CHARACTERIZATION OF ADIPOKINETIC HORMONE RECEPTOR IN STICK INSECT, *CARAUSIUS MOROSUS*

by

Merve Gizem Sınmaz

B.S., Molecular Biology and Genetics, Boğaziçi University, 2014

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2018

CHARACTERIZATION OF ADIPOKINETIC HORMONE RECEPTOR IN STICK INSECT, *CARAUSIUS MOROSUS*

APPROVED BY:

Assist. Prof. Necla Birgül İyison .....
(Thesis Supervisor)

Assist. Prof. Igor Kryvoruchko .....

Assoc. Prof. Serdar Durdağı .....

DATE OF APPROVAL: 02.08.2018

ACKNOWLEDGEMENTS

This thesis project is a collaborative product of Necla Birgül İyison's Cancer Signaling Laboratory (CSL) in Boğaziçi University with Serdar Durdağı from Bahçeşehir University. Transcriptomic analyses were performed in CSL and computational experiments were performed with Serdar Durdağı.

First and foremost, I would like to offer thanks to my supervisor Necla Birgül İyison for providing me with the opportunity to work in this project that expanded my knowledge in computational field. She made a great leading effort, showed patience and referred me to outstanding computational scientists whenever I asked help. I am very grateful to perform this project as one of her CSL team members.

I would also thank to Serdar Durdağı for his efforts on training me about computational tools, for sparing his time on analyses of this project and for sharing his valuable ideas with me. He definitely contributed my knowledge so much. His Ph.D. student Busecan Aksoydan spent so much time and effort on my training in using tools. She solved my biggest problems and supported me all the time in both academic and private life. I would not be able to come to this point without her. Working on this project let me meet Busecan and I am grateful to have such a great friend in my life.

I would like to add my appreciations to all CSL members; Tolga Aslan for assisting me whenever I need and for his personal friendship, Burçin Duan Şahbaz for her helps in my understanding of notions in bioinformatics with unlimited patience, for sharing her transcriptome analysis of stick insect with me, Aida Shahraki for her efforts to support me with a great patience about my ignorance in biophysics and Ayşe Nur Kayabaşı for making my lab enjoyable together. I am also grateful to have Efe Elbeyli for his assistance in bioinformatics, Aylin Alkan and İpek Selcen for their supports and friendships even outside my academic life.

Finally, I would like to admit that I could not have done this without love and support of my entire family. I am very lucky to have my little brother Mert who supported me all the time. He is a unique ally for my whole life. Also, I am grateful to Nalan Yıldız for being a sister, a friend and a colleague for me. She supported me, cheered me up and encouraged me to work hard all the time.

Thanks for all.

ABSTRACT

CHARACTERIZATION OF ADIPOKINETIC HORMONE RECEPTOR OF STICK INSECT, *CARAUSIUS MOROSUS*

Adipokinetic hormone (AKH) is an insect neuropeptide and known as gonadotropin-releasing hormone, which is orthologue to humans. It is released from corpora cardiaca and acts generally on fat body for energy mobilization. It is also involved in several different physiological functions such as increasing blood hemolymph trehalose levels, heart beat frequency, and protein synthesis. Furthermore it functions in inhibiting fatty acid and RNA synthesis in different insect species. It activates its corresponding G-Protein Coupled Receptors (GPCR), which are Adipokinetic Hormone receptors (AKHR). Many AKHRs were studied so far, but not in stick insect, *Carausius morosus*. Utilizing transcriptome data of stick insect, its AKHR (CamAKHR) and AKH (CamAKH) were identified. It was aimed to characterize AKHR of stick insect by homology modeling, docking analysis and molecular dynamics within this project contributing to the production of a novel neuropeptide-based next-generation pesticide. At the end, thirteen important residues mostly located in transmembrane 6, transmembrane 7 and extracellular loop 2 regions of CamAKHR model were observed contributing to ligand binding in CamAKHR- CamAKH complex. Six of them were found as the most conserved residues in different insect species, therefore proposed binding pocket includes these Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 residues corresponding to TM2, TM3, TM6 and TM7 of CamAKHR.

