also, an overlay of the Rhodflf-based model with its template (see **Figure 3.11 b)** gave an RMSD of 3.3Å for superimposition of Cα atoms of the whole molecule and 2.03Å for the helix bundles. Overall, both target templates have longer TM helices when compared to the two constructed models, with β2AR having 392 amino acid residue [331] as opposed to the constructed β2flf-based model which has 338 amino acids. Also, the rhodopsin template model has 442 amino acid residues [224] as against the constructed Rhodflf – based-AKHR model, which has 340 amino acid residues.

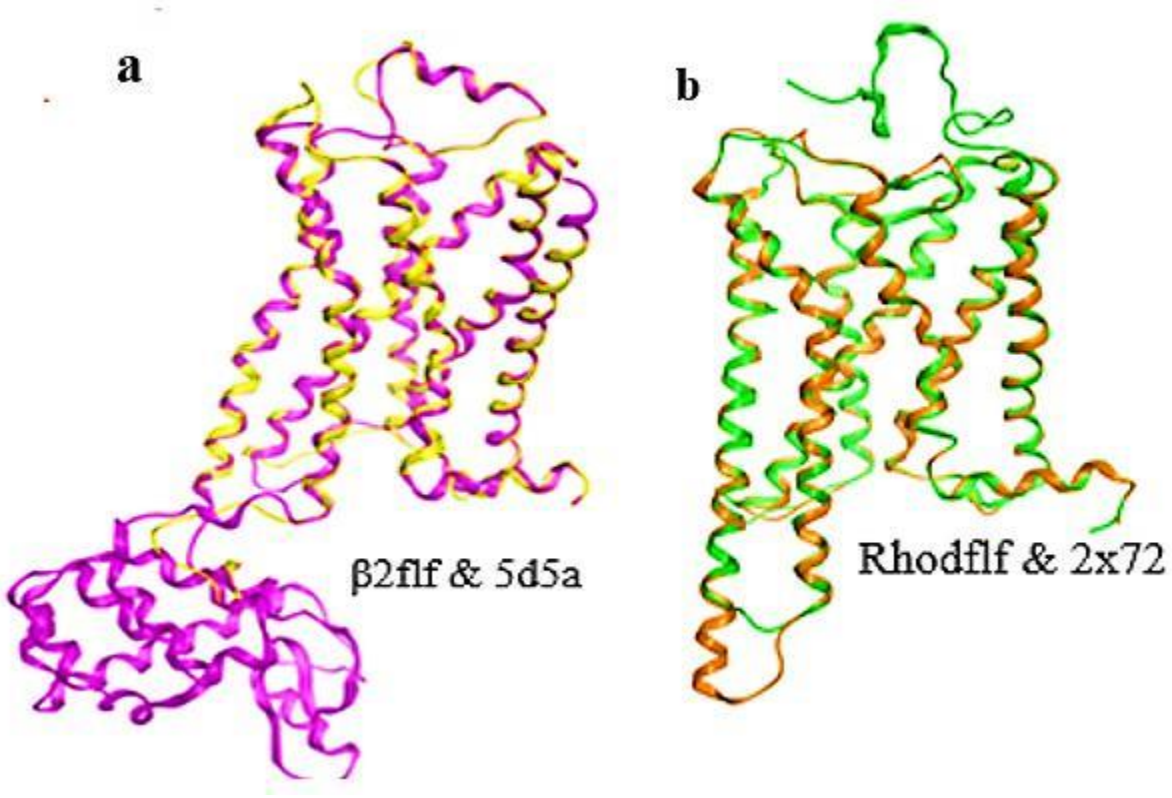


Figure 3.11: An overlay of β2flf-based-AKHR (the constructed model, in yellow) and 5d5a (the template used for constructing the model, in purple). b) shows an overlay of Rhodflf-based-AKHR (the constructed model, in green and 2x72 (the template used for constructing the model, in brown).

Both constructed models have the CWxP(Y/F) (CWTPY) motif in TM6. For β2AR, the CWxP(Y/F) motif is CWLPF, while for rhodopsin, it is CWLPY. This is the only significant difference noticed as regards the conserved residue. This motif is significant to receptor activations and will be discussed later.

3.4.1 Evaluation of the helix bundles

In both models, the transmembrane helices have an anticlockwise helical arrangement, when viewed from the extracellular side. This anticlockwise arrangement is found in most class A GPCRs [359]. Experimental data supported this anticlockwise arrangement, which is consistent

with the arrangement found in neurokinin 1 (NK-1) and κ-opioid receptor [359]. The helical bundles are stabilised by interhelical interactions comprising the highly conserved residues. [223]

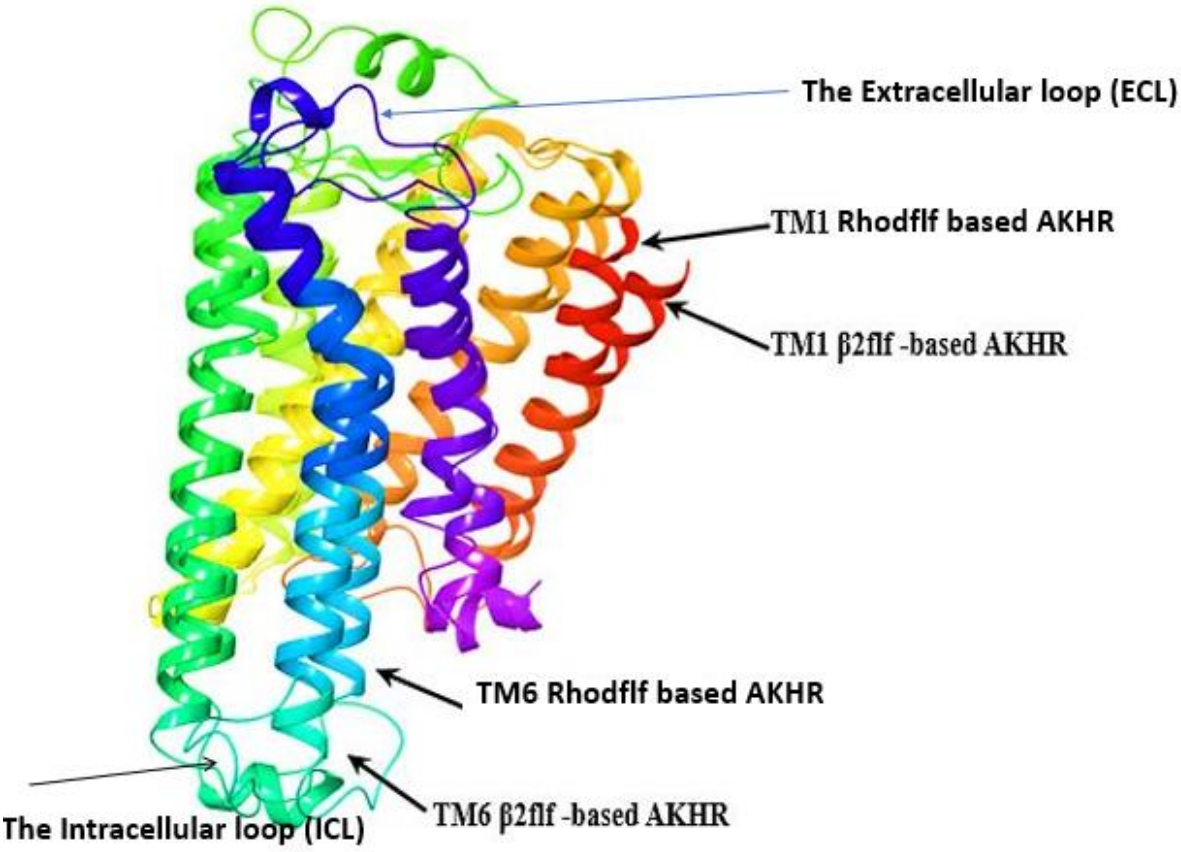


Figure 3.12: The superimposition of the β2flf-model structure and Rhodflf-model structure of Sarcr-AKHR.

Superimposition of the two constructed 3D models gave a root mean square deviation (rmsd) of 3.18 Å for superimposition of Cα atoms of the whole molecule and 2.1 Å for the helix bundles. TM6 and TM5 differ from each other with respect to length with both helices extending by 4-3 residues into the cytoplasm. This observation is consistent with studies conducted on the accessibility of nitroxide labels fixed to ECL3 joining TM6 and TM5 [360]. Both structures possess a tilted TM3 conformation. This tilting of TM3 is also found other class A GPCRs model[93,98,195,361,362]. The orientation of the seven transmembrane helices is subtly different in the two models, which leads to differences in the intracellular and extracellular

domains. Both the β2flf-based and Rhodflf-models have a disulphide bridge, commonly found in GPCRs, between Cys116$^{3-25}$ and Cys192.

The extracellular end of TM1 of β2flf-model bends away from the helical bundle (**Figure 3.12)** with an angle of deviation 38.9º with a distance of 4.35 Å between TM1 of β2flf-model and Rhodflf-model. This gives a more open structure to the extracellular domain of the β2flf-model as compared to the Rhodflf-model. At the same time, the intracellular end of TM6 a distance of 6.8 Å (**Figure 3.12**) was recorded for both β2flf-model and Rhodflf-model. This gives a more closed structure to the intracellular domain of the β2flf-model. Thus, the β2flf-model is described as the open or inactive model, while the Rhodfrf-model is termed the closed or active model[239]. The closed conformation of the Rhodflf-model is similar to the rhodopsin crystal structure, where the closed nature protects cis-retinal from hydrolysis [96,98]. Since ligand binding takes place on the extracellular domain [363], the closed nature of the Rhodflf-model prevents the ingress of the ligand. On the other hand, the open nature of the β2flf-model allows the ligand-free access to the receptor-binding pocket.

3.4.1.0 Overview of transmembrane helical interaction and molecular switches

A critical feature of most class A GPCRs is the existence of substantially preserved molecular switch motifs. The presence of these switches is essential for stabilising any GPCR either in the close (active) or open (inactive) state. To properly explain the switch the Ballesteros–Weinstein numbering scheme will be used (where the most conserved residue from each helix is given high priority and is pined at 50 that is Helix 1Asn (1-50) Helix 2Asp (2-50) Helix 3Arg (3-50) Helix 4Trp (4-50) Helix 5Pro (5-50) Helix 6 Pro (6-50) and Helix 8 Pro (8-50). Hence, the numbering of any amino acid from its primary sequence shall be noted as follows if the

counting is towards N terminal from the Most conserve residue of each Helix, one is expected to subtract from the fixed 50 and if the counting leads towards the C terminal addition is made. In summary, Thr305$^{4\text{-}56}$ means Thr305 (primary-AA sequence number from the alignment table) 4-56 (Helix 4 and it is 6 places after the most conserved residue Trp)[364,365]. In Sarcr-AKHR, a 3-6 lock exists between Arg262 $^{6\text{-}31}$ on TM6 and Tyr 142$^{3\text{-}51}$ on TM3. In the inactive state, Arg262$^{6\text{-}31}$ points away from TM3, and so no interaction is possible (**Figure 3.13 a**). Upon activation, however, TM6 twists moving Arg262$^{6\text{-}31}$ so that it now points toward Tyr142$^{3\text{-}51}$ (**Figure 3.13 b**). This interaction locks the receptor in the active state. **Figure 3.13 b** also shows that, in the active state, Arg262$^{6\text{-}31}$ on TM3 interacts with Glu 246 of ECL2. This loop closes over the binding pocket after ligand binding.

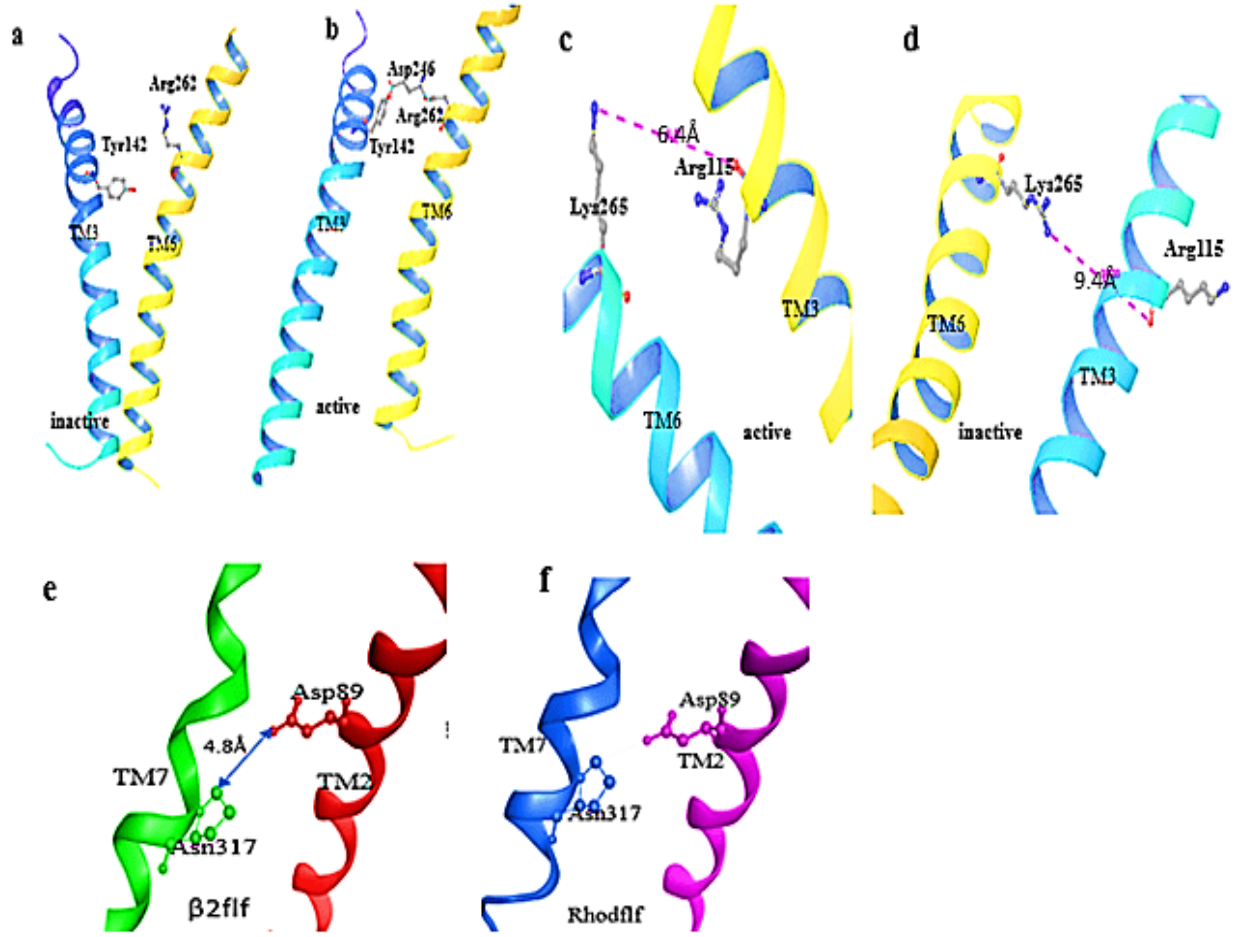


Figure 3.13: a) Molecular switch interactions (inactive) in the β2flf based between Tyr 142 in TM3 and Arg262 in TM6. b) molecular switch interactions (active) in the Rhodflf model between Tyr 142 in TM3 and Arg262 in TM6. c) molecular switch interactions (active) in the Rhodflf model between Lys265 in TM 6 and Arg115 in TM3. d) molecular switch interactions (inactive) in the β2flf model between Lys265 in TM 6 and Arg115 in TM3. e) molecular switch, interactions (inactive) in the β2flf model between Asp89 in TM2 and Asn317 in TM7. f) molecular switch interactions (active) in the Rhodflf model between Asp89 in TM2 and Asn317 in TM7.

In the Rhodflf-model, there is an ionic interaction between Lys265$^{6\text{-}35}$ in TM6 and the positively charged side chain nitrogen atom of Arg115$^{3\text{-}50}$ in TM3 (**Figure 3.13 c**). This interaction does not exist in the β2flf-model (**Figure 3.13 d**). These switches are found to be the same as those described for the AKH-receptor of *A. gambiae*, *Rhodnius prolixus*, and *Tribolium castaneum* [231,366].

Class A GPCRs have a conserved aspartate on TM2 and an asparagine on TM7. In the Rhodflf-model structure, there is an intramolecular interaction between Asp89 and Asn317 (**Figure 3.13 f**). In the β2flf-model (**Figure 3.13 e**), the helices shifted such that these two residues are now 4.8Å apart and can no longer interact [239]. Experimental data backs the importance of these intramolecular interactions because of the vital role they play during receptor activation, basal activity and conformation flexibility reduction [239].

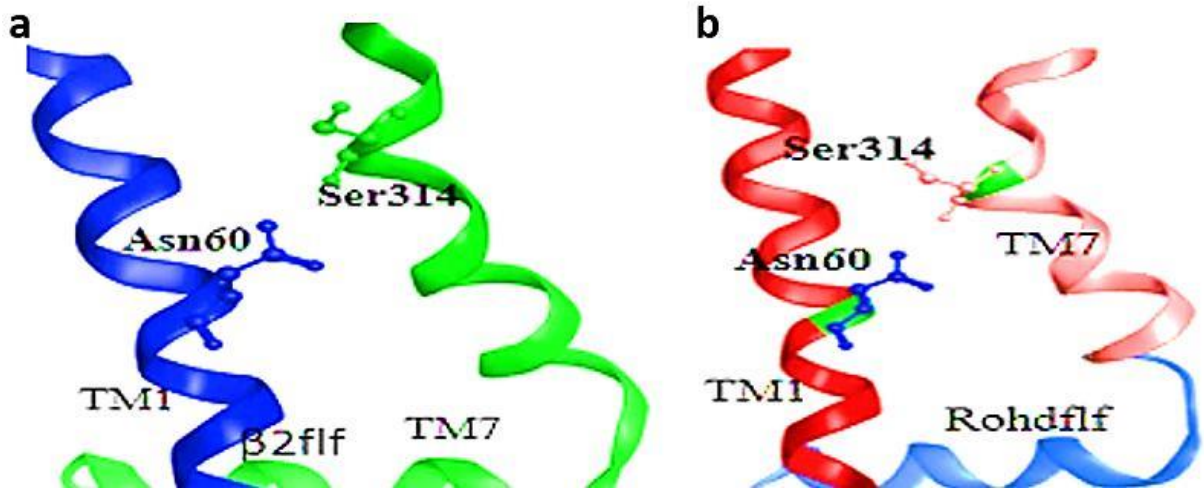


Figure 3.14: Polar interactions a) Rhodflf-based AKHR, the molecules are far apart making it impossible for the polar interactions between TM1 Asn60 and TM7 Ser314 to occur. b) β2flf-based AKHR showing presence of polar interactions between TM1 Asn60 and TM7 Ser314 because of the proximity between the two residues.

Talking about stability, in both β2flf-model AKHR and Rhodflf-model AKHR, there is polar interaction between TM1 Asn60$^{1\text{-}50}$ and TM7 Ser314 $^{7\text{-}45}$ (**Figure 3.14)**, which stabilises both structures. The β2flf-model AKHR TM6(Ser 238$^{6\text{-}70}$ and Tyr325) H-bond TM5 Lys 260$^{5\text{-}29}$ (**Figure 3.15 a**), while in Rhodflf-model this interaction is absent (**Figure 3.15 b**)

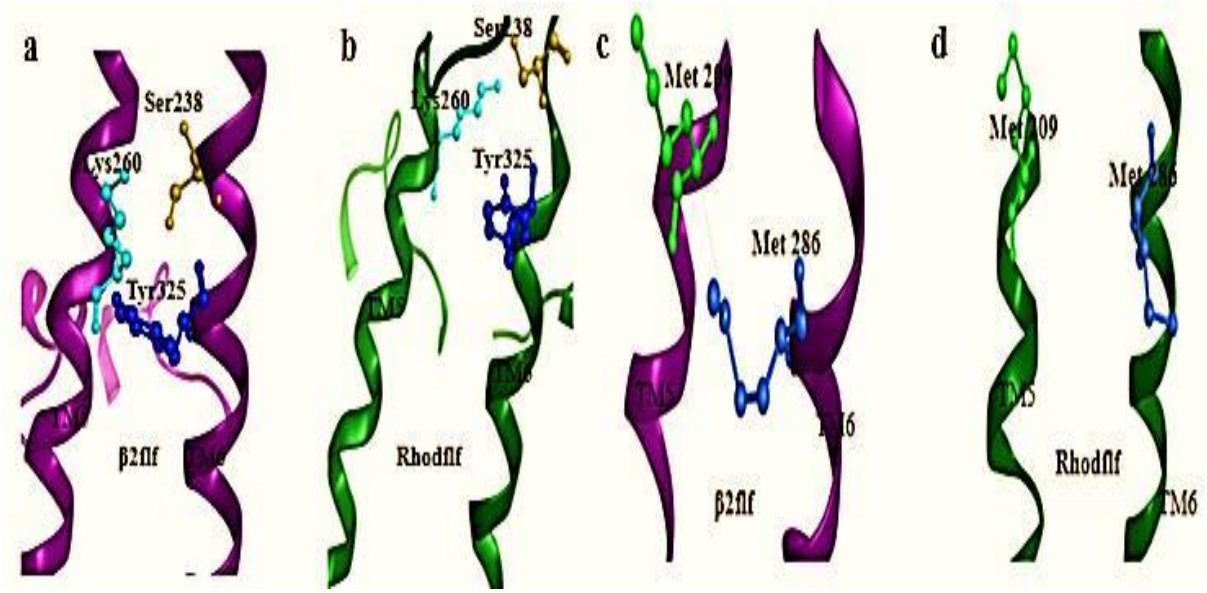


Figure 3.15: a) H-bond interaction between TM5 Lys260 and TM6 (Ser238 and Tyr325) for the open structure (β2flf-model AKHR). b) the absence of H-bond interaction between TM5 Lys260 and TM6(Ser238 and Tyr325) for the Rhodflf-model AKHR. c) helical interaction between TM5 Met 209 and TM6 Met286. d) absence of helical interaction between TM5 Met 209 and TM6 Met286

Another intramolecular interaction between Met 209[5-39] in helix 5 and Met 286 [6-56] in helix 6 (see **Figure 3.15 c & d**) was also identified in the open structure but absent in the closed structure. Dror et al., 2009 discovered that the inactive β2AR exists in equilibrium between conformations with the ionic lock 'on and off'[367]. These intramolecular interactions are responsible for the reduction in conformational flexibility, basal activity and play a vital role in receptor activation[239]

3.4.1.1 Evaluation of Kinks

Experimental data backs the existence of kinks that initiate the helical twist (bend) at almost the same locations where Tyr, Thr or Met come before Pro[195,368–370]. The TM helices in both open and closed AKHR models have these proline-generated bends or kinks (**Figure 3.16)** [371]. TM2 has a kink between Met 97 [2-58] and Pro 98[2-52] (**Figure 3.16 a**), TM5 has a kink between Tyr216 [5-47]and Pro219 [5-50] (**Figure 3.16 b**). For the β2flf-model, another turn was noticed in

TM4 at Phe168 [4-54] (**Figure 3.16 d**), although not found on the Rhodflf-model structure (**Figure 3.16 e**). From the extracellular region, the kink in TM 6 instigates a twist near TM 5. An overlay of the β2flf-model and Rhodflf-model in **Figure 3.16 a, b and c** shows the different effect of these kinks or bends on the two models.

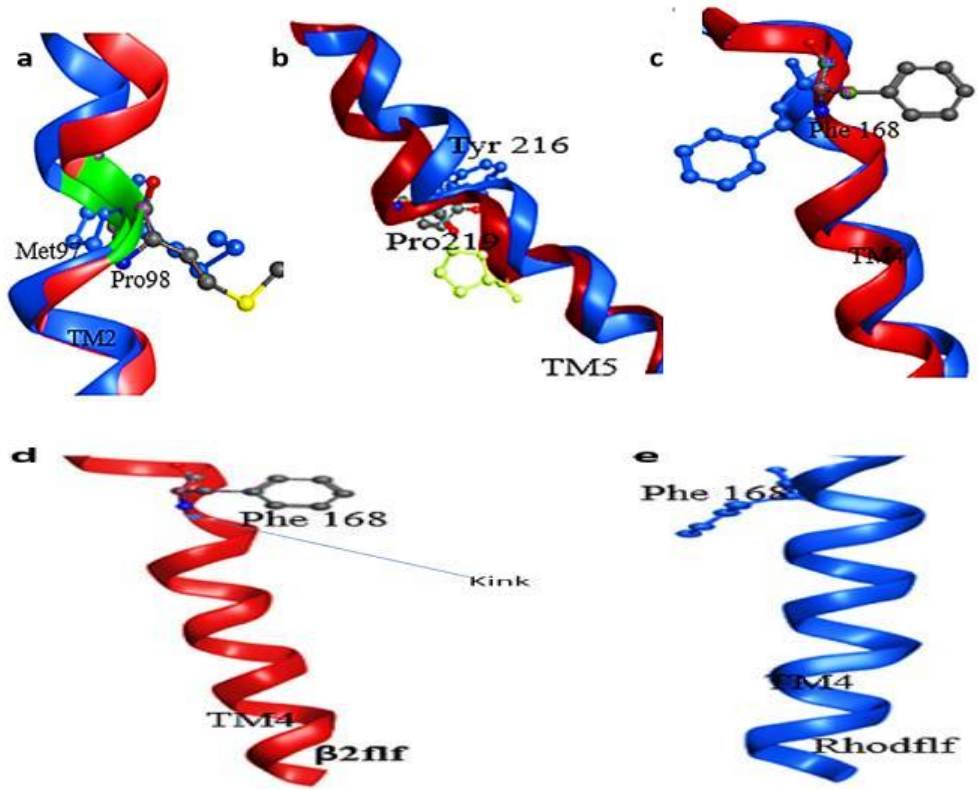


Figure 3.16: a) overlay of 3D models of the β2flf-model and Rhodflf-model structure, showing the presence of kinks in TM2. b) overlay of 3D models of the β2flf-model and Rhodflf-model structure showing the presence of kinks in TM5 and c) overlay of 3D models of the β2flf-model and Rhodflf-model showing the presence of kink in TM4. d) The β2flf-model, showing the presence of kink in TM4. e) The Rhodflf-model structure, showing the absence of kink in TM4. * β2flf = β2flf-model, Rhodflf =Rhodflf-model.

3.4.1.2 Evaluation of the Loop Regions of the Models

Besides the orientation of the transmembrane helices, there are other differences between the two models (**Figure 3.12**). In the extracellular loop I (ECL I) β2flf-model has a β-strand between Ala104 and Gln108, and a small helical turn at the N-terminus, while in the Rhodflf-

model AKHR, there is a β-strand between Glu181 and Glu182, and a small helical turn between His183 and Cys192. ECL2 of β2flf-model has a helical turn between Ala175 and Tyr190. This moves ECL2 away from the binding pocket. At the same time, the N-terminus also moves to give any potential ligand access to the receptor-binding pocket. However, ECL2 of the Rhodflf-model AKHR does not have the helical turn like β2flf-model AKHR. This keeps ECL2 of the Rhodflf-model AKHR in an upright position above the helical bundle. In this way, the N-terminus and ECL2 extend over the transmembrane helices, shielding all possible binding sites and preventing the ligand from gaining access to or leaving the binding pocket.

ICL3 is commonly not characterised in the crystal structures of GPCR. In both the β2flf-model and Rhodflf-model, ICL3 has a helical turn between Ile290 and Lys292. Dror *et al.* identified this turn in β2AR[367]. Moreover, the β2flf-model has a helical turn between Ser293-Ser294, while the Rhodfrf-model has a β-strand between Val295 and Asp296 and a helical turn between Asp297 and Ser300.

Most Class A GPCRs have an eighth helix, which lies parallel to the intracellular membrane surface. This helix is thought to be essential for receptor activation and binding of a G-protein [239]. Most GPCRs intracellular region are stabilised by palmitoylation of Cys, Ser, or Thr on the 8th helix [372–374]. Both the β2flf-model and Rhodflf-models have this 8th helix (the tail) see **Figure 3.9**, which has Cys311, which could be stabilised by palmitoylation.

Since GPCRs are highly flexible, the different structural features, particularly in the extra and intracellular region of the two models, could represent features in different conformations during molecular dynamics and in the active and inactive state of the receptor.

3.4.0 Conclusion

We have successfully constructed two 3D models for the flesh fly AKH receptor, Sarcr-*AKHR.* One model is in an open conformation as the inactive state, while the other has a closed

conformation as the active state of the AKHR. The elucidated structures possess the same features as most class A GPCRs.

Some structural differences can be found around the highly conserved amino acid Asp, Arg and Tyr (DRY) residues. Wherein the Rhodflf-model, posses an ionic interaction between TM3 Arg141$^{3-50}$ and the backbone C=O of Lys264$^{6-35}$ in TM6, this was not found in the open model.

H-bond was identified between TM6(Ser 238$^{6-70}$ and Tyr325) and TM5 Lys 260$^{5-29}$ for the β2flf-model, but no salt bridge was found between TM5 and TM6 for the Rhodflf-model.

A polar interaction was also identified between TM1 Asn60$^{1-50}$ and TM7 Ser314$^{7-45}$ for the closed model but not found in the open model.

In both models, the presence of kinks is found in TM2 (Met 97$^{2-5}$ and Pro 98$^{2-52}$), TM5 (Tyr216$^{5-47}$-Pro219$^{5-50}$) and TM4 Phe168$^{4-54}$

The general conformation and structural features of the β2flf-model AKHR and Rhodflf-model AKHR models can be used to study the binding and activation of Sarcr-*AKH* receptor by Phote-HrTH.

4

Constructing a 3D Model of the Fruit Fly Receptor, *Drome-AKHR*

Summary

The genome of the fruit fly was used to predict and build the three-dimensional (3D) model of the fruit fly adipokinetic hormone receptor (AKHR). Computational homology modelling technique was employed to achieve two standard 3D models. Beta2-adrenergic receptor (β2AR) and rhodopsin crystal structures were used as the target model. Essential for this regard, is to be able to explain the active and inactive conformation of a receptor. The β2AR structure has an open conformation meaning the receptor is inactive, while the rhodopsin structure possesses a closed (active) conformation.

It is understood that that active rhodopsin does not allow the diffusion of a ligand into the binding receptor binding pocket, but the β2AR dose.

The constructed flesh fly AKHR (chapter 3). When compared to the fruit fly AKHR is very similar but different.

4.1.0 Introduction.

Over the years, some pest insects like the fruit fly have gained resistance to the widely-used insecticides [37,38], and the more disturbing, massive decline in beneficial insects like the honey bee has been recorded of late [39,40]. Hence, the need for active biological pesticides that are selective in their activity, negatively affecting the aimed pest species without causing damage

to valuable insect species. The best idea for developing a biological insecticide that will be selective in its action is to target the GPCRs of pest insects[29–33,41,42]. For this to be successful and applicable, research into ligand-GPCR interactions is needed for pest and beneficial insects to find those that are specific to pest insects for further drug development. Thus, AKH and their AKHR/GPCR are considered suitable targets for new insect control agents. However, in order to target the GPCR binding site, the 3D structure of both ligand and receptor needs to be known. Unfortunately only very few 3D structures of most AKHRs are available as data on the constructed 3D molecules and the binding sites will allow the design of novel, species-specific, non-peptidic mimetics, which can block the binding site and hinder the activities of the insect. The published primary sequence of Drome-AKHR was extracted from the Genbank database (accession number AAN10047.1) and used to predict the 3D structure of the fruit fly receptor. The experimental procedure for the construction of the fruit fly AKHR is similar to those in chapter 3, so only results will be presented in this chapter.

4.2.0 Results and Discussion

4.2.1 Results

4.2.2 Analysis of primary sequence

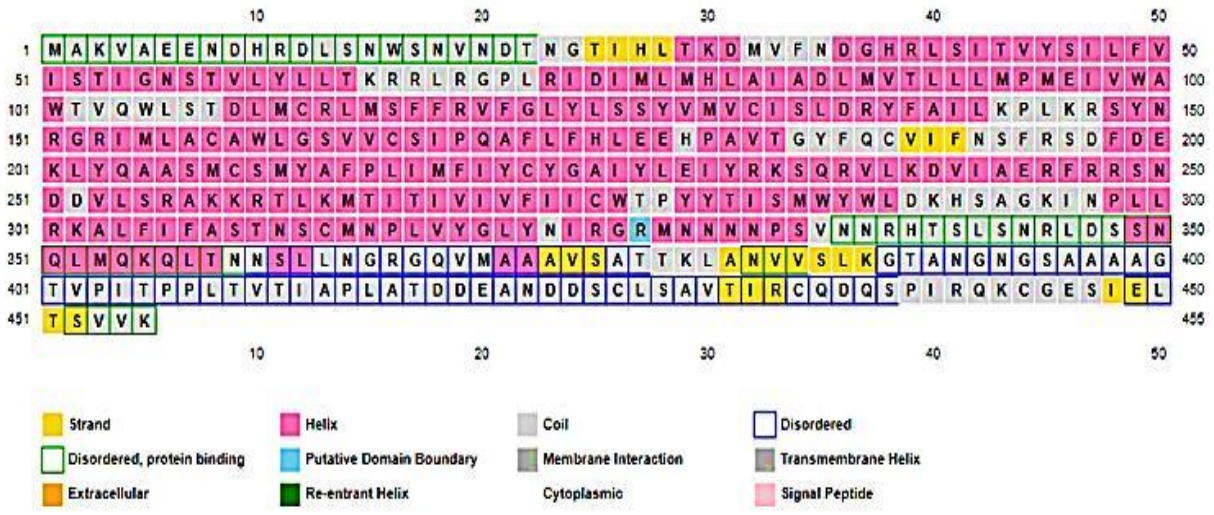


Figure 4. 1: Primary amino acid sequence submitted for secondary structure prediction of Drome-AKHR to PSIPRED.

The Drome-AKHR possess 443 primary amino acid residues (see **Figure 4.1**). The PSIPRED primary structure analysis indicated that not all the amino acid sequence residue could be used for the construction of the 3D model. A 2D Protter diagram (see **Figure 4.2**) was used to further analysed the transmembrane topology, and the interactive receptor features [351].

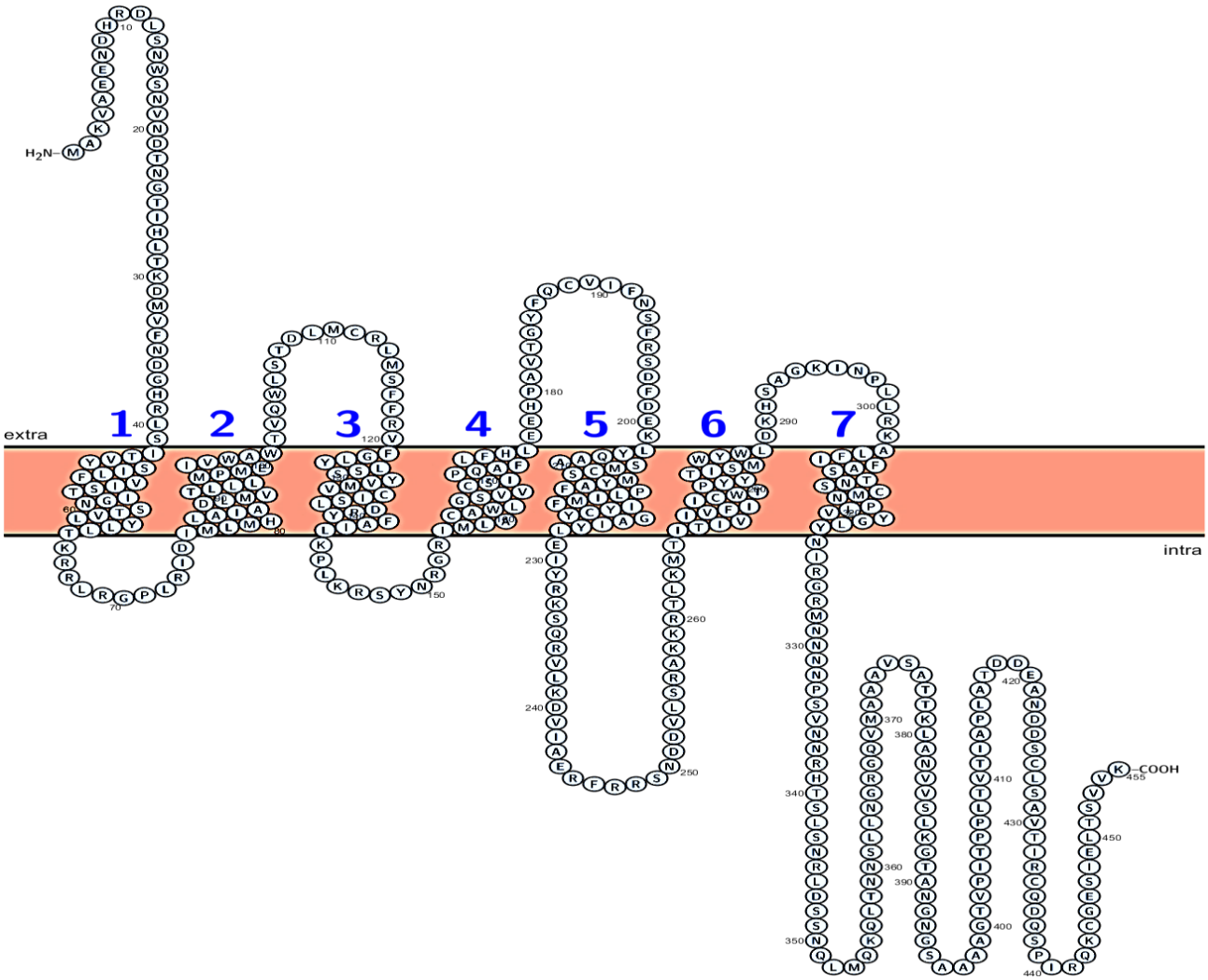


Figure 4. 2: A 2D interactive receptor diagram showing the overall transmembrane topology of Drome-AKHR. The numbering 1-7 indicates the transmembrane helices. In dirty brown, is the core and water-lipid interface regions of the lipid membrane.

The primary importance of the 2D interactive receptor diagram of Drome-AKHR is to enable a proper analysis of Drome-AKHR, and it also helps to avoid the use of any the disordered part of Drome-AKHR during receptor construction. The 2D interactive receptor diagram also helps to ensure the presence of conserved amino acid residues. As the presence of conserved amino acid residues, patterns, and motifs generally found in all Class A GPCRs is vital when any 3D model is constructed. These conserved residues play a vital role via helical interactions to stabilise the helices. Also, they are essential for receptor activation and ligand binding[204,375,376].

Having understood the 2D interactive receptor diagram and the primary structure analysis, three programs, PSIPRED 4.0 (**Figure 4.3**)[327,341,377,378], MEMSAT (**Figure 4.4**)[342,343,353], MEMSAT-SVM topology predictions (**Figure 4.5**)[341,353] were employed to predict the secondary structure and transmembrane region of Drome-AKHR accurately. The result indicated that only 329 residues make up the seven transmembrane domains (7TMs).

Figure 4. 3: PSIPRED predicted secondary structure of Drome-AKHR. H= helix (in yellow), S = strand (in pink) and C =coil (ash) Pred: Predicted secondary structure (H=helix, E=strand, C=coil), AA: Target sequence.

The predicted results from PSIPRED 4.0 (**Figure 4.3**)[327,341,377,378] and MEMSAT (**Figure 4.4**)[342,343,353] further established that the amino acid sequences are distributed into seven transmembrane (TMs) helical bundle. Also, they possess three loops connected to the helices from the extracellular region, making up the N-terminus and another three loops connected to the helices from the intracellular region making up the C-terminus. To further confirm this prediction, topology analysis was conducted using the MEMSAT-SVM topology predictions program(**Figure 4.5**)[341,353].

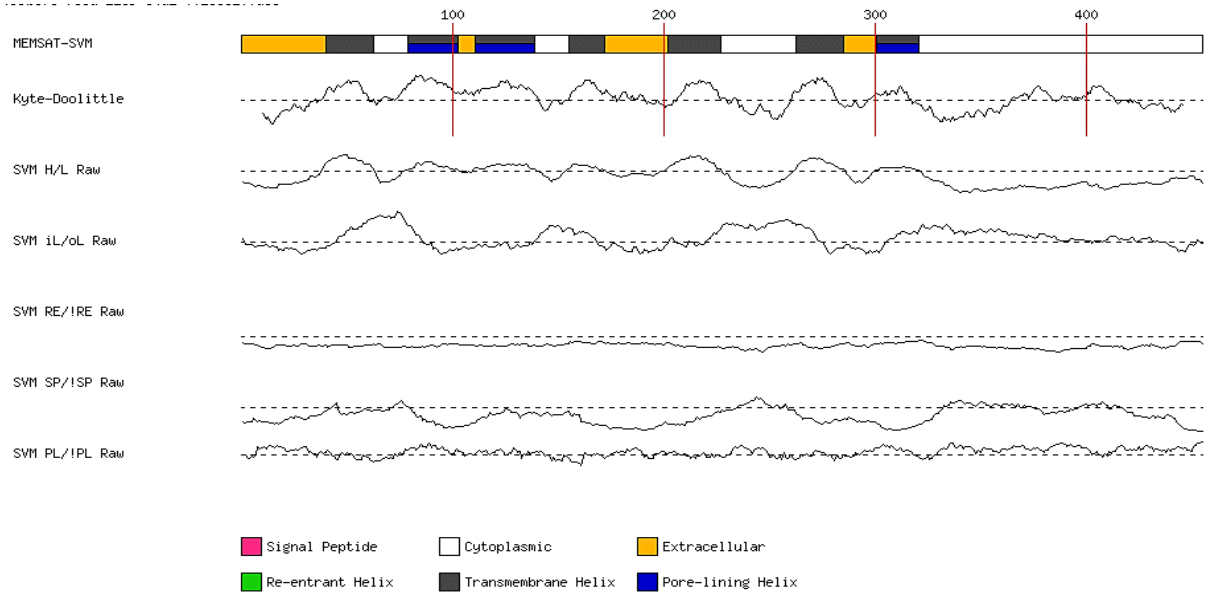


Figure 4.4: Schematic diagram of the MEMSAT-SVM predictions for the query sequence of Drome-AKHR. Traces indicate the RAW outputs for the prediction SVMs. Dashed lines indicate the prediction threshold. PL: Pore lining residue SP: Signal peptide residue RE: Re-entrant helix residue iL/oL & H/L: Helix prediction.

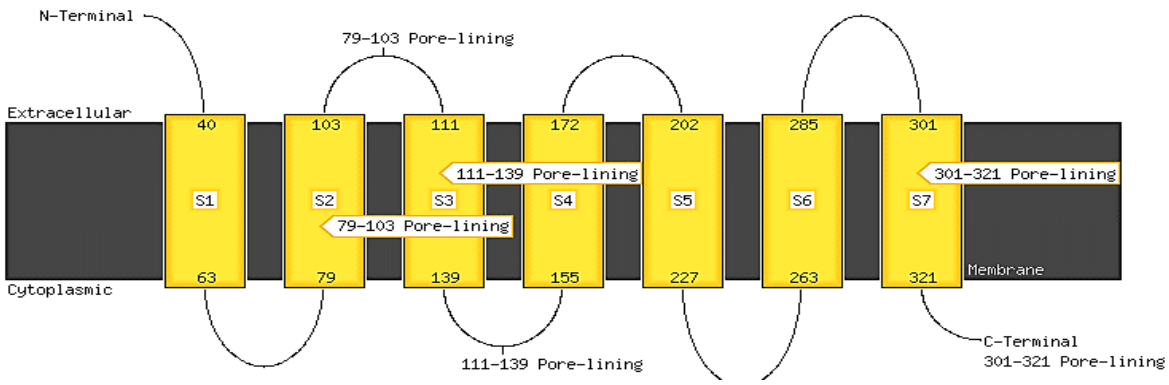


Figure 4.5: Prediction of transmembrane helices and the topology analysis of Drome-AKHR. The helices are represented in yellow and labelled S1-S7, the membrane (black) and the loops (thin black line) starting from the N-terminus (the extracellular region) and terminating at the C-terminus (the intracellular region)

The topology analysis predicted the helices as follows, H1(40-63), H2 (79-103), H3 (111-139), H4 (155-172), H5 (202-227), H6 (263-285) and H7 (301-321) see **Figure 4.5**. The pore-lining was found to be at residues 79-103, 111-139 and 301 -321. The extracellular region comprises the primary amino acid residue at 40, 103, 111, 172, 202, 285 and 301 while the cytoplasmic region comprises residues 63, 79, 139, 155, 227, 263 and 321.

4.2.3 Homology modelling

The same approach used in chapter 3 was applied to build two models (active and inactive). Also, the same template crystal structure of beta 2 adrenergic receptor (5D5A.1A) and the crystal structure of rhodopsin (2X72.1A) used in chapter 3 were used for the construction of the models. Clustalw2 was used to align the Drome-AKHR sequence to the templates, as shown in **Figure 4.6.** The constructed models are identified as β2frf-model (based on β2AR AKHR) and rhodfrf (based on rhodopsin AKHR).

Figure 4.6: Clustalw2 sequence alignment illustrating Drome-AKHR and the two selected target templates β2AR, (5D5A.1A / P07550), and rhodopsin (2X72.1A / P04695). The sequence similarity is highlighted in a black and white colour while the transmembrane domain highlighted in yellow.

The alignment (**Table 4.1**) showed a sequence identity ranging from 17% to 45% and sequence similarity between 51% and 89% for the different helices. The two target templates, β2AR, (5D5A.1A) and rhodopsin (2X72.1A), displayed a relatively low sequence percentage identity but possessed high sequence similarity (see **Table 4.1**). Low sequence percentage identity and

high sequence similarity are not new to class A GPCRs superfamily, because, with high percentage similarity, GPCRs / AKHRs do possess similar structural and functional features[298].

Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	SCOP Codes	Structure	CATH Entry
CERT	96.782	9e-09	-639.6	-3.1	372.0	284	304	429	5d5a		CATH Summary
CERT	93.886	2e-08	-469.0	13.5	461.0	321	348	429	Zx72		CATH Summary
CERT	90.390	4e-08	-460.8	11.1	428.0	325	350	429	4bu0		CATH Summary
CERT	90.259	4e-08	-553.5	12.6	410.0	320	363	429	1u19		CATH Summary
CERT	88.522	6e-08	-544.4	13.3	412.0	265	278	429	2z73		CATH Summary
CERT	82.355	2e-07	-544.8	24.8	412.0	299	454	429	4ea3		CATH Summary
CERT	78.793	6e-07	-515.1	4.7	314.0	277	429	429	4grv		CATH Summary
CERT	77.213	8e-07	-550.8	14.7	337.0	269	289	429	4ej4		CATH Summary
CERT	76.940	9e-07	-458.0	4.3	313.0	286	390	429	2rh1		CATH Summary
CERT	76.642	9e-07	-619.1	7.4	284.0	271	303	429	4ib4		CATH Summary

Figure 4.7: The statistical result from Psi-blast of the Drome-AKHR illustrating the common ancestry. The best two templates with net score 96.782 (β2AR, 5d5a) and 93.886 (rhodopsin 2x72) were selected for the construction of the active and inactive model of Drome- AKHR.

Having established the similarities and alignments (**Figure 4.6** and **Figure 4.7**); homology modelling was then conducted using Swiss-Model [330]. The alignments were used to build two homology models of the 7 TM bundles of AKHR. These models were labelled β2frf-model and Rhodfrf-model.

Table 4. 1. Comparison of sequences of AKHR, Rhodopsin and β2AR helices.

Helix	% Sequence identity with		% Sequence similarity with	
	β2AR (β2frf)	rhodopsin (Rhodfrf)	β2AR(β2frf)	rhodopsin (Rhodfrf)
1	16	20	60	61
2	25	16	58	52
3	29	30	68	60
4	28	40	76	80
5	23	17	65	60
6	34	23	83	74
7	43	45	88	88

High sequence similarity was recorded for the TM helix segments of both β2AR and rhodopsin and could be the reason for their prominent structural similarity[354,355].

The presence of conserved amino acid residues, patterns, and motifs commonly found in all Class A GPCRs [190,223,356], are also present in the constructed models see **Figure 4.8**

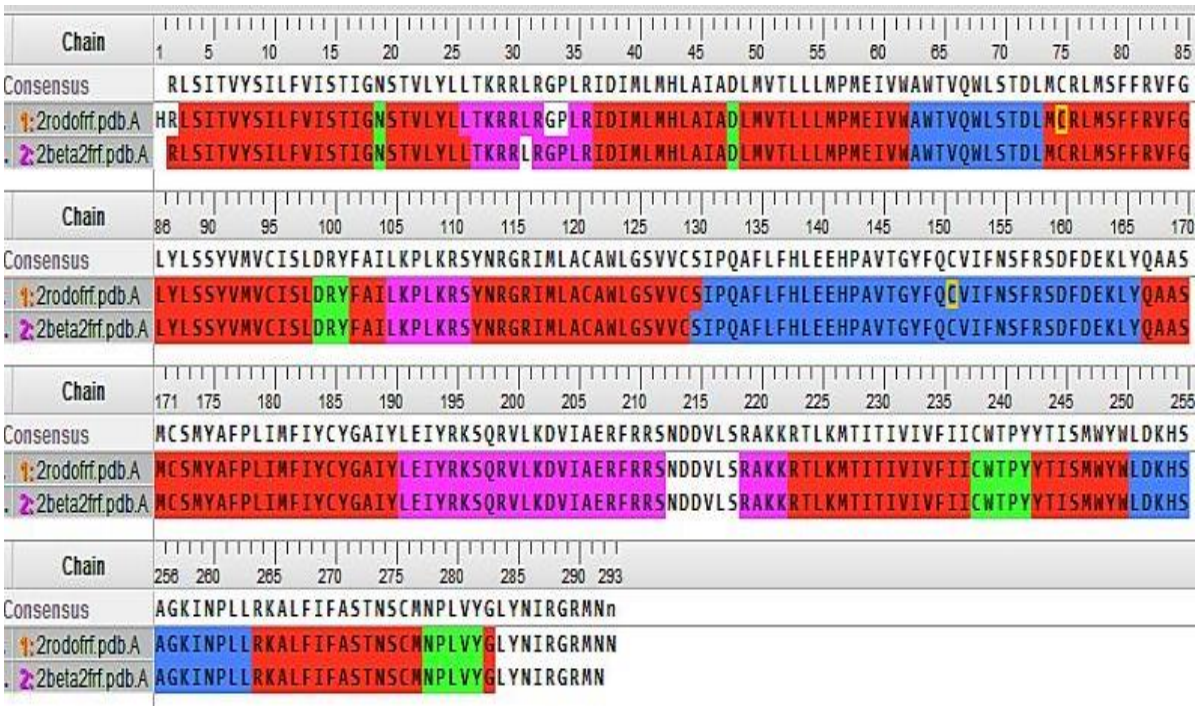


Figure 4. 8: Schematic diagram representing an alignment of both models constructed from both rhodopsin and beta2 androgenic receptor of Drome-AKHR. Red indicates residues in the seven transmembrane helices, blue shows the extracellular regions, while the intracellular regions are represented in purple, yellow denotes region cystine ionic lock, and green shows the highly conserved residues.

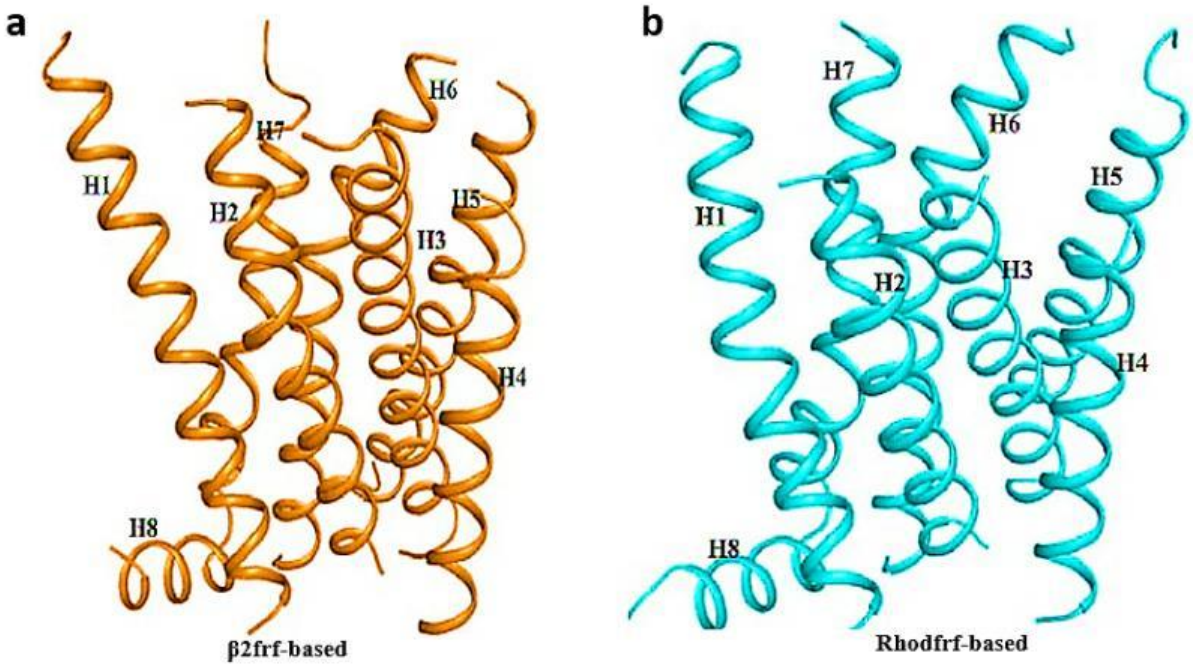


Figure 4. 9: The constructed seven transmembrane helices of the adipokinetic hormone receptor. a) β2frf- model seven transmembrane helices of AKHR and b) Rhodfrf-model seven transmembrane helices of AKHR.

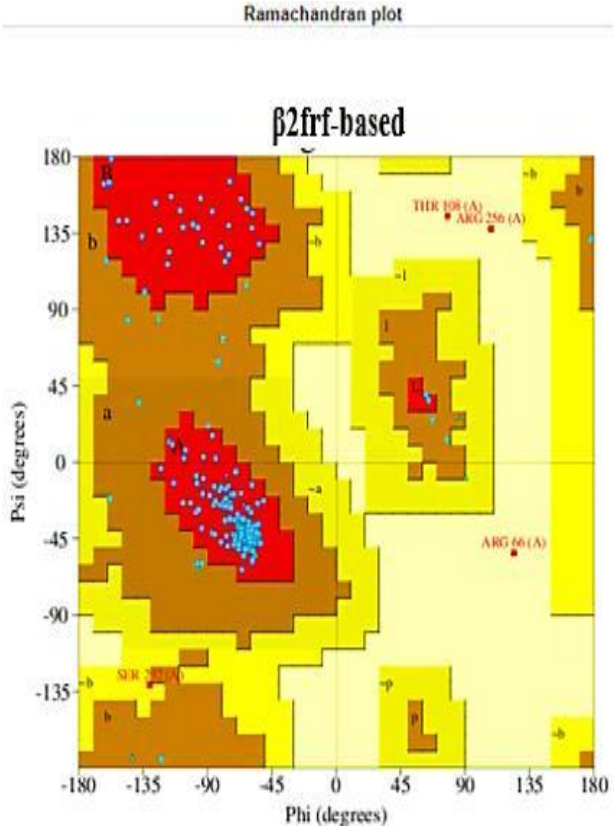
Overall, the constructed models (**Figure 4.9**) are alike, especially the anticlockwise arrangement of the transmembrane helices, and they both possess the eighth intracellular helix that is parallel to the cell membrane. Also, both structures have a tilted TM3, and the shortest transmembrane helix in both models is TM4.

(a)

1. Ramachandran Plot statistics for β2frf-based

		No. of residues	%-tage
Most favoured regions	[A,B,L]	251	93.0%
Additional allowed regions	[a,b,l,p]	15	5.6%
Generously allowed regions	[~a,~b,~l,~p]	1	0.4%
Disallowed regions	[XX]	3	1.1%*
Non-glycine and non-proline residues		270	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		10	
Proline residues		9	
Total number of residues		291	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].



(b)

2. Ramachandran Plot statistics for Rhodfrf-based

		No. of residues	%-tage
Most favoured regions	[A,B,L]	244	89.7%*
Additional allowed regions	[a,b,l,p]	19	7.0%
Generously allowed regions	[~a,~b,~l,~p]	7	2.6%
Disallowed regions	[XX]	2	0.7%*
Non-glycine and non-proline residues		272	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		10	
Proline residues		9	
Total number of residues		293	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R*-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

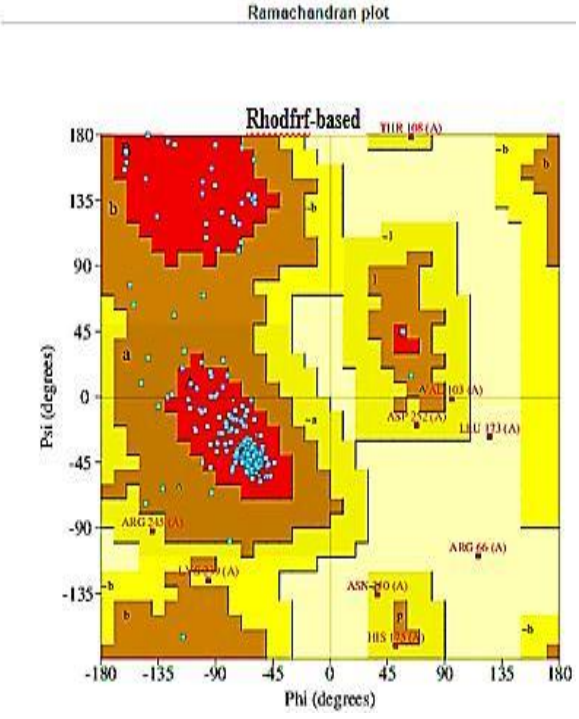


Figure 4. 10: Ramachandran plots of a) β2frf- model and b) Rhodfrf-model.

ERRAT[357] and PROCHECK[358] evaluation programs were used to check the quality of the constructed models. Also, Ramachandran plots of the two models see **Figure 4.10**. The result of both models shows that the phi and psi, which represent the backbone torsion angle, are acceptable and most of the primary amino acid residue of the constructed models are either in allowed or favoured regions. Also, the β2frf-model, Arg66, Arg256 and Thr108 had a φ angle 10^0 outside the generously allowed region, while the Rhodfrf-model had two residues Arg66 and Leu123, had ψ torsion angles 20^0 outside the allowed region. These problems were corrected during the molecular dynamics of the models. The ERRAT program score of 97.73 and 93.63 was calculated for the β2frf-model and Rhodfrf-models, as the acceptable score range for a high-quality model is > 50 [357]. As such, we conclude that these two models are of a high-quality model and are acceptable.

4.2.3.1 Structural comparison of models

The constructed models of both β2frf-model and Rhodfrf-model can be compared to the two-class A GPCR structures, that were used as a template to build them. An overlay of the β2frf-model with its templates (see **Figure 4.11 a)** gave an RMSD of 3.3Å for superimposition of Cα atoms of the whole molecule and 2.07Å for the helix bundles. Also, an overlay of the Rhodfrf-model with its templates (see **Figure 4.11 b)** gave an RMSD of 3.5Å for superimposition of Cα atoms of the whole molecule and 2.05Å for the helix bundles. Overall, the target templates have longer TM helices when compared to the two constructed models, with β2AR having 392 amino acid residue[331] as opposed to the constructed β2frf- model which has 330 amino acids. Also, the rhodopsin template model has 442 amino acid residue[224] as against the constructed Rhodfrf – based- model, which has 327 amino acid residues.

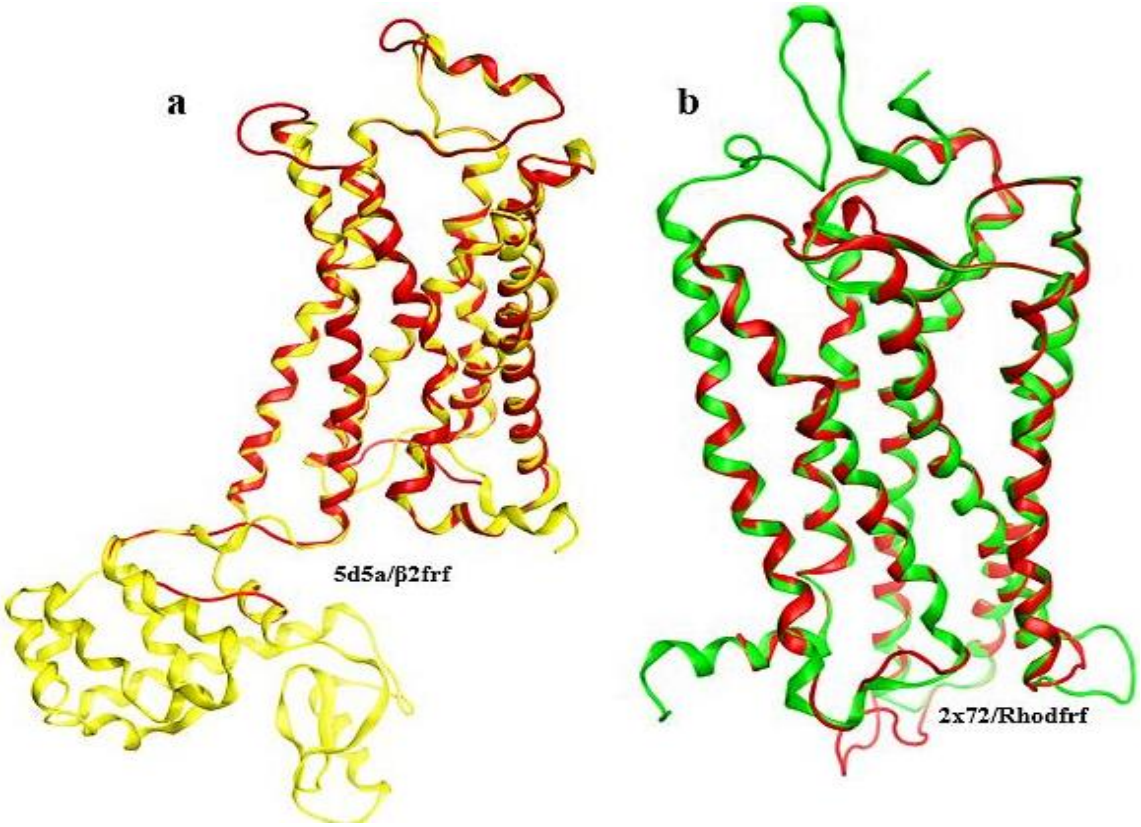


Figure 4.11: An overlay of β2frf- model-AKHR (the constructed model, in red) and 5d5a (the template used for constructing the model, in yellow) with RMSD 3.6Å. b) shows an overlay of Rhodfrf (the constructed model, in red) and 2x72 (the template used for constructing the model, in green) with RMSD 3.6Å.

Both constructed models have the CWxP(Y/F) (CWTPY) motif in TM6; when compared with the templates, a slight difference in the amino acid residues was noticed. For β2AR, the CWxP(Y/F) motif is CWLPF while for rhodopsin, it is CWLPY. This is the only significant difference noticed as regards the conserved residue. This motif is significant to receptor activations and will be discussed later.

4.2.3.2 Evaluation of the helix bundles

In both models, the transmembrane helices have an anticlockwise helical arrangement, when viewed from the extracellular side. This anticlockwise arrangement is found in most class A GPCRs [359]. Literature supported this anticlockwise arrangement and have linked the cause to distance restraints from Zinc (II) binding cavity in neurokinin 1 (NK-1) and κ-opioid receptor

[359]. The helical bundles are stabilised by interhelical interactions comprising the highly conserved residues [223]

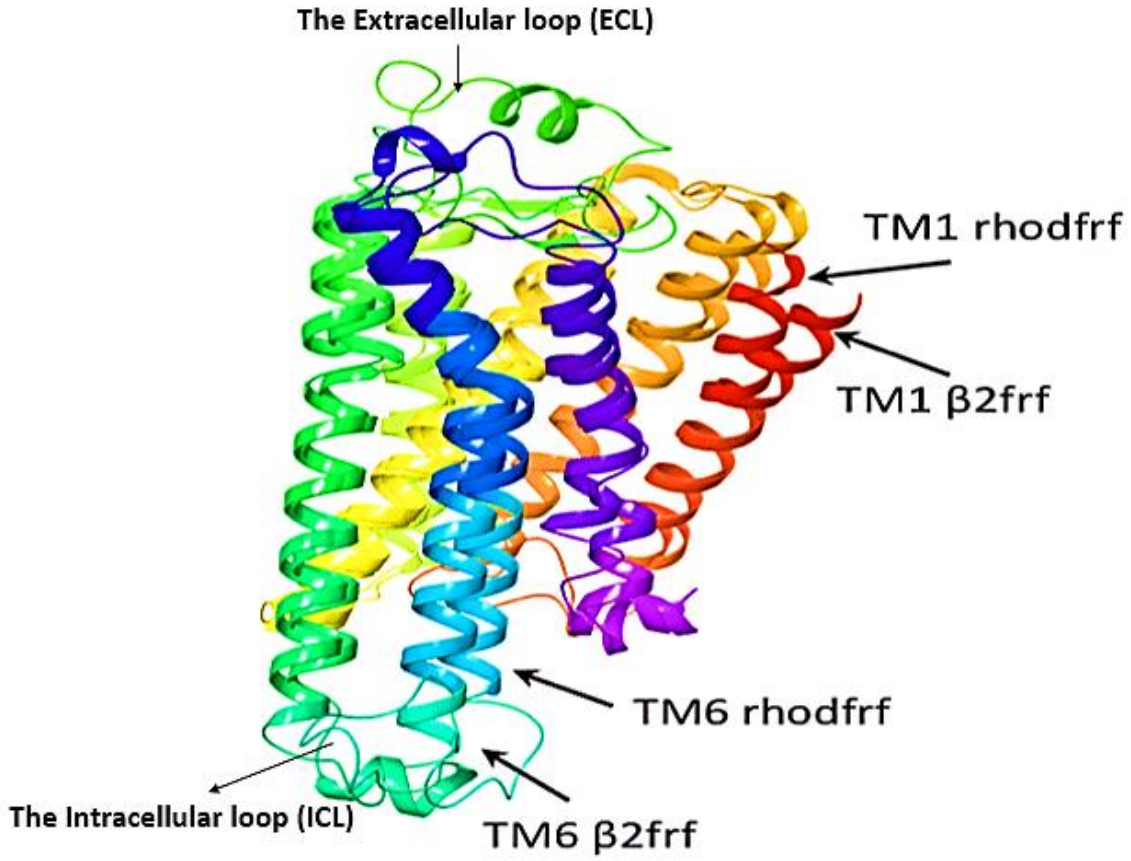


Figure 4. 12: The superimposition of β2frf-model and Rhodfrf-model of Drome-AKHR.

Superimposition of the two constructed 3D models gave a root means square deviation (rmsd) of 3.6 Å, for superimposition of Cα atoms of the whole molecule and 2.12 Å for the helix bundles. TM6 and TM5 differ from each other with respect to length with both helices extending by 4-3 residues into the cytoplasm. This observation is consistent with studies conducted on the accessibility of nitroxide labels fixed to ECL3 joining TM6 and TM5 [360]. Both structures possess a tilted TM3 conformation. This tilting of TM3 is also found other class A GPCRs model[93,98,195,361,362]. The orientation of the seven transmembrane helices is subtly different in the two models, which leads to differences in the intracellular and extracellular

domains. Also, both β2frf- model-AKHR and Rhodfrf-models have a disulphide bridge between Cys112 and Cys188. This disulphide bridge is commonly found in GPCRs.

In the β2frf- model, the extracellular end of TM1 bends away from the helical bundle (**Figure 4.12**) with a distance of 5.42 Å from TM1 of Rhodfrf-model. This gives a more open structure to the extracellular domain of the β2frf-model as compared to the Rhodfrf model. At the same time, a distance of 8.7 Å at the intracellular end of TM6 between both model (**Figure 4.12**) was recorded. This gives a more closed structure to the intracellular domain of the β2frf- model. Thus, the β2frf- model is described as the open or inactive model, while the Rhodfrf-model is termed the closed or active model[239]. The closed conformation of the Rhodfrf-model is similar to the rhodopsin crystal structure where the close nature protects cis-retinal from hydrolysis [96,98]. Since ligand binding takes place on the extracellular domain [363], the closed nature of the Rhodfrf-model prevents the ingress of the ligand. On the other hand, the open nature of the β2frf- model allows ligand-free access to the receptor-binding pocket.

4.2.3.3 Overview of transmembrane helical interaction and molecular switches

A crucial feature of most class A GPCRs is the presence of primarily preserved molecular switch motifs. The existence of these switches is vital for stabilising any GPCR either in the close (active) or open (inactive) state. In Drome-AKHR a transmembrane 3-6 lock occurs between Arg260 on TM6 and Tyr138$^{3\text{-}51}$ on TM3. In the inactive state, Arg260$^{6\text{-}29}$ points away from TM3, and so no interaction is possible (**Figure 4.13 a**). Upon activation, however, TM6 twists moving Arg260$^{6\text{-}29}$ so that it now points toward Tyr138$^{3\text{-}51}$ (**Figure 4.13 b**). This interaction locks the receptor in the active state. **Figure 4.13 b** also indicates that, in the active state, Arg260$^{6\text{-}29}$ on TM3 interacts with Glu 244 of ECL2. This loop closes over the binding pocket after ligand binding.

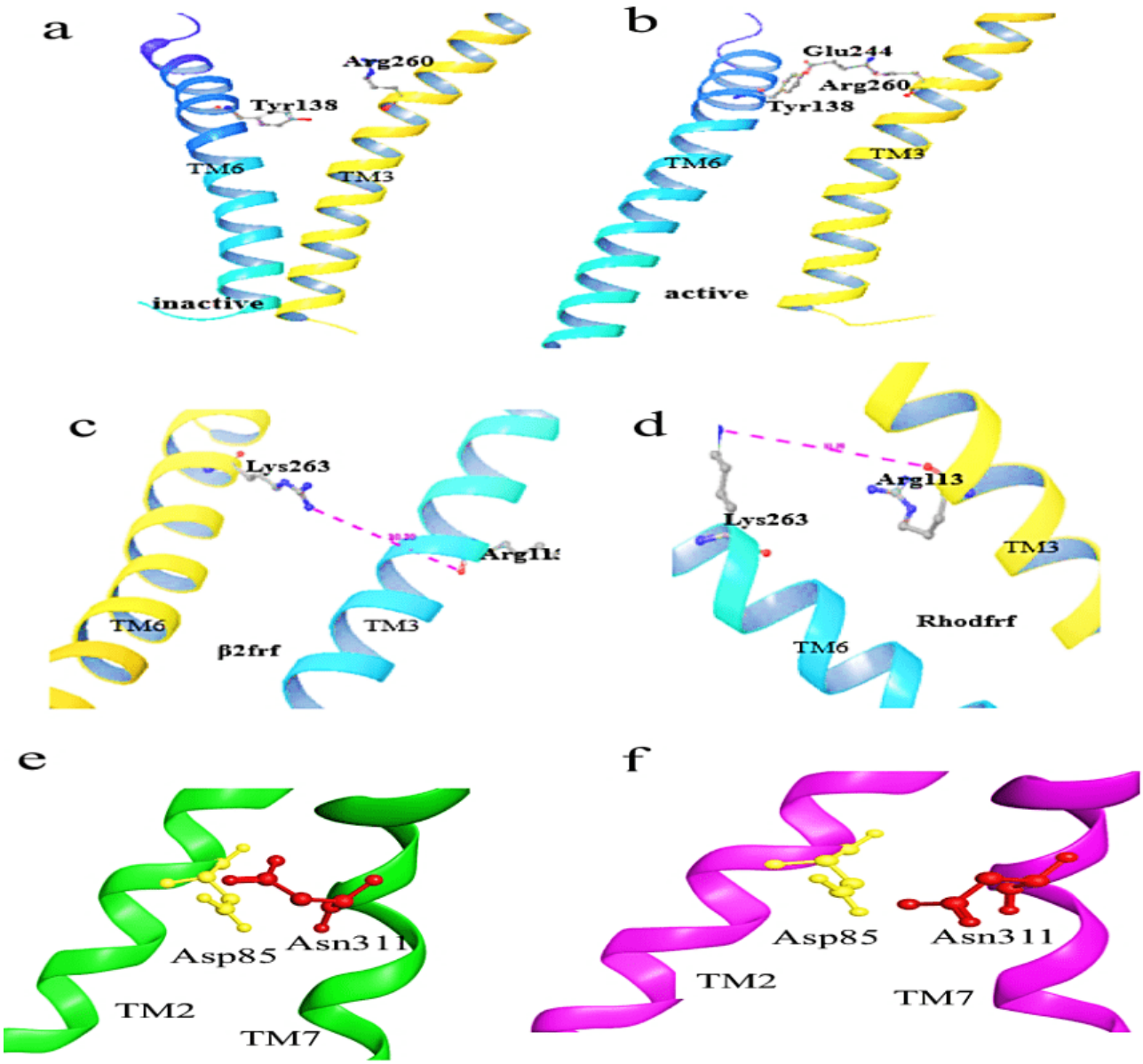


Figure 4. 13: a) molecular switch interactions (inactive) in the β2frf-model between Tyr 138 in TM6 and Arg260 in TM3. b) molecular switch interactions (active) in the Rhodfrf-model between Tyr 138 in TM6 and Arg260 in TM3. c) molecular switch interactions (active) in the Rhodfrf-model between Lys263 in TM 6 and Arg113 in TM3. d) molecular switch interactions (inactive) in the β2frf-model between Lys263 in TM6 and Arg113 in TM3. e) molecular switch, interactions (inactive) in the β2frf-model between Asp85 in TM2 and Asn311 in TM7. f) molecular switch interactions (active) in the Rhodfrf-model between Asp85 in TM2 and Asn311 in TM7.

In the Rhodfrf-model, there is an ionic interaction between Lys263$^{6\text{-}35}$ in TM 6 and the positively charged side chain nitrogen atom of Arg113$^{3\text{-}50}$ in TM 3 (**Figure 4.13 c**). This interaction does not exist in the β2frf- model (**Figure 4.13 d**). These switches are found to be

the same as those described for the AKH- receptor of *A. gambiae*, *Rhodnius prolixus* and *Tribolium castaneum* [231,366].

Class A GPCRs have a conserved aspartate on TM2 and an asparagine on TM7. In the Rhodfrf model, there is an intramolecular interaction between Asp85 and Asn311 (**Figure 4.13 f**). In the β2frf-model (**Figure 4.13 e**), the helices are shifted such that these two residues are now 4.8Å apart and can no longer interact [239]. Experimental data backs the importance of these intramolecular interactions because of the vital role they play during receptor activation, basal activity and conformation flexibility reduction [239].

The stability of the AKHRs is sometimes achieved by the presence of anionic / cationic-π interaction [379–381]. β2frf-model forms a π-cationic interaction between TM5 Tyr $231^{5\text{-}66}$ and TM 6 ($Asp251^{6\text{-}22}$ and Lys $259^{6\text{-}29}$) (see **Figure 4.14 a**), this interaction was not found with Rhodfrf-model instead an H-bond interaction was identified between TM 5 (Ser $234^{5\text{-}69)}$ and TM 6 (Lys $259^{6\text{-}29}$) (see **Figure 4.14 b**).

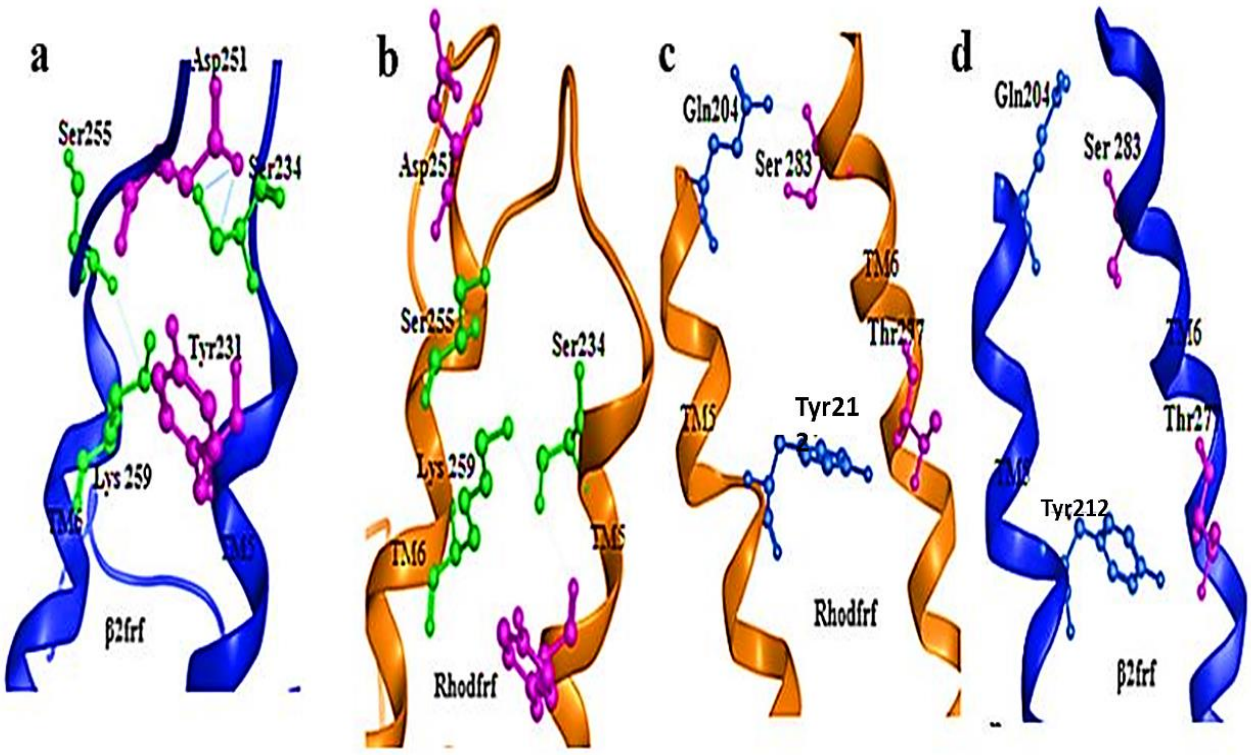


Figure 4.14: (a) β2frf-model have a π-cationic interaction between TM5 Tyr 231[5-66] and TM 6 (Asp251[6-22] and Lys 259[6-29]) (b) Rhodfrf-model H-bond interactions between TM 5 (Ser 234[5-69]) and TM 6 (Lys 259[6-29]). (c) the Rhodfrf-model have a molecular interaction between Tyr 212 [5-46] in helix 5 and Thr 277 [6-49] in helix 6 (d) the β2frf model have molecular interactions in between Gln 204[5-39] in helix 5 and Ser 283[6-55] in helix 6.

Intramolecular interaction was identified in Rhodfrf-model between Gln 204[5-39] in helix 5 and Ser 283[6-55] in helix 6 (see **Figure 4.14 c)**. This interaction was not found in β2frf-model see **Figure 4.14 d**. Also, another interaction, which was found in the Rhodfrf-model is an interaction between Tyr212[5-46] in helix 5 and one of the conserved residues Thr277 [6-49] in helix 6, see **Figure 4.14 c**. Thus, in the β2frf -model, due to helical shifts, the Tyr212 [5-46] in helix 5 and Thr277 [6-49] in helix 6 residues are 4.1Å apart, thereby decreasing any chances of interaction.

Class A GPCRs have a conserved aspartate on TM2 and an asparagine on TM7 [239]. In the β2frf-model (**Figure 4.15** a), the side chains between TM7 Asn311[7-49] and TM2 Asp85[2-50] are orientated away from each other. This decreases the possibilities of interactions between them.

In the Rhodfrf-model (**Figure 4.15 b**), the side chains of TM 7 Asn311[7-49]and TM 2 Asp 85[2-50] are orientated close to each other as such; it increases the chances of interactions between them.

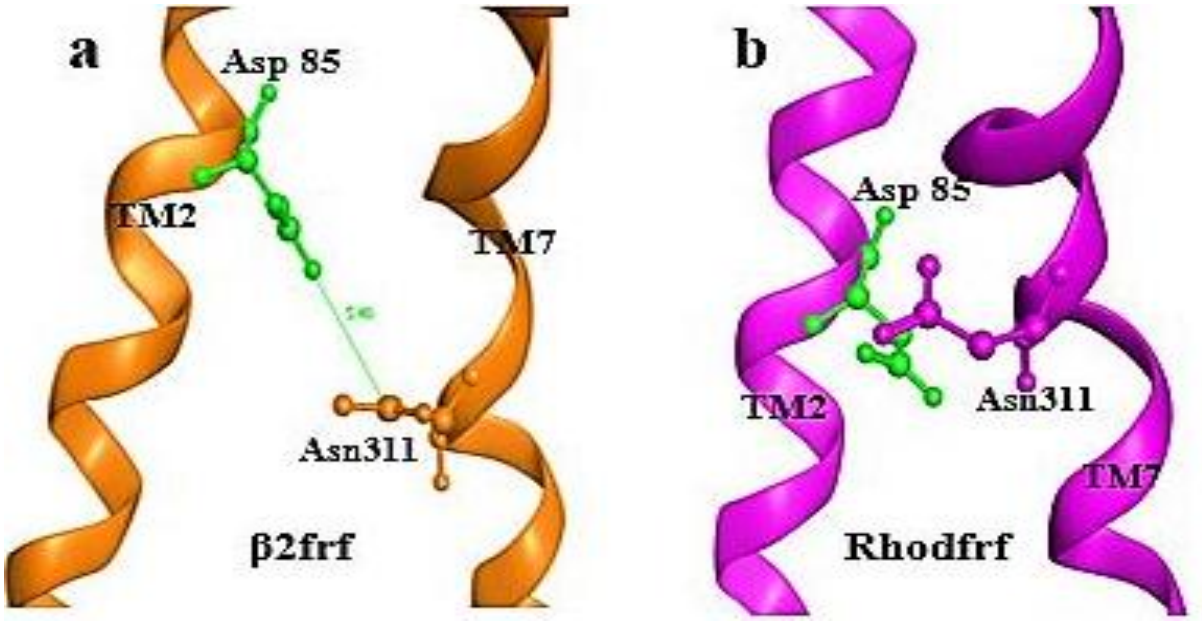


Figure 4.15: a) Absence of molecular interactions between TM1 Asp85 and TM7 Asn311 (β2frf -model). b) presence of molecular interactions between TM1 Asp85 and TM7 Asn311 (Rhodfrf-model).

In the Rhodfrf structures, Asp 85 [2-50] in helix 2 forms an ionic interaction with Ser126 [3-39] in helix 3, though this interaction is absent in the β2frf- model (**Figure 4.16 a & b**). Also, in both open and closed structures, there is polar interaction between Asn 56 [1-50] (helix 1) and Ser312 [7-46] (helix 7) (**Figure 4.16 d**), which stabilises the structure. Dror *et al*., 2009 discovered the 'on and off'[367] ionic lock in different conformations of an inactive β2AR while in equilibrium.

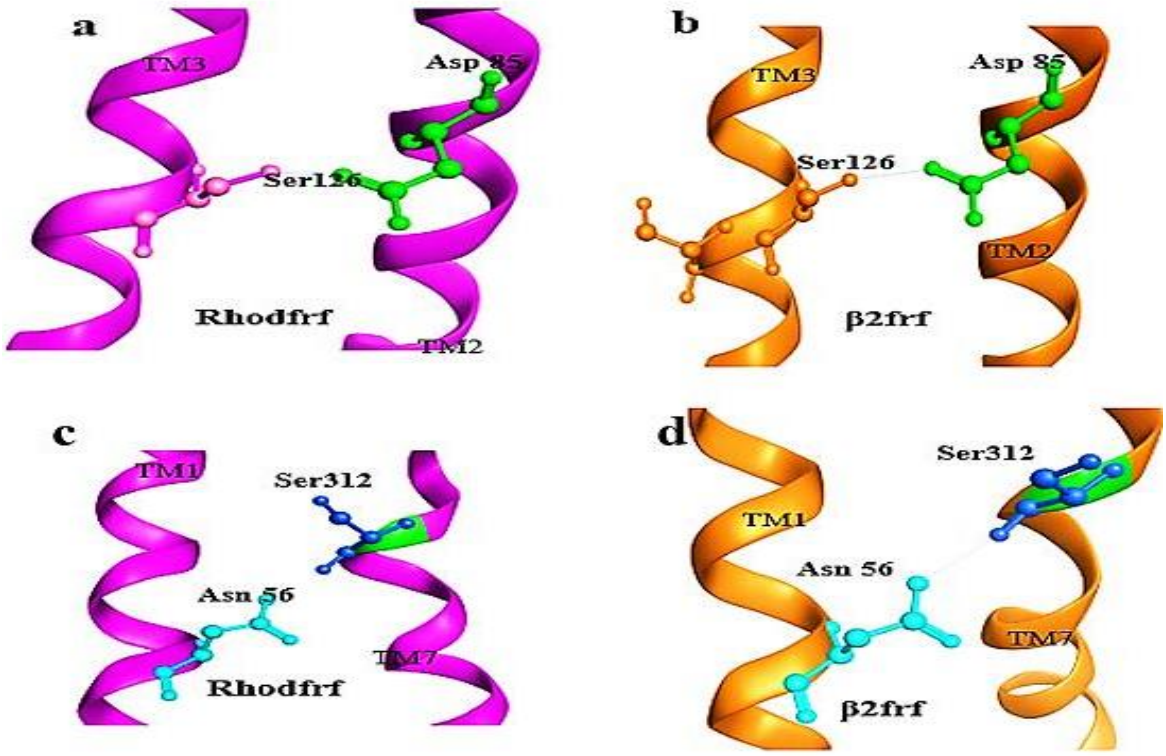


Figure 4. 16: (a) Rhodfrf model showing a salt-bridge inter interaction between Asp 85 in helix 2 and Ser 126 helix 3 . (b) The β2frf- model is showing the same interaction as in (a). (c) absence of polar interaction between TM 1 and TM 7. (d) presence of polar interaction between TM 1 and TM 7.

These interactions (**Figure 4.16)** involving helices 7,2 and 1 are structurally essential for all Class A GPCRs molecule stabilisation [382].

4.2.3.4 Evaluation of Kinks

The existence of kinks that initiate the helical twist (bend) at almost the same locations where Tyr, Thr or Met come before Pro is also backed by experimental data[195,368–370]. The TM helices in both open and closed AKHR models have these proline-generated bends or kinks (**Figure 4.17**). TM2 has a kink between Met $93^{2\text{-}58}$ and Pro $94^{2\text{-}59}$ (**Figure 4.17 a**). TM5 has a kink between $Tyr212^{5\text{-}47}$ and $Pro215^{5\text{-}50}$ (**Figure 4.17 b**). A turn was noted on TM4 between $Gly162^{4\text{-}52}$ and $Val64^{4\text{-}54}$ for the β2frf-model (**Figure 4.17 d**). This was not found in the Rhodfrf structure (**Figure 4.17 e**). From the extracellular region, the kink in TM 6 instigates a twist

near TM 5. An overlay of the β2frf- model Rhodfrf-models in **Figure 4.17** a, d and g show the different effect of these kinks or bends have on the two models.

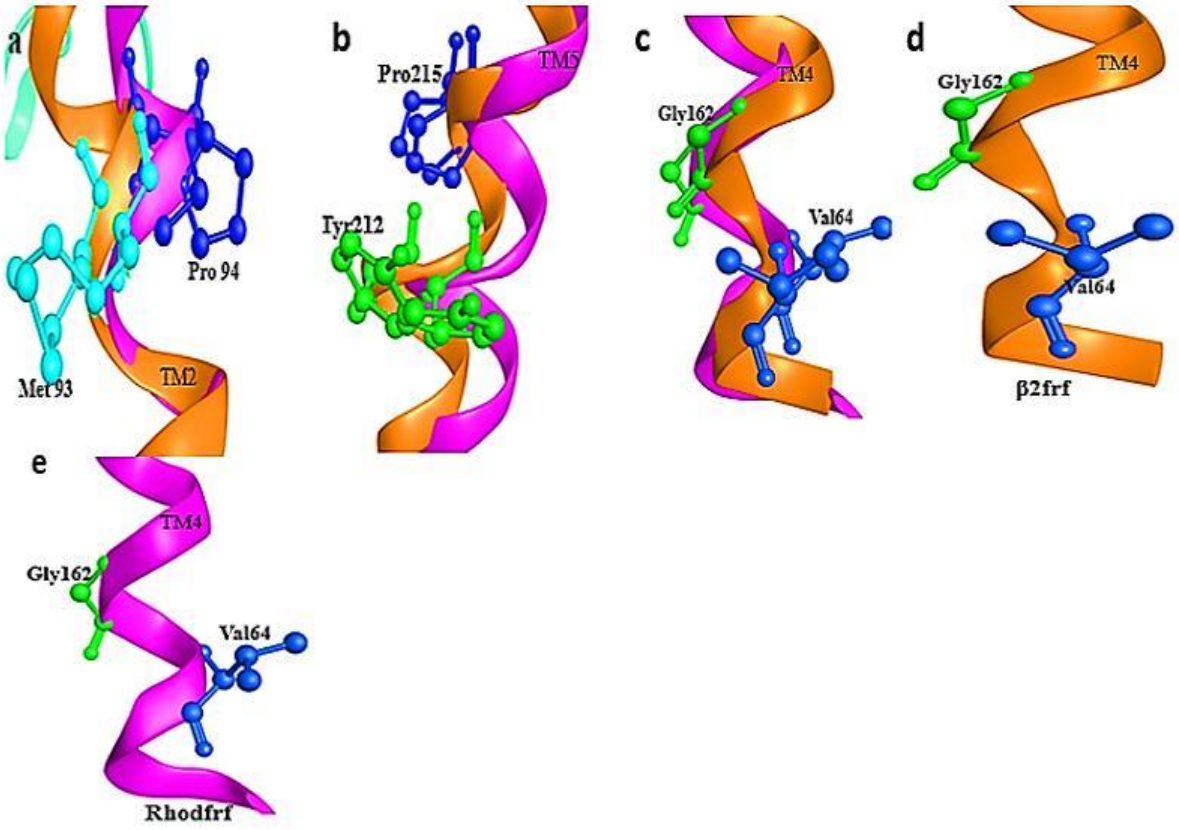


Figure 4. 17: Overlay of 3D models of the β2frf-model and Rhodfrf-model structure, showing the presence of kinks in TM2. b) overlay of 3D models of the β2frf-model and Rhodfrf-model structure showing the presence of kinks in TM5 and c) overlay of 3D models of the β2frf-model and Rhodfrf-model structure showing the presence of kink in TM4. d) The β2frf-model showing the presence of kink in TM4. e) The Rhodfrf-model showing the absence of kink in TM4. * β2flf = β2frf-model, Rhodfrf =Rhodfrf-model.

4.2.3.5 Evaluation of the loop regions of the constructed models

Apart from the orientation of the transmembrane helices, there are other dissimilarities between the two models (**Figure 4.12**). In the extracellular loop I (ECL I), the open structure has no β-strand, while in the Rhodfrf-model there is a β-strand between Ala100 and Gln104, a γ-turn between Trp105 and Leu106 and a small helical turn between Ser107 and Thr108. In the β2frf-model, there is a helical turn between Ala171 and Tyr185. This moves ECL2 away from the binding pocket. At the same time, the N-terminus also moves to give any potential ligand

access to the receptor-binding pocket. In the Rhodfrf-model, there is a β-strand between Glu177 and Gln184, and a small helical turn from His179 to Cys188. This keeps ECL2 in an upright position above the helical bundle. In this way the N-terminus and ECL2 extend over the transmembrane helices, shielding all possible binding sites and preventing the ligand from gaining access to or leaving the binding pocket.

ICL3 is usually not characterised in the crystal structures of GPCR. In both the β2frf-model and Rhodfrf-model, ICL3 has a helical turn between Ser234 and Ser249. Dror *et al*. identified this turn in β2AR[367]. In both the open and closed models, ECL3 has a β-strand involving Ile293 and Pro298. Moreover, the β2frf-model has a helical turn between Leu299 - Leu300, while the Rhodfrf-model has a β-strand between Leu288 and Lys290 and a helical turn between His291 and Gly294.

The eighth helix is familiar with most Class A GPCRs and lies parallel to the intracellular membrane surface. This helix is thought to be essential for receptor activation and binding of a G-protein [239]. Most GPCRs intracellular region are stabilised by palmitoylation of Cys, Ser or Thr on the 8th helix [372–374]. Both the β2frf-model and Rhodfrf-model have this 8th helix, which has Cys313, which could be stabilised by palmitoylation.

Since ligands prefer to binding at the extracellular part of a receptor [363], the closed nature of the Rhodfrf-model stops the ingress of the ligand that way proper binding of a ligand to the extracellular part is impossible. On the other hand, the open nature of the β2frf-model allows ligands free access to the receptor-binding pocket.

Also, since GPCRs are highly flexible, the different structural features, particularly in the extra and intracellular region of the two models, could represent features in different conformations during molecular dynamics and in the active and inactive state of the receptor.

4.3.0 Conclusion

We have successfully constructed 3D AKHR models for the fruit fly, *Drosophila melanogaster,* AKH-receptor and we present the open conformation as the inactive state and the closed conformation as the active state of the AKHR. The elucidated structures possess the same features as most class A GPCRs.

Structural differences can be found around the highly conserved amino acid DRY residues, where the Rhodfrf-model posses an ionic interaction linking the positively charged side chain nitrogen atom of Arg113$^{3-50}$ in TM 3 and the backbone C=O of Lys263$^{6-35}$ in TM 6. This was not found in the open model.

A π-cationic interaction between TM5 Tyr 231$^{5-66}$ and TM 6 (Asp251$^{6-22}$ and Lys 259$^{6-29}$) was identified in β2frf-model but absent in Rhodfrf-model instead an H-bond interaction was identified between TM 5 (Ser 234$^{5-69)}$ and TM 6 (Lys 259$^{6-29}$).

In the closed model, a polar interaction was identified between Asn 56 $^{1-50}$ on TM1 and Ser312 $^{7-46}$ on TM7, but not found in the open model. In the closed structure, an intramolecular interaction exists between TM5 (Gln204$^{5-39}$) and TM6 (Ser 283$^{6-55}$). Also, in the closed structure, another intramolecular interaction was identified between TM5 (Tyr212$^{5-46}$) and one of the conserved residues, Thr 277 $^{6-49}$ on TM6.

In both models, kinks were also identified in TM2 (Met 93$^{2-58}$-Pro 94$^{2-59}$), TM5 (Tyr212$^{5-47}$ - Pro215$^{5-50}$) and TM4 (Gly162 $^{4-52,}$ - Val64$^{4-54}$).

The general conformation and structural features of the β2frf-model and Rhodfrf-models can be used to study the binding and activation of the Drome-AKH receptor by Phote-HrTH.

5

Constructing a 3D Model of the AKH-receptor of the Oriental Fruit Fly, Bacdo-AKHR.

Summary

The newly available oriental fruit fly (Bactrocera dorsalis) AKHR was utilised in the construction of a three dimensional (3D) model of the receptor [120,190,330]*, doing so will enable the prediction and development of new insect control agent that are selective in their actions. Using homology modelling, two target crystal structures were found, namely the beta2-adrenergic receptor (β2AR) and Rhodopsin. The β2AR structure has an open conformation meaning the receptor is inactive, while the rhodopsin structure possesses a closed (active) conformation. The construction of the open and closed conformation was proposed to enable adequate explanation of the active and inactive model. The constructed AKHR from rhodopsin template does not allow the diffusion of the ligand into the receptor binding site, meaning the AKHR is active while the AKHR constructed from the β2AR template allows the diffusion of the ligand into the receptor-binding site. This study is consistent with other experimental studies conducted on both Rhodopsin and β2AR* [224,331]*. Both constructed models for Bacdo-AKHR are identified as β2off and Rhodoff.*

Both models when compared with those constructed for the flesh fly AKHR (chapter 3) and the fruit fly (Chapter 4) are very much alike (they possess almost the same molecular interaction) and can be compared to other class A GPCRs.

5.1.0 Introduction.

The oriental fruit fly *Bactrocera dorsalis* is a highly invasive insect pest and is a significant threat to food security. The oriental fruit fly not only feeds on ripe fruit but also on ripening fruit. This action can cause severe loss to farmers. Controlling the invasion of this pest insect has not been comfortable as they continue to gain resistance to the widely-used insecticides [37,38,44]. The widely used insecticides pose severe challenges as when used, both beneficial and

pest insects are killed [39,40]. This has prompted the search for "green insecticides", biological effective pesticides that are selective in their activity, negatively affecting the targeted pest species without causing damage to beneficial insect species. Since neuropeptides and their G-protein-coupled receptors are considered suitable targets for new insect control agents, similar to the bases used in the enhancement of drugs for the treatment of human disease [29–33,41]. Using the same method of targeting the GPCR/ AKHR, a new insect control agent that is selective in there could be produced. This cannot be achieved without the presence of the insect GPCR/AKHR 3D structure. The presence of 3D structures will help molecular explanation and understanding of ligand binding and receptor activations of the insect. The absence of adequate literature and elucidated 3D models hinders template target virtual screening[332,333]. Also, the lack of 3D models and inadequate ligand information has hindered ligand-based approaches to drug discovery [92,334]. Here to provide a possible solution by providing an adequate molecular explanation of the pest insect, the recently published primary sequence of the oriental fruit fly AKH-receptor (Bacdo-AKHR) was extracted from the GenBank database (accession number AQX83416) and used to predict the 3D structure of the oriental fruit fly AKH receptor. The experimental procedure for the construction of the oriental fruit fly AKHR is similar to that in chapter 3, so only results will be presented in this chapter.

5.2.0 Results and Discussion

5.2.1 Analysis of primary sequence

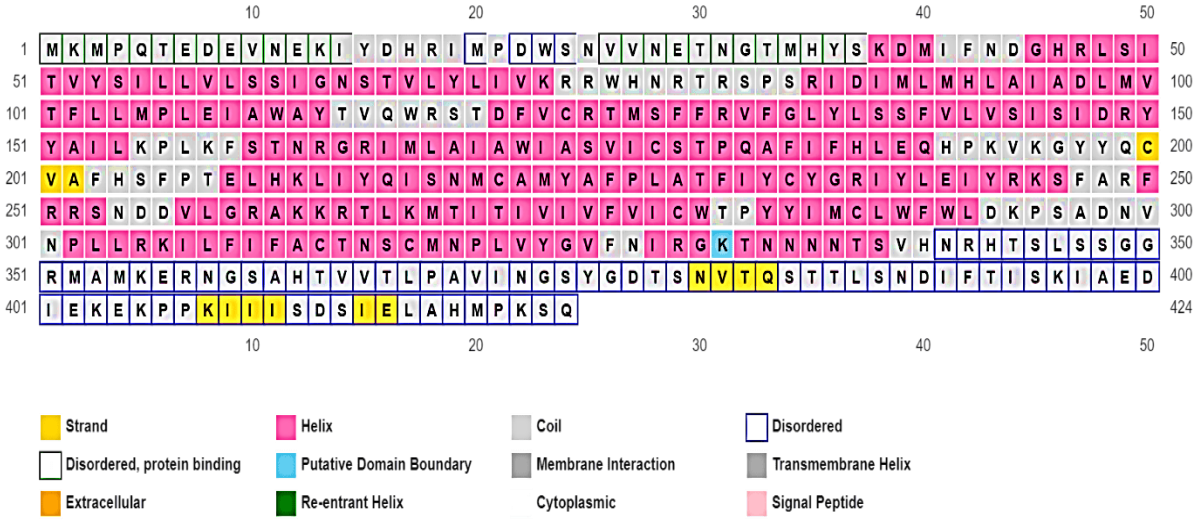


Figure 5.1: Primary amino acid sequence submitted for secondary structure prediction of Bacdo-AKHR to PSIPRED

Primary structure analysis of the 424 amino acid residues (see **Figure 5.1**) indicated that not all the amino acid sequence residue could be used for the construction of the 3D model. To understand how the transmembrane topology and the interactive receptor features might look, a 2D Protter diagram was employed to further analyse the transmembrane topology and the interactive receptor features see **Figure 5.2** [351].

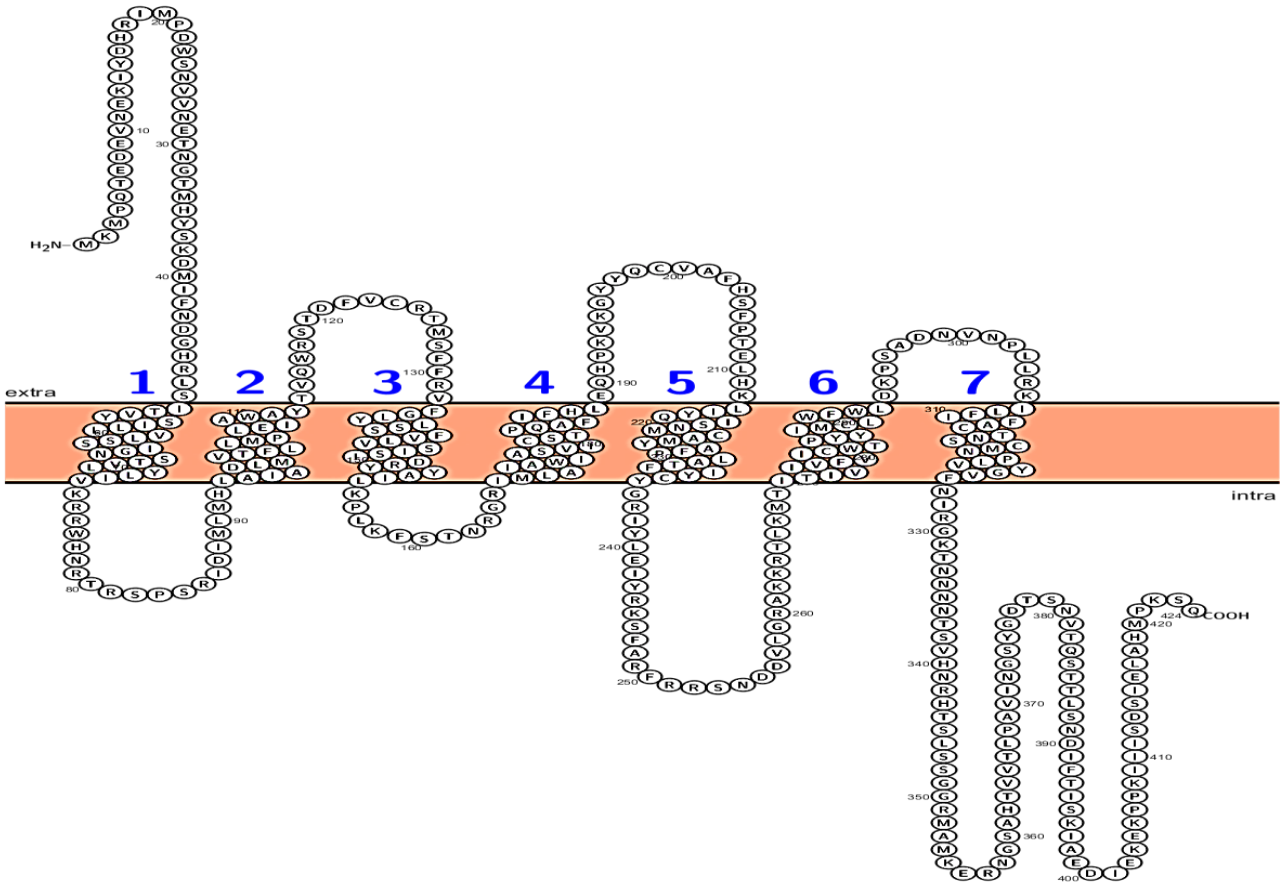


Figure 5.2: A 2D interactive receptor diagram showing the overall transmembrane topology of Bacdo-AKHR

Having confirmed the presence of the conserved residues motif in the Bacdo-AKHR, three secondary structure prediction programs were employed to help accurately predict the secondary structure and transmembrane region of Bacdo-AKHR. The programs are PSIPRED 4.0 (**Figure 5.3**), MEMSAT-SVM (**Figure 5.4**) and MEMSAT-SVM topology predictions (**Figure 5.5**)[342,353]. The result from the programs indicated that only 325 residues make up seven transmembrane (7TMs).