ÖZET

DEĞNEK BÖCEĞİ *CARAUSIUS MOROSUS'* TA ADİPOKİNETİK HORMON RESEPTÖRÜNÜN KARAKTERİZASYONU

Adipokinetik hormone (AKH) bir böcek nöropeptiti olup, insanlardaki gonadotropin-salgılayan hormonun ortoloğu olarak da bilinir. Corpora cardiaca'dan salgılanır ve enerji mobilizasyonu için genellikle depo lipitleri üzerinde etkisini gösterir. Kan hemolenfindeki trehaloz seviyesinin, kalp atışı sıklığının ve protein sentezinin artması gibi birçok çeşitli fizyolojik fonksiyonlarda da yer alırlar. Ayrıca bazı böcek türlerinde yağ asitlerinin ve RNA sentezinin inhibe edilmesinde de görev yaparlar. Kendilerine karşılık gelen adipokinetik hormone reseptörü (AKHR) olarak isimlendirilen G-protein kenetli reseptörleri (GPCR) aktive ederler. Şimdiye kadar birçok AKHR çalışılmış, ancak değnek böceği *C.morosus*'takiler çalışılmamıştır. Değnek böceğinin transkriptom verisini kullanılarak, bu böceğe ait adipokinetik hormone reseptör (CamAKHR) ve adipokinetik hormon (CamAKH) tanımlanmıştır. Değnek böceğine ait AKHR'nin homoloji modellemesi, kenetlenme analizi ve moleküler dinamik yöntemleri ile karakterize edilmesi ve bu sayede yeni bir nöropeptit-esaslı yeni-jenerasyon pestisit üretimine katkı sağlaması amaçlanmıştır. Çalışmanın sonunda ligandın bağlanmasında önem taşıyan on üç rezidünün, CamAKHR modelinin çoğunlukla transmembran 6, transmembran 7 ve hücre dışı ilmek 2 bölgelerinde yer aldığı gözlemlenmiştir. Bu on üç rezidüden altısı farklı böceklerde de korunmuş halde bulunduğundan, önerilen bağlanma cebi CamAKHR'nin TM2, TM3, TM6 ve TM7 bölgelerine denk gelen Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 rezidülerini içermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Å	Angstrom
Ψ	Psi angle
Φ	Phi angle
A, Ala	Alanine
C, Cys	Cysteine
Cα	Carbon-alpha atom
Cβ	Carbon-beta atom
D, Asp	Aspartic Acid
E, Glu	Glutamic acid
F, Phe	Phenylalanine
G, Gly	Glycine
H, His	Histidine
Hie	Epsilon-protonated His
I, Ile	Isoleucine
K, Lys	Lysine
L, Leu	Leucine
M, Met	Methionine
N, Asn	Asparagine
NH_2	Amide group
P, Pro	Proline
pGlu	Pyroglutamate
Q, Gln	Glutamine
R, Arg	Arginine
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
S, Ser	Serine
T, Thr	Threonine
V, Val	Valine
W, Trp	Tryptophan
X	Any amino acid

Y, Tyr	Tyrosine

LIST OF ACRONYMS/ ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
AKH	Adipokinetic hormone
AKHR	Adipokinetic hormone receptor
bp	Basepair
CamAKH	*C.morosus* Adipokinetic hormone
CamAKHR	*C.morosus* Adipokinetic hormone receptor
CC	*Corpus cardiaca*
CNS	Central nervous system
CSL	Cancer Signaling Laboratory
C-terminus	Carboxy-terminus
ECL	Extracellular loop
fs	femtosecond
GBSA	Generalized Born Solvent Accessible Surface Area
GnRH	Gonadotropin-Releasing Hormone
GPCR	G-protein-coupled receptor
HoTH	Hypotrehalosemic Hormone
HrTH	Hypertrehalosemic Hormone
ICL	Intracellular loop
JH	Juvenile Hormone
M	Molarity
MD	Molecular dynamics
MM	Molecular mechanics
MM-GBSA	Molecular mechanics generalized born surface area
NaCl	Sodium chloride
NPT	Isothermal-isobaric ensemble
ns	Nanosecond
N-terminus	Amino-terminus
PDB	Protein Data Bank
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

RNA	Ribonucleic acid
RPCH	Red Pigment Concentrating Hormone
TM	Transmembrane

1. INTRODUCTION

1.1. Insects

1.1.1. Importance of the Insect Existence

Insects are one of the largest classes of invertebrates within the phylum Arthropoda containing nearly 200 times more species than mammals [1]. They also constitute more than 58% of known global diversity, so this class can be considered as the most diverse group in animals. These organisms inhabit all habitat types on the world that is essential for the function and stabilization of terrestrial and aquatic ecosystems [2]. They facilitate pollination by carrying sticky polens from one plant to another when they are flying for food and recycling of compounds found in dead bodies of organisms and in organic wastes by feeding on them; thus, contributing to the maintenance of food chain [3,4]. Besides their essential ecological roles, they are utilized in food sector such as beetle and locusts, in medical treatments such as the use of fly larvae in maggot therapy and for aesthetic reasons such as the use of butterflies' colors and patterns in goods for societies in different locations of the world [3]. There are also materials produced by using insects such as silk from silkworms, lac and dye from scale insects and beeswax from bees [5].