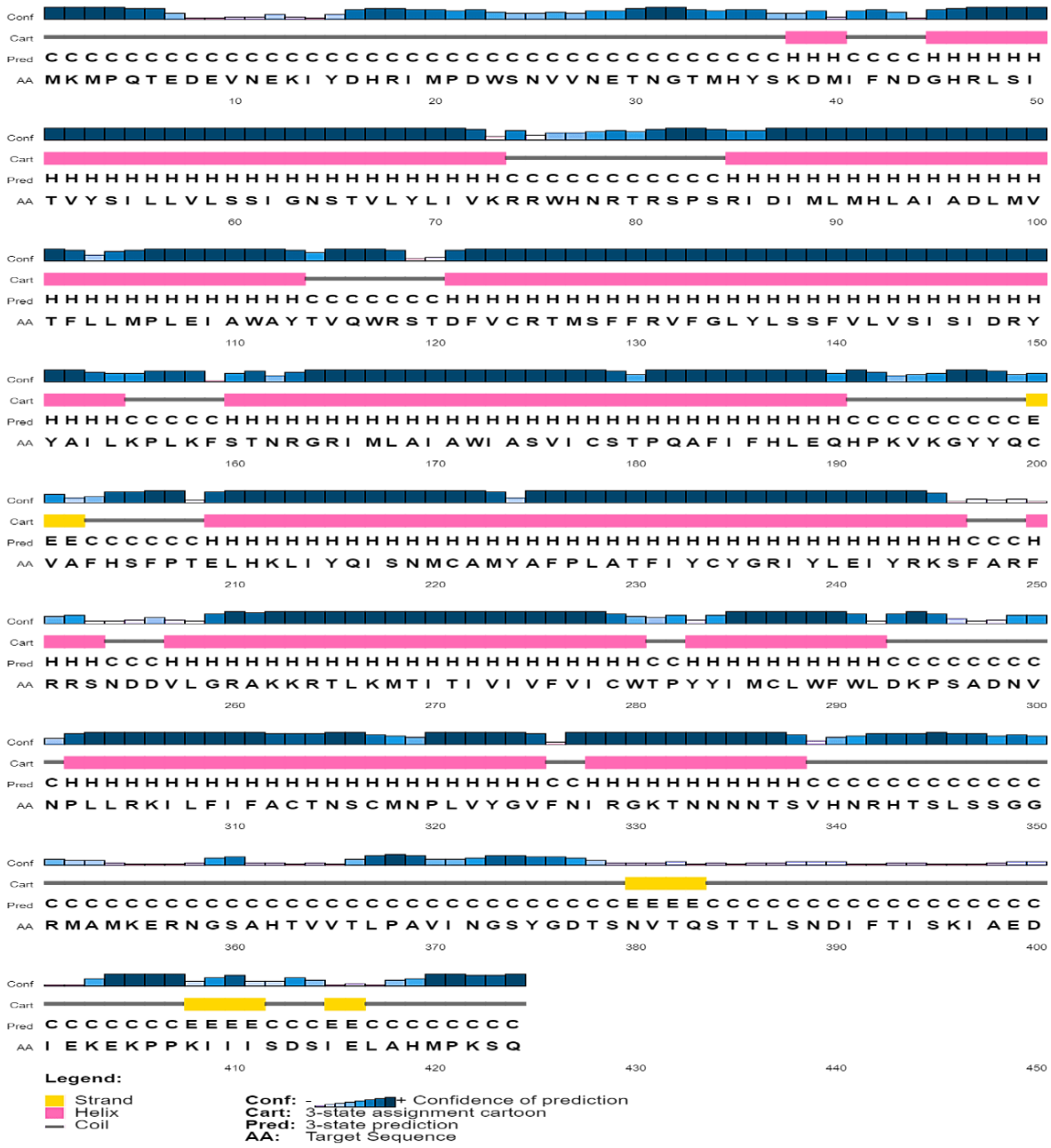


Figure 5.3: PSIPRED predicted secondary structure of Bacdo-AKHR H= helix (in yellow), S = strand (in pink) and C =coil (ash) Pred: Predicted secondary structure (H=helix, E=strand, C=coil), AA: Target sequence.

The predicted result indicated that the primary amino acid of Bacdo-AKHR when feed into the PSIPRED (**Figure 5.3**) *and* MEMSAT-SVM (**Figure 5.4**) programs. This further established that the amino acid sequences are distributed into seven transmembrane (TMs) helical bundles. Also, they possess three loops connected to the helices from the extracellular region, making up the N-terminus and another three loops connected to the helices from the intracellular region

making up the C-terminus. To further confirm this prediction, topology analysis was conducted see **Figure 5.4.**

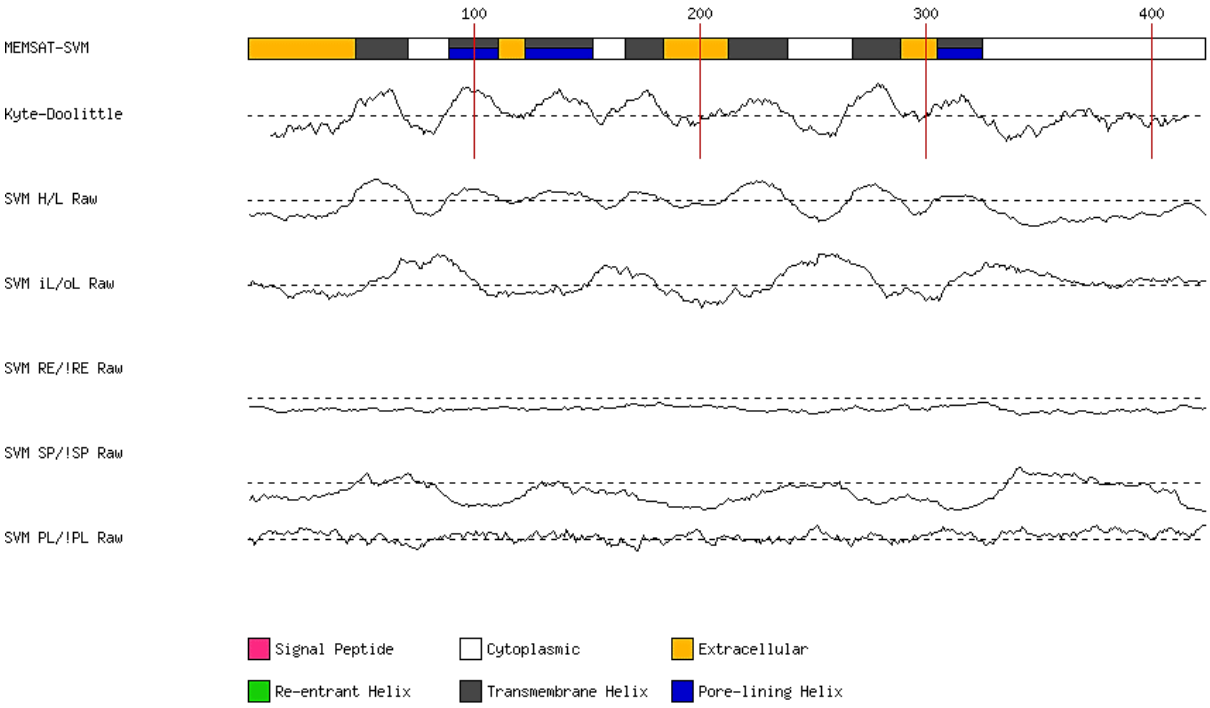


Figure 5.4: Schematic diagram of the MEMSAT-SVM predictions for the query sequence of Bacdo-AKHR. Traces indicate the RAW outputs for the prediction SVMs. Dashed lines indicate the prediction threshold. PL: Pore lining residue SP: Signal peptide residue RE: Re-entrant helix residue iL/oL & H/L: Helix prediction.

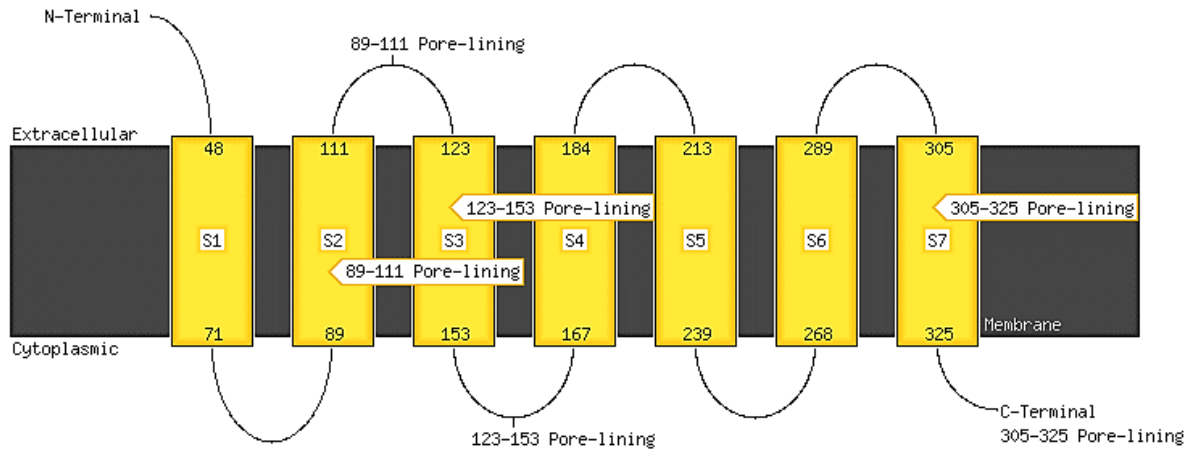


Figure 5.5: Prediction of transmembrane helices and the topology analysis of Bacdo-AKHR. The helices are represented in yellow and labelled S1-S7, the membrane (black) and the loops (thin black line) starting from the N-terminus (the extracellular region) and terminating at the C-terminus (the intracellular region)

The topology analysis predicted the helices as follows, H1(48-71), H2 (89-111), H3 (123-153), H4 (167-184), H5 (213-239), H6 (268-289) and H7 (305-325) see **Figure 5.5** The pore-lining was found to be at residues 89-111, 123-153 and 305 -325. The extracellular region comprises the primary amino acid residue at 44, 111, 123, 184, 213, 268, and 305 while the cytoplasmic region comprises residues 71, 89, 153, 167, 239, 268 and 325.

5.2.2 Homology modelling

Two templates were selected (same as those in chapter 3) for homology modelling, the crystal structure of beta 2 adrenergic receptor (5D5A.1A) and Rhodopsin (2X72.1A). These two templates were selected because the rhodopsin crystal structure has a covalently bound ligand and, as such, represents the active state of the receptor. The beta-2-adrenergic receptor (β2AR), on the other hand, does not have a bound ligand, and so represents the inactive state of the receptor. Clustalw2 was used to align the Bacdo-AKHR sequence to the templates as shown in **Figure 5.6**

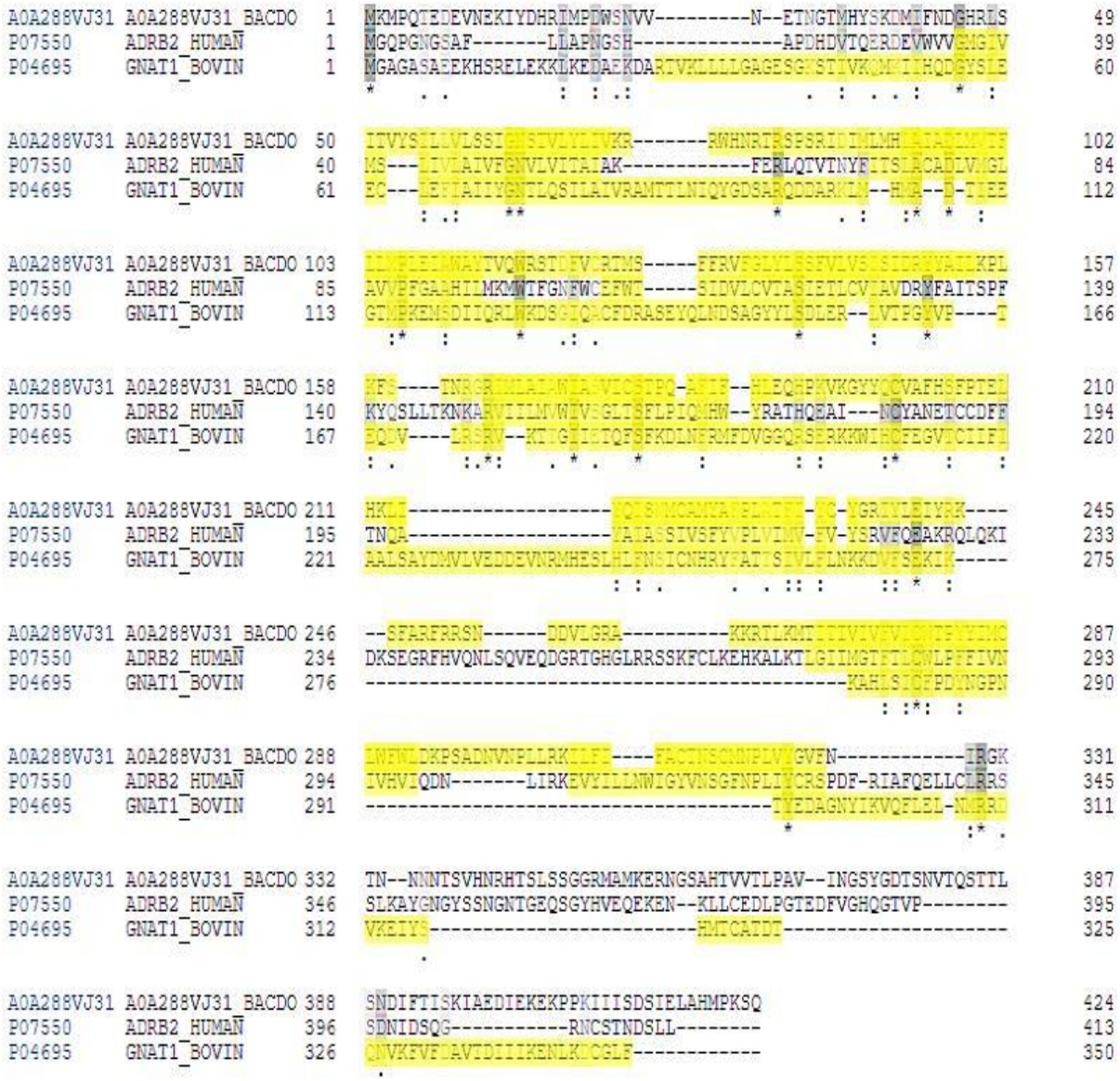
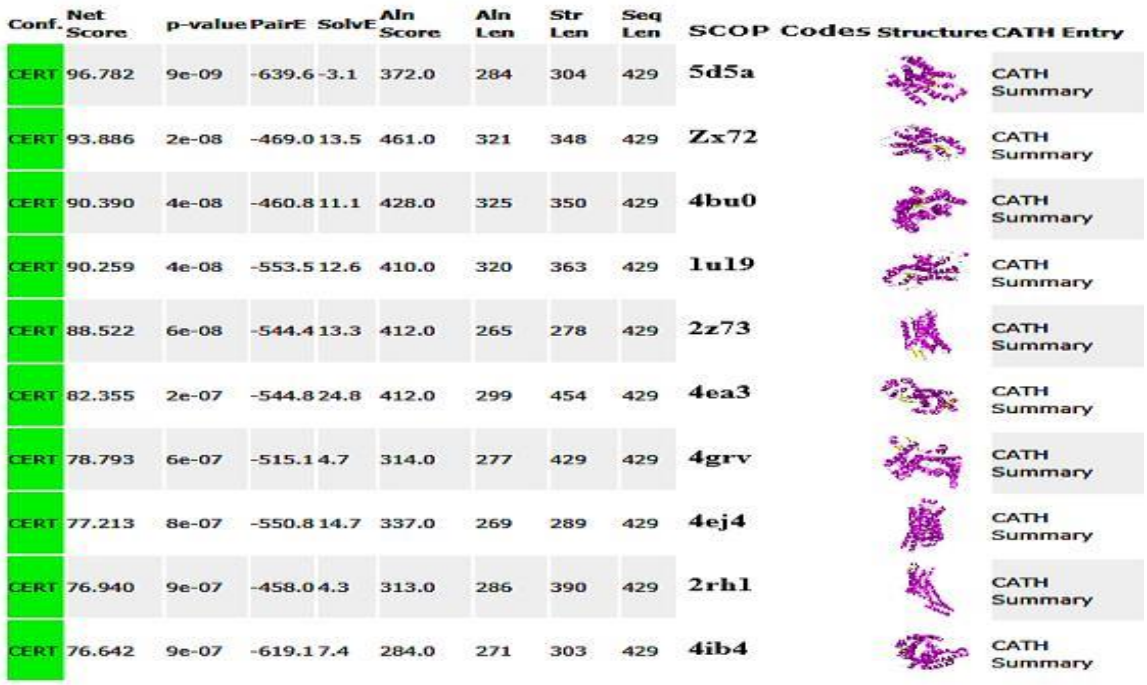


Figure 5.6: Clustalw2 sequence alignment illustrating Bacdo-AKHR and the two selected target templates β2AR, (5D5A.1A / P07550), and Rhodopsin (2X72.1A / P04695). The sequence similarity is highlighted in a black and white colour while the transmembrane domain highlighted in yellow.

The alignment (**Table 5.1**) showed a sequence identity ranging from 16% to 45% and sequence similarity between 52% and 88% for the different helices. Sequence alignments of the transmembrane residues of Bacdo-AKHR, with those of the target templates, β2AR, (5D5A.1A) and Rhodopsin (2X72.1A) (see **Figure 5.7**), Appendix A and B), displayed relatively low sequence percentage identity but possessed high sequence similarity (see **Table 5.1**). High sequence similarity and low in sequence percentage identity are not new to class A

GPCRs superfamily, because, with high percentage similarity, GPCRs / AKHRs do possess similar structural and functional features[298].



Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	SCOP Codes	Structure	CATH Entry
CERT	96.782	9e-09	-639.6	-3.1	372.0	284	304	429	5d5a		CATH Summary
CERT	93.886	2e-08	-469.0	13.5	461.0	321	348	429	Zx72		CATH Summary
CERT	90.390	4e-08	-460.8	11.1	428.0	325	350	429	4bu0		CATH Summary
CERT	90.259	4e-08	-553.5	12.6	410.0	320	363	429	1u19		CATH Summary
CERT	88.522	6e-08	-544.4	13.3	412.0	265	278	429	2z73		CATH Summary
CERT	82.355	2e-07	-544.8	24.8	412.0	299	454	429	4ea3		CATH Summary
CERT	78.793	6e-07	-515.1	4.7	314.0	277	429	429	4grv		CATH Summary
CERT	77.213	8e-07	-550.8	14.7	337.0	269	289	429	4ej4		CATH Summary
CERT	76.940	9e-07	-458.0	4.3	313.0	286	390	429	2rh1		CATH Summary
CERT	76.642	9e-07	-619.1	7.4	284.0	271	303	429	4ib4		CATH Summary

Figure 5.7: The statistical result from Psi-blast of the Bacdo-AKHR illustrating the common ancestry. The best two templates with net score 96.782 (β2AR, 5d5a) and 93.886 (rhodopsin 2x72) were selected for the construction of the active and inactive model of Bacdo-AKHR.

Table 5.1: Comparison of sequences of AKHR, Rhodopsin, and β2AR helices.

Helix	% Sequence identity with		% Sequence similarity with	
	β_2AR (β2off)	rhodopsin (Rhodoff)	β_2AR(β2off)	rhodopsin (Rhodoff)
1	17	22	65	63
2	25	16	58	52
3	29	30	68	60
4	28	40	76	80
5	23	17	65	60
6	34	23	83	74
7	44	45	88	88

The TM helix segments of both β2AR and Rhodopsin show high sequence similarity; these could be the reason for their prominent structural similarity[354,355]. The highly conserved amino

acid residues that are common to Class A GPCRs [190,223,356], are also present in the constructed models see **Figure 5.8.**

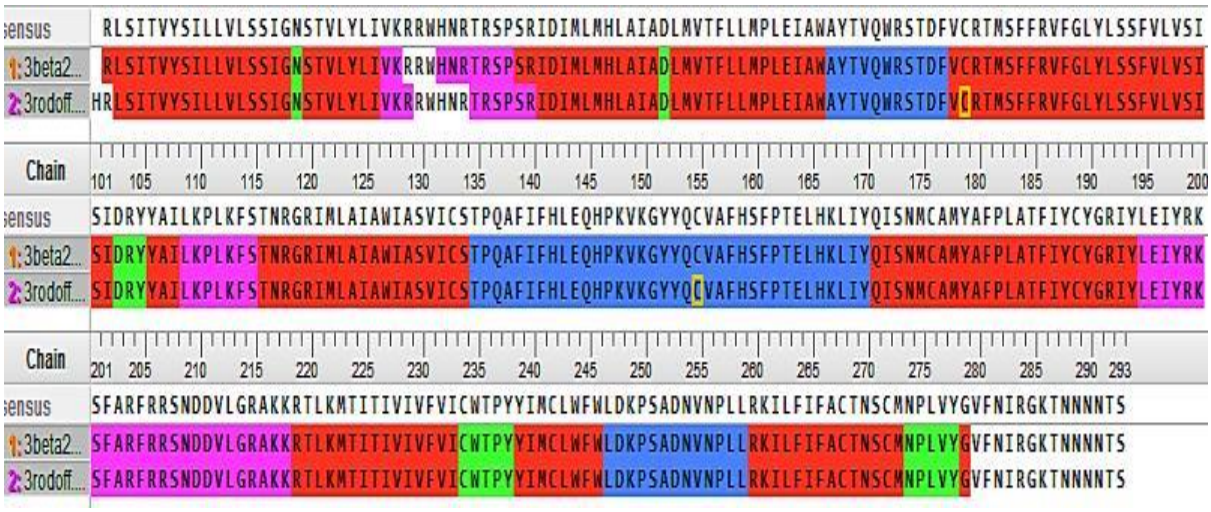


Figure 5.8: Schematic diagram representing an alignment of both models constructed from both Rhodopsin and beta2 adrenergic receptor of Bacdo-AKHR. Red indicates residues in the seven transmembrane helices, blue shows the extracellular regions, while the intracellular regions are represented in purple, yellow denotes region cystine ionic lock, and green shows the highly conserved residues.

Having established the similarities and alignments; homology modelling was then conducted with the two recognised template (β2AR 5d5a and rhodopsin 2x72) using the Swiss-Model server[330]. The models built are labelled β2off-model and Rhodoff-model (see **Figure 5.9)**.

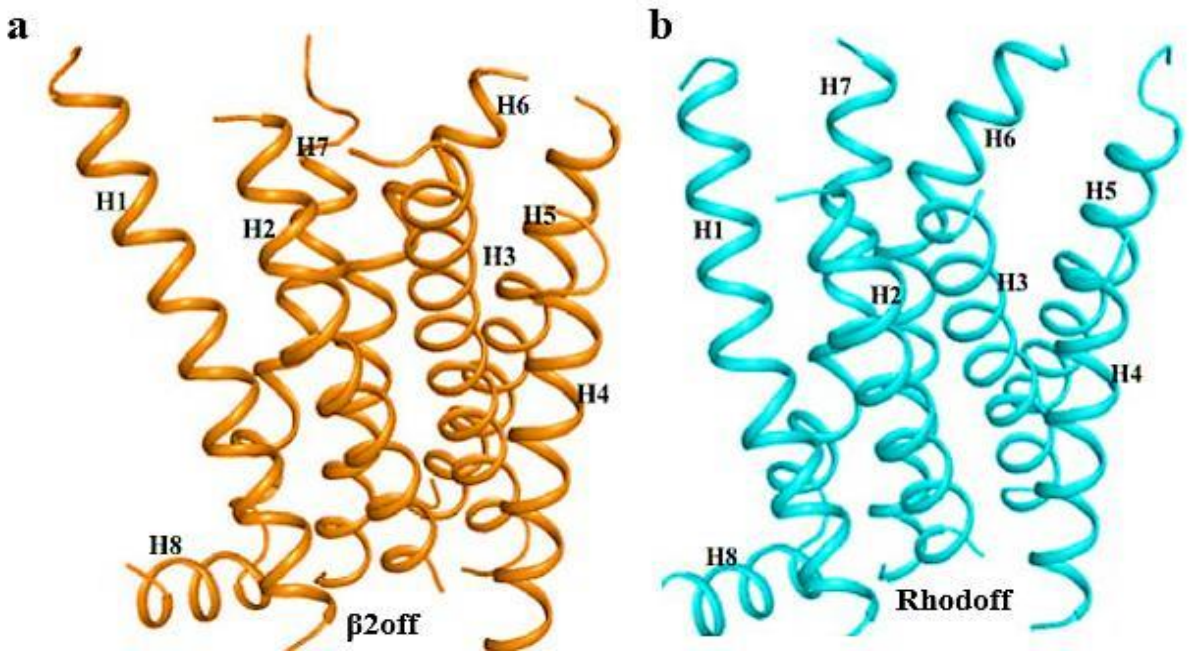
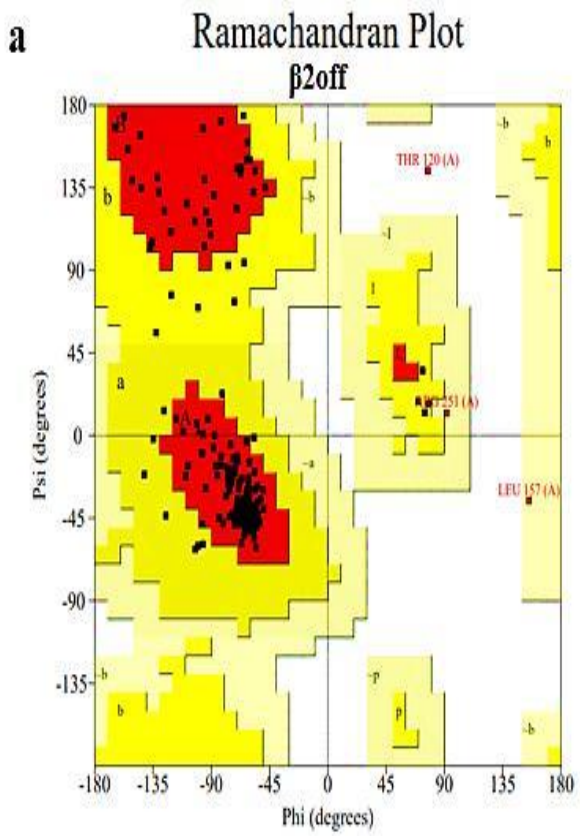


Figure 5.9: The constructed seven transmembrane helices of the Bacdo adipokinetic hormone receptor. a) β2off-model seven transmembrane helices of AKHR and b) Rhodoff-model seven transmembrane helices of AKHR.

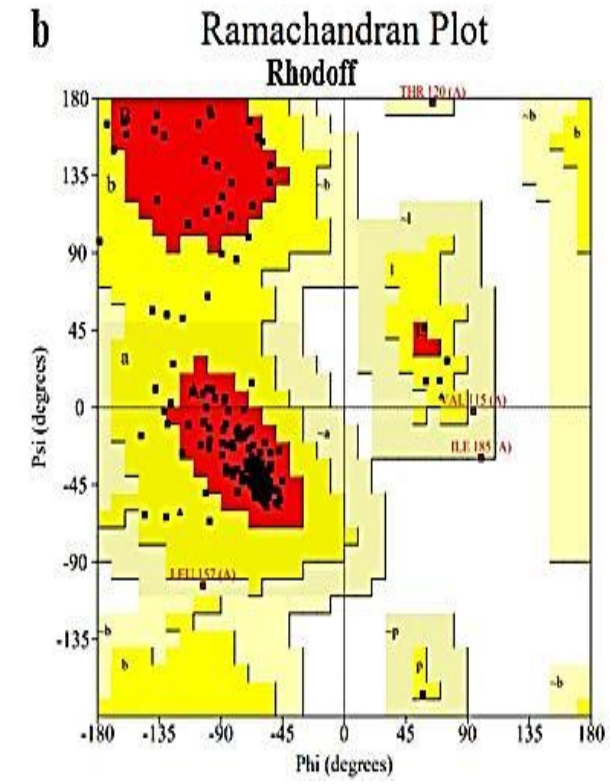
The overall structure of the constructed models (**Figure 5.9**) is alike, especially the anticlockwise arrangement of the transmembrane helices, and they both possess the eighth intracellular helix that is parallel to the cell membrane. Also, both structures have a tilted TM3, and the shortest transmembrane in both models is TM4.



Plot statistics

Residues in most favoured regions [A,B,L]	250	92.3%
Residues in additional allowed regions [a,b,l,p]	18	6.6%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.7%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	271	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	11	
Total number of residues	292	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.



Plot statistics

Residues in most favoured regions [A,B,L]	244	89.7%
Residues in additional allowed regions [a,b,l,p]	24	8.8%
Residues in generously allowed regions [~a,~b,~l,~p]	4	1.5%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	272	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	11	
Total number of residues	293	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 5.10: Ramachandran plots of a) β2off-model and b) Rhodoff-model.

The quality of the constructed models was assessed with evaluation programs, ERRAT[357] and PROCHECK[358]. Ramachandran plots of the two models are shown in (**Figure 5.10**). The result shows that most of the primary amino acid residue of the constructed models are either in allowed or favoured regions; also, phi and psi, which represent the backbone torsion angle, are acceptable. For the β2off-model, Thr120 had a φ angle of 14^0 outside the generously allowed region. In contrast, the ψ torsion angles 21^0 outside the allowed region. The ERRAT program was used to check the quality of constructed models (Appendix C and D). A score of 92.37 and 89.93 was calculated for the β2off-model and Rhodoff-model models, respectively, as the acceptable score range for a high-quality model is > 50 [357]. As such, we conclude that these two models are of a high-quality model and are acceptable.

5.2.2.1 Structural comparison of models

The constructed models of both β2off-model and Rhodoff-model can be compared to the two-class A GPCR structures, that were used as a template to build them. An overlay of the β2off -model with its template given in **Figure 5.11a** gave an RMSD of 3.3Å for superimposition of Cα atoms of the whole molecule and 2.01Å for the helix bundles. Also, an overlay of the Rhodoff-model with its template (see **Figure 5.11 b)** gave an RMSD of 3.1Å for superimposition of Cα atoms of the whole molecule and 2.03Å for the helix bundles. Overall, the target templates have longer TM helices when compared to the two constructed AKHR models, with β2AR having 392 amino acid residue [331] as opposed to the constructed β2off-model which has 325 amino acids. Also, the rhodopsin template model has 442 amino acid residues [224] as against the constructed Rhodoff –model, which has 330 amino acid residues.

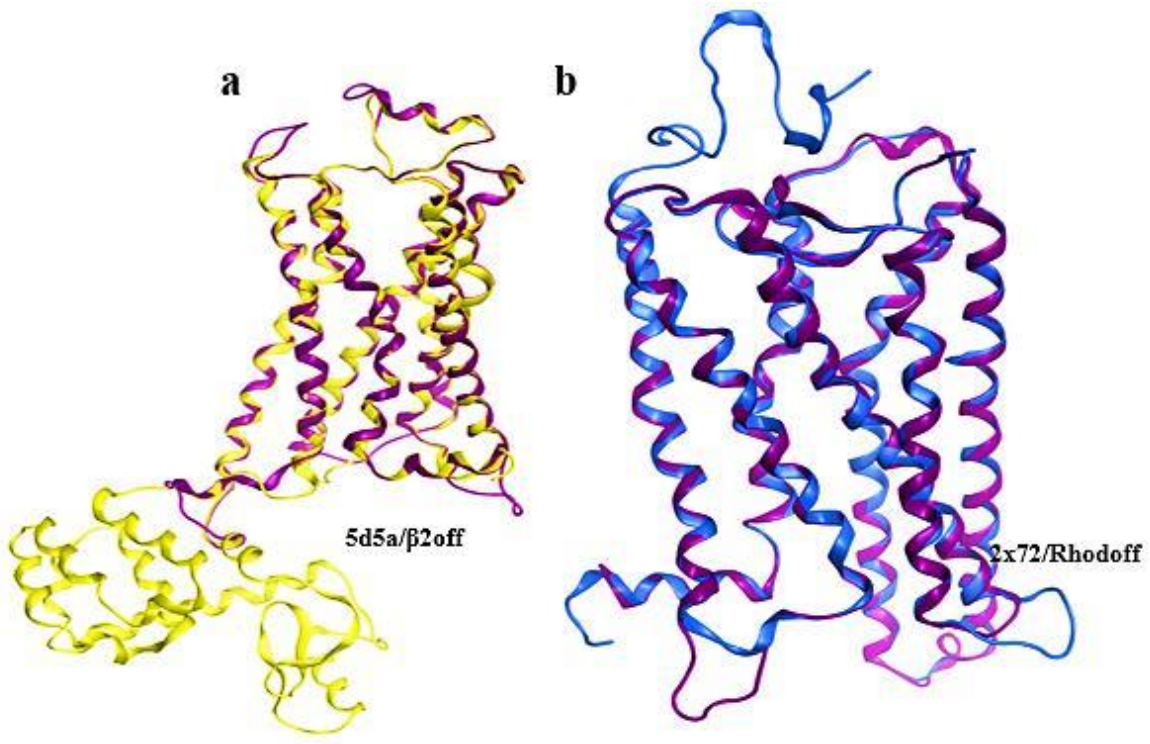


Figure 5.11. An overlay of β2off-model (the constructed model, in yellow) and 5d5a (the template used for constructing the model, in purple). b) shows an overlay of Rhodoff-model (the constructed model, in blue and 2x72 (the template used for constructing the model, in purple).

Both constructed models have the CWxP(Y/F) (CWTPY) motif in TM6; when compared with the templates, a slight difference in the amino acid residues was noticed. For β2AR, the CWxP(Y/F) motif is CWLPF while for Rhodopsin, it is CWLPY. This is the only significant difference noticed as regards the conserved residue. This motif is significant to receptor activations and will be discussed later.

5.2.2.2 Evaluation of the helix bundles

In both models, the transmembrane helices have an anticlockwise helical arrangement, when viewed from the extracellular side. This anticlockwise arrangement is found in most class A GPCRs [359]. Experimental data supported this anticlockwise arrangement of neurokinin 1 (NK-1) and κ-opioid receptor [359]. The helical bundles are stabilised by interhelical interactions comprising the highly conserved residues [223]

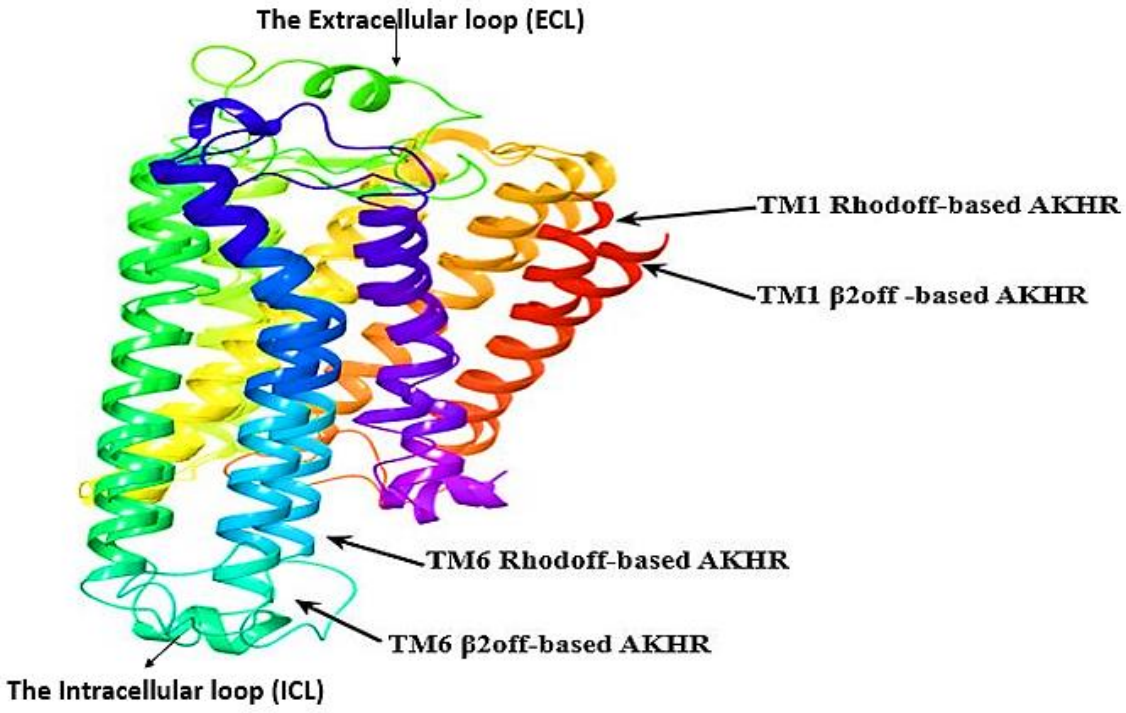


Figure 5.12: The superimposition of the β2off-model structure and Rhodoff-model structure of Bacdo-AKHR.

Superimposition of the two constructed 3D models gave a root mean square deviation (rmsd) of 3.36 Å for superimposition of Cα atoms of the whole molecule and 2.4 Å for the helix bundles. TM6 and TM5 differ from each other with respect to length with both helices extending by 4-3 residues into the cytoplasm. This observation is consistent with studies conducted on the accessibility of nitroxide labels fixed to ECL3 joining TM6 and TM5 [360]. Both structures possess a tilted TM3 conformation. This tilting of TM3 is also found other class A GPCRs model[93,98,195,361,362] The orientation of the seven transmembrane helices is subtly different in the two models, which leads to differences in the intracellular and extracellular domains. Also, both β2off-model and Rhodoff-model has a disulphide bridge, between Cys124[3-25] and Cys200, commonly found in GPCRs.

In β2off-model, the extracellular end of TM1 bends away with an angle of deviation 39.4º from the helical bundle (**Figure 5.12**). This deflection is not present in the Rhodoff-model.

Due to the bend of TM1 of β2off-model, a distance of 4.25Å was recorded between the end of TM1 helices of the superimposed structure. This gives a more open structure to the extracellular domain of the β2off-model when compared to the Rhodoff-model. At the same time, the intracellular end of TM6, a distance of 4.7 Å was recorded (**Figure 5.12**) relative to its position in the Rhodoff-model towards the helical bundle. This gives a more closed structure to the intracellular domain of the β2off-model. Thus, the β2off-model is described as the open or inactive model, while the Rhodoff-model is termed the closed or active model [239]. The closed conformation of the Rhodoff-model is similar to the rhodopsin crystal structure, where the close nature protects cis-retinal from hydrolysis [96,98]. Since ligand binding takes place on the extracellular domain [363], the closed nature of the Rhodoff-model prevents the ingress of the ligand. On the other hand, the open nature of the β2off-model allows ligand-free access to the receptor-binding pocket.

5.2.2.3 Overview of transmembrane helical interaction and molecular switches

A critical feature of most class A GPCRs is the existence of substantially preserved molecular switch motifs. The presence of these switches is essential for stabilising any GPCR either in the close (active) or open (inactive) state. In Bacdo-AKHR, a 3-6 lock exists between Tyr150$^{3\text{-}51}$ on TM3 and Arg 264$^{6\text{-}29}$ on TM6. In the inactive model, Arg 264$^{6\text{-}29}$ points away from TM3, such that no interaction is possible (**Figure 5.13 a**). Upon activation, however, TM6 twists moving Arg 264$^{6\text{-}29}$ so that it now points toward Tyr150$^{3\text{-}51}$ (**Figure 5.13 b**). This interaction locks the receptor in the active state. **Figure 5.13 b** also shows that, in the active state, Arg 264$^{6\text{-}29}$ on TM3 interacts with Glu 248 of ECL2. This loop closes over the binding pocket after ligand binding.

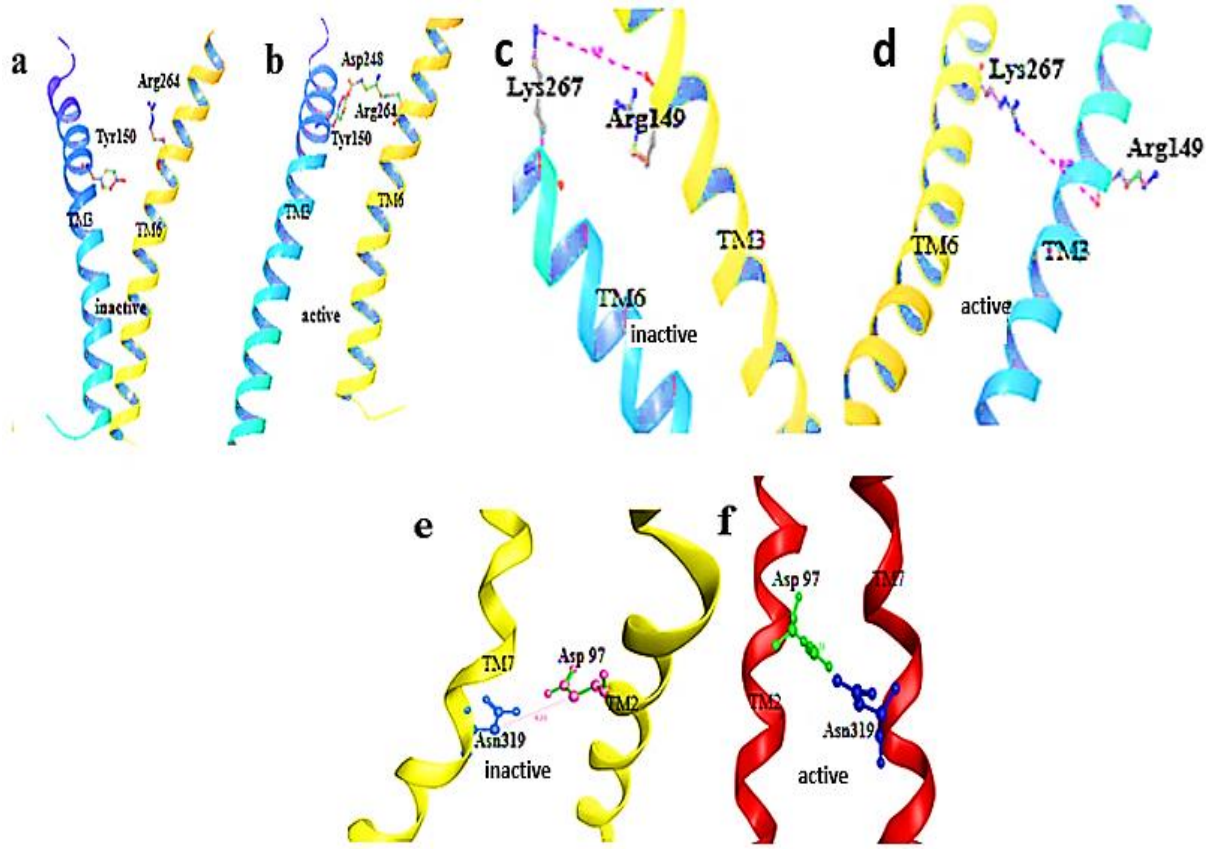


Figure 5.13: a) Open molecular switch in the β2off (inactive) model between Tyr 150 in TM3 and Arg264 in TM6. b) Closed molecular switch in the Rhodoff model (active) between Tyr 150 in TM3 and Arg264 in TM6. c) molecular switch interactions (inactive) in the β2off model between Lys267 in TM 6 and Arg149 in TM3. d) molecular switch interactions (active) in the Rhodoff model between Lys267 in TM 6 and Arg149 in TM3. e) molecular switch, interactions (inactive) in the β2off model between Asp97 in TM2 and Asn319 in TM7. f) molecular switch interactions (active) in the Rhodoff based between Asp97 in TM2 and Asn319 in TM7.

In the Rhodoff-model, there is an ionic interaction between Lys267$^{6\text{-}35}$ in TM6 and the positively charged side chain nitrogen atom of Arg149$^{3\text{-}50}$ in TM3 (**Figure 5.13 d**). This interaction does not exist in the β2off-model AKHR (**Figure 5.13 c**). These switches are found to be the same as those described for the AKH-R receptor of *A. gambiae, Rhodnius prolixus, and Tribolium castaneum* [231,366]. Class A GPCRs have a conserved aspartate on TM2 and an asparagine on TM7. In the Rhodoff-model structure, there is an intramolecular interaction between Asp97 and Asn319 (**Figure 5.13 f**). In the β2off-model(**Figure 5.13 e**), the helices shifted such that these two residues are now 3.4Å apart and can no longer interact [239].

Experimental data backs the importance of these intramolecular interactions because of the vital role they play during receptor activation, basal activity and conformation flexibility reduction [239].

In the Rhodoff-model (**Figure 5.14a**), Asp 97 [2-50] has a strong ionic interaction of 2.58 Å with Ser138 [3-39]. However, in the β2off-model this interaction is absent as the side chains of Asp 97 [2-50] and Ser138 [3-39] are orientated away from each other leading to an internuclear distance of 5.11 Å. Also, in the closed, Rhodoff-model, there is polar interaction between Asn60 [1-50] and Ser316 [7-45] (**Figure 5.14 b**), while in the open, β2off-model the interaction is between Ser316 [7-45] and Asn 64 [1-54].

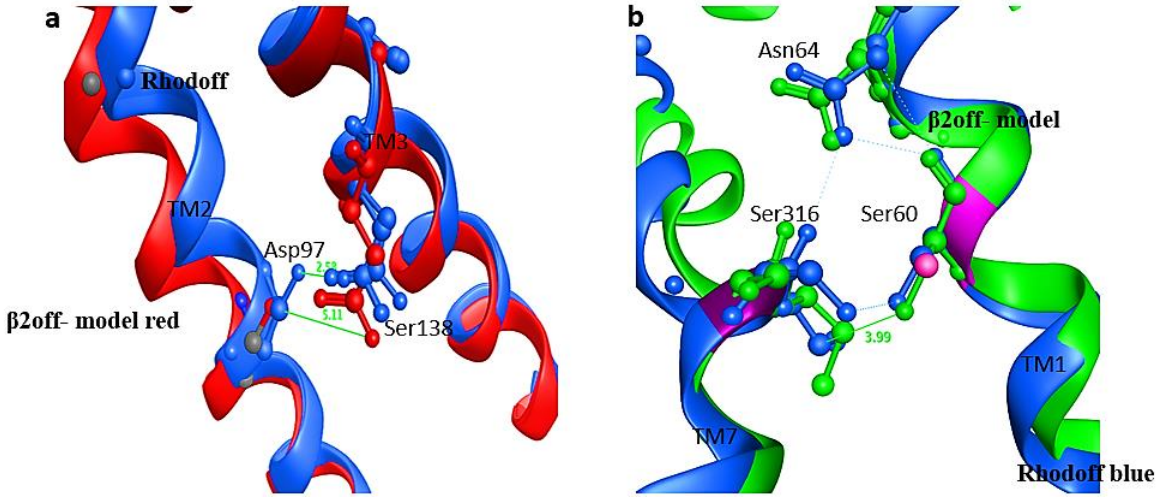


Figure 5.14: a) Rhodoff model (blue) with a salt-bridge interaction between Asp 97 and Ser138. In the β2off model (blue) this interaction is absent. (b) Rhodoff model (blue) a polar interaction is identified between Asn60 and Ser136, while in the β2off model (green) the polar interaction is identified between Asn64 and Ser136.

In the β2off-model there is a π-cationic interaction between Tyr239[5-70] and Lys263[6-29] (**Figure 5.15 a**). In the Rhodoff-model the π-cationic interaction is between Lys262[6-28] and Tyr243[5-74] (**Figure 5.15 b**)

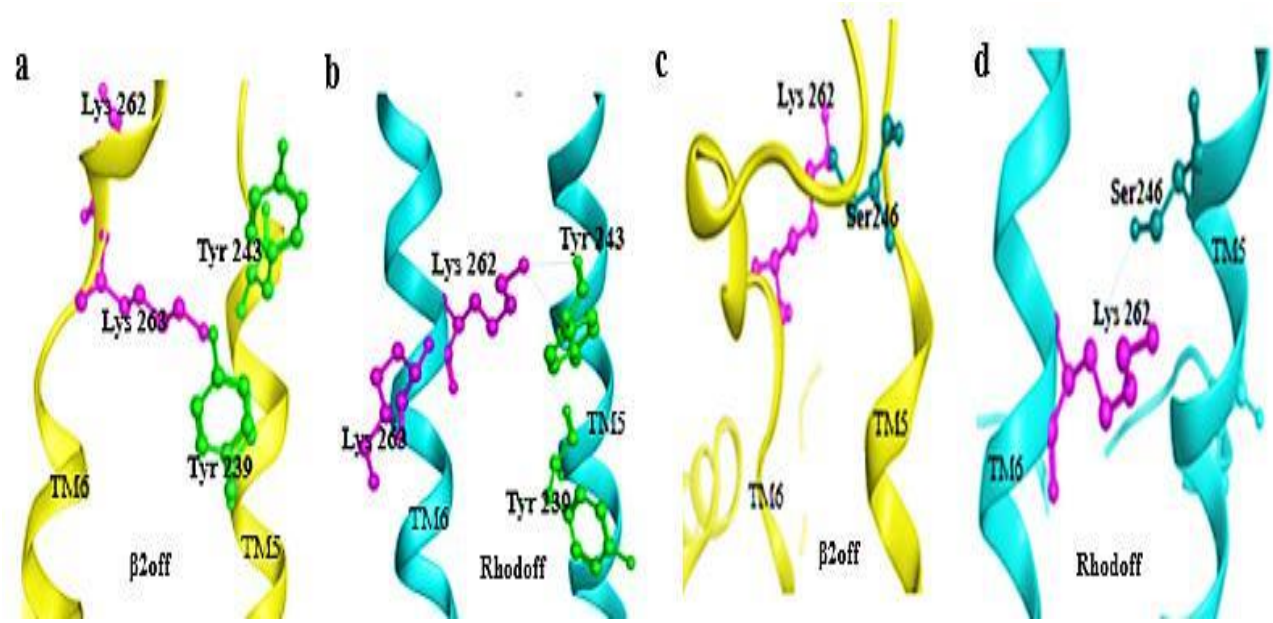


Figure 5. 15: a) π-cationic interactions between TM6 Lys 263 and TM5Tyr 239. b) π-cationic interactions between TM6 Lys 262 and TM5 Tyr 243. c) H-bond interactions betweenTM6 Lys262 and TM5 Ser246. d) absence of H-bond interactions between TM6 Lys262 and TM5 Ser246.