1.1.2. Detrimental Impacts of Insects

Besides several benefits of insects, they do not always show favorable impacts on society. There are also many detrimental impacts of insects affecting humans indirectly or directly. For instance, they can cause economical troubles by being pests of stored products including food and crops and by being parasites on farm animals. Furthermore, they can have more direct impacts on humans by facilitating the spread of several diseases like malaria [3,4,6].

1.1.3. Stick Insect

Stick insect, *Carausius morosus,* is the insect species in the center of the study presented in this thesis (Figure 1.1.). It belongs to Phasmatidae family within Phasmatodea (=Phasmatoptera, Cheleutoptera) order. It is herbivore and feeds on the leaves of deciduous trees and shrubs such as oak, eucalyptus and strawberry. Despite its pet status, this insect species is parthenogenetic, polyphagous and also can be easily grown compared to other species, so it can be a good candidate to be a research object [3,7,8]. It is called walking stick insect or Indian stick insect because it resembles a stick and is native to India. However, it is found in several other places such as South Africa, Great Britain and United States on the world because it invaded those places that is supposed to be a result of uncontrolled pet trade and possibly disposal of eggs carelessly [7,8]. It mostly invades tropical and subtropical regions and when it invades, it causes detrimental environmental impact through severe defoliation of a great number of trees resulting in decreased number of trees, destruction of animal habitats and ultimately economic loss [9].

Figure 1.1. Stick insect, *Carausius morosus*

1.1.4. Methods for Controlling Pests

Taking the harmful effects of insects mentioned above into consideration, there must be some cautions to be taken in order to regulate and control the populations of these insects. There are three commonly known methods to be applied. First one is physical method that uses electromagnetic radiation, thermal and mechanical shock and some sorts of physical

barriers. Second one is chemical method that uses chemicals. The other one is biological method that aims to eradicate or to control the number of targeted pest species by introducing other organisms acting against the target one [10]. At the first glance, the latter one seems a sensible alternative way. However, it is not always feasible because newly introduced organisms can become pest with increased numbers if there are not enough predators feeding on them.

Insecticides are also harmful, non-toxic and species-specific ones are needed. Traditional ways such as chemical insecticides are the most used way to control pests in agriculture [11]. They act by disrupting various physiological processes in insects such as inhibition of neuronal transmission and increasing oxidative stress regardless of species [12, 13]. They cause a threat to health by contaminating environment with their residues that can be toxic to humans and to ecosystem. When these neurotoxic insecticides are applied unconsciously, it leads to dramatic problems such as developments in insect resistance to insecticides and environmental contamination due to their toxic residues [11]. These are the reasons behind why more environmentally friendly and discriminative only against specific species insecticides are investigated for modern pest management [14]. Insect neuropeptides which regulates most critical metabolic, homeostatic, developmental, reproductive and behavioral processes in insects are found out to be good candidates reducing the amount of classical pesticides deployed against insect pests of crops, livestock, pets and people in environmentally friendly way [13,14]. Characterization of these neuropeptides is crucial in order to be utilized as pest control agents because the characterization of neuropeptides provides valuable insights into their structural information and functioning. With the advancements in structural biology, structural identifications of these neuropeptides as well as adipokinetic hormone (AKH) that controls energy metabolism, locomotion activity and immune response were provided [14]. Furthermore, it was also noted that insecticide efficiency was accelerated when co-applied with AKH [15]. Other than these, due to the fact that AKH is present in all insects, plays role in large number of vital processes and some AKHs are species-specific, use of AKH in next-generation pesticide production can be considered quite promising [14].

In order to understand the action mechanism of AKH based antagonist that can be designed as a novel next generation pesticide, AKH binding to its corresponding receptor

(AKHR) to activate related physiological processes in stick insect *C. morosus* should be investigated. Therefore, this study focused on the identification of important residues of AKH binding cavity in AKHR in stick insect *C. morosus.*

1.2. Neuropeptides

1.2.1. Functions of Neuropeptides

In insects, endocrine system possesses a high importance due to its regulatory roles in various physiological, developmental and behavioral functions. This system executes its roles by producing and secreting different hormones mainly from its two constitutive parts the brain and the prothoracic gland. There are some other cells that support the brain and the prothoracic gland in conducting their roles such as brain-associated specialized secretory neurons-containing (neurosecretory cells) *corpora allata* and *corpora cardiaca* and some endocrine cells located in the neural ganglia, gut and gonads [16, 17].

Insect endocrine system performs its functions through three mainly classified groups named as ecdysteroids, juvenile hormones (JH) and neuropeptides [16, 17]. Whereas the first group ecdysteroids are steroid hormones, which are produced primarily by prothoracic glands regulating several processes like molting, metamorphosis and reproduction, the second group juvenile hormones are produced primarily by *corpora allata* modulating metamorphosis and reproductive processes. The last group neuropeptides (so-called peptide hormones) are small peptidergic substances produced and released by specialized secretory neurons found in brain, *corpora cardiaca* and throughout the nervous system [15]. This last group, named neuropeptides, is responsible for the regulation of diverse behavioral, physiological and biochemical processes in insects. For instance, myotropic peptides regulate muscle activity, allatostatins and allatotropins regulate reproduction, growth and development, and, as in the concern of this thesis, adipokinetic peptides regulate metabolism in insect body [16].

1.2.2. Neuropeptide Family of Adipokinetic Hormone and its Structure

AKHs are short neuropeptides constituting AKH/RPCH (red pigment concentrating hormone) family. While peptides found in insects are named as AKHs, the ones found in crustaceans are named as RPCH. Because there are high structural similarity between AKH peptides and their cognate receptors with gonadotropin-releasing hormone (GnRH), AKH/RPCH peptide family has been considered as a member of GnRH family containing more members as shown in Table 1.1. [18, 19].

Table 1.1. Amino acid sequences of some of the AKH and AKH-like peptide hormones of the AKH/RPCH family (adapted from [19])

Neuropeptide name	**Amino acid sequence**	**Phylum**
Locusta migratoria AKH-1	pQLNFTPNWGTamide	Arthropoda
Globodera rostochiensis AKH	pQMTFSDGWamide	Nematoda
Hypsibius dujardini AKH	pQLSFSTGWGHamide	Tardigrada
Priapulus caudatus AKH-like	pQIFFSKGWRGamide	Priapulida
Crassostrea gigas AKH	pQVSFSTNWGSamide	Mollusca
Lottia gigantea AKH	pQIHFSPTWGSamide	Mollusca
Capitella teleta GnRH-2	pQFSFSLPGKWGNamide	Annelida
Platynereis dumerilii proto-AKH-1	pQFSFSLPGKWGNamide	Annelida
Helobdella robusta GnRH	pQFSFTPPGKWPFGTamide	Annelida
Lingula anatine proto-AKH	pQWHQTLGWGAAGMVamide	Brachiopodia

All AKH family members including AKHs, hypertrehalosemic (HrTH), hypotrehalosemic (HoTH) and hyperprolinemic hormones (or *corpus cardiacum* factors) are found in insects [18]. These peptide hormones belonged to AKH family and have many structural and functional similarities. The reason of their different denotation is that they affect different substrates. Whereas adipokinetic hormone peptide acts on lipid metabolism, hypertrehalosemic hormone and hyperprolinemic hormone peptides act on carbohydrate and on proline metabolism for energy production in insects, respectively [20]. For the sake of simplicity, all members in this family will be named as adipokinetic hormone for the rest of the work.

Beginning from the elucidation of the structure of the first AKH from 3000 locust *corpora cardiaca* in 1976, with the help of advancements in technology, 60 different AKHs were identified by sequencing of individuals from mostly insect orders by now as shown in Table 1.2. [18,21].

Table 1.2. Amino acid sequences of some of the AKH and AKH-like peptide hormones of insects (adapted from [18])

Insecta order	**Hormone code**	**Amino acid sequence**
Archaeognatha	Anaim-AKH	pGlu-Val-Asn-Phe-Ser-Pro-Ser-Trp NH_2
Zygentoma	Peram-CAH-I	pGlu-Val-Asn-Phe-Ser-Pro-Asn-Trp NH_2
Odonata	Libau-AKH	pGlu-Val-Asn-Phe-Thr-Pro-Ser-Trp NH_2
Plecoptera	Panbo-RPCH	pGlu-Leu-Asn-Phe-Ser-Pro-Gly-Trp NH_2
Dermaptera	Grybi-AKH	pGlu-Val-Asn-Phe-Ser-Thr-Gly-Trp NH_2
Blattodea	Bladi-HrTH	pGlu-Val-Asn-Phe-Ser-Pro-Gly-Trp-Gly-Thr NH_2
Mantodea	Emppe-AKH	pGlu-Val-Asn-Phe-Thr-Pro-Asn-Trp NH_2
Grylloblattodea	Galyu-AKH	pGlu-Val-Asn-Phe-Ser-Pro-Thr-Trp NH_2
Mantophasmatodea	Manto-CC	pGlu-Val-Asn-Phe-Ser-Pro-Gly-Trp NH_2
Phasmatodea	Carmo-AKH-II	pGlu-Leu-Thr-Phe-Thr-Pro-Asn-Trp-Gly-Thr NH_2
Orthoptera	Locmi-AKH-I	pGlu-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-Thr NH_2
Psocodea	Emppe-AKH	pGlu-Val-Asn-Phe-Thr-Pro-Asn-Trp NH_2
Hemiptera	Rhopr-AKH	pGlu-Leu-Thr-Phe-Ser-Thr-Asp-Trp NH_2
Neuroptera	Grybi-AKH	pGlu-Val-Asn-Phe-Ser-Thr-Gly-Trp NH_2
Megaloptera	Aedae-AKH	pGlu-Leu-Thr-Phe-Thr-Pro-Ser-Trp NH_2
Coleoptera	Melme-CC	pGlu-Leu-Asn-Tyr-Ser-Pro-Asp-Trp NH_2
Diptera	Phote-HrTH	pGlu-Leu-Thr-Phe-Ser-Pro-Ser-Trp NH_2
Lepidoptera	Manse-AKH	pGlu-Leu-Thr-Phe-Thr-Ser-Ser-Trp-Gly NH_2