An intramolecular interaction was also identified in the open structure, between TM6 Lys 262 and TM5 Ser246 (**Figure 5.16 c**). This interaction was not found in the closed structure (see **Figure 5.15 d**). Dror *et al.,* 2009, discovered that the inactive β2AR exists in equilibrium between conformations with the ionic lock 'on and off'[367]. This could be the case with the β2off-model AKHR structure, which does not have the ionic lock between helices 3 and 2 but between helices 5 and 6. These intramolecular interactions are responsible for the reduction in conformational flexibility, basal activity and play a vital role in receptor activation [239]

5.2.2.4 Evaluation of kinks

Experimental data backs the existence of kinks that initiate the helical twist (bend) at almost the same locations where Tyr, Thr or Met come before Pro [195,368–370]. The TM helices in both open and closed AKHR models have these proline-generated bends or kinks (**Figure 5.16**). TM2 a kink between Met 105 [2-58] and Pro106 [2-59] (**Figure 5.16 a**), TM5 has a kink between Tyr224 [5-47] and Pro227 [5-50] (**Figure 5.16 b**). Another turn was noticed in TM4 in Val 176 [4-54]

for the β2off-model (**Figure 5.16 c**) although not found on the Rhodoff-model structure (**Figure 5.16 d**). from the extracellular region, the kink in TM 6 instigates a twist near TM 5. An overlay of the β2off-model and Rhodoff-model **Figure 5.16** shows the different effect of these kinks or bends on the two models.

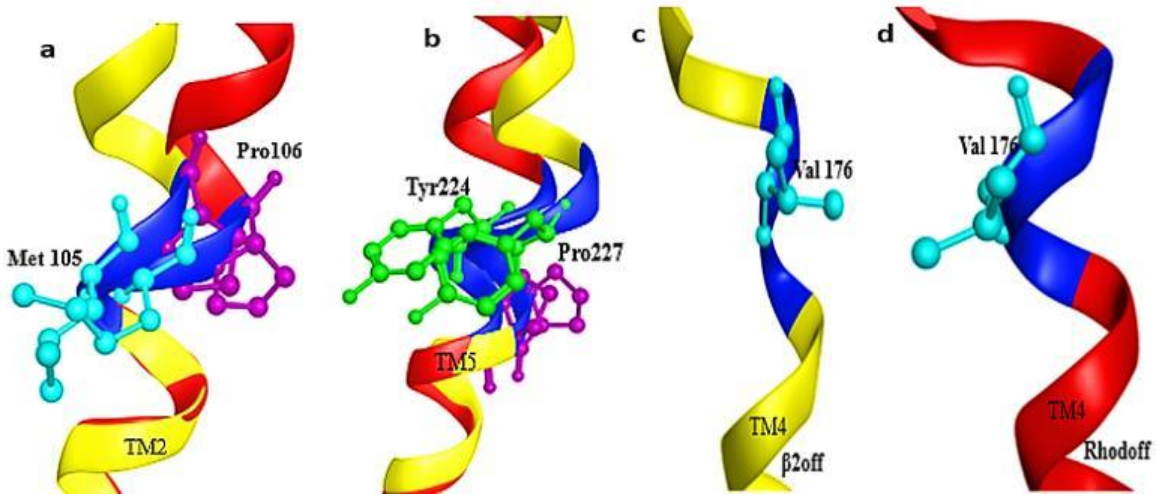


Figure 5.16: a) Overlay of 3D models of the β2off-model and Rhodoff-model, showing the presence of kinks in TM2. b) overlay of 3D models of the β2off-model and Rhodoff-model showing the presence of kinks in TM5. c) β2off-model showing the presence of kink in TM4. d) Rhodoff-model, showing the absence of kink in TM4. * β2off = β2off-model, Rhodoff =Rhodoff-model.

5.2.2.5 Analysis of the loop regions

Besides the orientation of the transmembrane helices, there are other differences between the two models (**Figure 5.12**). In the extracellular loop I (ECL I), the β2off-model has no β-strand, while the Rhodoff-model has a β-strand between Ala112 and Gln116 and a small helical turn at the N-terminus. In the ECL2 of the Rhodoff-model there is a β-strand between Ser119 and Phe112, and a γ -turn between Trp117 and Arg118. This keeps ECL2 in an upright position above the helical bundle. In this way, the N-terminus and ECL2 extend over the transmembrane helices, shielding all possible binding sites and preventing the ligand from gaining access to or leaving the binding pocket. In the ECL2 of the β2off-model, there is a β-strand involving Glu182 and 190 and a small helical turn between His192 and Cys200. This

moves the ECL2 away from the binding pocket. At the same time, the N-terminus also moves to give any potential ligand access to the receptor. This opens access to any potential binding pocket.

ICL3 is commonly not characterised in the crystal structures of GPCR. In both the β2off-model and Rhodoff-model, ICL3 has a helical turn between Ile292 and Lys294. Dror *et al.* identified this turn in β2AR [367]. Moreover, the β2off-model has a helical turn between Ser247 and Ser 253, while the Rhodoff-model has a β-strand between Val300 and Asn301 and a helical turn between Pro302 and Leu304. This findings are consistent with those identified and reported by Dror *et al.* [367]

Most Class A GPCRs have an eighth helix, which lies parallel to the intracellular membrane surface. This helix is thought to be essential for receptor activation and binding of a G-protein [239]. Most GPCRs intracellular region are stabilised by palmitoylation of Cys, Ser, or Thr on the 8th helix [372–374]. Both the β2off-model and Rhodoff-models have this 8th helix, which has Cys313, which could be stabilised by palmitoylation.

Since GPCRs are highly flexible, the different structural features, particularly in the extra and intracellular region of the two models, could represent features in different conformations during molecular dynamics and in the active and inactive state of the receptor.

5.3.0 A general overview of the constructed models

The constructed AKHRs from chapters 3, 4 and 5 all have similar properties. Alignment of their primary amino acid sequence along with their templates, shows a high similarity, particularly with their transmembrane helices, see **Figure 5.17**.

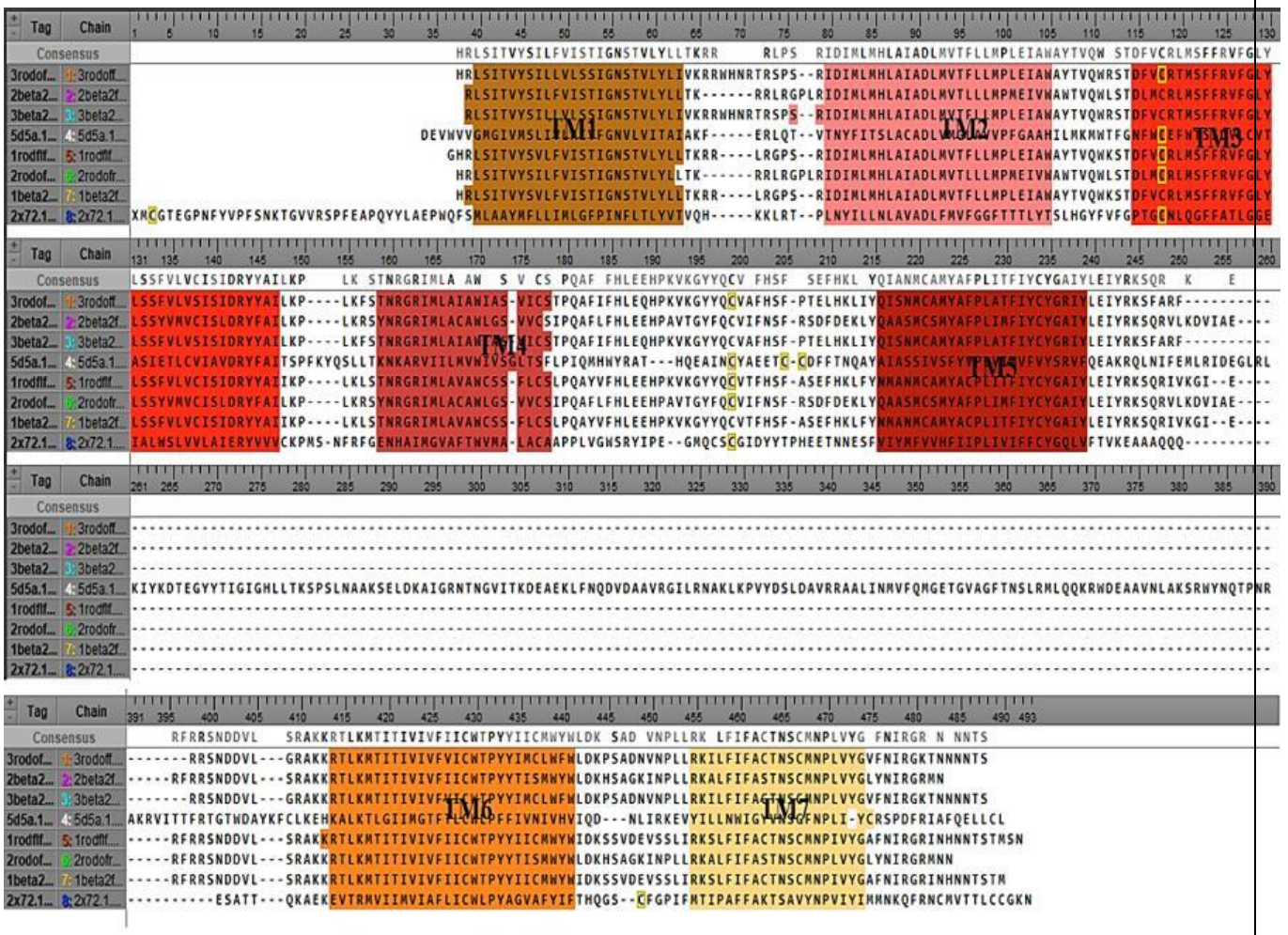


Figure 5.17: An alignment of the three constructed models and their template structure is showing helical similarities * 1rhodflf= rhodopsin based model of the flesh fly, 1β2flf = beta 2AR based model of the flesh fly, 2rhodfrf = rhodopsin based model of the fruit fly, 2β2frf= beta2 AR-based model of the fruit fly, 3rhodoff = rhodopsin based model of the oriental fruit fly, β2ofrf = β2AR based model of the oriental fruit fly, 5d5a = template β2AR and 2x27 = template rhodopsin model.

Further structural alignment of the constructed β2AR models (**Figure 5.18 a**) alongside the β2AR template (**Figure 5.18 b**) shows a perfect similarity. Also, the structural alignment of all three constructed Rhodopsin models (**Figure 5.18 c**) shows a good similarity when aligned with their template structure (**Figure 5.19 d**).

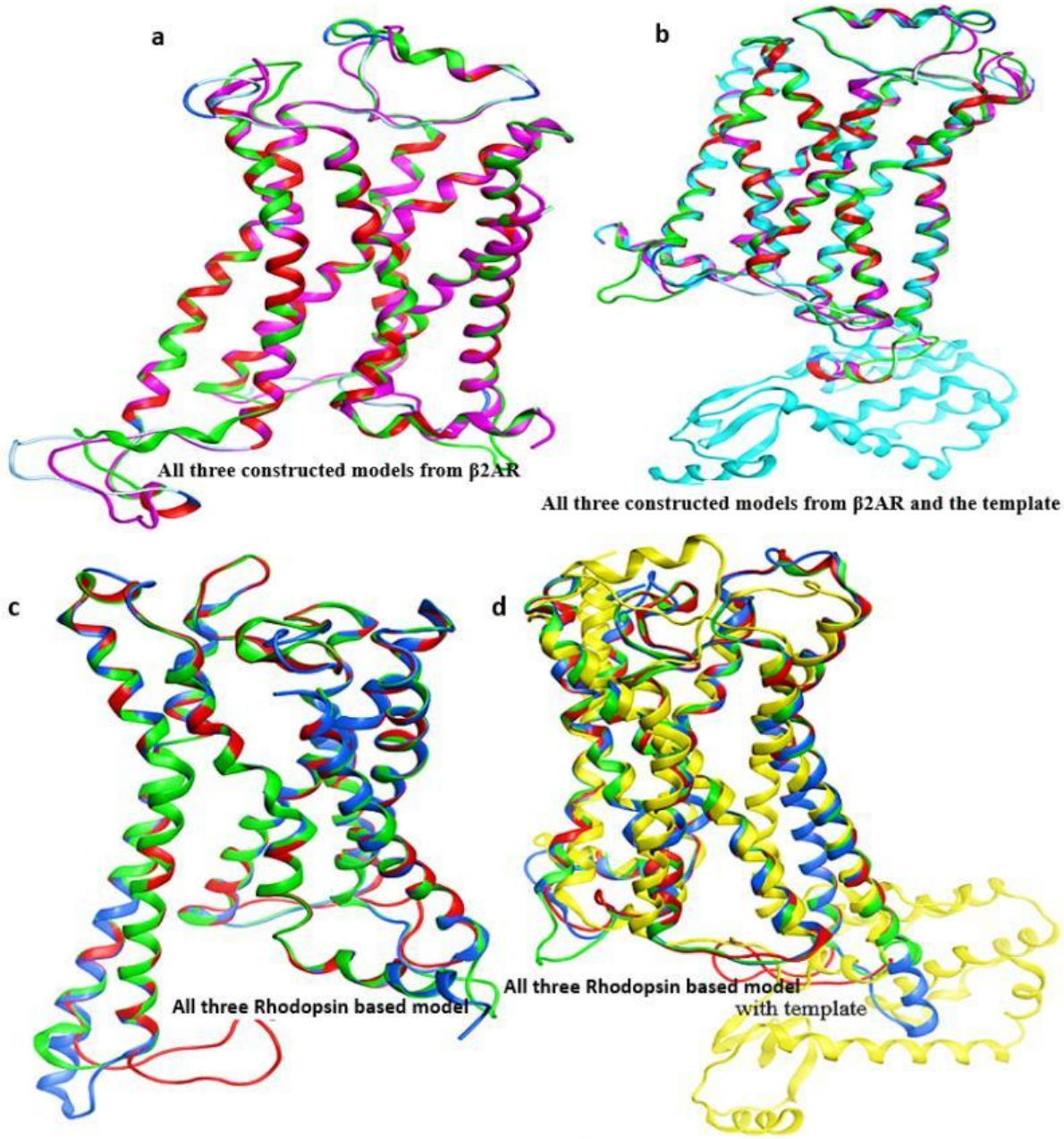


Figure 5. 18: a) An overlay of all three constructed model from β2AR. b) an overlay of all three constructed model from β2AR and the template β2AR. c) an overlay of all three constructed model from the Rhodopsin model. d) an overlay of all three constructed model from the Rhodopsin model and the Rhodopsin template model.

The alignment and similarities of these models could mean a close similarity in the point of activation during molecular docking calculations. Understanding the similarities and differences of the models could give proper insight into other AKHRs that their 3D models are yet to be elucidated.

5.4.0 Conclusion

We have successfully constructed 3D AKHR models for the oriental fruit fly AKH-receptor, Bacdo-AKHR and we present the open conformation as the inactive state and the closed conformation as the active state of the AKHR. The elucidated structures possess the same features as most class A GPCRs.

The structural differences can be found around the highly conserved amino acid residue Asp, Arg and Tyr (DRY) motif. Where the Rhodoff-model posses an ionic interaction between TM3 Arg149$^{3-50}$ and the backbone C=O of TM6 Lys267$^{6-35}$ but, this was not found in the open model.

π-cationic interaction was noticed in the closed model between TM2 Asp 97 $^{2-50}$ and TM3 Ser138$^{3-38}$. Also, the closed model has ionic interaction between TM6 Lys262$^{6-28}$ and TM5 Tyr243$^{5-74}$. While the open model has two, the ionic interaction is between TM5 Ser 246$^{5-70}$ and TM6 Lys 262$^{6-29}$ and the other between TM5 Tyr239$^{5-70}$ and TM6 Lys263$^{6-29}$

The closed model has a polar interaction between TM1 Asn60 $^{1-50}$ and TM7 Ser314 $^{7-45}$, but no polar interaction between TM1 Asn60 $^{1-50}$ and TM7 Ser314 $^{7-45}$ was found in the open model.

In both models, kinks were also identified in TM2 (Met 105 $^{2-58}$ and Pro106 $^{2-59}$), TM5 (Tyr224 $^{5-47}$and Pro227 $^{5-50}$) and TM4 Val 176 $^{4-54}$.

The general conformation and structural features of the β2off-model and Rhodoff-model could be used to study the binding and activation of the Bacdo-AKH receptor by Phote-HrTH.

The constructed AKHRs from chapters 3, 4 and 5 all have similar properties. Alignment of their primary amino acid sequence along with their templates, shows a high similarity, particularly with their transmembrane helices. The identified similarities of these models could mean a close similarity in the point of activation during molecular docking calculations.

6

Binding Pocket Identification of AKHRs, Molecular Docking Calculations, and MD Simulations of AKHRs in a Membrane

Summary

Using blind docking protocols, similar binding sites were identified at the extracellular half of the fruit fly AKHR, the oriental fruit fly and the flesh fly. Phote-HrTH (pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2) was used as the ligand during molecular docking. As predicted, the binding cavity was at the extracellular side of the receptor. It was easy for the ligand to access the binding cavity of the inactive β-2AR models. However, the Rhodopsin-based model generated no poses.

To determine the mode and actual changes during receptor-ligand binding for all three receptors, molecular dynamics simulation was executed, after the molecular docking calculation. Separately, the docked ligand-receptor complex for each fly species was placed into a membrane mimetic (POPC).

During the molecular dynamic simulations, the ligand was observed to sink deep into the receptor. After 1.5 ns of MD simulation, conformal changes were observed in each of the AKHR models. The helices of the receptors moved toward each other on the extracellular side. At the same time, they moved away from each other on the intracellular sides in a scissor-like action.

These changes were much pronounced in TM3, TM5, and TM6, respectively.

For all three AKHR-models, the AKHR-ligand complexes are stabilised by a strong network of H-bonds, π-stacking / π-cationic interaction and salt bridges. In each of the three models, the conserved residue motif CWTPYY in TM6 was identified to interact with the ligand during simulation binding; this could be vital for receptor activation.

6.1.0 Introduction

Molecular docking is a vital tool in computer-aided drug design (CADD) and structural biology. The primary aim of molecular docking is to predict the binding mode or binding site in a receptor molecule[187,202,383–385]. In drug design and discovery, molecular docking is executed as a means to cut down time, and to predict the best conformation of specific molecules, thereby reducing the number of molecules to be analysed during experimental procedures [302,386–389].

The first stage involves three processes. The first stage involves the cavity/ site identification, generation of pose around the binding pocket and the selection of best pose/ conformation at the binding site to mention but a few[196,390]. The other stages are examining[391,392] and scoring of the poses[383,393,394]. In situations where a homology modelled molecule binding cavity/ site is not known, it first has to be identified using other programs[191,395,396]. Hence, programs like Glide protocol from the Schrodinger package [397,398], AutoDock [399–403] program or the molecular environment (MOE)[404] program could be employed for both molecular docking and binding cavity/ site identification.

Molecular docking uses the three-dimensional structure of the target protein.[384,402] During free association of receptor-ligand complex, identification of the receptor cavity (active pocket) is dynamic due to conformational changes (poses) of both ligand and receptor, which is the only way to exploit total free energy by both ligand and receptor[187,206]. Molecular docking algorithms are designed such that a large number of poses can be generated within a short time. Example of molecular docking programs with such features include DOCK[390], FTDOCK [128,405], FlexX and QSDOCK [405,406]. DOCK, FTDOCK [128,405], FlexX, and QSDOCK programs can only be used when both receptor and ligand are rigid [390]. There are other programs designed for cases where both ligand and receptor are flexible or in a case where one is flexible, and the other is rigid, these programs include AutoDock [399,407,408], DARWIN[409] and GOLD [405,410].

A growing number of docking protocols have incorporated protein flexibility [196,383]. Determination of protein flexibility using energy minimisation during post-docking refinement helps to optimise dihedral angles. Studies have shown that receptor-ligand binding stimulates numerous conformational changes in the receptor which comprises side-chain rearrangement and pivot movement of the C-domain in some cases [196,410–412]. The assessment of relative binding affinity follows post-docking calculation of binding energy, and this is using Molecular Mechanics terms such as Generalised Born Surface Area (MM-GBSA) [413–418].

In this thesis, Prime MM-GBSA[413,417] was applied, using an implicit solvent model with a protein-ligand complex. The following equation (**2.13**) defines the binding affinity of the protein-ligand complex:

$$\Delta G_{bind} = G_{(PL)} - G_{(P)} - G_{(L)} \qquad \textbf{2.13}$$

Where the binding free energy is represented by ΔG_{bind}, protein-ligand complex is represented by $G_{(PL)}$, protein $G_{(P)}$ and ligand $G_{(L)}$. Relative binding affinities are calculated and are recognised to relate to the free energy of binding. The more negative the binding affinity is, the stronger the binding of the ligand to the receptor protein. The MM-GBSA free energy of binding is calculated according to equation **2.14**[414,416,419–421].

$$\Delta G\ bind = \Delta G_{solv} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib} \qquad \textbf{2.14}$$

Where ΔG_{solv} represents the impact of the solvent due to solvent effects, the simulated conformational changes in the receptor/ligand is represented by ΔG_{conf}. ΔGint represents the calculated free energy from the protein-ligand interactions. ΔGrot represents the free energy lost by freezing internal rotations of the protein and ligand complex. ΔG t/r denotes the loss in

rotational and translational free energy of the ligand and receptor as they interact to give the bound complex, and the vibrational mode free energy is denoted by ΔG_{vib}.

AKHRs/ GPCRs are highly flexible; as a result, they display multiple and dynamic structural properties [296]. GPCR models have been generated using MD simulation in vacuum for drug-related designs, but this does not represent the actual environment of the protein. To represent the actual biological environments of these GPCRs, one could depend on recent literature on several membrane-embedded human GPCR models [297,422–424]. Phospholipid bilayers represent the fundamental structures of most biomembranes [425]. Periodic lipid bilayers involving, for example,1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphocholine (POPC), dipalmitoylphospha-tidylcholine molecules (DPPC), dodecyl phosphocholine (DPC) and solvated with water, have been simulated and equilibrated [422]. POPC, DPC and DPPC are universally used in the MD simulation of GPCRs [426]. To that effect, MD simulations of both free and bound AKHR inserted in a preequilibrated POPC membrane model were performed.

In this study, the homology models of the AKH receptors of three fly species are presented in Chapter 3, 4 and 5, and the 3D conformation of the ligand is discussed before that.

6.2.0 Experimental Methods:

6.2.1 Molecular docking Calculations

6.2.2 Identification of the binding cavities for the receptors

To identify the binding cavities and the best binding pocket, Blind Docking (BD) [396,400,427] was performed using *AutoDock4.2*[428] and *MOE* version 2014.0901,[429] and confirmed using *SiteMap*.[430] these tools are highly recommended for both binding pockets identification[191,396,400,430–435] molecular docking[187,202,383–385].

The ligand used for this study is Phote-HrTH (pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2). Phote-HrTH have been reported to regulate the supply of fuel metabolites for oxidation in the flight muscles of insects (fruit fly, flesh fly and the oriental fruit fly) [43–45,436]. The solution structure of Phote-HrTH was discussed in Chapter two. Also, the receptors used are of the fruit fly, flesh fly and the oriental fruit fly. The elucidation of the 3D models of the three fly AKHRs are reported in chapters three, four and five.

The ligand was prepared by setting the number of active torsions and degrees of freedom. Docking parameters for the ligand and macromolecules were set, as described by Hetényi *et al.* [191,396,437,438]. Only the β2AR-based receptor models were used in the determination of the binding sites of each constructed receptor model. After that, the accessibility of the binding cavities from the extracellular domain was conducted as reported in the binding cavity of class A GPCRs[46,78,188,438]. Using the Auto Grid 4 program, grid maps from the extracellular region were generated [396,408,437]. In the form of 67 x 72 x 67 x, y and z position with a grid spacing of 0.387nm, having generated the grid map, a grid box in the same x y z position was created with the following parameters 36.12 x 70.11 x 36.11. All other parameters were set to default. To search for the binding cavity, the ligand was allowed to undergo 100 trials of the blind dock. After which cluster analysis was conducted for the protein-ligand molecule based on a cut-off value of <0.2 nm for protein-ligand backbone atom and the best conformer with the least energy collected with the lowest estimated free energy of binding [427]

Upon identifying the binding site, molecular docking calculations of the ligand-receptor were executed with Glide (version 2019-3, Schrödinger, LLC, New York, NY, 2019) molecular docking protocol.

6.2.3 Glide Receptor Preparation

Glide receptor preparation program (prep wizard version 2019-3, Schrödinger, LLC, New York, NY, 2019) [439], from Schrödinger package, for each of the AKHR model, prep wizard was executed for the preparation of the receptors. For pre-docking refinement, the AKHRs were energy minimised. The OPLS_2005 force field was used for the further refinement and energy minimisation to a 0.3Å RMSD of non-hydrogen atoms [438,439]. To execute the Glide docking program, receptor grids were generated for each of AKHRs identified binding sites. After which docking calculations were executed. During grid generation, no restraint was applied, and allowance was made for the peptide-docking grid. Docking calculations were executed using Glide S-peptide docking available in the Schrödinger Suite version 2019-3[398,412]. The protocols used are, the force field-based scoring methods of Glide and the empirical scoring for docking results [396,398] The best binding poses were selected and rescored and filtered based on MM-GBSA free energy binding calculations using Prime [413,416,418].

6.2.4 Preparation of Ligand

The modelled structure, presented in Chapter 3, was used. The charge of the ligand was Glide Docking calculations.

Docking was performed using the β2AR-based homology models constructed in chapters 3, 4 and 5. Glide version 6.3, Schrödinger, LLC, New York, NY, 2019 docking protocol [398] was used for the execution of all the docking calculations [440]. Grid-based SP-Peptide docking was performed using the determined peptide (see chapter 2) and the constructed β2AR-based models in chapter 3, 4 and 5. For each scenario, the top 50% docking conformations obtained from the docking calculations were refined, using post-docking MM-GBSA to calculate the protein-ligand binding affinity. The 30 poses generated per receptor were visually inspected.

The best poses were selected and optimised, and the free energy of binding calculated using MM-GBSA (Prime version 2019, Schrödinger, LLC, New York, NY, 2019) and the total binding free energy calculated ($\Delta \boldsymbol{G_{bind}}$).

$$\Delta \boldsymbol{G_{bind}} = \Delta \boldsymbol{G_{Phote/Drome}} - (\Delta \boldsymbol{G_{Phote}} + \Delta \boldsymbol{G_{Drome}})$$

$$\boldsymbol{G} = \boldsymbol{G_{SGB}} + \boldsymbol{E_{MM}} + \boldsymbol{G_{SA}}$$

$\boldsymbol{G_{SGB}}$ is the surface-generalised born model polar solvation energy, $\boldsymbol{E_{MM}}$ is the potential energy and $\boldsymbol{G_{SA}}$ represents the non-polar solvation energy.

6.3.0 MD Simulation in a Mimetic Membrane

For further refinement after the molecular docking calculations where conducted, the three docked models (β2flf, βfrf and βoff) each was inserted together with bond Phote-HrTH into a palmitoyl-oleoyl-phosphatidylcholine (POPC) membrane, (see **Figure 6.1**). MD simulation was achieved via GROMACS software package tools [307]. The simulation was executed in three phases, namely minimisation, equilibration and MD simulation.

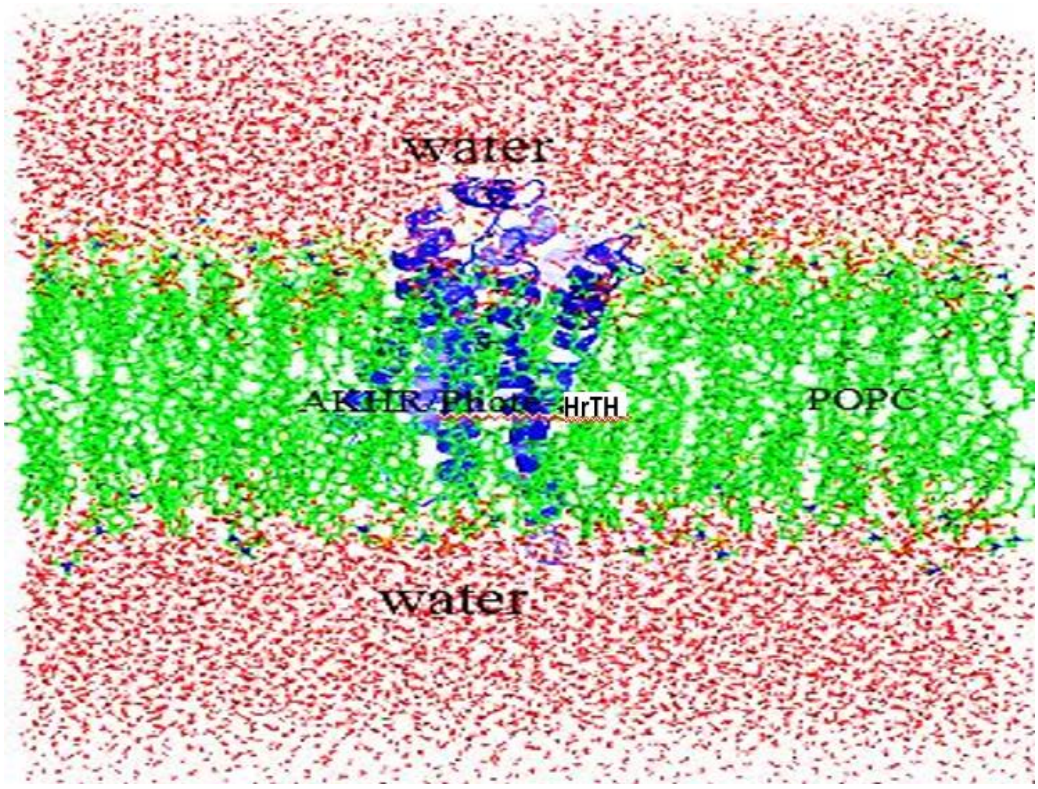


Figure 6. 1. Example of AKHR in a POPC membrane with its ligand docked at the extracellular part.

The MD simulation was performed for 10 ns under NPT condition to allow the lipid to arrange itself around the complex and the water to soak the construct. Cluster evaluation was then performed using the linkage algorithm of GROMACS and a cut-off value of < 0.1 nm for superimposing backbone atoms, and the best lowest structure saved. Further MD simulation was conducted for 100 ns under NPT condition, applying the same procedure.

6.4.0 Results and Discussion

6.4.1 Identifying the binding pocket of the AKHRs

The three different protocols (AutoDock4.2, MOE version 2014.0901 and Sitemap) identified similar binding sites. Four binding sites were identified for each AKHR. The possible binding site from the constructed models is given in **Figure 6.2**

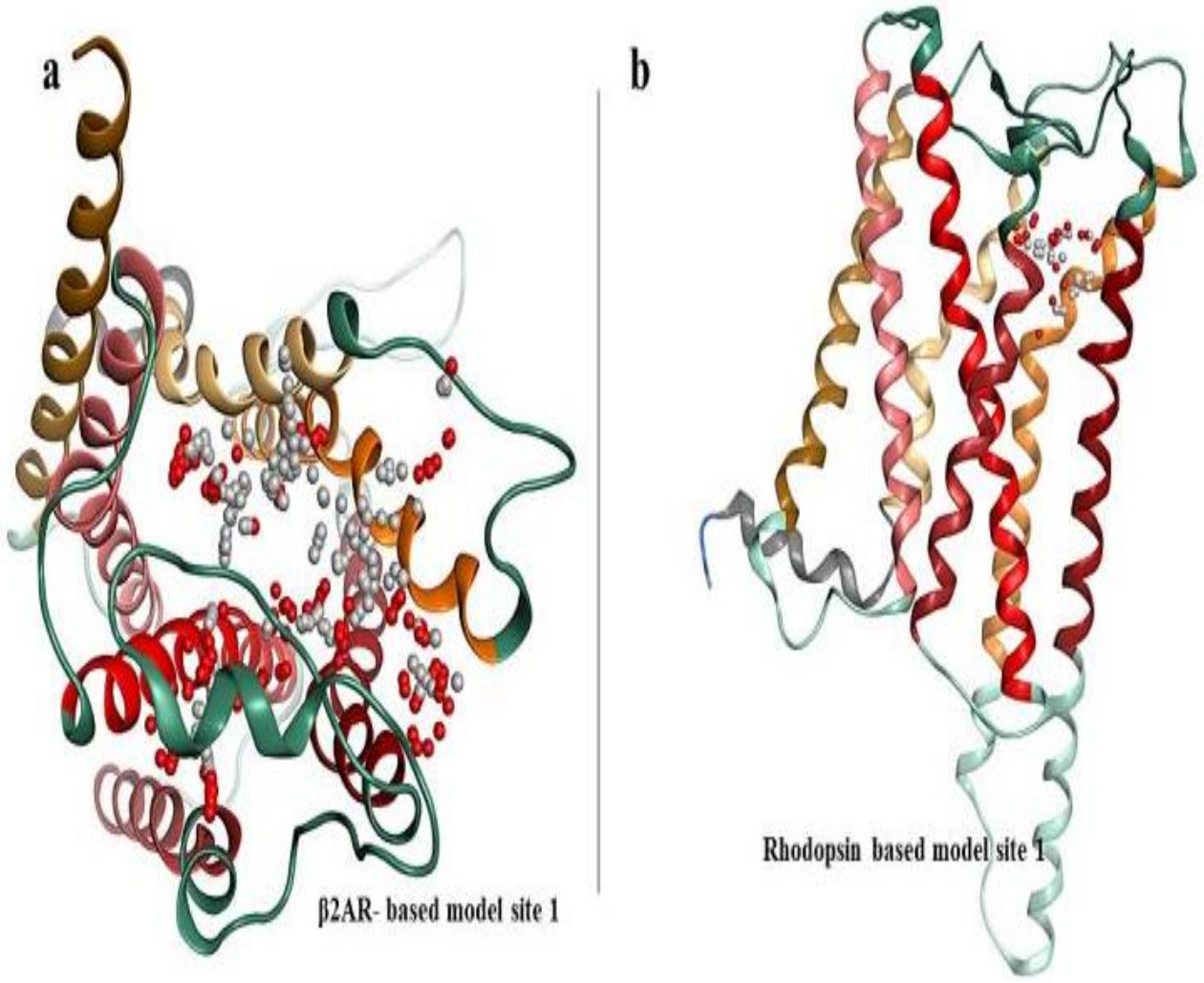


Figure 6. 2. 3D possible binding sites identified using blind docking program. a) the binding site for the β2AR based model b) binding site for the Rhodopsin based model β2AR model. * The white balls represent the binding site, and the red balls represent hydrophilic contact area.

The possible binding site could be seen in **Figure 6.2** a and b (tiny white balls). The potential water sites are omitted. Where a hydrophilic contact is probable, the little tiny clusters marked points are coloured red. To further explain the tiny clusters on each receptor, see **Table 6.1** for the Rhodopsin models and **Table 6.2** for the β2AR models. For clarity, since the results for all the AKHRs are similar, only that of the fruit fly will be displayed. For the rest of the fly AKHRs, see appendix (C & D).

Table 6. 1. Ranking of potential binding pockets in Rhodfrf receptor found using MOE * Size- Number of contributing spheres # PLB- predicted ligand-binding site # Hyd- Number of hydrophilic points # Side - side-chain contact atoms Residues - Residues at the local surface

Site	Size	PLB	Hyd	Side	Residues
1	122	2.41	48	85	ARG77 ILE78 ASP79 MET81 LEU82 ILE137 ASP140 ARG141 ILE145 TYR227 ILE230 VAL255 ARG258 ALA259 ARG262 THR263 LYS265 MET266 THR267 ILE270 VAL320 TYR321 GLY322 PHE324 ASN325 ILE326 ARG327 ARG329
2	54	1.74	28	55	ARG123 VAL124 TYR128 SER131 VAL193 THR194 HIS196 TYR207 ASN208 ASN211 MET212 MET215 TYR216 TRP278 TYR281 TYR282 CYS285
3	60	1.41	24	41	GLY59 THR62 VAL63 LEU66 LEU67 ARG70 GLY74 ASP79 LEU82 MET83 PRO318 ILE319 TYR321 GLY322 ALA323 ASN325 ARG327 GLY328 ARG329 ASN331 HIS332
4	33	0.84	28	41	PRO184 LYS185 TYR189 GLN191 VAL193 THR194 PHE195 HIS196 SER197 TYR281 ILE284 ARG303 LYS304 LEU306 PHE307

A thorough evaluation of the site map results of the Rhodfrf, reveals that the receptor is closed (active) and will not allow the ingress of ligand, while the β2AR models are open and inactive as such, they should allow free access of the ligand see **Table 6.2** for data representation. These finding further confirms the quality of our model.

Table 6. 2. Ranking of potential binding pockets in β2frf receptor found using MOE * Size- Number of contributing spheres # PLB- predicted ligand-binding site # Hyd- Number of hydrophilic points # Side - side-chain contact atoms Residues - Residues at the local surface

Site	Size	PLB	Hyd	Side	Residues
1	295	4.15	93	180	MET95 GLU96 TRP99 VAL103 GLN104 TRP105 ASP109 CYS112 ARG113 SER116 PHE117 ARG119 VAL120 LEU123 GLN170 LEU173 HIS175 LEU176 GLU177 ALA181 VAL182 THR183 TYR185 CYS188 VAL189 PHE191 SER193 PHE194 ARG195 SER196 ASP197 PHE198 ASP199 GLU200 LYS201 TYR203 GLN204 MET208 MET211 TRP276 TYR279 TYR280 ILE282 SER283 TYR286 TRP287 LYS290 HIS291 LYS295 ILE296 ASN297 PRO298 LEU299 ARG301 LYS302 PHE305
2	195	1.73	53	101	LEU63 THR64 ARG67 LEU68 ARG69 GLY70 PRO71 LEU72 ARG73 ILE74 ASP75 ILE76 MET79 ARG137 ALA140 ILE141 PRO144 ARG147 ARG153 ILE230 ALA257 LYS258 ARG260 THR261 LEU262 LYS263 MET264 TYR319 LEU321 TYR322 ASN323 ILE324 ARG325 GLY326 ARG327
3	54	0.31	15	22	PRO71 LEU72 ARG73 MET77 CYS132 LEU135 ASP136 PHE139 LYS143 PRO144 LEU145 LYS146 ARG147 SER148 ARG151 GLY152 MET155
4	22	-0.01	18	19	ILE220 TYR223 GLY224 TYR227 LEU254 LYS263 ILE266 VAL269 ILE270 ILE273

A thorough evaluation of the site map results of the β2AR-model, reveals that the site ranked 1 (see **Table 6.2**) has 295 residues as against the Rhodopsin-model which has 122 residues; this indicates that the ligand will fit more into the β2AR-model. Also, both models have four selected binding sites, labelled site 1 to 4. For Rhodopsin-model the best site ranked 1 is from the extracellular region, and the ECL2 covers the binding site, the reason for that could be because Rhodopsin-model is already activated. In contrast, β2AR-model has the extracellular

region as the best binding site and ranked 1, and the ECL2 does not cover the binding site. The best explanation for this is the receptor is inactive. The ranked extracellular region as 1 is expected because the location of the binding cavity and the participation of the entire extracellular domain conform with experimental data[239,441]. The same situation of ligand binding to the extracellular domain was reported for the human gonadotropin-releasing hormone receptor [442].

Having obtained the position of the preferred binding sites for the AKHRs, more focused docking was carried out to determine the binding mode of the AKHRs.

6.4.2 Accessibility of the AKHRs binding pocket, and determination of binding mode during docking calculations.

AKHs are biologically active peptides that are produced in modified neurons and are released into the circulatory system of the insect to trigger action through a ligand-receptor [14,87]. To trigger this said action at the fat body(where insects store energy reserves in the form of glycogen and triglycerides), the extracellular portion of the GPCR needs to be accessible; for the ligand to diffuse through the binding cavity[24,87]. Here in our observation, when the ligand (Phote-HrTH), was introduced from the extracellular side into the binding cavity, of both β2AR-models and Rhodopsin- models for the three-fly species only the β2AR-model binding cavity of all three modes were accessible (**Figure 6.3**). Meaning it was easy for the ligand to occupy the putative binding sites. In contrast, the Rhodopsin model produced no pose though this was expected, as the Rhodopsin AKH-receptors were already active and will not allow ligand penetration.

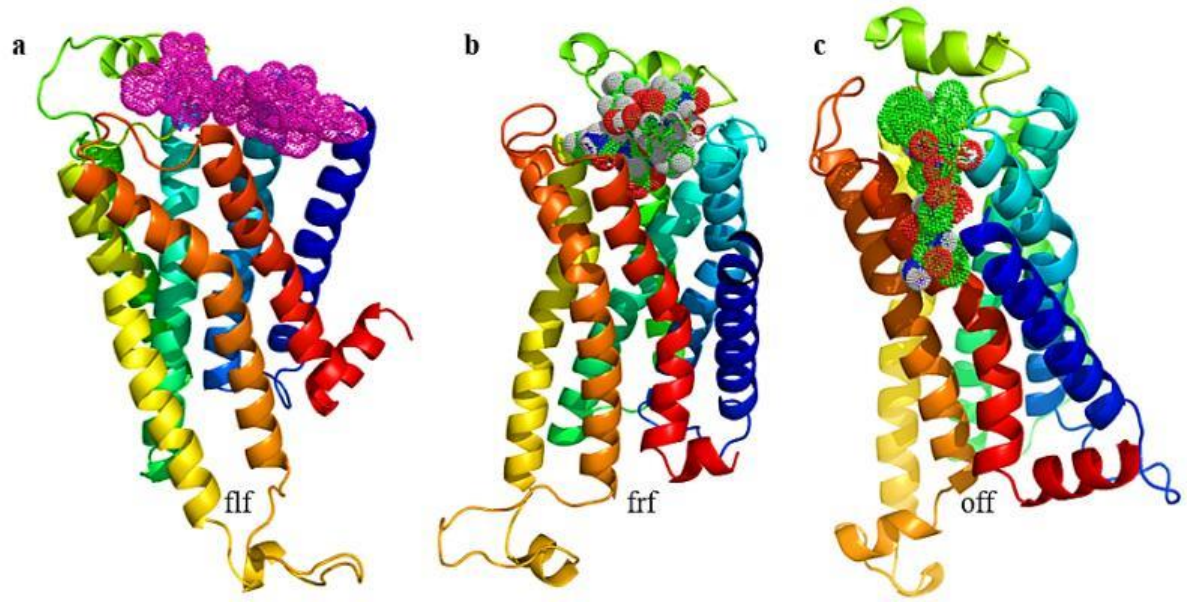


Figure 6. 3. The binding of Phote-HrTH to all three β2AR-based AKHRs. * a) flf = The flesh fly AKHR b) frf = The fruit fly AKHR c) off = oriental fruit fly AKHR. The colours represent H1-H6.

In contrast, in Rhodopsin models, even when the ligand was forced into the identified binding site from the extracellular region, the ECL2 had already covered the binding site leaving a tiny space. Unfortunately, the available space between the loops and the helices was too small for the ligand to access the binding cavity. Also, for the Rhodopsin models, when the Glide S-peptide docking command was initiated, no poses were generated, indicating the absence of cavities. While in the β2AR-based AKHRs, because of their open conformation, it was easy for proper diffusion of the peptide in and out of the binding site access. Since the Rhodopsin AKHR models are closed and would not allow further probing, by allowing the peptide to sink in, one could postulate that the β_2AR-models could represent the conformation of the receptor/inactive AKHR. At the same time, the Rhodopsin-model is used as a control. Hence, provide information on the active AKHR state of each receptor.

6.4.2.1 Molecular docking calculations for the β2flf-AKHR

Crucial interactions between Phote-HrTH peptide and β2flf AKHR were observed after the S-peptide docking. The docking scores were -13.20 kJmol^{-1}, and the $\mathbf{\Delta G_{bind}}$ (MMGBSA) was -

75 kJmol^{-1}. The most critical interaction is a salt bridge experienced between the Asp7 of Phote-HrTH and Arg303 of the receptor. To provide a significant amount of binding enthalpy and stability to the system, a π-π stacking interaction was experienced between Trp8 of Phote-HrTH and His 203 of the ECL2. This interaction with ECL2 help closes the loop over the binding site. These observations were similar to the studies on cation-pi interaction reported by Dougherty [379,443]. The ligand interaction maps are presented in **Figure 6.4**. The π-cationic interaction falls at 4.64Å, which is within the confines of < 6 Å proposed by Gallivan and Dougherty [380].

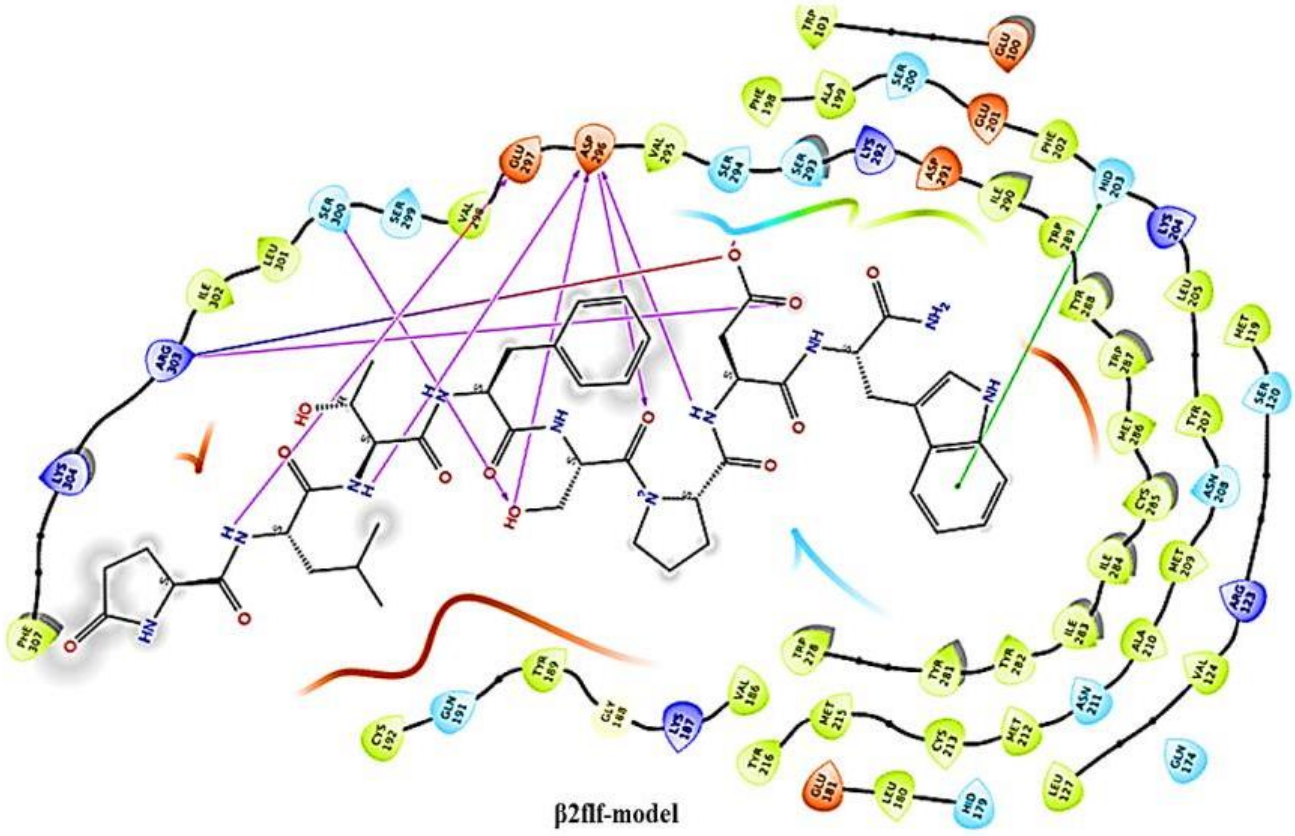


Figure 6.4: Ligand interaction diagram of Phote-HrTH bound to the β2flf model of Sarcr-AKHR. Residues are represented as coloured spheres, labelled with the residue name and residue number, and coloured according to their properties, green for hydrophobic, blue for positive charge and red for a negative charge. The ligand is displayed as a 2D structure. Interactions between the residues and the ligand are drawn as lines, coloured by interaction type, purple for H-bonding and green for pi-pi stacking. The binding pocket is indicated by a line drawn around the ligand, in the

same colour as the nearest residue. Solvent exposure is indicated on the ligand atoms, and by the break in the line drawn around the pocket.

6.4.2.2 Molecular docking calculations for the β2frf-AKHR

A β2frf putative binding site with Phote-HrTH is similar to those reported for *Anopheles gambiae* by Jackson et al., [324]. The docking scores were -14.33 kJ $mol^{-1,}$ and the ΔG_{bind} (MMGBSA) was -79 kJ mol^{-1}. The most critical interaction is a salt bridge between Asp7 of Phote-HrTH and Arg301. There is H-bonding between Phote-HrTH Ser5 and Val103 of ECL1 and Trp99 of TM2. The NH of Phote-HrTH Ser5 and C=O Asp7 H-bond to Ser196 on ECL2. Also, Phote-HrTH Trp8 NH interacts with Asp197 of ECL2. These interactions with ECL2 are interesting as this loop closes over the binding site after the ligand binds. The terminal amide of ligand H-bonds to Ile296 of ECL3. To provide proper stability and a significant amount of binding enthalpy to the system, π-stacking interactions were experienced between Trp8 and Phe4 of Phote-HrTH and Phe198 of ECL2. Hence, the π-π stacking interaction of 3.94 Å was noticed, which falls within the confines of < 6 Å as proposed by Gallivan and Dougherty [380] The ligand interaction is presented in **Figure 6.5.**

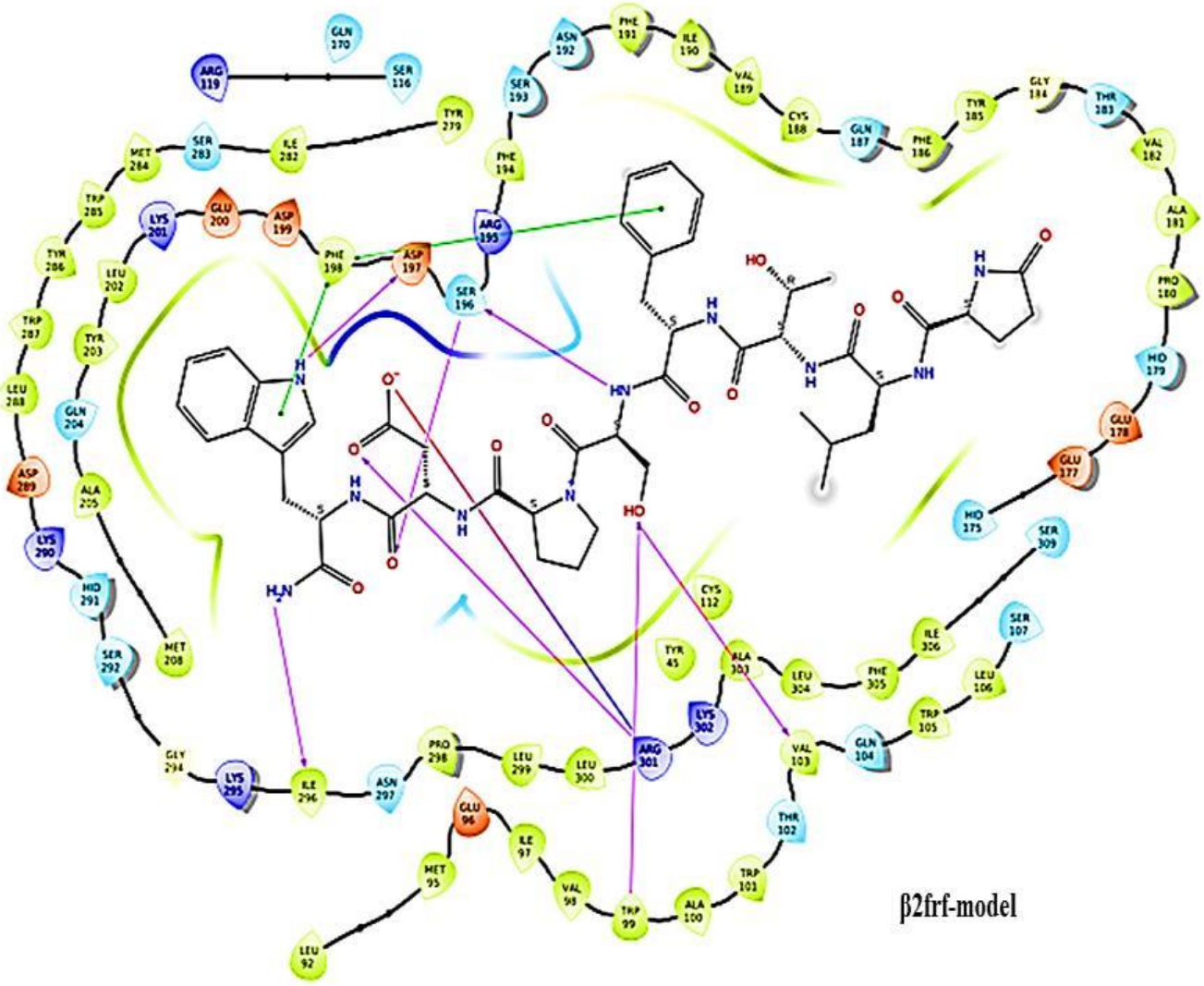


Figure 6. 5: Ligand interaction diagram of Phote-HrTH bound to the β2frf model of Drome-AKHR. Residues are represented as coloured spheres, labelled with the residue name and residue number, and coloured according to their properties, green for hydrophobic, blue for positive charge and red for a negative charge. The ligand is displayed as a 2D structure. Interactions between the residues and the ligand are drawn as lines, coloured by interaction type, purple for H-bonding and green for pi-pi stacking. The binding pocket is indicated by a line drawn around the ligand, in the same colour as the nearest residue. Solvent exposure is indicated on the ligand atoms, and by the break in the line drawn around the pocket.

6.4.2.3 Molecular docking calculations for the β2off-AKHR

β2off putative binding site with Phote-HrTH is similar to those reported for *Anopheles gambiae* by Jackson et al., [324], frf-model and flf-model. The docking scores were -14.03 kJmol$^{-1,}$ and the

ΔG$_{bind}$ (MMGBSA) was -82 kJmol^{-1}. Lys 306 was also discovered to act as an anchor in the binding pocket, forming a π-cation and hydrogen bond interaction with Phe4 of Phote-HrTH. A π-cation interaction of 4.96 Å was observed, which falls within the confines of < 6 Å as proposed by Gallivan and Dougherty. Also, a salt bridge was identified between Arg 306 and Asp 7 Phote-HrTH (see **Figure 6.6**). These interactions are similar to those shown for β2frf and β2flf.

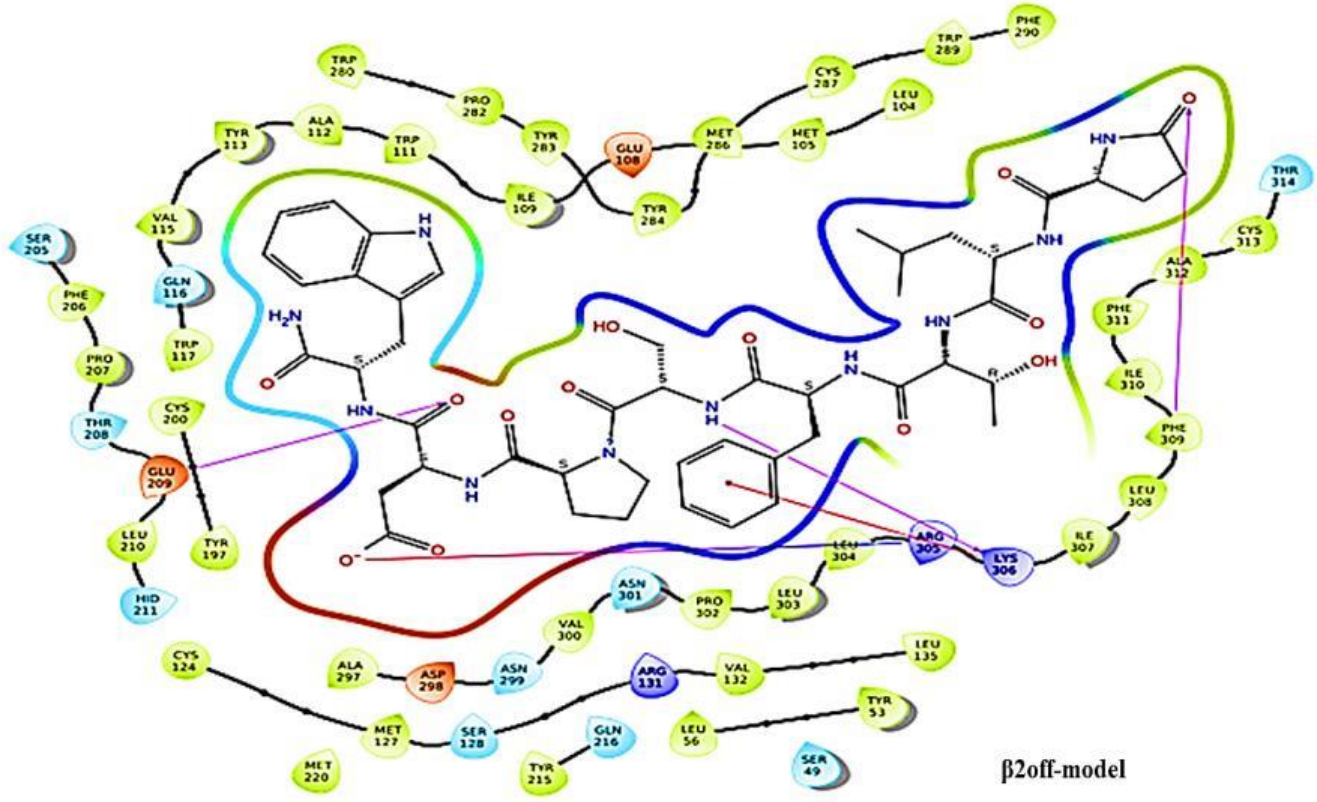


Figure 6. 6: Ligand interaction diagram of Phote-HrTH bound to the β2off model of Bacdo-AKHR. Residues are represented as coloured spheres, labelled with the residue name and residue number, and coloured according to their properties, green for hydrophobic, blue for positive charge and red for a negative charge. The ligand is displayed as a 2D structure. Interactions between the residues and the ligand are drawn as lines, coloured by interaction type, purple for H-bonding and green for pi-pi stacking. The binding pocket is indicated by a line drawn around the ligand, in the same colour as the nearest residue. Solvent exposure is indicated on the ligand atoms, and by the break in the line drawn around the pocket.

The ligand interaction diagram is comparable to one another since they all have hydrogen interactions, salt bridge, ionic interactions and π-π stacking interactions, though β2frf has two π-π stacking interactions originating from same molecule Phe4 and Trp8 of Phote-HrTH and Phe198 of the receptor, while β2flf has a π-π stacking interactions between Trp8 of Phote-HrTH and His203 and β2off possess a π-cationic interaction between Lys 306 and Phe4. The π-π stacking and π-cationic interactions experienced by these AKHRs and the ligand signifies discrepancies in the strength at which the molecules bind.

6.5.0 Molecular Dynamics simulation of the AKHRs-Phote-HrTH (ligand) complex

Having performed docking calculations for the AKHRs, to further refine and optimise the β2AR-based AKHRs-ligand complexes of the lowest ΔGb, MD simulations were performed [427]. For each system, the docked Phote-HrTH and β2flf or β2frf or β2off based AKHR was inserted in a mimetic POPC membrane and MD simulations executed for 100 ns in a water box see **Figure 6.1**. Since the receptors are open (inactive), and literature available pinpoint Phote-HrTH to be responsible for the activation of these receptors [43–45]. To understand the molecular explanation of how these activations occur by a single neuropeptide, we conducted these simulations in order to establish the real-life situation of binding conformations and hydrogen bond occupancy as well as evaluate whether the peptide will further sink into the receptors beyond its current position anchor during free MD simulations (see **Figure 6.7**).

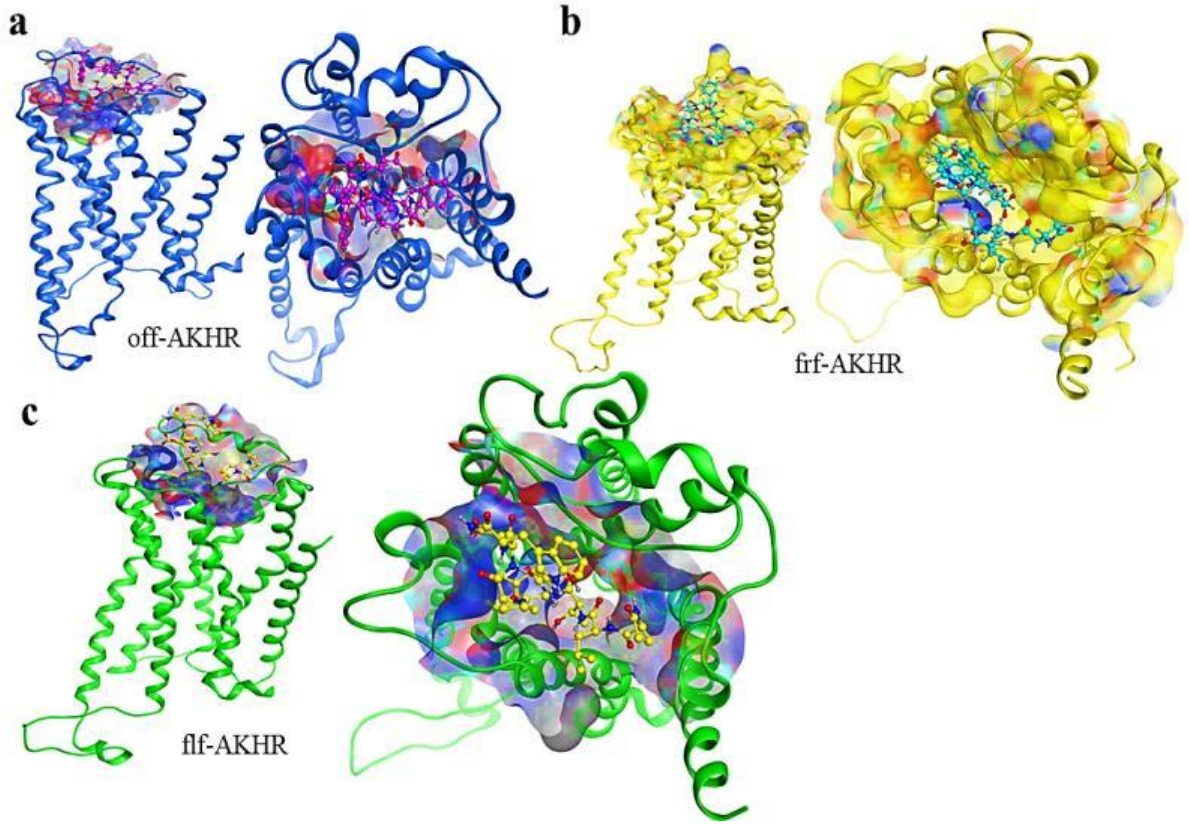


Figure 6. 7: Refined structure of Phote-HrTH /Phote AKHR after MD simulation. * a) off = oriental fruit fly AKHR. b) frf = The fruit fly AKHR. c) flf = The flesh fly AKHR.

Observations from the MD simulations were amazing, as the refined structures of the Phote-HrTH folded and sank deep into their AKHR, thereby fitting correctly into the binding pocket see **Figure 6.7**. Hence, causing a severe conformational change to the receptors.

6.5.1 Conformational Changes AKHRs during activation

For the specific receptors, after 1.5 ns, a drastic conformational change in the entire receptor molecule was seen. The helices of the individual receptors move towards each other on the extracellular side, at the same time, the move away from each other at the intracellular sides [367,444]. The best explanation for this is the molecule converted from an open (inactive) to a closed (active) conformation see **Figure 6.8**.

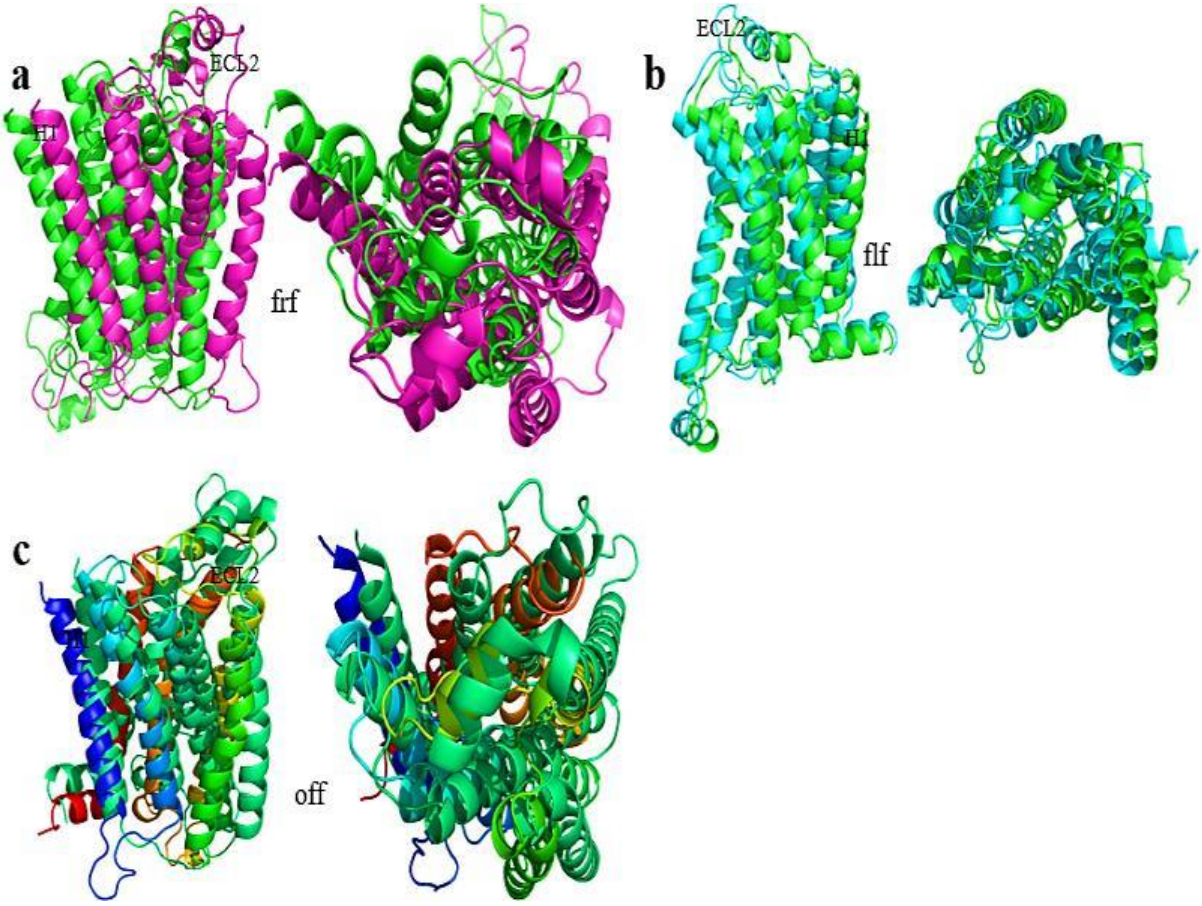


Figure 6.8: (a) Overlay of the β2frf receptor model of Drome-AKHR at the start (purple) and end (green) of a 100 ns MD simulation in a POPC membrane: a side view and top view. (b) Overlay of the β2flf receptor model of Sarcr-AKHR at the start (cyan) and end (green) of a 100 ns MD simulation in a POPC membrane: a side view and top view. Overlay of the β2off receptor model of Bacdo-AKHR at the start (blue and red) and end (green) of a 100 ns MD simulation in a POPC membrane: a side view and top view. * frf = fruit fly-AKHR, flf = flesh fly AKHR, off = oriental fruit fly.

An overlay of unbound and bound of the individual AKHRs helices is given in **Figure 6.8**. In each AKHR, both structures, most residues maintained α- helicity and the helix bundles are very stable. An inward movement of the individual AKHRs around their extracellular domain, of the receptor molecule, results in a closed conformation, like their Rhodopsin-based model. The Root Mean Square Deviation (RMSD) [445] of the amino acid backbone atoms between the starting structure and structure after 1.5 ns was calculated to be 2.37 Å for the β2frf receptor model, 2.17 Å for the β2flf receptor model and 2.46 Å for the β2off receptor model. Also, all three intracellular loops were pushed outwards, and the ECL2 loop closes over the binding

pocket with the ligand inside the binding site. These movements are shown in **Figure 6.8**. The best explanation for this is that the molecules converted from an open (inactive) to a closed (active) conformation. The same observation was recorded when epinephrine (agonist) was introduced to its receptor[89]. However, the rapid realisation of the active conformation is consistent with the activation of fluorescently labelled β2AR in a detergent by noradrenaline, which happens within approximately 40ms[231,446–448].

In the extracellular region of all three AKHR model, three helices H1, H3, and H6 had a significant tilt towards the centre of the helix bundle. This tilting leads to the breaking of the Asp, Arg and Tyr (DRY) ionic lock due to the movement of TM6 and TM3 from the intracellular regions of these AKHRs. For a proper understanding of these movements, the reader is referred to chapter 3, 4 and 5. After each MD simulation, each AKHR model is compared to its starting structure, and the distance between helices H1, H3, and H6 recorded. All three models behave the same (see **Table 6.3**). Their helices were drawn towards the centre of the individual helix bundle by an approximate 4 Å each from their initial positions, moreover, after 70ns approximately of MD simulation, all models were seen to have stabilised with no further movement. These individual distances when compared with their Rhodopsin-based models, the distances were found to be approximately the same. Also, the starting structure of all three models when compared to the template β2AR model were found to be approximately the same with respect to distance between the helices.

Table 6. 3: Change in conformational distance with time during MD simulation. * frf = fruit fly-AKHR, flf = flesh fly AKHR, off = oriental fruit fly. H1, H2 and H3 = helix 1-3

Time (ns)	β2frf H1	β-2frf H3	β-2frf H6	β-2flf H1	β-2flf H3	β-2flf H6	β-off H1	β-2off H3	β-2off H6
1.5	3Å	1Å	2Å	2Å	1Å	1Å	1Å	2Å	3Å
10	4Å	2Å	3Å	4Å	2Å	2Å	2Å	3Å	4Å
30	5Å	3Å	4Å	6Å	3Å	3Å	6Å	5Å	5Å
50	6Å	4Å	5Å	6Å	4Å	3Å	12Å	7Å	6Å
70	8Å	6Å	6Å	6Å	5Å	3Å	12Å	7Å	7Å
90	9Å	6Å	7Å	6Å	5Å	3Å	12Å	7Å	7Å
100	9Å	6Å	7Å	6Å	5Å	3Å	12Å	7Å	7Å

Finally, the conformational changes of all three AKHR models, when Phote-HrTH is introduced, has established Phote-HrTH peptide to be the agonist of these three AKHRs. Also, the braking of the DRY ionic lock due to the movement of TM6 and TM3 from the intracellular regions of these AKHRs and the easy diffusion of Phote-HrTH into these receptors is a clear indication that Phote-HrTH activated the receptor. Hence, this information could be utilised as a piece of firm supporting information relating to the binding of ligands to AKHR/GCPRs[449]. Hence, these observations support the idea that the results obtained in Chapters 3, 4 and 5, that upon receptor activation, the open β_2AR-based AKHR model represents the inactive state. In contrast, the closed state represents the activated state.

6.6.0 Conclusion

The binding sites of the three AKHRs were effectively identified with the aid of molecular docking calculations using the same neurohormone (Phote-HrTH) that is known to activate

each of them. The identified binding sites of these AKHRs were not different from most already identified GPCRs, where their helix bundles, the extracellular loop domains, and the N-terminus define their active sites. The docking score of the models are β2flf-model has -13.20 kJ/ mol, β2frf-model is - 14.33 kJ/mol and β2off-model is - 14.03 kJ/mol. The binding scores (**ΔG_{bind}**) calculated using MMGBSA, are -75 kJ/mol for β2flf-model, -79 kJ/mol for β2frf-model and – 82 kJ/ mol for β2off-model.

Only the β2AR models were used for the docking calculations because the Rhodopsin models did not give any pose as such, they served as a control for our model. The best-docked models were placed into a membrane box to mimic the actual environment of the receptor-ligand and further refine the docked structure; molecular dynamics simulations were conducted.

During MD simulations, simultaneously from both the extracellular and intracellular regions of the AKHRs, rapid opening and closing respectively of these regions were noticed. These opening and closing of receptors have been described as activation in GCPRs since AKHRs are also GCPRs. After 100 ns simulation, the ligand folded and sank into the individual receptors and their binding energy was calculated and found to be -95 kJ/ mol for β2flf, -102 kJ/ mol β2frf and – 107 kJ/ mol. In each model, it was noticed that the agonist Phote-HrTH was stabilised by polar and π-stacking / π-cationic contacts with residues in Helices 2, 5, 6 and the second extracellular loop. Also, Tyr in the conserved residue CWTPY in H6 was noted to interact continuously with the agonist in each model.

Since these AKH-receptors when activated are involved in the production of energy during insect flight, data on the constructed 3D molecules and the binding sites will allow the design of novel, species-specific, non-peptidic mimetics which can block the binding site and hence prevent insect flight (dispersal of pest insects).

7

Structure-activity studies of Phote-HrTH on the fruit fly (*Drosophila melanogaster*) AKHR (Drome-AKHR)

Summary

For the first time, an in-silico structure-activity relationship study (SAR) of an AKHR was conducted using the computational technique. The technique involves a series of mutation of the primary amino acid sequence of Phote-HrTH. At the already identified and docked Phote-HrTH/Drome-AKHR, the Phote-HrTH was mutated such that the N-terminal was substituted with alanine, and the amino acid residues at the C-terminal replaced with glycine. The mutation was done with the aim of identifying the amino acid residue position that is vital for receptor activation. Also, the SAR was conducted to validate the molecular docking process of Drome-AKHR/Phote-HrTH.

The aromatic Phe ring in position 4 and that of the indole ring Trp8, which have been reported in the previous chapter as providing stability to the ligand during ligand-receptor binding is crucial for receptor activation. Leu in position 2, Thr in position 3, Ser in position 5 and Pro in position 6 are also crucial in receptor binding but not necessary activation. Asp in position 7 does not seem to participate in either binding or activation of the receptor, as when exchanged with Gly the ligand gained an additional negative charge and the receptor was activated. Knowledge from these studies could help identify the most critical molecule necessary for ligand-receptor activation.

7.1.0 Introduction

Since the earlier discovery of an AKH in *Locusta migratoria*, there have been over 60 AKHs discovered in insects.[62] Although these AKHs differ in terms of primary sequence and length (octa-, nona- or decapeptides), they all have a blocked pyroglutamate at the N-terminus, a Phe

or Tyr at position four, Trp at position eight, and a blocked carboxy amidation at the C-terminus.[63] These specific residues appear to be essential for AKH peptide activity.

AKHs work by binding to their cognate GPCR as demonstrated via *in vitro* receptor assays and *in vivo* biological assays of a variety of insects[450,451]. For receptor activation to take place, the AKH must have a turn secondary structure, Phote-HrTH also demonstrates the turn structure in solution [436,452,453].

Previously, numerous structure-activity relationship (SAR) studies have been conducted via experimental procedures (*in vitro* and *in vivo* assays) to identify critical primary amino acids present in AKH peptides. Some of the insects that were studied include cockroaches, *Blaberus discoidalis* and *Periplaneta americana,* the moth, *Manduca sexta,* and the migratory locusts, *Locusta migratoria* as well as the desert locust, *Schistocerca gregaria*. Generally, the outcome from these studies showed that the amino acids at positions 2, 3, 4, 5 and 8 are vital for receptor activation. Also, position 4 (usually Phe) and 8 (Trp) are the most critical. Amino acids at positions 6 and 7 are either non-reactive or play a minor role in receptor activation[454–459]. Although *D. melanogaster* AKHR was characterised in 2002[45] and the SAR study conducted in 2012 [76], no molecular modelling (*in silico*) studies have been done on any AKHR/GPCR.

In this thesis, we report the first SAR *in silico* study on *D. melanogaster* AKH peptide using computational techniques. The study involves a sequential mutation of the Phote-HrTH amino acids corresponding to the experimental mutation study on *Drome*-AKHR.[76] The mutated peptides were then docked to the *Drome-AKHR* developed in this thesis. The objectives of this study were two-fold. Firstly, if the computational results mirror the experimental results, this would lend confidence to the receptor model developed in Chapter four. Secondly, the agreement would afford the first opportunity to explain the conserved nature of the amino acids at a molecular level.

7.2.0 Experimental Methods

7.2.1 The ligand, protein preparation, and molecular docking.

The elucidated structure of the ligand Phote-HrTH was presented in Chapter Two, the 3D structure of the receptor in chapter four, and the molecular docking results in chapter six. The alanine/ glycine (*in vitro*) replacement experiment was first reported in 2012 by Caers et al, , where the Drome-AKHR was expressed in a cell line and the chemically synthesised peptide ligands were applied to the receptor and receptor activation was monitored. [76] The same series of Phote-HrTH mutations were made computationally as shown in **Table 7.1**

Table 7.1. The Phote-HrTH sequence mutations

Peptide sequence / modification	Name
pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH2	Phote-HrTH
[N-Ac-Ala]-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH2	[N-Ac-Ala^{1}] Phote-HrTH
pGlu-Ala-Thr-Phe-Ser-Pro-Asp-Trp-NH2	[Ala^{2}] Phote-HrTH
pGlu-Leu-Ala-Phe-Ser-Pro-Asp-Trp-NH2	[Ala^{3}] Phote-HrTH
pGlu-Leu-Thr-Ala-Ser-Pro-Asp-Trp-NH2	[Ala^{4}] Phote-HrTH
pGlu-Leu-Thr-Phe-Gly-Pro-Asp-Trp-NH2	[Gly^{5}] Phote-HrTH
pGlu-Leu-Thr-Phe-Ser-Gly-Asp-Trp-NH2	[Gly^{6}] Phote-HrTH
pGlu-Leu-Thr-Phe-Ser-Pro-Gly-Trp-NH2	[Gly^{7}] Phote-HrTH
pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Gly-NH2	[Gly^{8}] Phote-HrTH
pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Gly-OH	[Trp-OH] Phote-HrTH

7.2.2 Alanine and glycine scanning

The amino acid mutation was achieved using the residue scanning module of Glide (version 2019, Schrödinger, LLC, New York, NY, 2019). This module provides the option to specify the residue of interest to be mutated. The stability and affinity of the mutated peptide was then calculated. Backbone minimisation and side-chain prediction were set to 0. The calculated

binding affinity predicted for the system was achieved with the following equation (thermodynamic cycle)

$$\Delta\Delta G_{bind} = \Delta G_{bind}^{MUT} - \Delta G_{bind}^{NT}$$

$$\Delta\Delta G_{bind} = (E_{A.B}^{MUT} - E_{A}^{MUT} - E_{B}^{MUT}) - (E_{A.B}^{NT} - E_{A}^{NT} - E_{B}^{NT})$$

$$R_A + L_B \xrightarrow{\Delta G_{bind}^{NT}} R_A * L_B$$

$$\Delta G_{Calc}^{inactive} \downarrow \qquad\qquad \downarrow \Delta G_{Calc}^{inactive}$$

$$R_A + L'_B \xrightarrow{\Delta G_{bind}^{Mut}} R_A * L'_B$$

Where ΔΔ**G** represents free energy while **E** is the calculated energy of the ligand before or after refinement, **L'** represents the mutated ligand, while **L** is for the unmutated ligand. **NT** represents the unmutated, and **R** is the receptor. Using these equations, the free energy difference, ΔΔG, was calculated by computing the free energy variations (vertical arrows). The vertical arrows symbolise variations that are simpler to simulate experimentally while the horizontal arrows symbolise recognisable procedures.

With Maestro, we were able to achieve the mutation of Phote-HrTH residues by positioning the chi-angles to those of the residue that was being substituted. In situations, where there is no chi- angles, the first residue is expanded to an extensive value (say above 180°). Hence, in

circumstances of refinement, where the residue radius happens to be higher than zero (> 0), that region or the residue is defined based on the size of arginine and minimised [460].

7.3.0 Results

7.3.1 Ligand modification and structure-activity relationship (SAR)

The Phote-HrTH, solution structure was elucidated and identified to possess a turn structure (see chapter two) and the 3D structure of the fruit fly AKHR has also been determined (see chapter four). Molecular docking calculations on the fruit fly AKHR were conducted using Phote-HrTH as the ligand, (see chapter six). Phote-HrTH in the docked position of the constructed fruit fly AKHR-model was modified such that position 1-4 primary amino acid was interchanged with alanine while glycine was used to replace positions 5-8, as seen in **Table 7.1.** The reason for choosing Ala or Gly is because of the availability of the experimental data of the fruit fly (Drome-AKHR)[76].

There are various experimental data available on the structure-activity of both AKHs/RPCHs to their cognate receptors[454,457–459,461,462]. However, no molecular data or computational comparison of any kind has been conducted, despite its importance. With the Drome-AKHR experimental data[76], we were able to conduct SAR studies at the molecular binding site to ascertain if there is any correlation between the experimental and molecular data (see **Figure 7.2**).

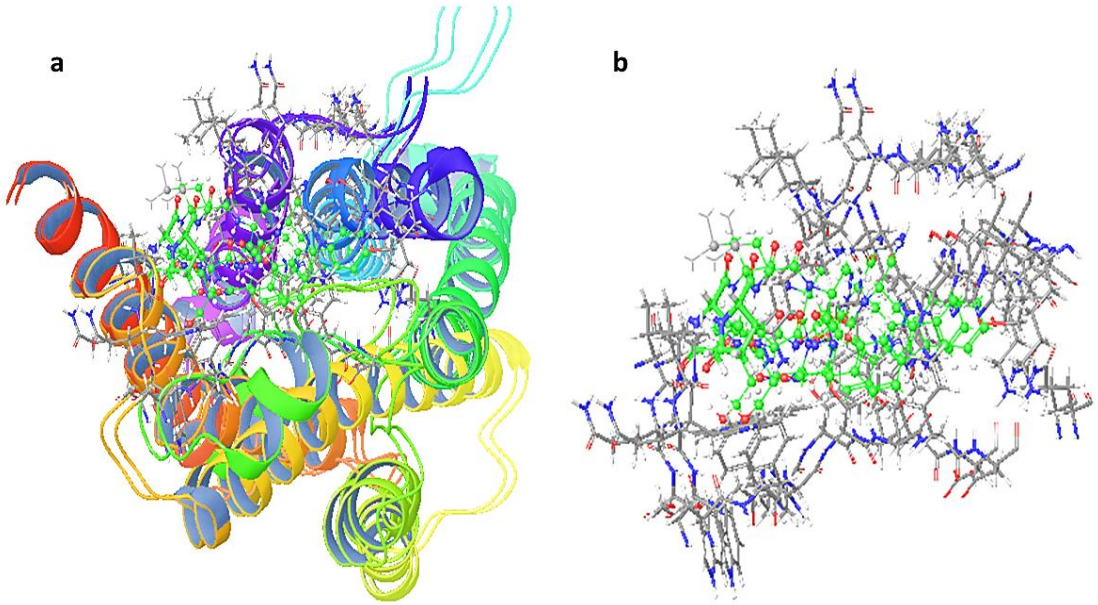


Figure 7. 1: a) Phote-HrTH docked to the extracellular part of Drome-AKHR alongside the mutated [Ala/Gly]Phote-HrTH. b) Phote-HrTH and the mutated [Ala/ Gly]Phote-HrTH showing the difference in conformation of both ligand.

7.3.2 Structure-activity of Phote-HrTH on its receptor binding site

The binding cavity of Drome-AKHR containing Phote-HrTH was identified in Chapter four. For this reason, it was possible to conduct a mutational study on the already docked ligand at the binding site. The Ala/Gly scanning for the N-terminus amino acids residues (Leu2, Thr3, Phe4) except for Thr3 shows a relatively low binding affinity. Also, the C-terminus residues (Ser5, Pro6 and Trp8) had much lower (more positive value) binding affinity when compared to the binding affinity of Leu2 and Phe4 of the C-terminal (see **Figure 7.2** B). The Ala/Gly scanning for Asp7 of Phote-HrTH reveals that both [Gly7] Phote-HrTH and Asp7 of Phote-HrTH have almost the same binding affinity (see **Figure 7.2 B**).

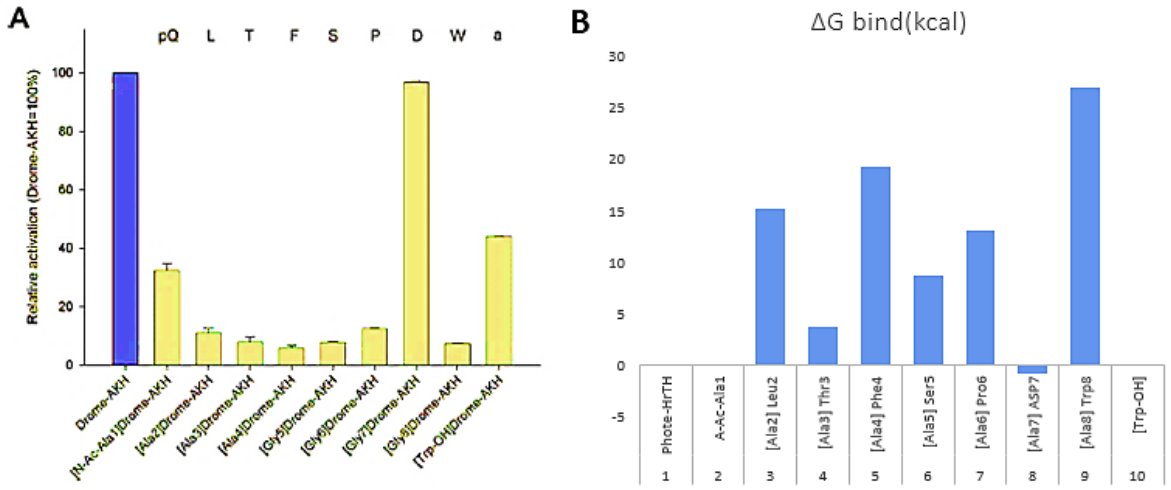


Figure 7. 2: A) Experimental Structure-activity relationship results from the Phote-HrTH [76]. B) Molecular data on the Relative binding activity of alanine/ glycine (KJ /mol) exchange against the Phote-HrTH peptide on the Drome-AKHR binding site. B) is also showing ([Gly^8] Phote-HrTH) and ([Ala^4] Phote-HrTH) as the most critical position with the lowest binding affinity (KJ /mol) when exchanged with Ala/ Gly.

([Ala^4] Phote-HrTH) and ([Gly^8] Phote-HrTH) had the highest (positive value)binding affinity (see **Figure 7.2 B**) when scanned with Ala/Gly. The higher positive value in binding affinity calls for attention. Since we are dealing with probability and inherent uncertainty, Monte Carlo simulation method was used for further optimisation of [Ala^4] Phote-HrTH and [Gly^8] Phote-HrTH molecule with the aim of understanding which residue, when exchanged, will have the most critical effect relative to Phote-HrTH at the receptor-binding site. The result shows that [Gly^8] Phote-HrTH has a higher (positive value) binding affinity when compared to [Ala^4] Phote-HrTH (see **Figure 7.3)**.

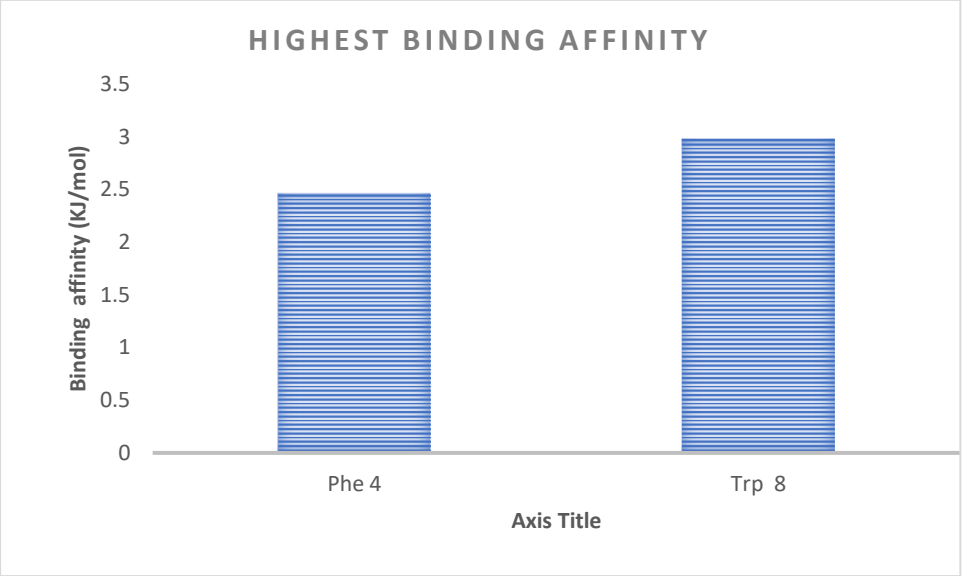


Figure 7.3 SAR scan of Phote-HrTH position 4 and position 8 via Monte Carlo simulation for the most critical ligand residue position

7.4.0 Discussion

Neuropeptides undergo conformational changes to obtain an "active conformation" during ligand-receptor binding[78]. The tendency to evaluate the activity of neuropeptides experimentally only based on their primary sequence and the charge of specific amino acids is known as structure-activity relationship. To understand the structure-activity relationship of a ligand, it is essential to consider the conformational changes that occur with the peptide analogue. Also, for conformational changes to take place during ligand binding, there must be a binding interaction between the ligand and the receptor. The strength of these interactions is known as binding affinity. For a ligand to properly bind to a receptor, the binding affinity must be high (in negative value). This implies that a positive relative binding affinity result designates a low binding while a negative relative binding affinity result represents a high binding. That is to say, the more negative the binding affinity data, the stronger the binding of the ligand to the receptor and the higher the activity response. Here we study the molecular structural relationship depending only on the strength of the relative binding affinity between the ligand-receptor.

Leu2, when compared to Ala2, has a higher binding affinity. Though, Leu2 does not show much interaction with its cognate receptor, the same as [Ala2]Phote-HrTH when scanned with Ala /Gly. Leu2 has the same hydrophobic properties as Ala, and the belief is that both molecules could play a similar role[463]. However, the reverse is the case when compared using the hydropathy index. The index helps to define the relative hydrophilicity or hydrophobicity of a molecule; the more positive the value, the more hydrophobic the residue is. Leu2 is 3.8 while is Ala 1.8 on the hydropathy index[464]. Also, Leu2 is more buried into the β-turn of the ligand in the receptor core when compared to Ala2 and distortion of this turn could lead to a breakdown instability of the ligand, hence lack of receptor response (see **Figure 7.1**). The presence of low stability caused by the Ala2 side chain could be the reason for the low binding affinity and the lack of response reported for the experimental SAR studies[76].

The binding affinity of Thr3, when compared to Ala3, was much higher. Thr3 interacts with the receptor by H-bonding to Gln187 of the receptor, but Ala3 does not have a proton to donate and instead was buried inside the receptor (see **Figure 7.1**). The absence of proton donor can have an adverse effect on the ligand stability in position 3 and hence, cause a reduction in the binding affinity. Almost the same response was reported in a biological experimental set-up when Ala3 was interchanged for Thr3 a nearly complete lack of response was recorded [76].

[Phe^4]Phote-HrTH of the native ligan is buried more into the receptor when compared to [Ala^4]Phote-HrTH (see **Figure 7.1**) and forms a pi-cationic interaction with Phe198. This pi-cationic interaction is crucial for ligand stability; no wonder the binding affinity is much higher when compared to another N-terminal molecule. Phe4, when replaced by Ala4, a sharp drop in the binding affinity was noticed. The best explanation for this is the lack of pi-interaction that

is supposed to create stability for the ligand. The absence of stability reduces the binding affinity. This claim is consistent with other response reported in a biological experimental set-up, where Phe4 was replaced by Ala4 lead to a complete lack of response by the receptor.[454,455,457,458]

Ser5 when compared to [Gly^5]Phote-HrTH there was a slight decrease in the binding affinity of the ligand. The decrease in the binding affinity could be because [Gly^5] Phote-HrTH is hydrophobic and is buried inside the protein (see **Figure 7.1**). The presence of β-strand at position5 is because of the Ser5 a polar molecule, and its absence could expose the backbone of the ligand molecule at position 5. The exposure of the backbone reduces the chance of attaining a turn structure and could lead to a drop in the binding affinity. This observation is consistent with the experimental data expressing the need for a turn structure at position 5; otherwise, the strand stability necessary for activation will be interrupted [76,458].

Pro6, when bonded to the receptor, provides rigidity to the peptide chain by imposing torsion angles on the segment of the structure of the ligand, thereby increasing the binding affinity. Pro6 when compared to Gly6, there was a sharp drop in the binding affinity (see **Figure 7.2 B**). Also, Pro6 and Gly6, are directly opposite of each other. And Gly6 when viewed from the docked site the molecule was found at the surface of the receptor, unlike Pro6 which was buried within the receptor molecule (see **Figure 7.1**). The presence of Gly6 rendered the backbone flexible hence a reduction in stability and Gly6 behaves more hydrophilic when compared to Pro6. The flexibility of the Gly6 could be the primary reason for the drop in the binding affinity of the ligand. This finding is consistent with other experimental studies where Pro6 when replaced with Gly6 lead to a reduction in the activity of the receptor.[454]

Asp7 forms a salt bridge with Arg301, H-bond with Arg301 and H-bond Ser196 of receptor molecule, and the presence of these interactions helps to stabilise the receptor. Asp7 when compared to [Gly^7] Phote-HrTH, Asp7 is understood to possess a nett negative charged sidechain, while Gly7 has no charge and no side chain. Gly7 has almost the same binding affinity when compared to Asp7. Gly7 due to its inert and flexible nature cannot form a salt bridge that is necessary for the stability of the ligand so a drop in binding would have been expected, but this was not the case as the inert Gly7 bonded almost the same as the charged Asp7 molecule. This observation is consistent with experimental data on SAR studies when Gly was exchanged with Asp7 showing no change in activation (see **Figure 7.2 A**) [76]. The possible explanation for the inadequate response of the system for the molecule in position Asp7 when interchanged Gly7, Asp is a negatively charged molecule, its presence causes the ligand to have a nett negative charge.

Trp8 forms a pi-cationic interaction with Phe198 and donates a proton to Asp197 of the receptor molecule. These interactions provide stability to the ligand. Trp8 when compared to Gly8, a serious drop in the binding affinity of the ligand was noticed. The Gly8 residue is buried inside the receptor molecule while Trp8 could be located at the surface of the receptor (see **Figure 7.1**). Trp8 analogue, due to its amphipathic orientation forms a β-strand; the presence of turn structure have been reported to be critical for receptor activation. [452,453,458,465]. So, the possible reason for the pronounced drop in binding affinity could be due to the flexible nature of the Gly8 molecule which distorts the stability and the amphipathic orientation of the ligand leading to virtually no response between the ligand-receptor molecule. The same observations were also confirmed in other SAR studies when Trp8 was replaced with Gly8 leading to a complete lack of response [454,455,457,458].

The Phote-HrTH analogues (Leu in position 2, Thr in position 3 and Ser in position 5) when scanned with the Ala/Gly, shows a reduced binding affinity when compared to Phote-HrTH. The reduction in binding could be as a result of the presence of turn structure between the N-terminus molecules of Phote-HrTH. The cause of the turn structure at the N-terminus residues is supported by the presence of alternating polar water-soluble group (pGlu1, Thr3 and Ser5) and the water-insoluble molecule (Leu2 and Phe4) causes an amphipathic orientation. Studies have proven the presence of amphipathic orientation as the primary cause of turn structure in a molecule [452,453,458,465]. So, the best explanation as to why there is a reduced binding affinity is that for proper binding to occur, there is a need for a turn structure. Phote-HrTH possess this turn structure. During Ala/Gly scanning, there is an alteration of the backbone amide protons by the side chains of Ala. The disruption of the side chains disrupts the amphiphilicity pattern necessary for providing stability to the ligand. So, the lower the stability, the lower the binding affinity.

The binding affinity of and [Gly8]Phote-HrTH is much higher when compared to the binding affinity of. [Ala4]Phote-HrTH. The best explanation for the higher binding affinity at [Ala4]Phote-HrTH and [Gly8]Phote-HrTH is that Ala molecule has a simple methyl side chain while the Gly molecule does not have a side chain. As such, it is expected for Phe4 when replaced with Ala, disrupts the backbone, thereby lowering the stability of the ligand and hence a low binding affinity. These specific residues at position four and eight appear to be essential for peptide activity [63] no wonder their activity response were much pronounced. An interchange of these amino acid residues in the said position with Ala in position 4 or Gly in position 8 results in the total shut down of Drome-AKHR [76] just like in other insects [454,455,458].

The investigation of the modified amino acid sequences and knowledge on which position is vital for receptor activations will go a long way in the design of agonist or antagonist or even species-specific pesticides [466–468].

7.5.0 Conclusion

The structure-activity studies of Phote-HrTH neuropeptide have been conducted using molecular binding data of the Drome-AKHR. This is the first time a molecular structure-activity study has been conducted on an AKH/RPCH family. In general, the result provides molecular binding data regarding Phote-HrTH/Drome-AKHR activation. The results correspond to the experimental data already available. This study further validates the constructed 3D model in previous chapters.

Since GPCRs are a frequent drug target for the treatment of human disease and given the physiological role of AKHs in energy metabolism in insects, AKHs and their GPCRs are considered suitable targets for new insect control agents. The binding of a neuropeptide to its receptor is vital for hormonal activity to take place. However, in order to target the GPCR, having an elucidated 3D structure and the identified ligand binding site, the position of the amino acid residue that is known to be most important for receptor activation could help in providing better information that could be used in the development of mimetic neuropeptide analogues that are selective in their activity, negatively affecting the aimed pest species without causing damage to valuable insect species.

8

Conclusions and Recommendations

Neuropeptides and their cognate G-protein-coupled receptors (GPCRs) engage in a significant role, particularly in the regulation of most crucial metabolic, reproductive, developmental, and behavioural processes in the life cycle of insects. G protein-coupled receptors (GPCRs) are the largest family of cell surface receptors. They are crucial human drug targets because bioactive peptides and proteins signal in the animal body via GPCRs, and if one interferes with this ligand-receptor binding, one could potentially influence or disrupt critical physiological processes in the animal body[29–33,41].

The AKHR is one such GPCR in insects that has the AKH as a ligand to regulate metabolism and energy availability for intense locomotion, like flying. AKH is functionally like the vertebrate hormone glucagon that causes a rise in the circulating glucose concentrations[15]. In insects, the sugar in the haemolymph is trehalose, and this is converted from stored glycogen in the fat body [15,18]. Another fuel for insect metabolism is in the form of the lipid diacylglycerol (DAG). DAG is obtained from the stored triacylglycerols (TAG) in the fat body of the insect[19,20]. The mobilisation of both fuel types for energy (cellular respiration) in animal cells is under the hormonal regulation of AKH[14–17].

Insects are the most abundant and most diverse group of animals on this planet. It is not surprising, therefore, that insects account for many pests as seen from a human perspective:

they compete for the same food (agricultural and stored food pests)[1], and they act as vectors of numerous, severe pathogens to both man and domesticated animals (medical pests and transmitters of disease) [469,470]. However, not all insects are pests; some are very beneficial, while some are both beneficial and pest [6,7]. Some beneficial insects serve as a source of pharmacologically active substances such as antimicrobial, antiviral and antitumor compounds. Some are decomposers of organic material [8,9,471]. Others are pollinators of flowers resulting in fruit and seed production.

Despite the difference in "ecological services" offered by different insect species, the current method of combatting insects is killing via commercially available non-specific toxic chemicals [472]. Over the years, some of the harmful insects have gained resistance to the widely used insecticides [34–36]. Moreover, more disturbing, a massive decline in insect biodiversity has been recorded of late; this includes the essential beneficial insects[37,38]. This has spurred on the search for "green insecticides" (species-specific pesticide that is selective in its action, negatively affecting the pest insect and causing no harm to the beneficial insect), to curb the massive, indiscriminate killing of insects, especially the beneficial ones. One idea for developing a green insecticide is to target the AKHRs of pest insects. For this to be successful and applicable, research into structure-based ligand-AKHR interactions is needed for pest and beneficial insects to find those that are specific to pest insects for further drug development. Essential for this is the availability of 3D AKHR models. Even though the relationship between physiological process and 3D structures are vital, minimal 3D structures of insect AKHRs are available. No 3D model of an AKHR of the flesh fly, the fruit fly, or the oriental fruit fly (*Bactrocera dorsalis, Drosophila melanogaster*, and *Sarcophaga crassipalpis*) is known. This means that no molecular data ranging from a 3D molecule of their AKHR and the binding sites of these insects are available. The absence of the 3D structures of these insect AKHRs serves as an obstacle to structure-based drug design.

In this context, this project was developed to focus on Phote-HrTH identified to have the same effect on the AKHR of three different fly species (*Drosophila melanogaster* (vinegar/fruit fly), *Bactrocera dorsalis* (oriental fruit fly) and *Sarcophaga crassipalpis* (flesh fly)), in which Phote-HrTH occurs naturally. Understanding the AKH/ AKHR interactions will provide us with enough molecular data in which we can investigate how an endogenous ligand (Phote-HrTH) may interact with its endogenous AKH receptor in beneficial, and pest fly species. The resulting models will shed light on whether the same amino acids are involved with ligand binding and whether the same conformation is maintained in all species. Hence, insight can ultimately be gained into whether a species-specific effect can be predicted between the three fly species and possibly others who also produce Phote-HrTH as a potential lead for peptide mimetics. Computational techniques were employed here to study AKH-AKHR interactions. Via nuclear magnetic resonance (NMR) restrained molecular dynamic (MD) the solution structure of the ligand (Phote-HrTH) was elucidated (see Chapter 2). Homology modelling techniques were used to construct two 3D models for each of the fly species, using β2AR and rhodopsin crystal structures as templates (see Chapter 3, 4 and 5 for more detail). The two receptor models built for each species differ in a few critical features, the most important of which is that the β2AR-based models have an open conformation, while the rhodopsin-based models have a closed structure[46]. Since ligand binding takes place on the extracellular domain [363], the closed nature of all the rhodopsin-like AKHR-models prevents the ingress of the ligand. This behaviour of the rhodopsin model is similar to the behaviour reported for the rhodopsin template crystal structure [96,98]. Also, the rhodopsin-models produced no pose when ligand docking was attempted, so only the β2AR models were used for molecular docking calculation. To enable us to imitate the ligand- receptor's precise nature, molecular docking calculation using Phote-HrTH as the ligand for each of the β2AR-models were conducted. The docking

calculation was followed by long MD simulations in a model membrane to determine stability and conformational changes during agonist binding (see Chapter 6 for more details).