All members belonged to AKH/RPCH peptide family show some structural properties in common. They are all octa-decapeptides, blocked at N-terminus with pyroglutamate, blocked at C-terminus with amide group, have generally uncharged like Trp and Gly amino acids at positions 8 and 9, and contain at least 2 aromatic amino acids (most have Phe at position 4, Trp at position 8) [16].

AKHs are typically 8-10 amino acid long peptide hormones and share a pyroglutamate (pGlu) residue as a post-translational modification at N-terminal and an amide group at C-terminal as other members in AKH/RPCH peptide family [16, 21, 22]. All members have conserved residues at position 1, 4, 8 as pGlu, an aromatic group such as Phe or Tyr and Trp, respectively. If the peptide is nona- or decapeptide, it also has a Gly at position 9. Other than these mostly and strictly conserved residues, other residues possess some conserved patterns. Most members possess a branched-chain amino acid residue such as Leu, Val, Ile, Phe and Tyr or aromatic residues such as Phe and Tyr at position 2, and either a Ser or a Thr at position 5. Positions 6, 7 and 10 are the most variable parts of the peptide. The peptide can also undergo various post-translational modifications [18]. For instance, a unique glycosylation (called as C-mannosylation) is present at Trp residue of one of two AKH peptides in stick insects, *Carausius morosus* [18,20]. Furthermore, there are some conformational restraints within Thr3-Phe4-Thr5 and Asn7-Trp8-Gly9 residues of the AKH peptide of stick insect. It was also noted that this glycosylation does not perturb conformational restraints within conserved residues of the peptide, so it can be considered that both glycosylated and not glycosylated forms of AKH peptide have this restraint [20]. Overall scheme of AKHs with conserved residues is shown in Table 1.3.

Table 1.3. Variations in the residues of AKHs (adapted from [18])

Variation per residue position										
Positions	1	2	3	4	5	6	7	8	9	10
	pGlu	Leu	Asn	Phe	Ser	Pro	Gly	Trp	Gly	Asn
		Val	Thr	Tyr	Thr	Ser	Ser			Thr
		Ile				Thr	Thr			Gln
		Phe				Ala	Asn			Gly
		Tyr				Asp	Asp			Ser
						Hyp	Trp			Tyr
							Ala			
							Val			

1.2.3. Functions of Adipokinetic Hormone

Adipokinetic hormone is a neuropeptide hormone in insects known as its function in the mobilization of substrates for energy generation required by contracting flight muscles [18,19]. It is also quite important to combat stress conditions such as insecticide application and energy stress by modulating homeostasis via metabolic control and its additional pleiotropic effects [14,16,23,24].

The functions of AKH can be classified into three groups, which are metabolic, biochemical and physiological effects. Its metabolic effects can be considered as the mobilization of lipids, carbohydrates and the stimulation of proline synthesis. Activation of adenyl cyclase and lipase [25], glycogen phosphorylase [26] and phospholipase C [27], stimulation of antioxidant mechanisms [28] and inhibition of lipid [29], and syntheses of protein [30] and RNA [31] can be evaluated as biochemical effects. Also, modulation of foraging behavior, which stems from starvation [32], prevention of egg maturation [33], stimulation of heart beat [34], and of processes in digestion [35], regulation of muscle tone [36] and locomotion activity [37], and accelerated immune response [38] indicate physiological roles of AKH.

Considering AKH role only in the mobilization of energy substrates for contracting flight muscles regardless of its other essential functions, most AKH studies have been undertaken using flying insects. However, adipokinetic hormone is still worth to investigate further because it is also found in non-flying insects (Table 1.4.) and there are additional significant adipokinetic hormone functions that are not associated with insect flights.