In order to validate the model built of the fruit fly AKH-receptor (Drome-AKHR), *in-silico* structure-activity studies were conducted and compared with previously published biological activity data.

8.1 Questions addressed by this study

The constructed 3D structures of all three fly species are based on the two-target template, namely β2AR (5D5A.1A) and rhodopsin (2X72.1A). The overall structure constructed AKHR models (**Figure 8.1**) are alike, especially the anticlockwise arrangement of the transmembrane (TM) helices, and they both possess the eighth intracellular helix that is parallel to the cell membrane. Also, all three structures have a tilted TM3, and the shortest transmembrane helix is TM4. In all the models, TM6 and TM5 differ with respect to length with both helices extending by approximately 4-7 residues into the cytoplasm. This observation is consistent with the studies conducted on the accessibility of nitroxide labels fixed to ECL3 joining TM6 and TM5 [360]. The AKHR models possess a disulphide bridge, commonly found in GPCRs.

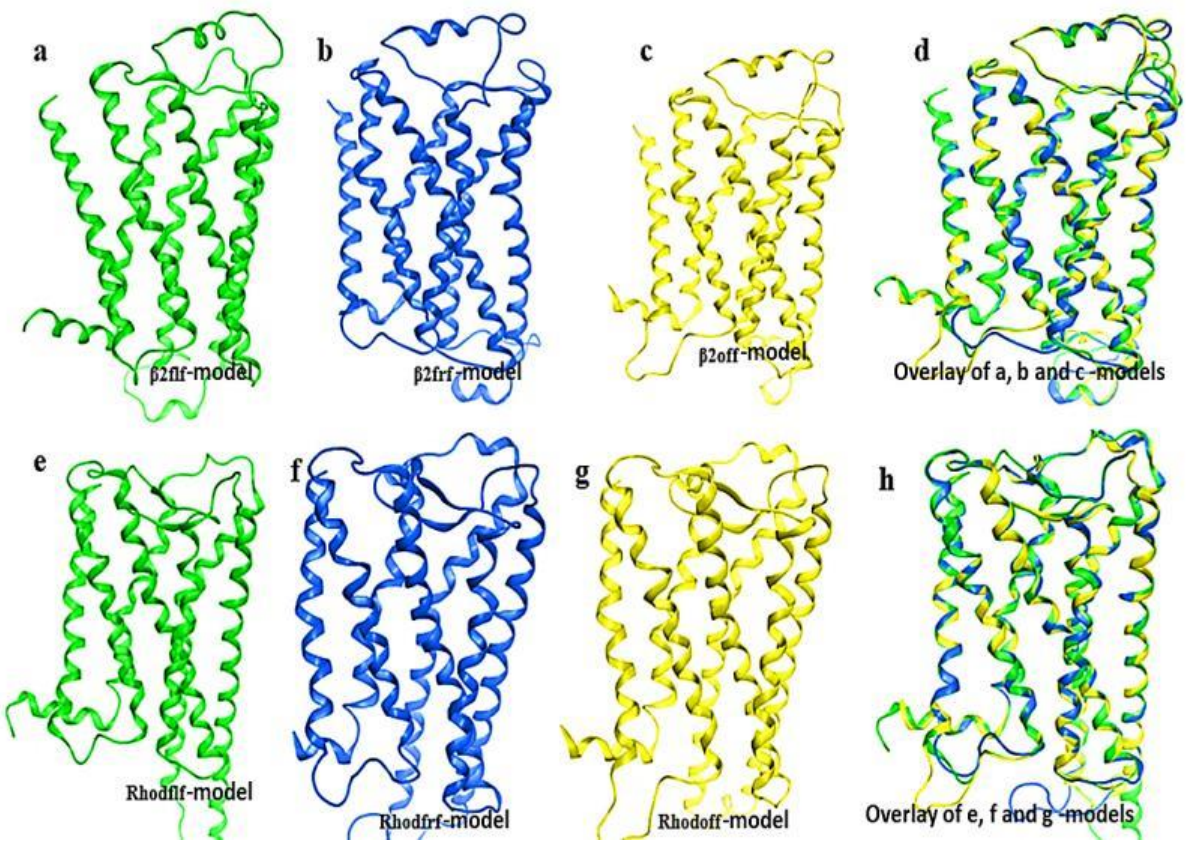


Figure 8.1: Constructed 3D model of the fly AKHR from the β2AR and rhodopsin templates. (a) β2AR Flesh fly model, (b) β2AR Fruit fly model and (c) β2AR Oriental fruit fly model. (d) Overlay of all three constructed 3D models from the β2AR template structure. (e) Rhodopsin flesh fly model, (f) Rhodopsin fruit fly model and (g) Rhodopsin oriental fruit fly model. (h) Overlay of all three rhodopsin models.

The two receptor models constructed for each species differ in a few critical features, the most important of which is that the β2AR-based models have an open conformation (inactive) (see **Figure 8.1 a-d**). In contrast, the rhodopsin-based models have a closed structure (active) (**Figure 8.1e-h**). The closed conformation of all the constructed rhodopsin-like AKHRs is similar to the template rhodopsin crystal structure [96,98]. Since ligand binding takes place on the extracellular domain [363], the closed nature of all the rhodopsin-like AKHR-models prevents the ingress of the ligand.

On the other hand, the open nature of all the β2AR-based AKHR-models allows the ligand-free access to the receptor-binding pocket. As is typical with most class A GPCRs, all the constructed AKHR models have flexible joints, known as kinks. The presence of kinks enables

the molecule to initiate a helical twist during receptor activation [371]. These kinks are usually identified with Tyr, Thr or Met proceeding a proline [368,370]. A critical feature of most class A GPCRs is the existence of substantially preserved molecular switch motifs. The presence of these switches is essential for stabilising any GPCR either in the close (active) or open (inactive) state. In all the constructed models, these switch motifs were identified. The famous 3-6 lock was identified between Tyr on TM3 and Arg on TM6. In the inactive state, Arg points away from TM3, and so no interaction is possible. Upon activation, however, TM6 twists moving Arg so that it now points toward Tyr. This interaction locks the receptor in the active state. Also, a molecular switch interaction between Lys in TM6 and the positively charged side chain nitrogen atom of Arg in TM3 was identified in all the rhodopsin based AKHR models. These switches are found to be the same as those described for the AKHR of *A. gambiae, Rhodnius prolixus*, and *Tribolium castaneum* [231,366].

The helices in all the constructed β2AR-based molecules (**Figure 8.1 a-d**) are longer when compared with that of the rhodopsin-based AKHR models (**Figure 8.1 e-h**). β2AR-based AKHR models all possess β-strands in ECL1. In all constructed β2AR-based AKHR models, a helix is identified in their N-terminus but is absent in all the rhodopsin AKHR models. Because of these differences, the rhodopsin models were proposed to serve as controls for the active models of GPCRs. This paragraph answers question one from Chapter 1 which asked about the difference and similarities of the constructed models.

During the MD simulations in a membrane (to mimic the actual receptors environment), all three receptors were seen to behave in the same manner. The opening on the extracellular side of all three AKHR/ligand complexes of the β2AR-based models gradually closed. Concurrently the intracellular side was observed to open up. The opening and closing of each receptor were evident from TM3 moving away from TM6. The inward movement of the extracellular domain of the receptor results in a closed conformation, similar to the rhodopsin-

based model. This movement or conformational structural change of these transmembrane helices is essential for receptor activation [29]. The RMSD of the amino acid backbone atoms between the starting structure and structure after 100 ns was calculated to be 2.37Å for Drome-AKHR, 2.47Å for Sarcr-AKHR and 2.31Å for Bacdo-AKHR. Also, for each of the three AKHR models, all three intracellular loops were pushed outwards. These movements are shown in **Figure 8.2**. The best explanation for this is that the molecule converted from an open (inactive) to a closed (active) conformation. The same observation was recorded when epinephrine (agonist) was introduced to its receptor. [89] The rapid realisation of the active conformation is consistent with the activation of fluorescently labelled β2AR in a detergent by noradrenaline which happens within approximately 40 ms. [447,448]. This paragraph answers the question concerning conformational changes of the constructed GPCRs / AKHR during MD simulations.

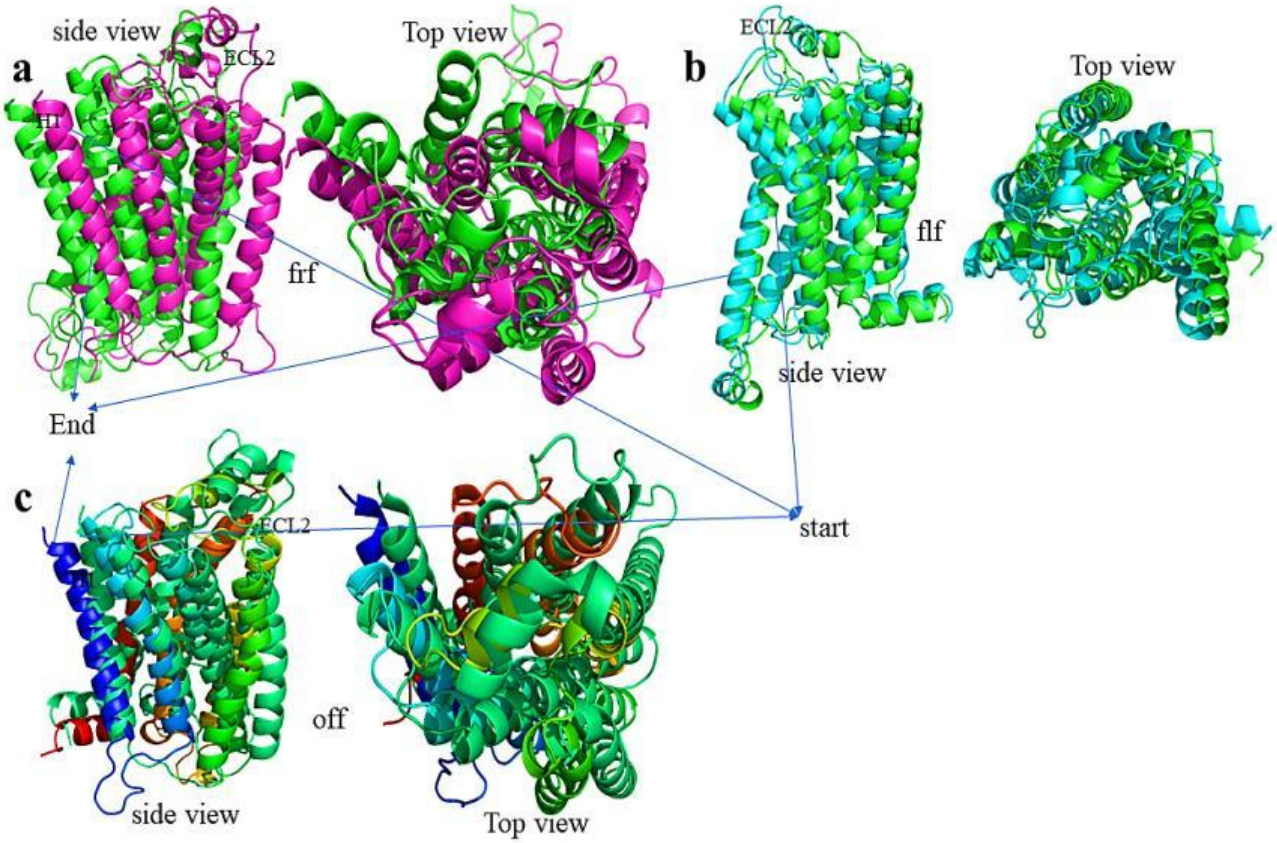


Figure 8.2: Overlay of β2AR models of Drome-AKHR (frf), Sarcr-AKHR (flf) and Bacdo-AKHR(off) at the start and end (blue lines) of a 100 ns MD simulation in a POPC membrane * flf -model = flesh fly, frf-model = fruit fly and off-model = oriental fruit fly.

The binding cavities of the constructed models are the same as their template structures. Rhodopsin models, even when the ligand was forced into the identified binding site from the extracellular region, the ECL2 had already covered the binding site leaving a tiny space. Unfortunately, the available space between the loops and the helices is too small for the ligand to access the binding cavity. Also, for the Rhodopsin models, when the Glide S-peptide docking command was initiated, no poses were generated, indicating the absence of cavities. While in the β2AR-based AKHRs, because of their open conformation, it was easy for proper diffusion of the peptide in and out of the binding site access. Since the Rhodopsin AKHR models are closed and would not allow further probing, by allowing the peptide to sink in, one could

postulate that the β2AR-models could represent the conformation of the receptor/inactive AKHR. At the same time, the Rhodopsin-model is used as a control. In each of the β2AR-based models, the extracellular loops (ECL) and transmembrane (TM) helices were involved in hormonal binding. Also, the identified binding cavity of these models is situated between ECL1, ECL2, the N-terminus and TM2-7. These finding of the binding site are consistent with those reported for *Anopheles gambiae* by Jackson et al.,[324].

The ligand binding to the individual constructed receptor (β2frf, β2off and β2flf-model) is comparable to each other since they all have hydrogen interactions, salt bridge, ionic interactions and π-π stacking interactions. Though, β2frf-model has two π-π stacking interactions originating from same molecule Phe4 and Trp8 of Phote-HrTH and Phe198 of the receptor, while β2flf-model has a π-π stacking interactions between Trp8 of Phote-HrTH and His203 and β2off-model possess a π-cationic interaction between Lys 306 and Phe4. The π-π stacking and π-cationic interactions are the only differences recorded, the presence of π-interaction provides a significant amount of binding enthalpy and stability to the system. The π-interaction with the ECL2 help closes the loop over the binding site; in other words, the receptor got activated. These observations were similar to those reported by Dougherty [379,443]. The question relating the location of the constructed receptors binding site and ligand binding of both active and inactive receptor is answered in this paragraph.

The docking score of the models are β2flf-model -13.20 kJ/ mol, β2frf-model is - 14.33 kJ/mol and β2off-model is - 14.03 kJ/mol. The binding scores (ΔG_{bind}) calculated using MMGBSA, are -75 kJ/mol for β2flf-model, -79 kJ/mol for β2frf-model and – 82 kJ/ mol for β2off-model.

During MD simulations of the beta2-adrenergic-based models, simultaneously from both the extracellular and intracellular regions of the AKHRs, rapid opening and closing respectively of these regions was noticed. These opening and closing of receptors have been described as activation in GCPRs since AKHRs are also GCPRs[141,224,329].

In order to validate our model of Drome-AKHR computational, mutation studies (structure-activity studies) were carried out. Experimental structural activity studies, conducted on Drome-AKHR, explains that most AKHs have Phe or Tyr at position four and Trp at position eight [63,450]. Hence, these specific residues appear to be essential for peptide activity [63]. Exchange of this amino acid residue with Ala at position 4 or Gly at position 8 results in total shut down of Drome-AKHR [76]. The same phenomenon is observed in other insect species [454,455,458]. These same observations were also confirmed in the molecular binding studies, see Chapter seven. The computational mutation studies were able to mimic the experimental activity data successfully. The aromatic Phe ring in position 4 and that of the indole ring Trp 8, which have been reported in the previous chapter as providing stability to the ligand during ligand-receptor binding is very much crucial for receptor binding and activation. Leu in position 2, Thr in position 3, Ser in position 5 and Pro in position 6 are also important in receptor binding but not necessary activation. Asp in position 7 does not seem to participate in either binding or activation of the receptor, as when exchanged with Gly the ligand gained an additional negative charge and the receptor was activated. Knowledge from these studies could hit the final nail in the coffin regarding fly species threatening food security and Agriculture. The correlation between the experimental data and molecular data on structure-activity is given in this section.

The constructed models have proven to be high standard as the data obtained is consistent with structures elucidated by the experimental method [125,201,331,361,444,473–478].

Finally, the aim of this project has been achieved. The project was able to demonstrate Phote-HrTH ligand activating receptors from three different insects, namely, the flesh fly *Sarcophaga crassipalpis*, the vinegar fly (also called fruit fly) *Drosophila melanogaster* and the oriental fruit fly *Bactrocera dorsalis*. Molecular data provided from this study could be exploited towards the design of biological active pesticides that are selective in their activity, negatively affecting the aimed pest species without causing damage to valuable insect species.

8.2 Future Work

An essential aspect of this research is the development of mimetic neuropeptide analogues that are selective in their activity, negatively affecting the targeted pest species without harm to beneficial insect species like the honeybee. Phote-HrTH could be modified, making sure it retains full efficacy on any of the constructed flies AKHR from this study. The modified Phote-HrTH should also be tested on the honeybee AKH-receptor (Apime-AKHR) to check selectivity.

Secondly, the tsetse fly, *Glossina morsitans morsitans*, has two AKHs code-named Glomo-AKH I (pGlu-Leu-Thr-Phe-Ser-Pro-Gly-Trp-NH_2) and Glomo-AKH II (Phote-HrTH: pGlu-Leu-Thr-Phe-Ser-Pro-Asp-Trp-NH_2). The binding of Glomo-AKH II/ Phote-HrTH to the extracellular binding site of a G protein-coupled receptor causes the tsetse fly receptor activation same as the Glomo-AKH I [43–45,70]. Since Phote-HrTH / Glomo-AKH II is also found to bind and activate Glomo-AKHR, Drome-AKHR, Sarcr-AKHR and Bacdo-AKHR, 3D Model of Glomo-AKHR and the peptide Glomo-AKH II should be elucidated. Molecular docking studies should be conducted on both Glomo-AKHR, Drome-AKHR, Sarcr-AKHR and Bacdo-AKHR using Glomo-AKH I as the ligand to see if the ligand will activate or show selectivity on the AKHRs.

Reference

1 S. Sela, D. Nestel, R. Pinto, E. Nemny-Lavy and M. Bar-Joseph, Mediterranean fruit fly as a potential vector of bacterial pathogens, *Appl. Environ. Microbiol.*, 2005, **71**, 4052–6.

2 F. Khamesipour, K. B. Lankarani, B. Honarvar and T. E. Kwenti, A systematic review of human pathogens carried by the housefly (Musca domestica L.), *BMC Public Health*, 2018, **18**, 1049.

3 Thomas Change, in *General Parasitology - 2nd Edition*, 2012, p. 660.

4 Grimaldi D. & Engel M., in *European Journal of Entomology*, European Journal of Entomology, 2006, vol. 103, pp. 273–275.

5 A. Ssymank, C. Kearns, P. T. Danish and F. Thompson, A major contribution to plant diversity and agricultural production, *Biodiversity*, 2008, 86–89.

6 A.-M. Klein, B. E. Vaissière, J. H. Cane, I. Steffan-Dewenter, S. A. Cunningham, C. Kremen and T. Tscharntke, Importance of pollinators in changing landscapes for world crops, *Proc. R. Soc. B Biol. Sci.*, 2007, **274**, 303–313.

7 B. B. Phillips, A. Williams, J. L. Osborne and R. F. Shaw, Shared traits make flies and bees effective pollinators of oilseed rape (Brassica napus L.), *Basic Appl. Ecol.*, 2018, **32**, 66–76.

8 A. C. Bhatia, T. E. Rohrer, C. A. Charles, A. F. Falabella and A. C. Fernández-Obregón, Leg Ulcer Management, *Surg. Ski.*, 2005, 743–766.

9 S. S. Kordestani and S. S. Kordestani, Wound Care Management, *Atlas Wound Heal.*, 2019, 31–47.

10 A. Y. Yakovlev, A. A. Kruglikova and S. I. Chernysh, Calliphoridae Flies in Medical Biotechnology, *Entomol. Rev.*, 2019, **99**, 292–301.

11 B. A. Rumpold and O. K. Schlüter, *Innov. Food Sci. Emerg. Technol.*, 2013, 17, 1–11.

12 D. A. Roff, The evolution of flightlessness in insects, *Ecol. Monogr.*, 1990, **60**, 389–421.

13 A. E. Kammer and B. Heinrich, Insect Flight Metabolism, *Adv. In Insect Phys.*, 1978, **13**, 133–228.

14 G. Gäde and H. G. Marco, in *Handbook of Biologically Active Peptides*, 2013, pp. 185–190.

15 G. Gäde and L. Auerswald, Insect neuropeptides regulating substrate mobilisation, *South African J. Zool.*, 1998, **33**, 65–70.

16 G. Gäde and L. Auerswald, Mode of action of neuropeptides from the adipokinetic hormone family, *Gen. Comp. Endocrinol.*, 2003, **132**, 10–20.

17 G. Gäde and H. G. Marco, Flight-related metabolism and its regulatory peptides in the spittle bug Locris arithmetica (Cicadomorpha: Cercopidae) and the stink bugs Nezara viridula (Heteroptera: Pentatomidae) and Encosternum delegorguei (Heteroptera: Tessaratomidae), *J. Insect Physiol.*, 2009, **55**, 1134–1144.

18 G. Gäde, The Explosion of Structural Information on Insect Neuropeptides, *Fortschr. Chem. Org. Naturst.*, 1997, **71**, 1–128.

19 M. Alves-Bezerra, I. F. De Paula, J. M. Medina, G. Silva-Oliveira, J. S. Medeiros, G. Gäde and K. C. Gondim, Adipokinetic hormone receptor gene identification and its

role in triacylglycerol metabolism in the blood-sucking insect Rhodnius prolixus, *Insect Biochem. Mol. Biol.*, 2016, **69**, 51–60.

20 E. L. Arrese and J. L. Soulages, Insect fat body: Energy, metabolism, and regulation, *Annu. Rev. Entomol.*, 2010, **55**, 207–225.

21 A. M. T. T. Beenakkers, D. J. Van der Horst and W. J. A. A. Van Marrewijk, *Insect Biochem.*, 1984, 14, 243–260.

22 C. M. Lok and D. J. van Der Horst, Chiral 1,2-diacylglycerols in the haemolymph of the locust, Locusta migratoria, *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.*, 1980, **618**, 80–87.

23 A. Tietz and H. Weintraub, The stereospecific structure of haemolymph and fat-body 1,2-diacylglycerol from Locusta migratoria, *Insect Biochem.*, 1980, **10**, 61–63.

24 D. J. Van Der Horst, in *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, Elsevier Inc., 2003, vol. 136, pp. 217–226.

25 G. Gäde, Adipokinetic hormones and the hormonal control of metabolic activity in Hemiptera, *Pestycydy*, 2005, 49–54.

26 J. Michitsch and J. E. Steele, Carbohydrate and lipid metabolism in cockroach (Periplaneta americana) fat body are both activated by low and similar concentrations of Peram-AKH II, *Peptides*, 2008, **29**, 226–234.

27 J. Caers, H. Verlinden, S. Zels, H. P. Vandersmissen, K. Vuerinckx and L. Schoofs, More than two decades of research on insect neuropeptide GPCRs: an overview., *Front. Endocrinol. (Lausanne).*, 2012, **3**, 151.

28 E. Marchal, S. Schellens, E. Monjon, E. Bruyninckx, G. H. Marco, G. Gäde, J. Vanden Broeck and H. Verlinden, Analysis of Peptide Ligand Specificity of Different Insect Adipokinetic Hormone Receptors, *Int. J. Mol. Sci.*, 2018, **19**, 542.

29 B. K. Kobilka, G protein coupled receptor structure and activation, *Biochim. Biophys. Acta - Biomembr.*, 2007, **1768**, 794–807.

30 N. Audsley and R. E. Down, G protein coupled receptors as targets for next generation pesticides, *Insect Biochem. Mol. Biol.*, 2015, **67**, 27–37.

31 H. Verlinden, R. Vleugels, S. Zels, S. Dillen, C. Lenaerts, K. Crabbé, J. Spit and J. Vanden Broeck, Receptors for neuronal or endocrine signalling molecules as potential targets for the control of insect pests, *Adv. In Insect Phys.*, 2014, **46**, 167–303.

32 T. G. Geary and T. M. Kubiak, Neuropeptide G-protein-coupled receptors, their cognate ligands and behavior in Caenorhabditis elegans., *Trends Pharmacol. Sci.*, 2005, **26**, 56–8.

33 R. Seifert and K. Wenzel-Seifert, Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors, *Naunyn. Schmiedebergs. Arch. Pharmacol.*, 2002, **366**, 381–416.

34 F. P. Carvalho, Agriculture, pesticides, food security and food safety, *Environ. Sci. Policy*, 2006, **9**, 685–692.

35 F. Geiger, J. Bengtsson, F. Berendse, W. W. Weisser, M. Emmerson, M. B. Morales, P. Ceryngier, J. Liira, T. Tscharntke, C. Winqvist, S. Eggers, R. Bommarco, T. Pärt, V. Bretagnolle, M. Plantegenest, L. W. Clement, C. Dennis, C. Palmer, J. J. Oñate, I. Guerrero, V. Hawro, T. Aavik, C. Thies, A. Flohre, S. Hänke, C. Fischer, P. W.

Goedhart and P. Inchausti, Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland, *Basic Appl. Ecol.*, 2010, **11**, 97–105.

36 G. Imfeld and S. Vuilleumier, Measuring the effects of pesticides on bacterial communities in soil: A critical review, *Eur. J. Soil Biol.*, 2012, **49**, 22–30.

37 G. Wang, Y. Dong, X. Liu, G. Yao, X. Yu and M. Yang, The Current Status and Development of Insect-Resistant Genetically Engineered Poplar in China., *Front. Plant Sci.*, 2018, **9**, 1408.

38 F. Gould, Z. S. Brown and J. Kuzma, Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance?, *Science*, 2018, **360**, 728–732.

39 W. E. Kunin, Robust evidence of declines in insect abundance and biodiversity, *Nature*, 2019, **574**, 641–642.

40 F. Sánchez-Bayo and K. A. G. Wyckhuys, Worldwide decline of the entomofauna: A review of its drivers, *Biol. Conserv.*, 2019, **232**, 8–27.

41 M. Bil, V. Broeckx, B. Landuyt and R. Huybrechts, Differential peptidomics highlights adipokinetic hormone as key player in regulating digestion in anautogenous flesh fly, Sarcophaga crassipalpis, *Gen. Comp. Endocrinol.*, 2014, **208**, 49–56.

42 A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth and D. E. Gloriam, Trends in GPCR drug discovery: new agents, targets and indications, *Nat. Rev. Drug Discov.*, 2017, **16**, 829–842.

43 M. Bil, I. Timmermans, H. Verlinden and R. Huybrechts, Characterization of the adipokinetic hormone receptor of the anautogenous flesh fly, Sarcophaga crassipalpis,

J. Insect Physiol., 2016, **89**, 52–59.

44 Q.-L. Hou, E.-H. Chen, H.-B. Jiang, D.-D. Wei, S.-H. Gui, J.-J. Wang and G. Smagghe, Adipokinetic hormone receptor gene identification and its role in triacylglycerol mobilization and sexual behavior in the oriental fruit fly (Bactrocera dorsalis), *Insect Biochem. Mol. Biol.*, 2017, **90**, 1–13.

45 F. Staubli, T. J. D. Jørgensen, G. Cazzamali, M. Williamson, C. Lenz, L. Søndergaard, P. Roepstorff and C. J. P. Grimmelikhuijzen, Molecular identification of the insect adipokinetic hormone receptors, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 3446–3451.

46 B. Kobilka and G. Schertler, New G-protein-coupled receptor crystal structures: insights and limitations, *Trends Pharmacol. Sci.*, 2008, **29**, 79–83.

47 M. Sarwar, *Typical Flies: Natural History, Lifestyle and Diversity of Diptera*, IntechOpen, 2020.

48 B. Larson, P. Kevan and D. Inouye, Flies and flowers: taxonomic diversity of anthophiles and pollinators, *Can. Entomol.*, 2001, **133**, 439–465.

49 Brake I, Flies –Pollinators on two wings | The Diptera Site, http://diptera.myspecies.info/diptera/content/flies-–pollinators-two-wings?citethispage=997ae, (accessed 22 January 2020).

50 M. S. Dharne, A. K. Gupta, A. Y. Rangrez, H. V. Ghate, M. S. Patole and Y. S. Shouche, Antibacterial activities of multi drug resistant Myroides odoratimimus bacteria isolated from adult flesh flies (Diptera: Sarcophagidae) are independent of metallo beta-lactamase gene, *Brazilian J. Microbiol.*, 2008, **39**, 394–404.

51 C. Paquette, K. H. Joplin, E. Seier, J. T. Peyton and D. Moore, Sex-specific differences

in spatial behaviour in the flesh fly Sarcophaga crassipalpis, *Physiol. Entomol.*, 2008, **33**, 382–388.

52 S. Sotiraki, R. Farkas and M. J. R. Hall, Fleshflies in the flesh: Epidemiology, population genetics and control of outbreaks of traumatic myiasis in the Mediterranean Basin, *Vet. Parasitol.*, 2010, **174**, 12–18.

53 N. Tolwinski, Introduction: Drosophila—A Model System for Developmental Biology, *J. Dev. Biol.*, 2017, **5**, 9.

54 A. E. A. Stephens, D. J. Kriticos and A. Leriche, The current and future potential geographical distribution of the oriental fruit fly, Bactrocera dorsalis (Diptera: Tephritidae), *Bull. Entomol. Res.*, 2007, **97**, 369–378.

55 R. D. Flannagan, S. P. Tammariello, K. H. Joplin, R. A. Cikra-Ireland, G. D. Yocum and D. L. Denlinger, Diapause-specific gene expression in pupae of the flesh fly Sarcophaga crassipalpis, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5616–5620.

56 H. Bänziger and T. Pape, Flowers, faeces and cadavers: natural feeding and laying habits of flesh flies in Thailand (Diptera: Sarcophagidae, *Sarcophaga* spp.), *J. Nat. Hist.*, 2004, **38**, 1677–1694.

57 S. Mansourian, A. Enjin, E. V. Jirle, V. Ramesh, G. Rehermann, P. G. Becher, J. E. Pool and M. C. Stensmyr, Wild African Drosophila melanogaster Are Seasonal Specialists on Marula Fruit, *Curr. Biol.*, 2018, **28**, 3960-3968.e3.

58 M. K. Schutze, N. Aketarawong, W. Amornsak, K. F. Armstrong, A. A. Augustinos, N. Barr, W. Bo, K. Bourtzis, L. M. Boykin, C. Cáceres, S. L. Cameron, T. A. Chapman, S. Chinvinijkul, A. Chomič, M. De Meyer, E. Drosopoulou, A. Englezou, S.

Ekesi, A. Gariou-Papalexiou, S. M. Geib, D. Hailstones, M. Hasanuzzaman, D. Haymer, A. K. W. Hee, J. Hendrichs, A. Jessup, Q. Ji, F. M. Khamis, M. N. Krosch, L. Leblanc, K. Mahmood, A. R. Malacrida, P. Mavragani-Tsipidou, M. Mwatawala, R. Nishida, H. Ono, J. Reyes, D. Rubinoff, M. San Jose, T. E. Shelly, S. Srikachar, K. H. Tan, S. Thanaphum, I. Haq, S. Vijaysegaran, S. L. Wee, F. Yesmin, A. Zacharopoulou and A. R. Clarke, Synonymization of key pest species within the Bactrocera dorsalis species complex (Diptera: Tephritidae): Taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data, *Syst. Entomol.*, 2015, **40**, 456–471.

59 M. F. Day, The Homologies of the Ring Gland of Diptera Brachycera, *Ann. Entomol. Soc. Am.*, 1943, **36**, 1–10.

60 J. Stone, W. Mordue, K. E. Batley and H. R. Morris, Structure of locust adipokinetic hormone, a neurohormone that regulates lipid utilisation during flight, *Nature*, 1976, **263**, 207–211.

61 H. Jaffe, A. K. Raina, C. T. Riley, B. A. Fraser, R. J. Nachman, V. W. Vogel, Y. S. Zhang and D. K. Hayes, Primary structure of two neuropeptide hormones with adipokinetic and hypotrehalosemic activity isolated from the corpora cardiaca of horse flies (Diptera)., *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 8161–8164.

62 G. Gäde, Peptides of the adipokinetic hormone/red pigment-concentrating hormone family: A new take on biodiversity, *Ann. N. Y. Acad. Sci.*, 2009, **1163**, 125–136.

63 G. Gäde and H. G. Marco, Structure, function and mode of action of select arthropod neuropeptides, *Stud. Nat. Prod. Chem.*, 2006, **33**, 69–139.

64 H. Wilps and G. Gäde, Hormonal regulation of carbohydrate metabolism in the

blowfly Phormia terraenovae, *J. Insect Physiol.*, 1990, **36**, 441–449.

65 G. Gäde, H. Wilps and R. Kellner, Isolation and structure of a novel charged member of the red–pigment-concentrating hormone-adipokinetic hormone family of peptides isolated from the corpora cardiaca of the blowfly *Phormia terraenovae* (Diptera), *Biochem. J.*, 1990, **269**, 309–313.

66 M. H. Schaffer, B. E. Noyes, C. A. Slaughter, G. C. Thorne and S. J. Gaskell, The fruitfly *Drosophila melanogaster* contains a novel charged adipokinetic-hormone-family peptide, *Biochem. J.*, 1990, **269**, 315–320.

67 J. G. Stoffolano, K. Croke, J. Chambers, G. Gäde, P. Solari and A. Liscia, Role of Phote-HrTH (Phormia terraenovae hypertrehalosemic hormone) in modulating the supercontractile muscles of the crop of adult Phormia regina Meigen, *J. Insect Physiol.*, 2014, **71**, 147–155.

68 G. M. Attardo, A. M. M. Abd-Alla, A. Acosta-Serrano, J. E. Allen, R. Bateta, J. B. Benoit, K. Bourtzis, J. Caers, G. Caljon, M. B. Christensen, D. W. Farrow, M. Friedrich, A. Hua-Van, E. C. Jennings, D. M. Larkin, D. Lawson, M. J. Lehane, V. P. Lenis, E. Lowy-Gallego, R. W. Macharia, A. R. Malacrida, H. G. Marco, D. Masiga, G. L. Maslen, I. Matetovici, R. P. Meisel, I. Meki, V. Michalkova, W. J. Miller, P. Minx, P. O. Mireji, L. Ometto, A. G. Parker, R. Rio, C. Rose, A. J. Rosendale, O. Rota-Stabelli, G. Savini, L. Schoofs, F. Scolari, M. T. Swain, P. Takáč, C. Tomlinson, G. Tsiamis, J. Van Den Abbeele, A. Vigneron, J. Wang, W. C. Warren, R. M. Waterhouse, M. T. Weirauch, B. L. Weiss, R. K. Wilson, X. Zhao and S. Aksoy, Comparative genomic analysis of six Glossina genomes, vectors of African trypanosomes, *Genome Biol.*, 2019, **20**, 187.

69 C. Kaufmann and M. R. Brown, Adipokinetic hormones in the African malaria mosquito, Anopheles gambiae: Identification and expression of genes for two peptides and a putative receptor, *Insect Biochem. Mol. Biol.*, 2006, **36**, 466–481.

70 J. Caers, T. Janssen, L. Van Rompay, V. Broeckx, J. Van Den Abbeele, G. Gäde, L. Schoofs and I. Beets, Characterization and pharmacological analysis of two adipokinetic hormone receptor variants of the tsetse fly, Glossina morsitans morsitans, *Insect Biochem. Mol. Biol.*, 2016, **70**, 73–84.

71 G. Gäde, P. Šimek and H. G. Marco, The Adipokinetic Peptides in Diptera: Structure, Function, and Evolutionary Trends, *Front. Endocrinol. (Lausanne).*, 2020, **11**, 153.

72 M. W. Lorenz and G. Gade, Hormonal regulation of energy metabolism in insects as a driving force for performance, *Integr. Comp. Biol.*, 2009, **49**, 380–392.

73 A. M. T. Beenakkers, R. E. B. Bloemen, T. A. De Vlieger, D. J. Van Der Horst and W. J. A. Van Marrewijk, Insect adipokinetic hormones, *Peptides*, 1985, **6**, 437–444.

74 H. G. Marco, H. Verlinden, J. Vanden Broeck, G. Gäde, J. Vanden Broeck and G. Gäde, Characterisation and pharmacological analysis of a crustacean G protein-coupled receptor: The red pigment-concentrating hormone receptor of Daphnia pulex, *Sci. Rep.*, 2017, **7**, 6851.

75 M. Lindemans, F. Liu, T. Janssen, S. J. Husson, I. Mertens, G. Gäde and L. Schoofs, Adipokinetic hormone signaling through the gonadotropin-releasing hormone receptor modulates egg-laying in *Caenorhabditis elegans*., *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 1642–7.

76 J. Caers, L. Peeters, T. Janssen, W. De Haes, G. Gäde and L. Schoofs, Structure-

activity studies of Drosophila adipokinetic hormone (AKH) by a 1 cellular expression system of dipteran AKH receptors, *Gen. Comp. Endocrinol.*, 2012, **177**, 332–337.

77 G. Gäde, L. Auerswald and H. G. Marco, Flight fuel and neuropeptidergic control of fuel mobilisation in the twig wilter, Holopterna alata (Hemiptera, Coreidae), *J. Insect Physiol.*, 2006, **52**, 1171–1181.

78 G. Mugumbate, G. E. Jackson, D. Van Der Spoel, K. E. Kövér and L. Szilágyi, Anopheles gambiae, Anoga-HrTH hormone, free and bound structure-A nuclear magnetic resonance experiment, *Peptides*, 2013, **41**, 94–100.

79 B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski and S. Filipek, Action of molecular switches in GPCRs--theoretical and experimental studies, *Curr. Med. Chem.*, 2012, **19**, 1090–109.

80 J. L. Alexander, A. Oliphant, D. C. Wilcockson, N. Audsley, R. E. Down, R. Lafont and S. G. Webster, Functional Characterization and Signaling Systems of Corazonin and Red Pigment Concentrating Hormone in the Green Shore Crab, Carcinus maenas., *Front. Neurosci.*, 2017, **11**, 752.

81 G. Cornilescu, F. Delaglio and A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR*, 1999, **13**, 289–302.

82 R. Koradi, M. Billeter and K. Wüthrich, MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graph.*, 1996, **14**, 51–55.

83 J. Caers, M. B. Van Hiel, K. Peymen, S. Zels, L. Van Rompay, J. Van Den Abbeele, L. Schoofs and I. Beets, Characterization of a neuropeptide F receptor in the tsetse fly,

Glossina morsitans morsitans, *J. Insect Physiol.*, 2016, **93**, 105–111.

84 J. Caers, K. Peymen, M. B. Van Hiel, L. Van Rompay, J. Van Den Abbeele, L. Schoofs and I. Beets, Molecular characterization of a short neuropeptide F signaling system in the tsetse fly, Glossina morsitans morsitans, *Gen. Comp. Endocrinol.*, 2016, **235**, 142–149.

85 Y. Park, Y.-J. Kim and M. E. Adams, Identification of G protein-coupled receptors for Drosophila PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution, *Proc. Natl. Acad. Sci.*, 2002, **99**, 11423–11428.

86 R. Nygaard, Y. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser and B. K. Kobilka, The dynamic process of β(2)-adrenergic receptor activation., *Cell*, 2013, **152**, 532–42.

87 Gäde and Auerswald, Flight substrates and their regulation by a member of the AKH/RPCH family of neuropeptides in Cerambycidae., *J. Insect Physiol.*, 2000, **46**, 1575–1584.

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

89 K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto and M. Miyano, Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor, *Science*, 2000, **289**, 739–745.

90 K. Kristiansen, Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: Molecular modeling and mutagenesis approaches to receptor structure and function, *Pharmacol. Ther.*, 2004, **103**, 21–80.

91 K. Lundstrom, Structural genomics of GPCRs, *Trends Biotechnol.*, 2005, **23**, 103–108.

92 C. Bissantz, A. Antoine Logean and D. Rognan, High-throughput modeling of human g-protein coupled receptors: Amino acid sequence alignment, three-dimensional model building, and receptor library screening, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1162–1176.

93 K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto and M. Miyano, Crystal structure of rhodopsin: A G protein-coupled receptor., *Science*, 2000, **289**, 739–45.

94 F. Fanelli and P. G. De Benedetti, Computational Modeling Approaches to Structure−Function Analysis of G Protein-Coupled Receptors, *Chem. Rev.*, 2005, **105**, 3297–3351.

95 F. Fanelli and P. G. De Benedetti, Update 1 of: Computational Modeling Approaches to Structure–Function Analysis of G Protein-Coupled Receptors, *Chem. Rev.*, 2011, **111**, PR438–PR535.

96 S. G. F. Rasmussen, H.-J. Choi, D. M. Rosenbaum, T. S. Kobilka, F. S. Thian, P. C. Edwards, M. Burghammer, V. R. P. Ratnala, R. Sanishvili, R. F. Fischetti, G. F. X. Schertler, W. I. Weis and B. K. Kobilka, Crystal structure of the human β2 adrenergic G-protein-coupled receptor, *Nature*, 2007, **450**, 383–387.

97 V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. H.-J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka and R. C. Stevens, High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein-Coupled Receptor, *Science*, 2007, **318**, 1258–1265.

98 V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka and R. C. Stevens, High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein-Coupled Receptor, *Science*, 2007, **318**, 1258–1265.

99 B. Mouillac, M. Caron, H. Bonin, M. Dennis and M. Bouvier, Agonist-modulated palmitoylation of beta 2-adrenergic receptor in Sf9 cells., *J. Biol. Chem.*, 1992, **267**, 21733–7.

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

101 D. C. Teller, T. Okada, C. A. Behnke, K. Palczewski and R. E. Stenkamp, Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs)., *Biochemistry*, 2001, **40**, 7761–72.

102 H. Khandelia and Y. N. Kaznessis, Molecular dynamics simulations of helical antimicrobial peptides in SDS micelles: what do point mutations achieve?, *Peptides*, 2005, **26**, 2037–49.

103 J. D. A. Tyndall, B. Pfeiffer, A. Giovanni Abbenante and D. P. Fairlie, Over one hundred peptide-activated g protein-coupled receptors recognize ligands with turn structure, , DOI:10.1021/CR040689G.

104 H. Leontiadou, A. E. Mark and S. J. Marrink, Antimicrobial Peptides in Action, *J. Am. Chem. Soc.*, 2006, **128**, 12156–12161.

105 R. S. Harrison, G. Ruiz-Gómez, T. A. Hill, S. Y. Chow, N. E. Shepherd, R.-J. Lohman, G. Abbenante, H. N. Hoang and D. P. Fairlie, Novel Helix-Constrained Nociceptin Derivatives Are Potent Agonists and Antagonists of ERK Phosphorylation and Thermal Analgesia in Mice, *J. Med. Chem.*, 2010, **53**, 8400–8408.

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

107 T. A. A. Cross and S. J. J. Opella, Solid-state NMR structural studies of peptides and proteins in membranes, *Curr. Opin. Struct. Biol.*, 1994, **4**, 574–581.

108 C. Chipot, F. Dehez, J. R. Schnell, N. Zitzmann, E. Pebay-Peyroula, L. J. Catoire, B. Miroux, E. R. S. Kunji, G. Veglia, T. A. Cross and P. Schanda, Perturbations of Native Membrane Protein Structure in Alkyl Phosphocholine Detergents: A Critical Assessment of NMR and Biophysical Studies, *Chem. Rev.*, 2018, **118**, 3559–3607.

109 K. Wüthrich, in *NMR in Structural Biology*, World Scientific, 1995, p. 738.

110 J.-S. Yu, H. Noda and M. Shibasaki, Quaternary β 2,2 -Amino Acids: Catalytic Asymmetric Synthesis and Incorporation into Peptides by Fmoc-Based Solid-Phase Peptide Synthesis, *Angew. Chemie Int. Ed.*, 2018, **57**, 818–822.

111 H. Yu, X. Daura and W. F. van Gunsteren, Molecular dynamics simulations of peptides containing an unnatural amino acid: Dimerization, folding, and protein binding, *Proteins Struct. Funct. Bioinforma.*, 2003, **54**, 116–127.

112 T. Cierpicki and J. Otlewski, Amide proton temperature coefficients as hydrogen bond indicators in proteins, *J. Biomol. NMR*, 2001, **21**, 249–261.

113 B. Kobilka, The structural basis of G-protein-coupled receptor signaling (Nobel Lecture)., *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 6380–8.

114 D. N. Langelaan, E. M. Bebbington, T. Reddy and J. K. Rainey, Structural Insight into G-Protein Coupled Receptor Binding by Apelin †, *Biochemistry*, 2009, **48**, 537–548.

115 D. N. Langelaan and J. K. Rainey, Headgroup-Dependent Membrane Catalysis of Apelin−Receptor Interactions Is Likely, *J. Phys. Chem. B*, 2009, **113**, 10465–10471.

116 K. Mauldin, B. L. Lee, M. Oleszczuk, B. D. Sykes and R. O. Ryan, The carboxyl-terminal segment of apolipoprotein A-V undergoes a lipid-induced conformational change., *Biochemistry*, 2010, **49**, 4821–6.

117 D. J. Chitwood, Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Service., *Pest Manag. Sci.*, 2003, **59**, 748–53.

118 F. Xu, H. Wu, V. Katritch, G. W. Han, K. A. Jacobson, Z.-G. Gao, V. Cherezov and R. C. Stevens, Structure of an agonist-bound human A2A adenosine receptor., *Science*, 2011, **332**, 322–7.

119 A. Evers and T. Klabunde, Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor, *J. Med. Chem.*, 2005, **48**, 1088–1097.

120 S. Jayanthi, S. W. Kang, D. Bingham, B. A. Tessaro, T. K. S. Kumar and W. J. Kuenzel, Identification of antagonists to the vasotocin receptor sub-type 4 (VT4R)

involved in stress by molecular modelling and verification using anterior pituitary cells., *J. Biomol. Struct. Dyn.*, 2014, **32**, 648–660.

121 E. van der Horst, J. E. Peironcely, A. P. IJzerman, M. W. Beukers, J. R. Lane, H. W. van Vlijmen, M. T. Emmerich, Y. Okuno and A. Bender, A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization, *BMC Bioinformatics*, 2010, **11**, 316.

122 V. Katritch, G. Fenalti, E. E. Abola, B. L. Roth, V. Cherezov and R. C. Stevens, Allosteric sodium in class A GPCR signaling, *Trends Biochem. Sci.*, 2014, **39**, 233–244.

123 C. W. Gruber, M. Muttenthaler and M. Freissmuth, Ligand-based peptide design and combinatorial peptide libraries to target G protein-coupled receptors., *Curr. Pharm. Des.*, 2010, **16**, 3071–88.

124 K. Hollenstein, C. de Graaf, A. Bortolato, M.-W. Wang, F. H. Marshall and R. C. Stevens, Insights into the structure of class B GPCRs, *Trends Pharmacol. Sci.*, 2014, **35**, 12–22.

125 R. P. Millar and C. L. Newton, The Year In G Protein-Coupled Receptor Research, *Mol. Endocrinol.*, 2010, **24**, 261–274.

126 H. Eishingdrelo and S. Kongsamut, Minireview: Targeting GPCR Activated ERK Pathways for Drug Discovery., *Curr. Chem. genomics Transl. Med.*, 2013, **7**, 9–15.

127 M. Kuhn, M. Campillos, P. González, L. J. Jensen and P. Bork, Large-scale prediction of drug-target relationships, *FEBS Lett.*, 2008, **582**, 1283–1290.

128 A. J. Orry and B. A. Wallace, Modeling and docking the endothelin G-protein-coupled

receptor., *Biophys. J.*, 2000, **79**, 3083–94.

129 R. O. Dror, D. H. Arlow, D. W. Borhani, M. O. Jensen, S. Piana and D. E. Shaw, Identification of two distinct inactive conformations of the 2-adrenergic receptor reconciles structural and biochemical observations, *Proc. Natl. Acad. Sci.*, 2009, **106**, 4689–4694.

130 A. Heifetz, G. F. X. Schertler, R. Seifert, C. G. Tate, P. M. Sexton, V. V. Gurevich, D. Fourmy, V. Cherezov, F. H. Marshall, R. I. Storer, I. Moraes, I. G. Tikhonova, C. S. Tautermann, P. Hunt, T. Ceska, S. Hodgson, M. J. Bodkin, S. Singh, R. J. Law and P. C. Biggin, GPCR structure, function, drug discovery and crystallography: report from Academia-Industry International Conference (UK Royal Society) Chicheley Hall, 1–2 September 2014, *Naunyn. Schmiedebergs. Arch. Pharmacol.*, 2015, **388**, 883–903.

131 A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit and A. R. Leach, The ChEMBL database in 2017., *Nucleic Acids Res.*, 2017, **45**, D945–D954.

132 A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos and J. P. Overington, The ChEMBL bioactivity database: an update, *Nucleic Acids Res.*, 2014, **42**, D1083–D1090.

133 N. Y. Mok and R. Brenk, Mining the ChEMBL database: an efficient chemoinformatics workflow for assembling an ion channel-focused screening library., *J. Chem. Inf. Model.*, 2011, **51**, 2449–54.

134 A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S.

McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, 2012, **40**, D1100–D1107.

135 S.-K. Park, S. R. Shanbhag, A. E. Dubin, M. de Bruyne, Q. Wang, P. Yu, N. Shimoni, S. D'Mello, J. R. Carlson, G. L. Harris, R. A. Steinbrecht and C. W. Pikielny, Inactivation of olfactory sensilla of a single morphological type differentially affects the response ofDrosophila to odors, *J. Neurobiol.*, 2002, **51**, 248–260.

136 C. Kaufmann, H. Merzendorfer and G. Gäde, The adipokinetic hormone system in Culicinae (Diptera: Culicidae): Molecular identification and characterization of two adipokinetic hormone (AKH) precursors from Aedes aegypti and Culex pipiens and two putative AKH receptor variants from A. aegypti, *Insect Biochem. Mol. Biol.*, 2009, **39**, 770–781.

137 G. M. Attardo, J. B. Benoit, V. Michalkova, G. Yang, L. Roller, J. Bohova, P. Takáč and S. Aksoy, Analysis of lipolysis underlying lactation in the tsetse fly, Glossina morsitans, *Insect Biochem. Mol. Biol.*, 2012, **42**, 360–370.

138 R. Ziegler, J. Isoe, W. Moore, M. A. Riehle and M. A. Wells, The putative AKH receptor of the tobacco hornworm, Manduca sexta, and its expression., *J. Insect Sci.*, 2011, **11**, 40.

139 D. Wicher, H.-J. Agricola, S. Söhler, M. Gundel, S. H. Heinemann, L. Wollweber, M. Stengl and C. Derst, Differential Receptor Activation by Cockroach Adipokinetic Hormones Produces Differential Effects on Ion Currents, Neuronal Activity, and Locomotion, *J. Neurophysiol.*, 2006, **95**, 2314–2325.

140 D. J. Van der Horst and K. W. Rodenburg, Locust flight activity as a model for

hormonal regulation of lipid mobilization and transport, *J. Insect Physiol.*, 2010, **56**, 844–853.

141 D. M. Rosenbaum, S. G. F. Rasmussen and B. K. Kobilka, The structure and function of G-protein-coupled receptors, *Nature*, 2009, **459**, 356–363.

142 A. Manglik and A. C. Kruse, Structural Basis for G Protein-Coupled Receptor Activation, *Biochemistry*, 2017, **56**, 5628–5634.

143 A. Terakita, *Genome Biol.*, 2005, 6, 213.

144 B. Borhan, M. L. Souto, H. Imai, Y. Shichida, K. Nakanishi, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto and M. Miyano, Movement of Retinal Along the Visual Transduction Path, *Science*, 2000, **288**, 2209–2212.

145 Y. Shichida and H. Imai, Visual pigment: G-protein-coupled receptor for light signals., *Cell. Mol. Life Sci.*, 1998, **54**, 1299–315.

146 H. Imai, D. Kojima, T. Oura, S. Tachibanaki, A. Terakita and Y. Shichida, Single amino acid residue as a functional determinant of rod and cone visual pigments, *Proc. Natl. Acad. Sci.*, 1997, **94**, 2322–2326.

147 S. Filipek, R. E. Stenkamp, D. C. Teller and K. Palczewski, G Protein-Coupled Receptor Rhodopsin: A Prospectus, *Annu. Rev. Physiol.*, 2003, **65**, 851–879.

148 M. Das, Y. Du, O. Ribeiro, P. Hariharan, J. S. Mortensen, D. Patra, G. Skiniotis, C. J. Loland, L. Guan, B. K. Kobilka, B. Byrne and P. S. Chae, Conformationally Preorganized Diastereomeric Norbornane-Based Maltosides for Membrane Protein Study: Implications of Detergent Kink for Micellar Properties, *J. Am. Chem. Soc.*, 2017, **139**, 3072–3081.

149 C. D. Strader, T. Gaffney, E. E. Sugg, M. R. Candelore, R. Keys, A. A. Patchett and R. A. Dixon, Allele-specific activation of genetically engineered receptors., *J. Biol. Chem.*, 1991, **266**, 5–8.

150 G. Liapakis, J. A. Ballesteros, S. Papachristou, W. C. Chan, X. Chen and J. A. Javitch, The Forgotten Serine, *J. Biol. Chem.*, 2000, **275**, 37779–37788.

151 Q. Jiang, B. X. Lee, M. Glashofer, A. M. van Rhee and K. A. Jacobson, Mutagenesis reveals structure-activity parallels between human A2A adenosine receptors and biogenic amine G protein-coupled receptors., *J. Med. Chem.*, 1997, **40**, 2588–95.

152 J. H. Perlman, L. J. Laakkonen, F. Guarnieri, R. Osman and M. C. Gershengorn, A Refined Model of the Thyrotropin-Releasing Hormone (TRH) Receptor Binding Pocket. Experimental Analysis and Energy Minimization of the Complex between TRH and TRH Receptor†, *Biochemistry*, 1996, **35**, 7643–7650.

153 J. Kim, Q. Jiang, M. Glashofer, S. Yehle, J. Wess and K. A. Jacobson, Glutamate residues in the second extracellular loop of the human A2a adenosine receptor are required for ligand recognition., *Mol. Pharmacol.*, 1996, **49**, 683–91.

154 W. K. Kroeze, K. Kristiansen and B. L. Roth, Molecular biology of serotonin receptors structure and function at the molecular level., *Curr. Top. Med. Chem.*, 2002, **2**, 507–28.

155 S. A. Schädel, M. Heck, D. Maretzki, S. Filipek, D. C. Teller, K. Palczewski and K. P. Hofmann, Ligand Channeling within a G-protein-coupled Receptor, *J. Biol. Chem.*, 2003, **278**, 24896–24903.

156 G. Mugumbate, Enhancing the fight against malaria : from genome to structure and

activity of a G-protein coupled receptor from the mosquito, Anopheles Gambiae, *Thesis., Univ. Cape T. ,Faculty Sci. ,Department Chem.*, 2010, **1**, 1–220.

157 Y. Yokota, C. Akazawa, H. Ohkubo and S. Nakanishi, Delineation of structural domains involved in the subtype specificity of tachykinin receptors through chimeric formation of substance P/substance K receptors., *EMBO J.*, 1992, **11**, 3585–91.

158 U. Gether, T. E. Johansenq and T. W. Schwartz, Chimeric NK1 (Substance P)/NK3 (Neurokinin B) Receptors, identification of domains determining the binding specificity of tachykinin agonists, *J. Biol. Chem.*, 1993, **268**, 7893.

159 T. M. Fong, H. Yu, R. R. C. Huang and C. D. Strader, The extracellular domain of the neurokinin-1 receptor is required for high-affinity binding of peptides, *Biochemistry*, 1992, **31**, 11806–11811.

160 H. Huang, X. Deng, X. He, W. Yang, G. Li, Y. Shi, L. Shi, L. Mei, J. Gao and N. Zhou, Identification of distinct c-terminal domains of the Bombyx adipokinetic hormone receptor that are essential for receptor export, phosphorylation and internalization, *Cell. Signal.*, 2011, **23**, 1455–1465.

161 N. Bhogal, F. E. Blaney, P. M. Ingley, J. Rees and J. B. C. Findlay, Evidence for the Proximity of the Extreme N-Terminus of the Neurokinin-2 (NK 2) Tachykinin Receptor to Cys 167 in the Putative Fourth Transmembrane Helix, *Biochemistry*, 2004, **43**, 3027–3038.

162 T. W. Schwartz, S. Perlman, M. M. Rosenkilde and S. A. Hjorth, How Receptor Mutagenesis May Confirm or Confuse Receptor Classification, *Ann. N. Y. Acad. Sci.*, 1997, **812**, 71–84.

163 P. Walker, M. Munoz, R. Martinez and M. C. Peitsch, Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding., *J. Biol. Chem.*, 1994, **269**, 2863–2869.

164 N. Cotte, M. N. Balestre, S. Phalipou, M. Hibert, M. Manning, C. Barberis and B. Mouillac, Identification of residues responsible for the selective binding of peptide antagonists and agonists in the V2 vasopressin receptor., *J. Biol. Chem.*, 1998, **273**, 29462–8.

165 B. Han and A. H. Tashjian, Importance of extracellular domains for ligand binding in the thyrotropin-releasing hormone receptor., *Mol. Endocrinol.*, 1995, **9**, 1708–1719.

166 J. S. Davidson, D. Assefa, A. Pawson, P. Davies, J. Hapgood, I. Becker, C. Flanagan, R. Roeske and R. Millar, Irreversible Activation of the Gonadotropin-Releasing Hormone Receptor by Photoaffinity Cross-Linking: Localization of Attachment Site to Cys Residue in N-Terminal Segment, *Biochemistry*, 1997, **36**, 12881–12889.

167 T. Kannoa, A. Kanatani, S. L. C. Keen, S. Arai-Otsuki, Y. Haga, T. Iwama, A. Ishihara, A. Sakuraba, H. Iwaasa, M. Hirose, H. Morishima, T. Fukami and M. Ihara, Different binding sites for the neuropeptide Y Y1 antagonists 1229U91 and J-104870 on human Y1 receptors, *Peptides*, 2001, **22**, 405–413.

168 M. Valiquette, H. K. Vu, S. Y. Yue, C. Wahlestedt and P. Walker, Involvement of Trp-284, Val-296, and Val-297 of the human delta-opioid receptor in binding of delta-selective ligands., *J. Biol. Chem.*, 1996, **271**, 18789–96.

169 J. A. Lee, J. A. Brinkmann, E. D. Longton, C. E. Peishoff, M. A. Lago, J. D. Leber, R. D. Cousins, A. Gao and J. M. Stadel, Lysine 182 of Endothelin B Receptor Modulates Agonist Selectivity and Antagonist Affinity: Evidence for the Overlap of Peptide and

Non-peptide Ligand Binding Sites, *Biochemistry*, 1994, **33**, 14543–14549.

170 P. S. Stern, L. Yu, M.-Y. Choi, R. A. Jurenka, L. Becker and A. Rafaeli, Molecular modeling of the binding of pheromone biosynthesis activating neuropeptide to its receptor, *J. Insect Physiol.*, 2007, **53**, 803–818.

171 M.-Y. Choi, E.-J. Fuerst, A. Rafaeli and R. Jurenka, Role of extracellular domains in PBAN/pyrokinin GPCRs from insects using chimera receptors., *Insect Biochem. Mol. Biol.*, 2007, **37**, 296–306.

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

173 P. Samama, S. Cotecchia, T. Costa and R. J. Lefkowitz, A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model., *J. Biol. Chem.*, 1993, **268**, 4625–36.

174 C. Xue, Y. Wang and Y.-P. Hsueh, Assessment of constitutive activity of a G protein-coupled receptor, CPR2, in Cryptococcus neoformans by heterologous and homologous methods., *Methods Enzymol.*, 2010, **484**, 397–412.

175 D. Wifling, K. Löffel, U. Nordemann, A. Strasser, G. Bernhardt, S. Dove, R. Seifert and A. Buschauer, Molecular determinants for the high constitutive activity of the human histamine H4 receptor: functional studies on orthologues and mutants., *Br. J. Pharmacol.*, 2015, **172**, 785–98.

176 A. Vanden broeck, Jozef Torfs, Herbert Poels, Jeroen Poyer, Wendy Swinnen, Elfriede Ferket, Kathelijne Loof, Tachykinin-like Peptides and Their Receptors: A Review, *Ann. N. Y. Acad. Sci.*, 1999, **897**, 374–387.

177 C. Lenaerts, D. Cools, R. Verdonck, L. Verbakel, J. Vanden Broeck and E. Marchal, The ecdysis triggering hormone system is essential for successful moulting of a major hemimetabolous pest insect, Schistocerca gregaria, *Sci. Rep.*, 2017, **7**, 650–652.

178 J. Vanden Broeck, Insect G protein-coupled receptors and signal transduction, *Arch. Insect Biochem. Physiol.*, 2001, **48**, 1–12.

179 L. A. Hu, W. Chen, N. P. Martin, E. J. Whalen, R. T. Premont and R. J. Lefkowitz, GIPC Interacts with the β_1-Adrenergic Receptor and Regulates β_1-Adrenergic Receptor-mediated ERK Activation, *J. Biol. Chem.*, 2003, **278**, 26295–26301.

180 A. L. Elefsinioti, P. G. Bagos, I. C. Spyropoulos and S. J. Hamodrakas, A database for G proteins and their interaction with GPCRs., *BMC Bioinformatics*, 2004, **5**, 208.

181 T. M. Cabrera-Vera, J. Vanhauwe, T. O. Thomas, M. Medkova, A. Preininger, M. R. Mazzoni and H. E. Hamm, Insights into G Protein Structure, Function, and Regulation, *Endocr. Rev.*, 2003, **24**, 765–781.

182 O. Tastan, J. Klein-Seetharaman and H. Meirovitch, The effect of loops on the structural organization of alpha-helical membrane proteins., *Biophys. J.*, 2009, **96**, 2299–312.

183 K. Sale, J.-L. Faulon, G. A. Gray, J. S. Schoeniger and M. M. Young, Optimal bundling of transmembrane helices using sparse distance constraints., *Protein Sci.*, 2004, **13**, 2613–27.

184 M. Tahir ul Q, S. Kiran, U. A. Ashfaq, M. R. Javed, F. Anwar, M. A. Ali and A. ul H. Gilani, Discovery of Novel Dengue NS2B/NS3 Protease Inhibitors Using Pharmacophore Modeling and Molecular Docking Based Virtual Screening of the

ZINC Database, *Int. J. Pharmacol.*, 2016, **12**, 621–632.

185 N. Kunishima, Y. Shimada, Y. Tsuji, T. Sato, M. Yamamoto, T. Kumasaka, S. Nakanishi, H. Jingami and K. Morikawa, Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor, *Nature*, 2000, **407**, 971–977.

186 S. Vohra, B. Taddese, A. C. Conner, D. R. Poyner, D. L. Hay, J. Barwell, P. J. Reeves, G. J. G. Upton and C. A. Reynolds, Similarity between class A and class B G-protein-coupled receptors exemplified through calcitonin gene-related peptide receptor modelling and mutagenesis studies., *J. R. Soc. Interface*, 2013, **10**, 20120846.

187 J. Skolnick, H. Zhou and M. Gao, Are predicted protein structures of any value for binding site prediction and virtual ligand screening?, *Curr. Opin. Struct. Biol.*, 2013, **23**, 191–7.

188 D. E. Gloriam, S. M. Foord, F. E. Blaney and S. L. Garland, Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design, *J. Med. Chem.*, 2009, **52**, 4429–4442.

189 M. Fridén-Saxin, T. Seifert, M. Malo, K. da Silva Andersson, N. Pemberton, C. Dyrager, A. Friberg, K. Dahlén, E. A. A. Wallén, M. Grøtli and K. Luthman, Chroman-4-one and chromone based somatostatin β-turn mimetics, *Eur. J. Med. Chem.*, 2016, **114**, 59–64.

190 N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos and W. A. Goddard, Prediction of structure and function of G protein-coupled receptors, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12622–12627.

191 C. Hetényi and D. van der Spoel, Toward prediction of functional protein pockets

using blind docking and pocket search algorithms, *Protein Sci.*, 2011, **20**, 880–893.

192 S. Mühlfeld, H.-P. Schmitt-Wrede, A. Harder and F. Wunderlich, FMRFamide-like neuropeptides as putative ligands of the latrophilin-like HC110-R from Haemonchus contortus, *Mol. Biochem. Parasitol.*, 2009, **164**, 162–164.

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

194 F. Wu, G. Song, C. de Graaf and R. C. Stevens, Structure and Function of Peptide-Binding G Protein-Coupled Receptors, *J. Mol. Biol.*, 2017, **429**, 2726–2745.

195 C. De Graaf and D. Rognan, Customizing G Protein-coupled receptor models for structure-based virtual screening., *Curr. Pharm. Des.*, 2009, **15**, 4026–48.

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

197 J.-U. Peters, J. Hert, C. Bissantz, A. Hillebrecht, G. Gerebtzoff, S. Bendels, F. Tillier, J. Migeon, H. Fischer, W. Guba and M. Kansy, Can we discover pharmacological promiscuity early in the drug discovery process?, *Drug Discov. Today*, 2012, **17**, 325–335.

198 C. Bissantz, A. Logean, D. Rognan, A. Antoine Logean and D. Rognan, High-throughput modeling of human g-protein coupled receptors: Amino acid sequence alignment, three-dimensional model building, and receptor library screening, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1162–1176.

199 A. Lavecchia and C. Di Giovanni, Virtual screening strategies in drug discovery: a critical review., *Curr. Med. Chem.*, 2013, **20**, 2839–60.

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

201 C. L. Worth, G. Kleinau and G. Krause, Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models, *PLoS One*, 2009, **4**, e7011.

202 T. Yarnitzky, A. Levit and M. Y. Niv, Homology modeling of G-protein-coupled receptors with X-ray structures on the rise., *Curr. Opin. Drug Discov. Devel.*, 2010, **13**, 317–25.

203 B. Schlegel, W. Sippl and H.-D. Höltje, Molecular dynamics simulations of bovine rhodopsin: influence of protonation states and different membrane-mimicking environments, *J. Mol. Model.*, 2005, **12**, 49–64.

204 J. M. Baldwin, The probable arrangement of the helices in G protein-coupled receptors., *EMBO J.*, 1993, **12**, 1693–703.

205 K. Bondensgaard, M. Ankersen, H. Thøgersen, B. S. Hansen, A. Birgitte S. Wulff and R. P. Bywater, Recognition of Privileged Structures by G-Protein Coupled Receptors, *J. Med. Chem.*, 2004, **47**, 888–899.

206 R. Adams, C. L. Worth, S. Guenther, M. Dunkel, R. Lehmann and R. Preissner, Binding sites in membrane proteins – Diversity, druggability and prospects, *Eur. J. Cell Biol.*, 2012, **91**, 326–339.

207 W. R. Pearson, Selecting the right similarity-scoring matrix, *Curr. Protoc. Bioinforma.*, 2013, **43**, 351–359.

208 T. Schmidt, A. Bergner and T. Schwede, Modelling three-dimensional protein structures for applications in drug design, *Drug Discov. Today*, 2014, **19**, 890–897.

209 T. Schwede, J. Kopp, N. Guex and M. C. Peitsch, SWISS-MODEL: An automated protein homology-modeling server., *Nucleic Acids Res.*, 2003, **31**, 3381–5.

210 C. Munk, V. Isberg, S. Mordalski, K. Harpsøe, K. Rataj, A. S. Hauser, P. Kolb, A. J. Bojarski, G. Vriend and D. E. Gloriam, GPCRdb: the G protein-coupled receptor database - an introduction., *Br. J. Pharmacol.*, 2016, **173**, 2195–207.

211 L. F. Kolakowski, GCRDb: a G-protein-coupled receptor database., *Receptors Channels*, 1994, **2**, 1–7.

212 Y. Zhang, M. E. DeVries and J. Skolnick, Structure Modeling of All Identified G Protein–Coupled Receptors in the Human Genome, *PLoS Comput. Biol.*, 2006, **2**, e13.

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

214 S. R. Krystek, S. R. Kimura and A. J. Tebben, Modeling and active site refinement for G protein-coupled receptors: application to the β-2 adrenergic receptor, *J. Comput. Aided. Mol. Des.*, 2006, **20**, 463–470.

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

216 M. Suwa, M. Sugihara and Y. Ono, Functional and Structural Overview of G-Protein-Coupled Receptors Comprehensively Obtained from Genome Sequences, *Pharmaceuticals*, 2011, **4**, 652–664.

217 M. Y. Niv, L. Skrabanek, M. Filizola and H. Weinstein, Modeling activated states of GPCRs: the rhodopsin template., *J. Comput. Aided. Mol. Des.*, 2006, **20**, 437–48.

218 P. A. Hargrave and J. H. McDowell, Rhodopsin and phototransduction: a model system for G protein-linked receptors., *FASEB J.*, 1992, **6**, 2323–31.

219 D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis, R. C. Stevens and B. K. Kobilka, GPCR Engineering Yields High-Resolution Structural Insights into 2-Adrenergic Receptor Function, *Science*, 2007, **318**, 1266–1273.

220 K. Lundstrom, Structural biology of G protein-coupled receptors, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3654–3657.

221 S. Bhattacharya, S. E. Hall, H. Li and N. Vaidehi, Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation., *Biophys. J.*, 2008, **94**, 2027–42.

222 D. M. Rosenbaum, S. G. F. Rasmussen and B. K. Kobilka, The structure and function of G-protein-coupled receptors., *Nature*, 2009, **459**, 356–63.

223 P. Chelikani, V. Hornak, M. Eilers, P. J. Reeves, S. O. Smith, U. L. Rajbhandary and H. Gobind Khorana, Role of group-conserved residues in the helical core of 2-adrenergic receptor, *PNAS April*, 2007, **24**, 7027–7032.

224 J. Standfuss, P. C. Edwards, A. D'Antona, M. Fransen, G. Xie, D. D. Oprian and

G. F. X. Schertler, The structural basis of agonist-induced activation in constitutively active rhodopsin, *Nature*, 2011, **471**, 656–660.

225 J. Popp, K. Pető and J. Nagy, *Agron. Sustain. Dev.*, 2013, 33, 243–255.

226 C. A. Hallmann, M. Sorg, E. Jongejans, H. Siepel, N. Hofland, H. Schwan, W. Stenmans, A. Müller, H. Sumser, T. Hörren, D. Goulson, H. De Kroon and H. de Kroon, More than 75 percent decline over 27 years in total flying insect biomass in protected areas, *PLoS One*, 2017, **12**, e0185809.

227 L. Saunders and R. Pezeshki, Glyphosate in Runoff Waters and in the Root-Zone: A Review, *Toxics*, 2015, **3**, 462–480.

228 A. H. C. Van Bruggen, M. M. He, K. Shin, V. Mai, K. C. Jeong, M. R. Finckh and J. G. Morris, *Sci. Total Environ.*, 2018, 616–617, 255–268.

229 G. Mugumbate, G. E. Jackson and D. Van Der Spoel, Open conformation of adipokinetic hormone receptor from the malaria mosquito facilitates hormone binding, *Peptides*, 2011, **32**, 553–559.

230 T. W. Schwartz, T. M. Frimurer, B. Holst, M. M. Rosenkilde and C. E. Elling, Molecular mechanism of 7tm Receptor activation—A global toggle switch model, *Annu. Rev. Pharmacol. Toxicol.*, 2006, **46**, 481–519.

231 G. Swaminath, X. Deupi, T. W. Lee, W. Zhu, F. S. Thian, T. S. Kobilka and B. Kobilka, Probing the β_2 Adrenoceptor Binding Site with Catechol Reveals Differences in Binding and Activation by Agonists and Partial Agonists, *J. Biol. Chem.*, 2005, **280**, 22165–22171.

232 C. G. Ravina, M. A. Seda, F. M. Pinto, M. Fernandez-Sanchez, O. C. Pintado and M.

L. Candenas, Characterization of tachykinin receptors in human sperm, *Fertil. Steril.*, 2007, **88**, S362.

233 E. Kellenberger, J.-Y. Springael, M. Parmentier, M. Hachet-Haas, J.-L. Galzi and D. Rognan, Identification of Nonpeptide CCR5 Receptor Agonists by Structure-based Virtual Screening, *J. Med. Chem.*, 2007, **50**, 1294–1303.

234 A. Singh and C. Haskell-Luevano, in *Peptide Chemistry and Drug Design*, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2015, pp. 75–112.

235 R. K. Mishra and J. Singh, Generation, Validation, and Utilization of a Three-Dimensional Pharmacophore Model for EP3 Antagonists, *J. Chem. Inf. Model.*, 2010, **50**, 1502–1509.

236 I. D. Pogozheva, B.-X. Chai, A. L. Lomize, T. M. Fong, D. H. Weinberg, R. P. Nargund, M. W. Mulholland, I. Gantz and H. I. Mosberg, Interactions of Human Melanocortin 4 Receptor with Nonpeptide and Peptide Agonists, *Biochemistry*, 2005, **44**, 11329–11341.

237 I. D. Pogozheva, A. L. Lomize and H. I. Mosberg, The transmembrane 7-alpha-bundle of rhodopsin: distance geometry calculations with hydrogen bonding constraints., *Biophys. J.*, 1997, **72**, 1963–85.

238 S. Raman, O. F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T. A. Ramelot, A. Eletsky, T. Szyperski, M. A. Kennedy, J. Prestegard, G. T. Montelione and D. Baker, NMR structure determination for larger proteins using backbone-only data, *Science*, 2010, **327**, 1014–1018.

239 U. Gether, Uncovering molecular mechanisms involved in activation of G protein-

coupled receptors, *Endocr. Rev.*, 2000, **21**, 90–113.

240 S. P. Skinner, B. T. Goult, R. H. Fogh, W. Boucher, T. J. Stevens, E. D. Laue, G. W. Vuister and IUCr, Structure calculation, refinement and validation using CcpNmr Analysis, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2015, **71**, 154–161.

241 K. Wuthrich, K. Wüthrich, K. Wuthrich, K. Wüthrich, K. Wuthrich and K. Wüthrich, Protein structure determination in solution by nuclear magnetic resonance spectroscopy, *Science*, 1989, **243**, 45–50.

242 D. Marion and K. Wüthrich, Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of1H-1H spin-spin coupling constants in proteins, *Biochem. Biophys. Res. Commun.*, 1983, **113**, 967–974.

243 A. P. Tonge, P. Murray-rust, W. A. Gibbons and L. K. McLachlan, Determination of the major solution conformation of tyrocidine A, using molecular mechanics energy minimization and NMR-derived distance and torsion angle constraints, *J. Comput. Chem.*, 1988, **9**, 522–538.

244 M. G. Hinds, H. Welsh, D. M. Brennand, J. Fisher, M. J. Glennie, N. G. J. Richards, L. Turner and A. Robinson, Synthesis, Conformational Properties, and Antibody Recognition of Peptides Containing beta-Turn Mimetics Based on alpha-Alkylproline Derivatives, *J. Med. Chem.*, 1991, **34**, 1777–1789.

245 G. C. K. (Gordon C. K. Roberts, *NMR of macromolecules : a practical approach*, IRL Press at Oxford University Press, 1993.

246 D. C. Fry, V. S. Madison, D. R. Bolin, D. N. Greeley, V. Toome and B. B. Wegrzynsk, Solution Structure of an Analogue of Vasoactive Intestinal Peptide As Determined by

Two-Dimensional NMR and Circular Dichroism Spectroscopies and Constrained Molecular Dynamics, *Biochemistry*, 1989, **28**, 2399–2409.

247 J. A. Barden and B. E. Kemp, NMR Solution Structure of Human Parathyroid Hormone(1–34), *Biochemistry*, 1993, **32**, 7126–7132.

248 K. Nagayama, A. Kumar, K. Wüthrich, R. E.-J. of Magnetic and U. 1980, Experimental techniques of two-dimensional correlated spectroscopy, *J. Magn. Reson.*, 1980, **40**, 321–334.

249 K. Nagayama, K. Wüthrich, P. Bachmann and R. R. Ernst, Two-dimensional J-resolved 1H n.m.r. spectroscopy for studies of biological macromolecules, *Biochem. Biophys. Res. Commun.*, 1977, **78**, 99–105.

250 S. Kosol, S. Contreras-Martos, C. Cedeño and P. Tompa, Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy, *Molecules*, 2013, **18**, 10802–10828.

251 M. S. Cheung, M. L. Maguire, T. J. Stevens and R. W. Broadhurst, DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure, *J. Magn. Reson.*, 2010, **202**, 223–233.

252 J. Jee and P. Güntert, Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment, *J. Struct. Funct. Genomics*, 2003, **4**, 179–189.

253 T. Herrmann, P. Güntert and K. Wüthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA., *J. Mol. Biol.*, 2002, **319**, 209–27.

254 P. Güntert, in *Protein NMR Techniques*, Humana Press, New Jersey, 2004, vol. 278, pp. 353–378.

255 K. J. (Keith J. Laidler, J. H. Meiser and B. C. Sanctuary, Physical chemistry, *J. Chem. Educ.*, 2003, **80**, 1060.

256 P. W. (Peter W. Atkins and J. De Paula, *Atkins' Physical chemistry*, 2013.

257 A. D. Kline, W. Braun and K. Wüthrich, Studies by 1H nuclear magnetic resonance and distance geometry of the solution conformation of the α-amylase inhibitor Tendamistat, *J. Mol. Biol.*, 1986, **189**, 377–382.

258 R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber and K. Wuthrich, NMR structure of the mouse prion protein domain PrP(121-231), *Nature*, 1996, **382**, 180–182.

259 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, 1986, vol. 32.

260 H. S. Haugen, H. Sophie, G. Fimland, J. Nissen-Meyer, P. E. Kristiansen and H. Sophie, The Three-Dimensional NMR-Structure of the Membrane-Permeabilizing Pediocin-Like Antimicrobial Peptide Curvacin A in Lipid Micelles, *Biochemistry*, 2005, **44**, 16149–16157.

261 M. V. Berjanskii and D. S. Wishart, Unraveling the meaning of chemical shifts in protein NMR, *Biochim. Biophys. Acta - Proteins Proteomics*, 2017, **1865**, 1564–1576.

262 D. S. Wishart, B. D. Sykes and F. M. Richards, The Chemical Shift Index : A Fast and Simple Method for the Assignment of Protein Secondary Structure Through NMR Spectroscopy, *Biochemistry*, 1992, **31**, 1647–1651.

263 D. S. Wishart and B. D. Sykes, [12] Chemical shifts as a tool for structure determination, *Methods Enzymol.*, 1994, **239**, 363–392.

264 D. S. S. Wishart, B. D. Sykes and F. M. Richards, Relationship between nuclear magnetic resonance chemical shift and protein secondary structure., *J. Mol. Biol.*, 1991, **222**, 311–33.

265 M. Berjanskii and D. S. Wishart, NMR: Prediction of protein flexibility, *Nat. Protoc.*, 2006, **1**, 683–688.

266 M. V Berjanskii and D. S. Wishart, Application of the random coil index to studying protein flexibility, *J Biomol NMR*, 2008, **40**, 3--38.

267 D. Wishart and B. Sykes, The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data, *J. Biomol. NMR*, 1994, **4**, 171–180.

268 A. G. Palmer, J. Cavanagh, P. E. Wright and M. Rance, Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy, *J. Magn. Reson.*, 1991, **93**, 151–170.

269 J. Cavanagh, W. Fairbrother, A. P. III and N. Skelton, *Protein NMR spectroscopy: principles*, 1995.

270 H. Friebolin and J. Becconsall, *Basic one-and two-dimensional NMR spectroscopy*, 1993.

271 H. Rattle, in *An NMR primer for life scientists*, Partnership Press, 1995, p. 124.

272 R. R. Ernst, G. Bodenhausen and A. Wokaun, Principles of Nuclear Magnetic

Resonance in One and Two Dimensions, *Magn. Reson. Imaging*, 1988, **42**, 75–76.

273 J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys.*, 1979, **71**, 4546–4553.

274 W. P. Aue, E. Bartholdi and R. R. Ernst, Two-dimensional spectroscopy. Application to nuclear magnetic resonance, *J. Chem. Phys.*, 1976, **64**, 2229–2246.

275 H. Kessler, C. Griesinger, R. Kerssebaum, K. Wagner and R. R. Ernst, Separation of Cross-Relaxation and J Cross-Peaks in 2D Rotating-Frame NMR Spectroscopy, *J. Am. Chem. Soc.*, 1987, **109**, 607–609.

276 R. E. Klevit, D. E. Wemmer and B. R. Reid, 1H NMR Studies on the Interaction between Distamycin a and a Symmetrical DNA Dodecamer, *Biochemistry*, 1986, **25**, 3296–3303.

277 D. E. Wemmer, N. V. Kumar, R. M. Metrione, M. Lazdunski, G. Drobny and N. R. Kallenbach, NMR Analysis and Sequence of Toxin II from the Sea Anemone Radianthus paumotensis, *Biochemistry*, 1986, **25**, 6842–6849.

278 P. Serrano, S. K. Dutta, A. Proudfoot, B. Mohanty, L. Susac, B. Martin, M. Geralt, L. Jaroszewski, A. Godzik, M. Elsliger, I. A. Wilson and K. Wüthrich, NMR in structural genomics to increase structural coverage of the protein universe., *FEBS J.*, 2016, **283**, 3870–3881.

279 I. R. Chandrashekar and S. M. Cowsik, Three-dimensional structure of the mammalian tachykinin peptide neurokinin A bound to lipid micelles., *Biophys. J.*, 2003, **85**, 4002–11.

280 W. Hu, L. T. Kakalis, L. Jiang, F. Jiang, X. Ye and A. Majumdar, 3D HCCH-COSY-TOCSY experiment for the assignment of ribose and amino acid side chains in 13C labeled RNA and protein., *J. Biomol. NMR*, 1998, **12**, 559–64.

281 H. S. Haugen, G. Fimland, J. Nissen-Meyer and P. E. Kristiansen, Three-Dimensional Structure in Lipid Micelles of the Pediocin-like Antimicrobial Peptide Curvacin A †, ‡, *Biochemistry*, 2005, **44**, 16149–16157.

282 S. Macura and R. R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy (Reprinted from Molecular Physics, vol 41, pg 95-117, 1980), *Mol. Phys.*, 2002, **100**, 135–147.

283 E. A. Mahrous and M. A. Farag, Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review, *J. Adv. Res.*, 2015, **6**, 3–15.

284 W. R. Croasmun and R. M. K. Carlson, *Two-dimensional NMR spectroscopy. Applications for chemists and biochemists*, 1987.

285 J. Cavanagh, *Protein NMR spectroscopy : principles and practice*, Academic Press, 2007.

286 P. Güntert, C. Mumenthaler and K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program Dyana, *J. Mol. Biol.*, 1997, **273**, 283–298.

287 J. P. Linge, M. Habeck, W. Rieping and M. Nilges, ARIA: automated NOE assignment and NMR structure calculation., *Bioinformatics*, 2003, **19**, 315–6.

288 T. Herrmann, P. Güntert and K. Wüthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, *J. Biomol. NMR*, 2002, **24**, 171–189.

289 B. Bardiaux, T. Malliavin and M. Nilges, *ARIA for Solution and Solid-State NMR*, 2012.

290 S. Nozinovic, B. Fürtig, H. R. A. Jonker, C. Richter and H. Schwalbe, High-resolution NMR structure of an RNA model system: The 14-mer cUUCGg tetraloop hairpin RNA, *Nucleic Acids Res.*, 2009, **38**, 683–694.

291 R. Born, M., and Oppenheimer, Zur Quantentheorie der Molekeln, *Ann. Phys.*, 1927, **389**, 457–484.

292 A. R. Leach, *moleculer modelling second edition*, Prentice Hall, 2001.

293 K. Zhu, M. R. Shirts, R. A. Friesner and M. P. Jacobson, Multiscale optimization of a truncated newton minimization algorithm and application to proteins and protein-ligand complexes, *J. Chem. Theory Comput.*, 2007, **3**, 640–648.

294 M. Abraham, B. Hess, D. van der Spoel and E. Lindahl, GROMACS User Manual version 5.0.7, *Www.Gromacs.Org*, 2015, **5.0.7**, 1–310.

295 D. Van Der Spoel, H. J. Vogel, H. J. C. Berendsen' and H. J. C. Berendsen, Molecular Dynamics Simulations of N-Terminal Peptides From a Nucleotide Binding Protein, *Proteins Struct. Funct. Genet.*, 1996, **24**, 450–466.

296 R. O. Dror, D. H. Arlow, D. W. Borhani, M. Ø. O. Ø. Jensen, S. Piana, D. E. Shaw, R. O. Dror, D. H. Arlow, S. Piana, D. E. Shaw, D. W. Borhani, M. Ø. O. Ø. Jensen, S. Piana and D. E. Shaw, Identification of two distinct inactive conformations of the 2-adrenergic receptor reconciles structural and biochemical observations, *Proc. Natl. Acad. Sci.*, 2009, **106**, 4689–4694.

297 A. Ivetac and M. S. P. P. Sansom, Molecular dynamics simulations and membrane

protein structure quality, *Eur. Biophys. J.*, 2008, **37**, 403–409.

298 A. R. Leach, in *Computers*, Prentice Hall, 2001, p. 744.

299 S. Kirkpatrick, ; C D Gelatt and ; M P Vecchi, Optimization by Simulated Annealing, *Science*, 1983, **220**, 671–680.

300 D. Bertsimas and J. Tsitsiklis, Simulated Annealing, *Stat. Sci.*, 1993, **8**, 10–15.

301 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, 26, 1701–1718.

302 J. D. Durrant and J. A. McCammon, Molecular dynamics simulations and drug discovery, *BMC Biol.*, 2011, **9**, 71.

303 R. W. W. Hockney, The potential calculation and some applications, *Methods in Comp. Phys.*, 1970, **13**, 57–60.

304 L. Verlet, Computer ‘experiments’ on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, *Phys. Rev.*, 1967, **159**, 98–103.

305 L. Verlet, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Phys. Rev.*, 1967, **159**, 98–103.

306 M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, P. Lindqvist and I. Vattulainen, Molecular dynamics simulations of lipid bilayers: Major artifacts due to truncating electrostatic interactions, *Biophys. J.*, 2003, **6**, 3636–45.

307 and the G. M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess and ، development team, *GROMACS Documentation*, Netherlands., 2018.

308 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1**–**2**, 19–25.

309 H. Bekker, E. J. Dijkstra, M. K. R. Renardus and H. J. C. Berendsen, An Efficient, Box Shape Independent Non-Bonded Force and Virial Algorithm for Molecular Dynamics, *Mol. Simul.*, 1995, **14**, 137–151.

310 W. Dzwinel, J. Kitowski and J. Mościński, "Checker Board" Periodic Boundary Conditions in Molecular Dynamics Codes, *Mol. Simul.*, 1991, **7**, 171–179.

311 D. van der Spoel, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, B. Hess, K. A. Feenstra, E. Lindahl, R. van Drunen and H. J. C. Berendsen, *GROMACS Reference Manual Contributions from*, 2005.

312 L. G. Mureddu, T. J. Ragan, E. J. Brooksbank and G. W. Vuister, CcpNmr AnalysisScreen, a new software programme with dedicated automated analysis tools for fragment-based drug discovery by NMR, *J. Biomol. NMR*, 2020, **74**, 565–577.

313 P. J. Kraulis, P. J. Domaille, S. L. Campbell-Burk, T. Van Aken, E. D. L. J, E. D. Laue, P. J. Domaille, S. L. Campbell-Burk and T. Van Aken, Solution Structure and Dynamics of Ras p21·GDP Determined by Heteronuclear Three- and Four-Dimensional NMR Spectroscopy, *Biochemistry*, 1994, **33**, 3515–3531.

314 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, The CCPN data model for NMR spectroscopy: Development of a software pipeline, *Proteins Struct. Funct. Genet.*, 2005, **59**, 687–696.

315 W. Rieping, B. Bardiaux, A. Bernard, T. E. Malliavin and M. Nilges, ARIA2: Automated NOE assignment and data integration in NMR structure calculation, *Bioinformatics*, 2007, **23**, 381–382.

316 V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 12–21.

317 E. Apol, R. Apostolov, H. J. C. Berendsen, A. Van Buuren, P. Bjelkmar, R. Van Drunen, A. Feenstra, S. Fritsch, G. Groenhof, C. Junghans, J. Hub, P. Kasson, C. Kutzner, B. Lambeth, P. Larsson, J. A. Lemkul, E. Marklund, P. P. Meulenhoff, T. Murtola, S. Páll, S. Pronk, R. Schulz, M. Shirts, A. Sijbers, P. Tieleman, C. Wennberg, M. W. M. Abraham, B. Hess, D. Van Der Spoel, E. Lindahl, V. Lindahl, M. Lundborg, E. Marklund, P. P. Meulenhoff, T. Murtola, S. Páll, S. Pronk, R. Schulz, M. Shirts, A. Sijbers, P. Tieleman, C. Wennberg, M. W. M. Abraham, B. Hess, D. Van Der Spoel and E. Lindahl, *GROMACS Reference Manual Contributions from*, 1991.

318 G. E. Jackson, A. N. Mabula, S. R. Stone, G. Gäde, K. E. Kövér, L. Szilágyi and D. van der Spoel, Solution conformations of an insect neuropeptide: Crustacean cardioactive peptide (CCAP), *Peptides*, 2009, **30**, 557–564.

319 D. P. Tieleman, D. van der Spoel, H. J. C. Berendsen, * D. P. Tieleman, ‡ and D. van der Spoel and H. J. C. Berendsen†, Molecular Dynamics Simulations of Dodecylphosphocholine Micelles at Three Different Aggregate Sizes: Micellar Structure and Chain Relaxation, *J. Phys. Chem. B*, 2000, **104**, 6380–6388.

320 L. Eidenschink, B. L. Kier, K. N. L. Huggins and N. H. Andersen, Very short peptides

with stable folds: Building on the interrelationship of Trp/Trp, Trp/cation, and Trp/backbone-amide interaction geometries, *Proteins Struct. Funct. Bioinforma.*, 2009, **75**, 308–322.

321 R. M. Fesinmeyer, F. M. Hudson and N. H. Andersen, Enhanced Hairpin Stability through Loop Design: The Case of the Protein G B1 Domain Hairpin, *J. Am. Chem. Soc.*, 2004, **126**, 7238–7243.

322 M.-L. L. Tremblay, A. W. Banks and J. K. Rainey, The predictive accuracy of secondary chemical shifts is more affected by protein secondary structure than solvent environment, *J. Biomol. NMR*, 2010, **46**, 257–270.

323 L. Szilágyi, Chemical shifts in proteins come of age, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1995, **27**, 325–442.

324 G. E. Jackson, R. Gamieldien, G. Mugumbate and G. Gäde, Structural studies of adipokinetic hormones in water and DPC micelle solution using NMR distance restrained molecular dynamics, *Peptides*, 2014, **53**, 270–277.

325 G. E. Jackson, E. Pavadai, G. Gäde, Z. Timol and N. H. Andersen, Interaction of the red pigment-concentrating hormone of the crustacean Daphnia pulex , with its cognate receptor, Dappu-RPCHR: A nuclear magnetic resonance and modeling study, *Int. J. Biol. Macromol.*, 2018, **106**, 969–978.

326 M. V. Berjanskii and D. S. Wishart, The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts, *Nucleic Acids Res.*, 2007, **35**, W531–W537.

327 P. F. J. Fuchs and A. J. P. Alix, High accuracy prediction of β-turns and their types using propensities and multiple alignments, *Proteins Struct. Funct. Bioinforma.*, 2005,

59, 828–839.

328 C. Giragossian, M. Pellegrini and D. F. Mierke, NMR studies of CCK-8/CCK1 complex support membrane-associated pathway for ligand-receptor interaction., *Can. J. Physiol. Pharmacol.*, 2002, **80**, 383–387.

329 R. Sankararamakrishnan, Recognition of GPCRs by Peptide Ligands and Membrane Compartments theory: Structural Studies of Endogenous Peptide Hormones in Membrane Environment, *Biosci. Rep.*, 2006, **26**, 131–158.

330 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes, *Nucleic Acids Res.*, 2018, **46**, W296–W303.

331 C. Y. Huang, V. Olieric, P. Ma, N. Howe, L. Vogeley, X. Liu, R. Warshamanage, T. Weinert, E. Panepucci, B. Kobilka, K. Diederichs, M. Wang and M. Caffrey, In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures, *Acta Crystallogr. Sect. D Struct. Biol.*, 2016, **72**, 93–112.

332 C. S. Reddy, K. Vijayasarathy, E. Srinivas, G. M. N. Sastry and G. M. N. Sastry, Homology modeling of membrane proteins: A critical assessment, *Comput. Biol. Chem.*, 2006, **30**, 120–126.

333 V. Sarramegna, F. Talmont, P. Demange and A. Milon, Heterologous expression of G-protein-coupled receptors: Comparison of expression systems from the standpoint of large-scale production and purification, *Cell. Mol. Life Sci.*, 2003, **60**, 1529–1546.

334 A. Tramontano, Homology Modeling with Low Sequence Identity, *Methods*, 1998, **14**, 293–300.

335 K. Arnold, L. Bordoli, J. Kopp and T. Schwede, The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics*, 2006, **22**, 195–201.

336 D. Baker and A. Sali, Protein Structure Prediction and Structural Genomics, *Science*, 2001, **294**, 93–96.

337 P. Benkert, M. Biasini and T. Schwede, Toward the estimation of the absolute quality of individual protein structure models, *Bioinformatics*, 2011, **27**, 343–350.

338 S. Bienert, A. Waterhouse, T. A. P. De Beer, G. Tauriello, G. Studer, L. Bordoli and T. Schwede, The SWISS-MODEL Repository-new features and functionality, *Nucleic Acids Res.*, 2017, **45**, D313–D319.

339 D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.*, 1999, **292**, 195–202.

340 D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson and D. T. Jones, Scalable web services for the PSIPRED Protein Analysis Workbench., *Nucleic Acids Res.*, 2013, **41**, W349-57.

341 L. J. McGuffin, K. Bryson and D. T. Jones, The PSIPRED protein structure prediction server., *Bioinformatics*, 2000, **16**, 404–5.

342 H. B. Kazemian, K. White and D. Palmer-Brown, Applications of evolutionary SVM to prediction of membrane alpha-helices, *Expert Syst. Appl.*, 2013, **40**, 3412–3420.

343 R. Ter Horst and J. S. Lolkema, Rapid screening of membrane topology of secondary transport proteins, *Biochim. Biophys. Acta - Biomembr.*, 2010, **178**, 672–680.

344 L. D. Altschul SF, Gish W, Miller W, Myers EW, Basic local alignment search tool, *J. Mol. Biol.*, 1990, **215**, 403–410.

345 H. M. Berman and IUCr, The Protein Data Bank: A historical perspective, *Acta Crystallogr. Sect. A Found. Crystallogr.*, 2008, **64**, 88–95.

346 H. M. Berman, G. J. Kleywegt, H. Nakamura and J. L. Markley, The Protein Data Bank archive as an open data resource, *J. Comput. Aided. Mol. Des.*, 2014, **28**, 1009–1014.

347 J. D. Thomopson, D. G. Higgins and T. J. Gibson, Clustal W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res*, 1994, **22**, 6473–4680.

348 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics*, 2007, **23**, 2947–2948.

349 N. Guex, M. C. Peitsch and T. Schwede, Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective, *Electrophoresis*, 2009, **30**, S162–S173.

350 W. L. DeLano, The PyMOL Molecular Graphics System, Version 1.1, *CCP4 Newsl. protein Crystallogr.*, 2002, **40**, 82–92.

351 U. Omasits, C. H. Ahrens, S. Müller and B. Wollscheid, Protter: Interactive protein feature visualization and integration with experimental proteomic data, *Bioinformatics*, 2014, **30**, 884–886.

352 D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics*, 2007, **23**, 538–544.

353 D. T. Jones, Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics*, 2007, **23**, 538–544.

354 N. Guex and M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis*, 1997, **18**, 2714–2723.

355 M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Šali, Comparative Protein Structure Modeling of Genes and Genomes, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 291–325.

356 S. Singh, B. K. Malik and D. K. Sharma, Targeting HIV-1 through molecular modeling and docking studies of CXCR4: Leads for therapeutic development, *Chem. Biol. Drug Des.*, 2007, **69**, 191–203.

357 C. Colovos and T. O. Yeates, Verification of protein structures: Patterns of nonbonded atomic interactions, *Protein Sci.*, 1993, **2**, 1511–1519.

358 R. A. Laskowski, J. A. C. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton, AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR*, 1996, **8**, 477–486.

359 C. E. Elling and T. W. Schwartz, Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering., *EMBO J.*,

2018, **15**, 6213–9.

360 Z. T. Farahbakhsh, K. D. Ridge, H. G. Khorana and W. L. Hubbell, Mapping light-dependent structural changes in the cytoplasmic loop connecting helices C and D in rhodopsin: a site-directed spin labeling study., *Biochemistry*, 1995, **34**, 8812–9.

361 B. Zare, A. Madadkar-Sobhani, S. Dastmalchi and M. Mahmoudian, Prediction of the Human EP1 Receptor Binding Site by Homology Modeling and Molecular Dynamics Simulation., *Sci. Pharm.*, 2011, **79**, 793–816.

362 R. J. Trabanino, S. E. Hall, N. Vaidehi, W. B. Floriano, V. W. T. Kam and W. A. Goddard, First Principles Predictions of the Structure and Function of G-Protein-Coupled Receptors: Validation for Bovine Rhodopsin, *Biophys. J.*, 2004, **86**, 1904–1921.

363 C. E. Elling, S. M. Nielsen and T. W. Schwartz, Conversion of antagonist-binding site to metal-ion site in the tachykinin NK-1 receptor, *Nature*, 1995, **374**, 74–77.

364 J. A. Ballesteros and H. Weinstein, Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors, *Methods Neurosci.*, , DOI:10.1016/S1043-9471(05)80049-7.

365 V. Isberg, C. de Graaf, A. Bortolato, V. Cherezov, V. Katritch, F. H. Marshall, S. Mordalski, J.-P. Pin, R. C. Stevens, G. Vriend and D. E. Gloriam, Generic GPCR residue numbers – aligning topology maps while minding the gaps, *Trends Pharmacol. Sci.*, 2015, **36**, 22–31.

366 M. Rasmussen, M. Leander, S. Ons and R. Nichols, Conserved molecular switch

interactions in modeled cardioactive RF-NH2 peptide receptors: Ligand binding and activation, *Peptides*, 2015, **71**, 259–267.

367 R. O. Dror, D. H. Arlow, D. W. Borhani, M. Ø. Jensen, S. Piana and D. E. Shaw, Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations., *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4689–94.

368 X. Deupi, M. Olivella, C. Govaerts, J. A. Ballesteros, M. Campillo and L. Pardo, Ser and Thr Residues Modulate the Conformation of Pro-Kinked Transmembrane α-Helices, *Biophys. J.*, 2004, **86**, 105–115.

369 D. Latek, P. Pasznik, T. Carlomagno and S. Filipek, Towards Improved Quality of GPCR Models by Usage of Multiple Templates and Profile-Profile Comparison, *PLoS One*, 2013, **8**, e56742.

370 M. Weber, L. Tome, D. Otzen and D. Schneider, A Ser residue influences the structure and stability of a Pro-kinked transmembrane helix dimer, *Biochim. Biophys. Acta - Biomembr.*, 2012, **1818**, 2103–2107.

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

372 J. Greaves and L. H. Chamberlain, Palmitoylation-dependent protein sorting, *J. Cell Biol.*, 2007, **176**, 249–254.

373 T. Chum, D. Glatzová, Z. Kvíčalová, J. Malínský, T. Brdička and M. Cebecauer, The role of palmitoylation and transmembrane domain in sorting of transmembrane adaptor

proteins, *J. Cell Sci.*, 2016, **129**, 95–107.

374 M. Hundt, Y. Harada, L. De Giorgio, N. Tanimura, W. Zhang and A. Altman, Palmitoylation-Dependent Plasma Membrane Transport but Lipid Raft-Independent Signaling by Linker for Activation of T Cells, *J. Immunol.*, 2009, **183**, 1685–1694.

375 S. Yuan, S. Filipek, K. Palczewski and H. Vogel, Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway, *Nat. Commun.*, 2014, **9**, 4733.

376 P. Chelikani, V. Hornak, M. Eilers, P. J. Reeves, S. O. Smith, U. L. Rajbhandary, H. Gobind Khorana, H. G. Khorana, V. Hornak, P. J. Reeves, S. O. Smith, U. L. Rajbhandary, H. G. Khorana, M. Eilers, P. J. Reeves, S. O. Smith, U. L. Rajbhandary and H. G. Khorana, Role of group-conserved residues in the helical core of beta2-adrenergic receptor, *Proc. Natl. Acad. Sci.*, 2007, **24**, 7027–7032.

377 T. Nugent and D. T. Jones, Transmembrane protein topology prediction using support vector machines, *BMC Bioinformatics*, 2009, **10**, 159.

378 L. J. McGuffin and D. T. Jones, Improvement of the GenTHREADER method for genomic fold recognition., *Bioinformatics*, 2003, **19**, 874–81.

379 D. A. Dougherty, Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp, *Science.*, 1996, **271**, 163–168.

380 J. P. Gallivan and D. A. Dougherty, Cation-pi interactions in structural biology., *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9459–9464.

381 D. S. Yakovlev and E. S. Boek, Molecular Dynamics Simulations of Mixed Cationic/Anionic Wormlike Micelles, *Langmuir*, 2007, **23**, 6588–6597.

382 J. H. Perlman, A. O. Colson, W. Wang, K. Bence, R. Osman and M. C. Gershengorn, Interactions between conserved residues in transmembrane helices 1, 2, and 7 of the thyrotropin-releasing hormone receptor, *J. Biol. Chem.*, 1997, **272**, 11937–42.

383 I. Halperin, B. Ma, H. Wolfson and R. Nussinov, Principles of docking: An overview of search algorithms and a guide to scoring functions, *Proteins Struct. Funct. Genet.*, 2002, **47**, 409–443.

384 S. Y. Huang and X. Zou, Advances and challenges in Protein-ligand docking, *Int. J. Mol. Sci.*, 2010, **11**, 3016–3034.

385 S. Lindert, I. Maslennikov, E. J. C. Chiu, L. C. Pierce, J. A. McCammon and S. Choe, Drug screening strategy for human membrane proteins: from NMR protein backbone structure to in silica- and NMR-screened hits., *Biochem. Biophys. Res. Commun.*, 2014, **445**, 724–33.