Table 1.4. Amino acid sequences of AKH peptides of walking and flying insect species (adapted from [23], h*:hexose)

	Species	Hormone code	Amino acid sequence
Walking Species	*Carausius morosus*	Carmo-HrTH-I Carmo-HrTH-II	pQLTFTPNWh*GTamide pQLTFTPNWGTamide
	Periplaneta americana	Peram-CAH-I Peram-CAH-II	pQVNFSPNWamide pQLTFTPNWamide
	Tenebrio molitor	Tenmo-HrTH	pQLNFSPNWamide
	Pyrrhocoris apterus	Pyrap-AKH	pQLNFTPNWamide
Flying species	*Locusta migratoria*	Locmi-I-AKH Locmi-II-AKH Locmi-III-AKH	pQLNFTPNWGTamide pQLNFSAGWamide pQLNFTPWWamide
	Schistocerca gregaria	Schgr-II-AKH	pQLNFSTGWamide
	Manduca sexta	Manse-AKH	pQLTFTSSWGamide
	Drosophila melanogaster	Drome-HrTH	pQLTFSPDWamide

Adipokinetic hormones help insects, which lost their flying ability or mostly use walking ability for locomotion such as *C.morosus*, to survive under metabolic stress conditions which consume huge amounts of energy such as diapause, starvation and molting by mobilizing energy storages [18, 23]. In addition to these functions, hypothetical metabolic and neuromodulatory roles of AKHs in locomotory activity in non-flying insects are still under investigation.

Furthermore, although insects are widely used in AKH studies, AKHs are also widespread in other Ecdysozoa and Lophotrochozoa including many species [19]. Due to its vital functions associated with flight, muscle activity, locomotion and combating stress situations and its prevalence in different animal species, AKH is worth to investigate.

1.2.4. Adipokinetic Hormone Biosynthesis

AKHs are synthesized and stored mostly in neuroendocrine cells (so-called adipokinetic cells) of the glandular lobes of the *corpus cardiacum* (CC) which is a

neuroendocrine gland connected to brain and physiologically equivalent to the pituitary of mammals [39]. Energy consuming processes such as the initiation of locomotion activity or stress factor induce the release of bioactive AKH peptide, synthesized from precursor polypeptides as other neuropeptides by proteolytic cleavage, from CC cells to fat body [18]. Fat body is an organ that mobilizes substrates like carbohydrates and lipids to generate energy [18]. It can be considered as combined version of vertebrate liver and adipose tissue because it integrates many properties of them [18]. After released from CC cells, they require their corresponding G-protein coupled receptors to transmit their signals through the cell membrane into the cell due to the fact that peptide hormones are not small enough to pass through cell membrane.

1.2.5. Adipokinetic Hormone Receptors

Adipokinetic hormones bind to their cognate adipokinetic hormone receptors that belong to the Class A of G-protein-coupled receptors located on the membrane of fat body cells and activate them to provide energy mobilization from carbohydates and lipids stored in fat body by triggering a number of coordinated signal transduction processes. Hence, mobilized carbohydrate and lipid reserves are ultimately converted into trehalose and free fatty acids at the end and these are released into circulatory system (called hemolymph) to be transported to the contracting muscles that can be a flight muscle or leg muscle providing locomotion activity of the insect [39].

1.3. G- Protein- Coupled Receptors

1.3.1. Introduction

G-protein-coupled receptors (GPCRs) constitute the largest superfamily of eukaryotic transmembrane receptors that mediate many physiological responses to a variety of environmental stimulants ranging from neurotransmitters to peptides, hormones, and even light across plasma membrane [40].

At their simplest, GPCRs are characterized by seven transmembrane α-helices that are connected by three intracellular and extracellular loops and are arranged around a hydrophobic core [41, 42]. On the extracellular side, there are N-terminus and three extracellular loops that facilitate stimuli recognition and on the intracellular side, there are C-terminus and three intracellular loops that mediate binding of G-proteins to GPCR for signal transduction. GPCR activation typically begins with ligand recognition in an orthosteric site which is the primary binding site of a ligand on a receptor located in extracellular part of the membrane and proceeds with a conformational rearrangement to enable intracellular partners (*e.g.* G-proteins) for further downstream signaling in the cell via second messengers (*e.g.* cAMP) [40].

GPCR superfamily in vertebrates are divided into five main families based on their sequence and structural similarities: Rhodopsin-like (Class A), Secretin (Class B), Glutamate (Class C), Adhesion, and Frizzled/Smoothened families [41]. Rhodopsin family which possesses 290-951 amino acid length is the largest of all and all its members share conserved sequence motifs implying common structural features and activation mechanisms [40, 43]. This conserved topology enables a direct comparison of equivalent positions across different receptors via using the Ballesteros-Weinstein residue numbering system [44].