386 S. Kalyaanamoorthy and Y.-P. P. P. P. Chen, Structure-based drug design to augment hit discovery, *Drug Discov. Today*, 2011, **16**, 831–839.

387 P. Śledź and A. Caflisch, Protein structure-based drug design: from docking to molecular dynamics, *Curr. Opin. Struct. Biol.*, 2018, **48**, 93–102.

388 D. W. Borhani and D. E. Shaw, The future of molecular dynamics simulations in drug discovery, *J. Comput. Aided. Mol. Des.*, 2012, **26**, 15–26.

389 X. Liu, D. Shi, S. Zhou, H. Liu, H. Liu and X. Yao, Molecular dynamics simulations and novel drug discovery, *Expert Opin. Drug Discov.*, 2018, **13**, 23–37.

390 D. T. Moustakas, P. T. Lang, S. Pegg, E. Pettersen, I. D. Kuntz, N. Brooijmans and R. C. Rizzo, Development and validation of a modular, extensible docking program:

DOCK 5, *J. Comput. Aided. Mol. Des.*, 2006, **20**, 601–619.

391 B. Nguyen, Ngoc Thanh Trawiński, R. Kosala and B. Trawiński, Intelligent Information and Database Systems12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings, Part II, *Intell. Inf. database Syst.*, 2015, **8398**, 590.

392 D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, *Nat. Rev. Drug Discov.*, 2004, **3**, 935–949.

393 N. Okimoto, N. Futatsugi, H. Fuji, A. Suenaga, G. Morimoto, R. Yanai, Y. Ohno, T. Narumi and M. Taiji, High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations, *PLoS Comput. Biol.*, 2009, **5**, 100052–100053.

394 H. A. Gabb, R. M. Jackson and M. J. E. Sternberg, Modelling protein docking using shape complementarity, electrostatics and biochemical information 1 1Edited by J. Thornton, *J. Mol. Biol.*, 1997, **272**, 106–120.

395 D. Ghersi and R. Sanchez, Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites, *Proteins Struct. Funct. Bioinforma.*, 2009, **74**, 417–424.

396 C. Hetényi and D. van der Spoel, Efficient docking of peptides to proteins without prior knowledge of the binding site, *Protein Sci.*, 2002, **11**, 1729–1737.

397 F. Ercolessi, *A molecular dynamics primer*, 1997.

398 T. Beuming, W. Sherman, B. T. and S. W., Current assessment of docking into GPCR

crystal structures and homology models: Successes, challenges, and guidelines, *J. Chem. Inf. Model.*, 2012, **52**, 3263–3277.

399 B. D. Bursulaya, M. Totrov, R. Abagyan and C. L. Brooks, Comparative study of several algorithms for flexible ligand docking., *J. Comput. Aided. Mol. Des.*, 2003, **17**, 755–63.

400 D. Van Der Spoel, C. Hetényi, D. Van Der Spoel, D. Van Der Spoel and C. Hetényi, Blind docking of drug-sized compounds to proteins with up to a thousand residues ´, *FEBS Lett.*, 2006, **580**, 1447–1450.

401 S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell and A. J. Olson, Virtual screening with AutoDock: theory and practice, *Expert Opin. Drug Discov.*, 2010, **5**, 597–607.

402 P. A. Ravindranath, S. Forli, D. S. Goodsell, A. J. Olson and M. F. Sanner, AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility, *PLoS Comput. Biol.*, 2015, **11**, e100458.

403 V. Tanchuk, V. Tanin, A. Vovk and G. Poda, A New Scoring Function for Molecular Docking Based on AutoDock and AutoDock Vina, *Curr. Drug Discov. Technol.*, 2015, **12**, 170–178.

404 G. P. Brady and P. F. Stouten, Fast prediction and visualization of protein binding pockets with PASS., *J. Comput. Aided. Mol. Des.*, 2000, **14**, 383–401.

405 C. Bissantz, P. Bernard, M. Hibert and D. Rognan, Protein-based virtual screening of chemical databases. II. Are homology models of g-protein coupled receptors suitable targets?, *Proteins Struct. Funct. Bioinforma.*, 2002, **50**, 5–25.

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

407 R. Huey and G. Morris, *Using AutoDock 4 with AutoDockTools: A Tutorial*, 2008.

408 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function, *J. Comput. Chem.*, 1998, **19**, 1639–1662.

409 J. S. Taylor and R. M. Burnett, DARWIN: a program for docking flexible molecules., *Proteins*, 2000, **41**, 173–91.

410 G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen, *J. Mol. Biol.*, 1997, **267**, 727–748.

411 B. Isin, G. Estiu, O. Wiest and Z. N. Oltvai, Identifying Ligand Binding Conformations of the β2-Adrenergic Receptor by Using Its Agonists as Computational Probes, *PLoS One*, 2012, **7**, e50186.

412 A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler and M. Madan Babu, Molecular signatures of G-protein-coupled receptors, *Nature*, 2013, **494**, 185–194.

413 K. Wichapong, A. Rohe, C. Platzer, I. Slynko, F. Erdmann, M. Schmidt and W. Sippl, Application of docking and QM/MM-GBSA rescoring to screen for novel Myt1 kinase inhibitors, *J. Chem. Inf. Model.*, 2014, **54**, 881–893.

414 F. Chen, H. Liu, H. Sun, P. Pan, Y. Li, D. Li and T. Hou, Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein

binding free energies and re-rank binding poses generated by protein-protein docking, *Phys. Chem. Chem. Phys.*, 2016, **18**, 22129–22139.

415 T. Hou, J. Wang, Y. Li and W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations., *J. Chem. Inf. Model.*, 2011, **51**, 69–82.

416 S. Genheden and U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert Opin. Drug Discov.*, 2015, **10**, 449–461.

417 F. Godschalk, S. Genheden, P. Söderhjelm and U. Ryde, Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations, *Phys. Chem. Chem. Phys.*, 2013, **15**, 7731–7739.

418 G. Rastelli, A. Del Rio, G. Degliesposti and M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, *J. Comput. Chem.*, 2010, **31**, 797–810.

419 B. Albert, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology Of The Cell 4th Ed*, 1989.

420 C. Mulakala and V. N. Viswanadhan, Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies?, *J. Mol. Graph. Model.*, 2013, **46**, 41–51.

421 B. J. Mcconkey, V. Sobolev and M. Edelman, The performance of current methods in ligand-protein docking, *Asp. Biotechnol. Curr. Sci.*, 2002, **83**, 845–856.

422 C. Bissantz, A. Logean and D. Rognan, High-Throughput Modeling of Human G-Protein Coupled Receptors: Amino Acid Sequence Alignment, Three-Dimensional

Model Building, and Receptor Library Screening, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1162–1176.

423 O. Dror, D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. J. Wolfson, Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications, *J. Chem. Inf. Model.*, 2009, **49**, 2333–2343.

424 R. J. Law, C. Capener, M. Baaden, P. J. Bond, J. Campbell, G. Patargias, Y. Arinaminpathy and M. S. P. Sansom, in *Journal of Molecular Graphics and Modelling*, 2005, vol. 24, pp. 157–165.

425 C. Anezo, A. H. De Vries, H. D. Holtje, D. P. Tieleman and S. J. Marrink, Methodological issues in lipid bilayer simulations, *J. Phys. Chem. B*, 2003, **107**, 9424–9433.

426 P. Spijker, N. Vaidehi, P. L. Freddolino, P. A. J. Hilbers and W. A. Goddard, Dynamic behavior of fully solvated beta2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist, *Proc. Natl. Acad. Sci.*, 2006, **103**, 4882–4887.

427 F. Manetti, C. Tintori, M. Armand-Ugón, I. Clotet-Codina, S. Massa, R. Ragno, J. A. Esté and M. Botta, 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.*, 2006, **46**, 1344–1351.

428 R. B. and A. J. O. Garrett M. Morris, David S. Goodsell, Michael E. Pique, William "Lindy" Lindstrom, Ruth Huey, Stefano Forli, William E. Hart, Scott Halliday, *User guide autodock version 4.2*, 2012.

429 C. R. Corbeil, C. I. Williams and P. Labute, Variability in docking success rates due to

dataset preparation, *J. Comput. Aided. Mol. Des.*, 2012, **26**, 775–786.

430 T. A. Halgren, Identifying and Characterizing Binding Sites and Assessing Druggability, *J. Chem. Inf. Model.*, 2009, **49**, 377–389.

431 M. Kovács, J. Tóth, C. Hetényi, A. Málnási-Csizmadia, J. R. Sellers and J. R. Seller, Mechanism of Blebbistatin Inhibition of Myosin II, *J. Biol. Chem.*, 2004, **279**, 35557–35563.

432 Z. Bikádi, E. Hazai, F. Zsila and S. F. Lockwood, Molecular modeling of non-covalent binding of homochiral (3S,3′S)-astaxanthin to matrix metalloproteinase-13 (MMP-13), *Bioorg. Med. Chem.*, 2006, **14**, 5451–5458.

433 E. Hazai, Z. Bikádi, F. Zsila and S. F. Lockwood, Molecular modeling of the non-covalent binding of the dietary tomato carotenoids lycopene and lycophyll, and selected oxidative metabolites with 5-lipoxygenase, *Bioorganic Med. Chem.*, 2006, **14**, 6859–6867.

434 B. Iorga, D. Herlem, E. Barré and C. Guillou, Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the “blind docking” approach, *J. Mol. Model.*, 2006, **12**, 366–372.

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

436 M. M. Nair, G. E. Jackson and G. Gäde, Conformational study of insect adipokinetic hormones using NMR constrained molecular dynamics., *J. Comput. Aided. Mol. Des.*, 2001, **15**, 259–270.

437 D. Van Der Spoel and C. Hetényi, Blind docking of drug-sized compounds to proteins with up to a thousand residues, *FEBS Lett.*, 2006, **580**, 1447–1450.

438 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy., *J. Med. Chem.*, 2004, **47**, 1739–1749.

439 G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments, *J. Comput. Aided. Mol. Des.*, 2013, **27**, 221–231.

440 S. S. L. Priya, R. P. Devi and A. Madeswaran, In Silico docking studies of RP2 (X-linked retinitis pigmentosa) protein using anthocyanins as potential inhibitors, *Bangladesh J. Pharmacol.*, 2013, **8**, 292–299.

441 M. Berthold, U. Kahl, A. Jureus, K. Kask, G. Nordvall, U. Langel and T. Bartfai, Mutagenesis and Ligand Modification Studies on Galanin Binding to its GTP-Binding-Protein-Coupled Receptor GalR1, *Eur. J. Biochem.*, 1997, **249**, 601–606.

442 S. C. Sealfon, H. Weinstein and R. P. Millar, Molecular Mechanisms of Ligand Interaction with the Gonadotropin-Releasing Hormone Receptor, *Endocr. Rev.*, 1997, **18**, 180–205.

443 R. A. Kumpf and D. A. Dougherty, A mechanism for ion selectivity in potassium channels: Computational studies of cation-π interactions, *Science.*, 1993, **261**, 1708–1710.

444 M. Schneider, S. Wolf, J. Schlitter and K. Gerwert, The structure of active opsin as a

basis for identification of GPCR agonists by dynamic homology modelling and virtual screening assays, *FEBS Lett.*, 2011, **585**, 3587–3592.

445 J. A. R. Dalton and R. M. Jackson, Homology-modelling protein-ligand interactions: Allowing for ligand-induced conformational change, *J. Mol. Biol.*, 2010, **399**, 645–661.

446 X. Yao, C. Parnot, X. Deupi, V. R. P. P. Ratnala, G. Swaminath, D. Farrens and B. Kobilka, Coupling ligand structure to specific conformational switches in the β2-adrenoceptor, *Nat. Chem. Biol.*, 2006, **2**, 417–422.

447 J. P. Vilardaga, M. Bünemann, C. Krasell, M. Castro and M. J. Lohse, Measurement of the millisecond activation switch of G protein-coupled receptors in living cells, *Nat. Biotechnol.*, 2003, **21**, 807–812.

448 M. J. Lohse, C. Hoffmann, V. O. Nikolaev, J. P. Vilardaga and M. BÜnemann, in *Advances in Protein Chemistry*, 2007, vol. 74, pp. 167–188.

449 A. R. Curran and D. M. Engelman, *Curr. Opin. Struct. Biol.*, 2003.

450 H. G. Marco and G. Gäde, Five Neuropeptide Ligands Meet One Receptor: How Does This Tally? A Structure-Activity Relationship Study Using Adipokinetic Bioassays With the Sphingid Moth, Hippotion eson., *Front. Endocrinol. (Lausanne).*, 2019, **10**, 231.

451 H. G. Marco and G. Gäde, Structure and function of adipokinetic hormones of the large white butterfly Pieris brassicae, *Physiol. Entomol.*, 2017, **42**, 103–112.

452 O. Cusinato, A. F. Drake, G. Gäde and G. J. Goldsworthy, The molecular conformations of representative arthropod adipokinetic peptides determined by circular

dichroism spectroscopy, *Insect Biochem. Mol. Biol.*, 1998, **28**, 43–50.

453 I. Z. Z. Zubrzycki and G. Gade, Conformational Study on an Insect Neuropeptide of the AKH/RPCH-Family by Combined ^{1}H-NMR Spectroscopy and Molecular Mechanics, *Biochem Biophys Res Commun*, 1994, **198**, 228–235.

454 A. Velentza, S. Spiliou, C. P. Poulos and G. J. Goldsworthy, Synthesis and biological activity of adipokinetic hormone analogues with modifications in the 4-8 region, *Peptides*, 2000, **21**, 631–637.

455 R. Ziegler, A. S. Cushing, P. Walpole, R. D. Jasensky and H. Morimoto, Analogs of Manduca adipokinetic hormone tested in a bioassay and in a receptor-binding assay, *Peptides*, 1998, **19**, 481–600.

456 M. J. Lee, O. Cusinato, R. Luswata, C. H. Wheeler and G. J. Goldsworthy, N-terminal modifications to AKH-I from Locusta migratoria: assessment of biological potencies in vivo and in vitro, *Regul. Pept.*, 1997, **69**, 69–76.

457 M. J. Lee, G. J. Goldsworthy, C. P. Poulos and A. Velentza, Synthesis and biological activity of adipokinetic hormone analogues modified at the C-terminus, *Peptides*, 1996, **17**, 1285–90.

458 G. Gäde and T. K. Hayes, Structure-activity relationships for Periplaneta americana hypertrehalosemic hormone I: The importance of side chains and termini, *Peptides*, 1995, **16**, 1173–1180.

459 G. GÄDE, Structure–function studies on hypertrehalosaemic and adipokinetic hormones: activity of naturally occurring analogues and some N- and C-terminal modified analogues, *Physiol. Entomol.*, 1990, **15**, 299–316.

460 D. W. Sammond, Z. M. Eletr, C. Purbeck, R. J. Kimple, D. P. Siderovski and B. Kuhlman, Structure-based Protocol for Identifying Mutations that Enhance Protein-Protein Binding Affinities, *J. Mol. Biol.*, 2007, **371**, 1392–1404.

461 G. J. Goldsworthy, M. J. Lee, R. Luswata, A. F. Drake and D. Hyde, Structures, assays and receptors for locust adipokinetic hormones, *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.*, 1997, **117**, 483–496.

462 T. K. Hayes and L. L. Keeley, Structure-activity relationships on hyperglycemia by representatives of the adipokinetic/hyperglycemic hormone family in Blaberus discoidalis cockroaches, *J. Comp. Physiol. B*, 1990, **160**, 187–194.

463 S. Scheiner, T. Kar and J. Pattanayak, Comparison of various types of hydrogen bonds involving aromatic amino acids, *J. Am. Chem. Soc.*, 2002, **124**, 13257–13264.

464 J. C. Biro, Amino acid size, charge, hydropathy indices and matrices for protein structure analysis, *Theor. Biol. Med. Model.*, 2006, **3**, 15.

465 E. T. Kaiser and F. J. Kézdy, Amphiphilic secondary structure: Design of peptide hormones, *Science.*, 1984, **223**, 249–255.

466 E. Clynen, A. Reumer, G. Baggerman, I. Mertens and L. Schoofs, Neuropeptide Biology in Drosophila, *Adv. Exp. Med. Biol.*, 2010, **692**, 192–210.

467 G. Gäde and G. J. Goldsworthy, Insect peptide hormones: a selective review of their physiology and potential application for pest control, *Pest Manag. Sci.*, 2003, **59**, 1063–1075.

468 T. Zdobinsky, Insect neuropeptides: Structures, chemical modifications and potential for insect control, *Bioorg. Med. Chem.*, 2009, **17**, 4071–4084.

469 R. Imes, *Beginner's guide to entomology*, Chancellor Press, 2000.

470 M. Chinery, *Insects of Britain & Northern Europe*, HarperCollins, 1993.

471 J. H. Diaz, Myiasis and Tungiasis, *Mand. Douglas, Bennett's Princ. Pract. Infect. Dis.*, 2015, 3255-3259.e1.

472 S. Mazid, A review on the use of biopesticides in insect pest management, *Int. J. Sci. Adv. Technol.*, 2011, **1**, 169–178.

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

474 D. K. Vassilatis, J. G. Hohmann, H. Zeng, F. Li, J. E. Ranchalis, M. T. Mortrud, A. Brown, S. S. Rodriguez, J. R. Weller, A. C. Wright, J. E. Bergmann and G. A. Gaitanaris, The G protein-coupled receptor repertoires of human and mouse, *Proc. Natl. Acad. Sci.*, 2003, **100**, 4903–4908.

475 A. J. Kooistra, S. Kuhne, I. J. P. de Esch, R. Leurs and C. de Graaf, A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design, *Br. J. Pharmacol.*, 2013, **170**, 101–126.

476 S. Wolf, M. Bö Ckmann, U. Hö Weler C, J. Rgen Schlitter and K. Gerwert, Simulations of a G protein-coupled receptor homology model predict dynamic features and a ligand binding site, *FEBS Lett.*, 2008, **582**, 3335–3342.

477 P. H. Reggio, Computational methods in drug design: modeling G protein-coupled receptor monomers, dimers, and oligomers., *AAPS J.*, 2006, **8**, E322-36.

478 T. L. Senkow, D. R. Woldring, M. Kruziki, B. J. Hackel and J. N. Sachs, Evaluation of High Affinity Scaffolds from High-Throughput Discovery and Targeting DR5 and TNFR1, *Biophys. J.*, 2017, **112**, 530a.

Appendix A

Target templates for the AKHRs

Trg-Tpl	Models	Name	Coverage	Range	Seq Id	Oligo-state	Description	Ligands	Method	Resolution	Seq Similarity	Found by	GMQE	QSQE	Score	Other	
		6do1.1.A	0.28	47 - 173	40.50	hetero-1-2-mer	Type-1 angiotensin II receptor,Soluble cytochrome b562 BRIL fusion protein	2 x OLC, 1 x NAG, 1 x SAR-ARG-VAL-TYR-ILE-HIS-PRO-ILE	X-ray	2.90	0.41	BLAST	0.18	NA	0.48	e_value=7.93633e-13, bit_score=70.8626, score=172	Build Model (Monomer)
		5xjm.1.A	0.31	39 - 176	31.11	hetero-1-1-1-mer	Type-2 angiotensin II receptor,Soluble cytochrome b562,Type-2 angiotensin II receptor	1 x SAR-ARG-VAL-TYR-ILE-HIS-PRO-ILE	X-ray	3.20	0.38	BLAST	0.18	NA	0.51	e_value=7.01371e-06, bit_score=48.9062, score=115	Build Model (Monomer)
		5ung.1.A	0.61	39 - 321	31.18	monomer	Chimera protein of Type-2 angiotensin II receptor and Soluble cytochrome b562	1 x OLA, 1 x OLC, 1 x 8ES	X-ray	2.80	0.37	BLAST	0.45	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Monomer)
		6c1q.1.A	0.62	44 - 344	31.34	monomer	Soluble cytochrome b562, C5a anaphylatoxin chemotactic receptor 1 chimera	1 x 9P2, 1 x ACE-PHE-ORN-PRO-ZAL-TRP-ARG	X-ray	2.90	0.37	BLAST	0.43	NA	1.00	e_value=1.58479e-12, bit_score=69.707, score=169	Build Model (Monomer)
		5unf.1.A	0.61	39 - 321	31.18	homo-dimer	Chimera protein of Type-2 angiotensin II receptor and Soluble cytochrome b562	2 x 8ES	X-ray	2.80	0.37	BLAST	0.43	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Homomer) Build Model (Monomer)
		6dde.1.E	0.60	46 - 322	28.91	hetero-1-1-1-1-mer	Mu-type opioid receptor	1 x TYR-DAL-GLY-MEA-ETA	EM	0.00	0.37	BLAST	0.42	NA	0.95	e_value=2.85356e-12, bit_score=68.5514, score=166	Build Model (Monomer)
		5c1m.1.A	0.60	46 - 322	29.30	hetero-oligomer	Mu-type opioid receptor	1 x 4VO, 2 x OLC, 1 x CLR	X-ray	2.10	0.37	BLAST	0.42	NA	0.95	e_value=8.12226e-12, bit_score=66.6254, score=161	Build Model (Monomer)
		5unh.1.A	0.61	39 - 321	31.18	homo-dimer	Soluble cytochrome b562,Type-2 angiotensin II receptor	2 x 8EM	X-ray	2.90	0.37	BLAST	0.41	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Homomer) Build Model (Monomer)
		6meo.1.C	0.59	43 - 322	28.35	hetero-1-1-1-mer	C-C chemokine receptor type 5	2 x MAN, 17 x NAG, 1 x A2G	EM	3.90	0.37	BLAST	0.41	NA	0.95	e_value=2.22682e-12, bit_score=68.5514, score=166	Build Model (Monomer)
		2x72.1.A	0.60	52 - 329	24.32	hetero-oligomer	RHODOPSIN	2 x NAG-NAG-BMA-MAN-MAN, 2 x RET, 2 x LPP, 2 x PEF, 4 x PLM, 2 x BOG	X-ray	3.00	0.35	BLAST	0.35	NA	0.94	e_value=4.82238e-13, bit_score=70.8626, score=172	Build Model (Monomer)
		5lwe.1.A	0.59	48 - 329	27.95	monomer	C-C chemokine receptor type 9	15 x OLA, 1 x 79K, 1 x CLR	X-ray	2.80	0.35	BLAST	0.35	NA	0.93	e_value=9.54245e-10, bit_score=60.8474, score=146	Build Model (Monomer)
		5lwe.2.A	0.59	48 - 329	27.95	monomer	C-C chemokine receptor type 9	13 x OLA, 1 x 79K	X-ray	2.80	0.35	BLAST	0.35	NA	0.93	e_value=9.54245e-10, bit_score=60.8474, score=146	Build Model (Monomer)
		5f8u.1.A	0.57	52 - 321	29.67	monomer	Beta-1 adrenergic receptor		X-ray	3.35	0.35	BLAST	0.35	NA	0.90	e_value=3.99876e-16, bit_score=79.7221, score=195	Build Model (Monomer)
		3c9m.1.A	0.60	52 - 329	24.32	monomer	Rhodopsin	1 x NAG, 1 x RET	X-ray	3.40	0.35	BLAST	0.34	NA	0.94	e_value=7.27779e-13, bit_score=70.4774, score=171	Build Model (Monomer)
		4a4m.1.A	0.60	52 - 329	24.32	hetero-oligomer	RHODOPSIN	2 x NAG-NAG, 2 x PLM, 2 x RET, 2 x BOG	X-ray	3.30	0.35	BLAST	0.34	NA	0.94	e_value=1.8513e-12, bit_score=69.3218, score=168	Build Model (Monomer)
		2j4y.1.A	0.60	52 - 329	24.32	homo-dimer	RHODOPSIN	2 x NAG, 2 x RET	X-ray	3.40	0.35	BLAST	0.33	NA	0.94	e_value=7.33076e-13, bit_score=70.4774, score=171	Build Model (Homomer) Build Model (Monomer)
		2ks9.1.A	0.60	66 - 335	24.90	hetero-oligomer	Substance-P receptor		NMR	0.00	0.35	BLAST	0.33	NA	0.94	e_value=2.19669e-13, bit_score=72.0182, score=175	Build Model (Monomer)
		3pxo.1.A	0.60	52 - 329	24.32	homo-dimer	Rhodopsin	2 x NAG-NAG, 2 x PLM, 2 x NAG-NAG-BMA-MAN, 2 x RET, 4 x BOG	X-ray	3.00	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Homomer) Build Model (Monomer)
		1ln6.1.A	0.60	52 - 329	24.32	monomer	RHODOPSIN	1 x RET	NMR	0.00	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Monomer)
		1f88.1.A	0.60	52 - 329	24.32	homo-dimer	RHODOPSIN	3 x NAG-NAG, 4 x ZN, 6 x HG, 2 x RET, 1 x NAG-NAG-MAN	X-ray	2.80	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Homomer) Build Model (Monomer)
								1 x ZN, 2 x NAG-									

Trg-Tpl	Models	Name	Coverage	Range	Seq Id	Oligo-state	Description	Ligands	Method	Resolution	Seq Similarity	Found by	GMQE	QSQE	Score	Other	
		6do1.1.A	0.28	47 - 173	40.50	hetero-1-2-mer	Type-1 angiotensin II receptor,Soluble cytochrome b562 BRIL fusion protein	2 x OLC, 1 x NAG, 1 x SAR-ARG-VAL-TYR-ILE-HIS-PRO-ILE	X-ray	2.90	0.41	BLAST	0.18	NA	0.48	e_value=7.93633e-13, bit_score=70.8626, score=172	Build Model (Monomer)
		5xjm.1.A	0.31	39 - 176	31.11	hetero-1-1-1-mer	Type-2 angiotensin II receptor,Soluble cytochrome b562,Type-2 angiotensin II receptor	1 x SAR-ARG-VAL-TYR-ILE-HIS-PRO-ILE	X-ray	3.20	0.38	BLAST	0.18	NA	0.51	e_value=7.01371e-06, bit_score=48.9062, score=115	Build Model (Monomer)
		5ung.1.A	0.61	39 - 321	31.18	monomer	Chimera protein of Type-2 angiotensin II receptor and Soluble cytochrome b562	1 x OLA, 1 x OLC, 1 x 8ES	X-ray	2.80	0.37	BLAST	0.45	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Monomer)
		6c1q.1.A	0.62	44 - 344	31.34	monomer	Soluble cytochrome b562, C5a anaphylatoxin chemotactic receptor 1 chimera	1 x 9P2, 1 x ACE-PHE-ORN-PRO-ZAL-TRP-ARG	X-ray	2.90	0.37	BLAST	0.43	NA	1.00	e_value=1.58479e-12, bit_score=69.707, score=169	Build Model (Monomer)
		5unf.1.A	0.61	39 - 321	31.18	homo-dimer	Chimera protein of Type-2 angiotensin II receptor and Soluble cytochrome b562	2 x 8ES	X-ray	2.80	0.37	BLAST	0.43	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Homomer) Build Model (Monomer)
		6dde.1.E	0.60	46 - 322	28.91	hetero-1-1-1-1-mer	Mu-type opioid receptor	1 x TYR-DAL-GLY-MEA-ETA	EM	0.00	0.37	BLAST	0.42	NA	0.95	e_value=2.85356e-12, bit_score=68.5514, score=166	Build Model (Monomer)
		5c1m.1.A	0.60	46 - 322	29.30	hetero-oligomer	Mu-type opioid receptor	1 x 4VO, 2 x OLC, 1 x CLR	X-ray	2.10	0.37	BLAST	0.42	NA	0.95	e_value=8.12226e-12, bit_score=66.6254, score=161	Build Model (Monomer)
		5unh.1.A	0.61	39 - 321	31.18	homo-dimer	Soluble cytochrome b562,Type-2 angiotensin II receptor	2 x 8EM	X-ray	2.90	0.37	BLAST	0.41	NA	0.99	e_value=5.46002e-14, bit_score=74.3294, score=181	Build Model (Homomer) Build Model (Monomer)
		6meo.1.C	0.59	43 - 322	28.35	hetero-1-1-1-mer	C-C chemokine receptor type 5	2 x MAN, 17 x NAG, 1 x A2G	EM	3.90	0.37	BLAST	0.41	NA	0.95	e_value=2.22682e-12, bit_score=68.5514, score=166	Build Model (Monomer)
		5d5a.1.A	0.44	41 - 237	27.81	monomer	Beta-2 adrenergic receptor,Endolysin,Beta-2 adrenergic receptor	1 x N9S, 1 x CAU, 1 x 12P, 1 x ACM, 1 x PLM, 3 x CLR	X-ray	2.48	0.35	BLAST	0.27	NA	0.68	e_value=4.78188e-09, bit_score=58.9214, score=141	Build Model (Monomer)
		5x7d.1.A	0.44	41 - 237	27.81	monomer	Chimera protein of Beta-2 adrenergic receptor and Lysozyme	1 x 8VS, 2 x CLR, 1 x CAU, 1 x ACM	X-ray	2.70	0.35	BLAST	0.27	NA	0.68	e_value=4.78188e-09, bit_score=58.9214, score=141	Build Model (Monomer)
		2r4s.1.A	0.62	41 - 321	29.81	hetero-1-1-1-mer	Beta-2 adrenergic receptor		X-ray	3.40	0.35	BLAST	0.26	NA	0.97	e_value=1.76248e-16, bit_score=81.2629, score=199	Build Model (Monomer)
		5glh.1.A	0.45	44 - 238	26.18	hetero-oligomer	Endothelin Receptor Subtype-B		X-ray	2.80	0.35	BLAST	0.26	NA	0.69	e_value=4.21508e-09, bit_score=59.3066, score=142	Build Model (Monomer)
		3p0g.1.A	0.44	41 - 237	27.81	hetero-oligomer	Beta-2 adrenergic receptor, Lysozyme	1 x P0G	X-ray	3.50	0.35	BLAST	0.26	NA	0.68	e_value=4.91705e-09, bit_score=58.9214, score=141	Build Model (Monomer)
		4dkl.1.A	0.42	46 - 234	24.31	homo-dimer	Mu-type opioid receptor, lysozyme chimera	4 x MPG, 2 x BF0, 2 x CLR	X-ray	2.80	0.35	BLAST	0.25	NA	0.66	e_value=7.96581e-07, bit_score=51.9878, score=123	Build Model (Homomer) Build Model (Monomer)
		4daj.2.A	0.42	41 - 233	28.57	monomer	Muscarinic acetylcholine receptor M3, Lysozyme	1 x 0HK	X-ray	3.40	0.35	BLAST	0.25	NA	0.66	e_value=2.26981e-05, bit_score=47.3654, score=111	Build Model (Monomer)
		4u14.1.A	0.42	41 - 233	28.57	monomer	Muscarinic acetylcholine receptor M3,Endolysin,Muscarinic acetylcholine receptor M3	1 x 0HK	X-ray	3.57	0.35	BLAST	0.25	NA	0.66	e_value=2.98846e-05, bit_score=46.9802, score=110	Build Model (Monomer)
		4u15.1.A	0.42	41 - 233	28.57	monomer	Muscarinic acetylcholine receptor M3,Lysozyme,Muscarinic acetylcholine receptor M3	1 x 0HK, 2 x TAR	X-ray	2.80	0.35	BLAST	0.25	NA	0.66	e_value=3.89986e-05, bit_score=46.595, score=109	Build Model (Monomer)
		4u15.2.A	0.42	41 - 233	28.57	monomer	Muscarinic acetylcholine receptor M3,Lysozyme,Muscarinic acetylcholine receptor M3	1 x 0HK, 2 x OLC, 2 x TAR	X-ray	2.80	0.35	BLAST	0.25	NA	0.66	e_value=3.89986e-05, bit_score=46.595, score=109	Build Model (Monomer)
		4daj.1.A	0.42	41 - 233	28.57	monomer	Muscarinic acetylcholine receptor M3, Lysozyme	1 x 0HK	X-ray	3.40	0.35	BLAST	0.24	NA	0.66	e_value=2.26981e-05, bit_score=47.3654,	Build Model (Monomer)
		5lwe.1.A	0.59	48 - 329	27.95	monomer	C-C chemokine receptor type 9	15 x OLA, 1 x 79K, 1 x CLR	X-ray	2.80	0.35	BLAST	0.35	NA	0.93	e_value=9.54245e-10, bit_score=60.8474, score=146	Build Model (Monomer)
		5lwe.2.A	0.59	48 - 329	27.95	monomer	C-C chemokine receptor type 9	13 x OLA, 1 x 79K	X-ray	2.80	0.35	BLAST	0.35	NA	0.93	e_value=9.54245e-10, bit_score=60.8474, score=146	Build Model (Monomer)
		5f8u.1.A	0.57	52 - 321	29.67	monomer	Beta-1 adrenergic receptor		X-ray	3.35	0.35	BLAST	0.35	NA	0.90	e_value=3.99876e-16, bit_score=79.7221, score=195	Build Model (Monomer)
		3c9m.1.A	0.60	52 - 329	24.32	monomer	Rhodopsin	1 x NAG, 1 x RET	X-ray	3.40	0.35	BLAST	0.34	NA	0.94	e_value=7.27779e-13, bit_score=70.4774, score=171	Build Model (Monomer)
		4a4m.1.A	0.60	52 - 329	24.32	hetero-oligomer	RHODOPSIN	2 x NAG-NAG, 2 x PLM, 2 x RET, 2 x BOG	X-ray	3.30	0.35	BLAST	0.34	NA	0.94	e_value=1.8513e-12, bit_score=69.3218, score=168	Build Model (Monomer)
		2j4y.1.A	0.60	52 - 329	24.32	homo-dimer	RHODOPSIN	2 x NAG, 2 x RET	X-ray	3.40	0.35	BLAST	0.33	NA	0.94	e_value=7.33076e-13, bit_score=70.4774, score=171	Build Model (Homomer) Build Model (Monomer)
		2ks9.1.A	0.60	66 - 335	24.90	hetero-oligomer	Substance-P receptor		NMR	0.00	0.35	BLAST	0.33	NA	0.94	e_value=2.19669e-13, bit_score=72.0182, score=175	Build Model (Monomer)
		3pxo.1.A	0.60	52 - 329	24.32	homo-dimer	Rhodopsin	2 x NAG-NAG, 2 x PLM, 2 x NAG-NAG-BMA-MAN, 2 x RET, 4 x BOG	X-ray	3.00	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Homomer) Build Model (Monomer)
		1ln6.1.A	0.60	52 - 329	24.32	monomer	RHODOPSIN	1 x RET	NMR	0.00	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Monomer)
		1f88.1.A	0.60	52 - 329	24.32	homo-dimer	RHODOPSIN	3 x NAG-NAG, 4 x ZN, 6 x HG, 2 x RET, 1 x NAG-NAG-MAN	X-ray	2.80	0.35	BLAST	0.32	NA	0.94	e_value=1.29547e-13, bit_score=72.7886, score=177	Build Model (Homomer) Build Model (Monomer)
								1 x ZN, 2 x NAG-									

Target templates from which the AKHRs where built

Appendix B

Ranking of potential binding pockets in β2AR and the rhodopsin models found using MOE * Size- Number of contributing spheres # PLB- predicted ligand-binding site # Hyd- Number of hydrophilic points # Side - sidechain contact atoms Residues - Residues at the local surface

Rhodflf (rhodopsin model from the flesh fly)

Site	Size	PLB	Hyd	Side	Residues
1	185	2.91	54	102	ARG73 ILE74 ASP75 MET77 LEU78 ILE133 ASP136 ARG137 ALA140 ILE141 LEU145 LYS146 ARG147 TYR223 ILE226 ILE230 LYS233 GLU244 ARG245 ALA257 ARG260 THR261 LYS263 MET264 THR265 ILE268 TYR322 ASN323 ILE324 ARG325 ARG327
2	70	1.99	38	57	ARG119 VAL120 LEU123 TYR124 VAL189 ILE190 ASN192 TYR203 GLN204 SER207 MET208 MET211 TYR212 TRP276 TYR279 TYR280 SER283 ARG301
3	50	1.29	25	45	HIS38 SER41 VAL103 GLN104 TRP105 THR108 ASP109 CYS112 GLU178 HIS179 ALA181 VAL182 THR183 GLY184 TYR185 PHE186 PRO298 LEU299 LYS302
4	42	0.55	15	28	ASN192 SER193 PHE194 ARG195 SER196 ASP199 GLU200 LYS201 GLN204 SER283 TYR286 TRP287 LYS290 ALA293

Beta2flf (Beta 2AR model from the flesh fly)

Site	Size	PLB	Hyd	Side	Residues
1	295	4.15	93	180	MET95 GLU96 TRP99 VAL103 GLN104 TRP105 ASP109 CYS112 ARG113 SER116 PHE117 ARG119 VAL120 LEU123 GLN170 LEU173 HIS175 LEU176 GLU177 ALA181 VAL182 THR183 TYR185 CYS188 VAL189 PHE191 SER193 PHE194 ARG195 SER196 ASP197 PHE198 ASP199 GLU200 LYS201 TYR203 GLN204 MET208 MET211 TRP276 TYR279 TYR280 ILE282 SER283 TYR286 TRP287 LYS290 HIS291 LYS295 ILE296 ASN297 PRO298 LEU299 ARG301 LYS302 PHE305
2	195	1.73	53	101	LEU63 THR64 ARG67 LEU68 ARG69 GLY70 PRO71 LEU72 ARG73 ILE74 ASP75 ILE76 MET79 ARG137 ALA140 ILE141 PRO144 ARG147 ARG153 ILE230 ALA257 LYS258 ARG260 THR261 LEU262 LYS263 MET264 TYR319 LEU321 TYR322 ASN323 ILE324 ARG325 GLY326 ARG327
3	54	0.31	15	22	PRO71 LEU72 ARG73 MET77 CYS132 LEU135 ASP136 PHE139 LYS143 PRO144 LEU145 LYS146 ARG147 SER148 ARG151 GLY152 MET155

4	22	-0.01	18	19	ILE220 TYR223 GLY224 TYR227 LEU254 LYS263 ILE266 VAL269 ILE270 ILE273

Rhodoff (rhodopsin model from the oriental fruit fly)

Site	Size	PLB	Hyd	Side	Residues
1	47	1.94	30	57	ARG131 VAL132 TYR136 SER139 LYS195 VAL201 ALA202 HIS204 GLN216 ASN219 MET220 MET223 TYR224 TRP280 TYR283 TYR284 CYS287 ARG305
2	108	1.92	41	67	ILE86 ASP87 LEU90 ILE145 ASP148 ARG149 ILE153 PHE159 TYR235 ILE238 ALA261 ARG264 THR265 LYS267 MET268 THR269 ILE272 PHE326 ASN327 ILE328 ARG329
3	54	1.27	21	41	ILE62 GLY63 THR66 VAL67 LEU70 ILE71 SER82 ASP87 LEU90 MET91 TYR323 GLY324 VAL325 ASN327 ARG329 GLY330 LYS331 ASN333 ASN334
4	48	0.90	25	40	LYS193 VAL194 HIS204 SER205 PHE206 PRO207 LYS212 LEU213 GLN216 CYS287 PHE290 TRP291 LYS294 ALA297 ASP298 VAL300

(Beta 2AR model from the oriental fruit fly)

Site	Size	PLB	Hyd	Side	Residues
1	190	4.38	70	136	LEU107 GLU108 TRP111 VAL115 GLN116 TRP117 ASP121 CYS124 ARG125 MET127 SER128 PHE129 ARG131 VAL132 LEU135 GLN182 ILE185 HIS187 LEU188 TYR197 CYS200 VAL201 PHE203 SER205 PHE206 PRO207 THR208 GLU209 LEU210 HIS211 TYR215 GLN216 ASN219 MET220 MET223 TYR283 TYR284 MET286 CYS287 PHE290 TRP291 PRO302 LEU303 ARG305 LYS306 PHE309
2	95	0.96	28	53	ARG81 SER82 SER84 ILE86 ASP87 TYR239 ARG260 ALA261 LYS263 ARG264 THR265 LYS267 MET268 TYR323 PHE326 ASN327 ILE328 ARG329 GLY330
3	64	0.79	23	40	GLU189 VAL194 TYR197 THR208 GLU209 LEU210 HIS211 LYS212 LEU213 GLN216 PHE290 TRP291 ASP293 LYS294 PRO295 SER296 ALA297 ASP298 VAL300 PRO302 ARG305
4	36	0.30	19	27	ILE62 GLY63 THR66 VAL67 LEU70 PRO320 LEU321 TYR323 GLY324 VAL325 GLY330 LYS331 ASN334 ASN335

Interproton distances from NMR (nm) used during Molecular dynamic simulations. Photo-HrTH 150 mM SDS micelles, pH 4.5, 2mM peptide, Temp = 310 K referenced to DSS according to Wüthrich, K., NMR of Proteins & Nucleic Acids, p. 178, Wiley 1986)

Number	Resonances	Value	Upper Limit	Lower Limit
1	4PheH-2LeuHg	3.72257	4.46709	2.97806
2	2LeuH-2LeuHg	3.40984	4.09181	2.72787
3	3ThrHg2*-2LeuHda*	5.91826	7.10191	4.73461
4	6ProHa-2LeuHda*	4.89370	5.87244	3.91496
5	3ThrHg2*-4PheHba	4.90055	5.88066	3.92044
6	4PheHba-4PheHbb	4.01776	4.82132	3.21421
7	4PheH-4PheHba	2.47042	2.96450	1.97633
8	6ProHga-4PheHba	4.18416	5.02100	3.34733
9	7AspHba-4PheHba	3.85542	4.62651	3.08434
10	4PheHba-4PheHe*	4.21901	5.06282	3.37521
11	5SerH-4PheHba	2.98967	3.58760	2.39173
12	3ThrHa-4PheHba	3.86036	4.63243	3.08828
13	5SerHa-4PheHba	2.46959	2.96351	1.97567
14	4PheHba-4PheHd*	2.94521	3.53426	2.35617
15	3ThrHg1-3ThrHa	3.04423	3.65308	2.43539
16	3ThrHa-4PheH	2.88589	3.46307	2.30871
17	3ThrHa-6ProHa	3.85641	4.62769	3.08513
18	3ThrHa-3ThrHb	2.68034	3.21641	2.14427
19	3ThrH-3ThrHa	3.29356	3.95227	2.63485
20	3ThrHg2*-1GluHga	4.35194	5.22233	3.48155
21	1GluHbb-1GluHga	3.97017	4.76421	3.17614
22	5SerHbb-5SerHba	4.04949	4.85939	3.23959
23	7AspHba-5SerHbb	3.76029	4.51235	3.00823
24	5SerH-5SerHbb	3.30984	3.97181	2.64787
25	5SerHa-5SerHbb	2.95008	3.54010	2.36006
26	3ThrHg2*-1GluHgb	4.95113	5.94136	3.96091
27	1GluHbb-1GluHgb	3.99956	4.79947	3.19965
28	8TrpH-8TrpHbb	3.56140	4.27368	2.84912
29	8TrpH-5SerHg	4.05864	4.87037	3.24691
30	8TrpHa-8TrpH	2.85458	3.42550	2.28366
31	7AspHa-8TrpH	2.51479	3.01775	2.01183
32	3ThrHb-8TrpH	3.77499	4.52998	3.01999
33	8TrpH-6ProHa	4.42037	5.30445	3.53630
34	8TrpH-7AspHbb	4.11011	4.93213	3.28809
35	8TrpH-8TrpHba	2.44787	2.93745	1.95830
36	3ThrHg2*-8TrpH	4.98005	5.97606	3.98404
37	3ThrHg1-4PheH	3.32954	3.99545	2.66363
38	3ThrHb-4PheH	3.28340	3.94007	2.62672
39	6ProHa-4PheH	3.88625	4.66350	3.10900
40	4PheH-2LeuHa	3.91957	4.70349	3.13566
41	5SerHa-4PheH	2.82402	3.38883	2.25922
42	1GluHba-4PheH	3.91823	4.70187	3.13458
43	[66]-4PheH	2.53358	3.04030	2.02686
44	[67]-6ProHa	4.06876	4.88251	3.25500
45	7AspH-3ThrHg1	3.21520	3.85824	2.57216

46	7AspH-8TrpHbb	2.90826	3.48992	2.32661
47	7AspH-3ThrHb	2.50124	3.00149	2.00100
48	7AspH-2LeuHbb	3.85558	4.62669	3.08446
49	7AspH-7AspHba	3.73586	4.48303	2.98869
50	7AspH-7AspHbb	3.25887	3.91064	2.60709
51	7AspH-7AspHa	2.73808	3.28569	2.19046
52	7AspHa-8TrpHbb	3.52949	4.23539	2.82359
53	7AspHa-5SerHg	3.85507	4.62609	3.08406
54	7AspHa-7AspHbb	3.13039	3.75647	2.50432
55	3ThrH-[6AspHa]	3.67547	4.41056	2.94038
56	6ProHa-6ProHbb	4.23142	5.07770	3.38513
57	[69]-6ProHbb	3.29083	3.94900	2.63267
58	5SerHg-6ProHbb	3.90086	4.68103	3.12069
59	6ProHa-6ProHba	4.26548	5.11858	3.41239
60	3ThrHg2*-6ProHba	4.91487	5.89784	3.93189
61	7AspHba-4PheHbb	3.09162	3.70995	2.47330
62	7AspHba-6ProHa	4.01229	4.81475	3.20983
63	7AspHba-5SerHba	2.26200	2.71439	1.80960
64	5SerH-7AspHba	4.16331	4.99597	3.33065
65	7AspHba-5SerHa	2.62832	3.15398	2.10265
66	7AspHba-2LeuHbb	3.81678	4.58014	3.05342
67	7AspHba-5SerHg	3.11629	3.73954	2.49303
68	8TrpHa-5SerH	2.57080	3.08496	2.05664
69	3ThrH-8TrpHa	3.43547	4.12256	2.74838
70	3ThrHb-4PheHd*	4.55469	5.46563	3.64375
71	1GluHba-4PheHd*	4.44558	5.33470	3.55647
72	3ThrHg2*-6ProHga	4.72130	5.66556	3.77704
73	6ProHga-6ProHa	3.92018	4.70422	3.13615
74	6ProHga-6ProHdb	3.46389	4.15666	2.77111
75	6ProHga-8TrpHba	3.98401	4.78082	3.18721
76	6ProHga-6ProHda	3.56610	4.27932	2.85288
77	1GluHa-[70]	3.87125	4.64550	3.09700
78	1GluH*-5SerHba	4.81621	5.77946	3.85297
79	1GluH*-1GluHba	4.71241	5.65490	3.76993
80	1GluHa-1GluH*	4.77503	5.73003	3.82002
81	3ThrHg2*-1GluHa	4.80218	5.76261	3.84174
82	1GluHa-6ProHa	4.07541	4.89049	3.26032
83	1GluHa-1GluHba	3.87562	4.65074	3.10049
84	3ThrHg1-4PheHbb	3.05079	3.66095	2.44063
85	4PheHbb-5SerHba	2.97483	3.56980	2.37986
86	1GluHba-4PheHbb	4.17370	5.00844	3.33896
87	3ThrHg2*-1GluHba	4.62476	5.54971	3.69981
88	1GluHba-5SerHg	4.05882	4.87059	3.24706
89	6ProHa-[71]	4.02729	4.83275	3.22183
90	8TrpHba-4PheHz	3.76779	4.52135	3.01423
91	8TrpHba-4PheHe*	3.41675	4.10010	2.73340
92	3ThrHg2*-2LeuHa	4.29059	5.14871	3.43248
93	3ThrHg1-2LeuHa	3.74595	4.49514	2.99676
94	1GluHbb-2LeuHa	4.09061	4.90873	3.27249
95	2LeuHa-2LeuHbb	3.17959	3.81550	2.54367
96	2LeuH-2LeuHa	3.46526	4.15832	2.77221
97	3ThrH-2LeuHa	2.45121	2.94145	1.96097
98	3ThrHg1-6ProHda	3.97850	4.77420	3.18280
99	7AspHbb-6ProHda	3.34034	4.00841	2.67228

100	6ProHa-4PheHz	3.98217	4.77860	3.18574
101	6ProHgb-6ProHa	4.12540	4.95048	3.30032
102	6ProHa-4PheHa	4.25684	5.10821	3.40547
103	3ThrHg1-3ThrHb	2.55894	3.07073	2.04715
104	3ThrHb-5SerHg	4.04079	4.84895	3.23263
105	3ThrHb-5SerHba	3.72490	4.46987	2.97992
106	3ThrHb-4PheHe*	4.58505	5.50205	3.66804
107	3ThrHb-6ProHa	4.06908	4.88290	3.25526
108	3ThrH-3ThrHb	3.26538	3.91845	2.61230
109	3ThrHb-2LeuHbb	2.82527	3.39032	2.26022
110	3ThrHg2*-3ThrHb	5.01528	6.01834	4.01223
111	5SerH-5SerHg	4.18477	5.02173	3.34782
112	5SerH-6ProHa	3.38206	4.05847	2.70565
113	5SerH-5SerHba	3.91947	4.70336	3.13558
114	5SerH-5SerHa	2.46415	2.95698	1.97132
115	5SerH-[72]	3.16133	3.79360	2.52906
116	6ProHa-2LeuHdb*	5.27082	6.32499	4.21666
117	5SerHa-5SerHba	2.73650	3.28379	2.18920
118	5SerHa-2LeuH	3.17229	3.80675	2.53783
119	5SerHa-6ProHa	2.80798	3.36957	2.24638
120	3ThrH-5SerHa	3.33047	3.99656	2.66437
121	1GluHbb-1GluHa	4.43330	5.31996	3.54664
122	6ProHa-3ThrHb	3.97354	4.76824	3.17883
123	2LeuH-3ThrH	3.41164	4.09397	2.72931
124	2LeuHba-6ProH	3.16112	3.79334	2.52890
125	5SerHg-5SerHba	3.17467	3.80961	2.53974
126	6ProHa-5SerHg	4.14568	4.97482	3.31654
127	3ThrHg1-5SerHg	2.91101	3.49322	2.32881
128	7AspHbb-5SerHg	4.02260	4.82712	3.21808
129	3ThrHg1-2LeuHbb	2.45663	2.94796	1.96531
130	3ThrHg1-5SerHba	4.01698	4.82037	3.21358
131	3ThrHg1-3ThrH	3.18049	3.81659	2.54439
132	3ThrHg2*-7AspHbb	4.90332	5.88399	3.92266
133	3ThrHg2*-5SerHba	5.02820	6.03384	4.02256
134	3ThrHg2*-1GluHbb	4.55216	5.46260	3.64173
135	3ThrHg2*-3ThrH	4.64014	5.56817	3.71211
136	3ThrHg2*-2LeuH	4.63556	5.56267	3.70845
137	3ThrHg2*-2LeuHba	4.69633	5.63559	3.75706
138	3ThrHg2*-6ProHdb	4.78140	5.73768	3.82512
139	3ThrHg2*-2LeuHbb	4.40272	5.28326	3.52218
140	6ProHa-5SerHba	3.91832	4.70198	3.13466
141	2LeuHbb-5SerHba	4.03364	4.84037	3.22691
142	1GluHbb-2LeuHba	2.77387	3.32864	2.21910
143	6ProHa-2LeuH	3.75188	4.50226	3.00150
144	3ThrH-2LeuHbb	3.51083	4.21299	2.80866
145	2LeuH-2LeuHbb	2.93991	3.52790	2.35193
146	2LeuHbb-6ProHdb	4.26364	5.11637	3.41091
147	2LeuHba-2LeuH	3.59980	4.31976	2.87984