1.3.2. Conserved Properties of GPCRs in Family A

When the sequences of members of class A are compared, several conserved motifs in atomic details have been found. These include NPXXY motif in transmembrane helix 7, DRY motif in intracellular loop 2, CWXP motif in transmembrane helix 6 which are considered in association with the activation of class A GPCRs [43]. Other than these commonly known properties, there are other conservations distributed among transmembrane helices in residue level. These conserved residues are Gly and Asn in TM1, Leu and Asp in TM2, Cys and AspArgTyr in TM3, Trp and Pro in TM4, Pro and Tyr in TM5, Phe, Trp and Pro in TM6 and Asn, Pro and Tyr of NPXXY motif in TM7 [43].

Besides these conserved residues, there are conserved Cys residues in first or second extracellular loops and top of transmembrane helix 3 forming a disulfide bridge that enables proper receptor folding in in most of the GPCRs [42].

When crystal structure of fully active state of beta-2 adrenergic receptor with a bound agonist and G protein was obtained, it was found out that Class A members also share similarity in conformational changes upon GPCR activation leading to G-protein binding [45]. It also demonstrated that major conformational changes occur in the part of receptor that is close to intracellular region, whereas extracellular half of the receptor undergo only subtle changes in comparison with the remaining parts of the receptor [45].

1.3.3. Structural/ Functional Relationship of GPCRs

Due to the fact that a GPCR's function is to transduce signal to the interior of the cell in the presence of a specific ligand bound to extracellular part of the cell across cell membrane, it is significant to determine its 3D structure in order to find its most possible conformation. It is believed that there are two conformations of GPCRs as active and inactive that are kept in equilibrium regardless of the absence of any ligand. When a ligand binds to the receptor, equilibrium is shifted [46-48].

1.3.4. Bioinformatics in Structural Studies

With the advancement in GPCR structural biology, atomic-level details of several receptors in different conformations within this superfamily are provided via experimental methods such as X-ray crystallography.

Rhodopsin was the first crystalized GPCR providing first insights into the structure of GPCRs [49,50]. However, some troubles appeared when others GPCRs within Family A were intended to crystalize in order to get more information about structure/ function relationships [51]. One of the major challenges is that GPCRs show low expression in their native expressed tissues [40]. This problem can be compensated by recombinant expression method which enables expression of natively folded receptor using different cell types belonged to other species [40]. Another major challenge is their inherent

structural flexibility. Due to their disordered loop regions, stability of these regions varies significantly under different thermal conditions [52]. In order to overcome this thermal stability problem, ligands and receptors can be stabilized by introducing different combinations of mutations, adding lipids in both purification and crystallization processes and increasing salt concentration [52-57]. Other than stability problem, they behave vulnerable to proteolysis due to possessing disordered loop regions [40]. These regions are required to be stabilized by protein engineering, truncating, or binding to an antibody fragment methods [40]. For instance, third intracellular loop (ICL3) connecting TM5 and TM6 of GPCRs is considered the most flexible region. It can be either bound to an antibody fragment like Fab5 in human beta-2 adrenergic receptor by recognizing N and C terminal of ICL3 or replaced with a well-folded soluble protein like T4L lysozyme by protein engineering methods [53-55, 57, 58]. As a third option, this disordered ICL3 region can be truncated [53-55, 57]. Other than these two difficulties, the purification of stable GPCRs in large quantities seems to be very hard due to the lack of exposed polar surface outside of micelle [40]. Besides some structural modifications, solutions such as binding to an antibody fragment or the fusion of soluble protein with flexible regions as mentioned above provide crystal lattice contacts, too [51, 53, 54, 58]. Even though high-resolution crystal structures of active (both ligand and G-protein coupled) and inactive forms of GPCRs are essential to enlighten molecular mechanisms of protein function, these limitations of X-ray crystallography should be taken into consideration [51].

Because of the challenges in crystallization of GPCRs mentioned above, it seems unlikely to fill the gap in experimental GPCR structural biology field in short time period. Even though insect GPCRs play essential roles in various physiological processes, there is no known crystal structure of insect GPCR recorded in Protein Data Bank (PDB). As an effective alternative way, 3D structural data of unknown GPCRs including the ones belonged to insects can be met to some extent by utilizing computational methods such as homology modeling [59]. Homology modeling enables to study the structure and function of GPCRs by providing to construct possible 3D structure of a GPCR based on another GPCR template that possess known crystal structure and highly similar sequence with the target [60]. In the work presented in thesis, adipokinetic hormone receptor of stick insect was studied to define its 3D structure and significant residues in ligand binding pocket via bioinformatics.

2. PURPOSE

The purpose of this thesis project is to characterize a gonadotropin-like receptor called as adipokinetic hormone receptor in stick insect *Carausius morosus* and to find significant residues in its ligand binding pocket via bioinformatics.

3. MATERIALS AND METHODS

3.1. RNA-Sequencing of *C. morosus*

100bp paired-end RNA-sequencing of *C.morosus* from total adult tissue was performed by Genewiz Inc. (South Plainfeld, NJ, USA). Its assembly was performed using Trinity [61] and cDNA library was built. Then, when its analysis was done, adipokinetic hormone (CamAKH) and its corresponding receptor called adipokinetic hormone receptor (CamAKHR) were found out (unpublished data, B.Duan Sahbaz).

3.2. Homology Modeling and Protein Preparation

Increasing number of crystallized proteins with high resolutions enable to translate genetic information into structural information thereby facilitating characterization of proteins. However, there are still many proteins including membrane proteins like GPCRs with unknown 3D structures [60]. Due to difficulties in GPCR crystallization, computational approaches such as homology modeling is utilized to study the structures and functions of these proteins by constructing 3D models of them. The main idea behind the homology modeling approach relies on the fact that proteins with similar sequences tend to share a common evolutionary structure, hereby they are considered to have similar 3D structures [62]. It consists of four main sequential steps: (i) template selection, (ii) sequence and structure alignment with template and target protein, (iii) homology model building and (iv) model refinement [63, 64].

Firstly, the sequences of templates with known 3D structures are aligned with the target sequence and the ones with high similarity to target are chosen as template(s). In case of low similarity (<30%) between the target and template, multiple templates can be chosen to get a better model [65, 66]. Then, structures of the chosen templates are also aligned to check structural similarity. After alignment is performed, homology model is constructed. Various softwares or web-servers like Swiss-model perform these steps automatically as used in this work. Then, the built model is refined to avoid errors like

steric clashes and atomic overlap of residues; thus, leading to the conformation with the lowest free energy [67]. These steps were performed in the work presented in this thesis as follows. There is no known crystal structure of insect adipokinetic hormone receptor. Because AKHR belongs to Class A GPCR family, including beta2- adrenergic receptor and share the same highly conserved residues with this protein, the homology model of CamAKHR was constructed based on two beta-2 adrenergic receptor (PDB ID: 3sn6 and 4qkx) via Swiss-model server [68, 69]. It was challenging to construct a homology model based on one of the mentioned receptors because they lacked some parts of transmembrane helices and also had low sequence similarity with the target when tried (details are explained in Results). That's why, two beta-2 adrenergic receptors which are coded as 3sn6 and 4qkx in PDB were utilized to build a model. Finally, the constructed homology model of CamAKHR was minimized for the further steps using Protein Preparation module of Maestro molecular modeling suite [70]. Bond orders were assigned, hydrogen bonds were added, and unnecessary water molecules were deleted as the preprocess of refinement. Then, the orientation of water molecules and pKa predictions were implemented at the pH 7.0 using PROPKA [71]. Energy minimization was performed by OPLS2005 force field [72, 73].

3.3. Ligand Preparation

The 2D structure of the CamAKHR was manually sketched including post-translational modifications including C-mannosylation at 8th Trp residue, C-term amidation and N-term pyroglutamate structures. Protonation state of the ligand was assigned by Epik module applied in Schrodinger molecular modeling package and pH was set to pH (pH=7.00) [74, 75]. Structural optimizations of the ligand were performed using Protein Preparation module of Maestro. Then, the energy minimization was performed via Macromodel module of Maestro [76]. To obtain the lowest energy conformation of the ligand, a conformational search method was used, and all possible conformations were generated via Macromodel module.

3.4. Molecular Docking

Molecular docking mainly aims to computationally predict the association between biologically relevant proteins that plays a central role in signal transduction. If there is any association between molecules such as receptors and ligands they must somehow recognize each other before binding to one another. During molecular recognition, molecules must come to optimized conformations and relative orientations with respect to each other in order to provide signal transduction. Molecular docking methods provide ligand placement, conformational search and scoring algorithms to obtain the best ligand binding pose that means the preferred orientation and conformation of a ligand in receptor binding pocket during docking with the lowest energy [77].

In order to identify ligand binding pocket in beta-2 adrenergic receptor-based model of AKHR, molecular docking studies were performed on Glide module of Maestro [78-80]. Grid box was generated including the coordinates of active sites in the core region of the receptor.

Positional contraints were applied to four residues which are involved in ligand binding of the model corresponding to those residues in the template. These four residues are Asn 356, Arg 269, Tyr 424, Phe 449. All positions are grouped into one group and max distance was set to 5 Angstrom. Standard precision (SP) protocol of Maestro molecular modeling package was applied to the ligand conformations.

3.5. Molecular Mechanics Generalized Born Surface Area (MM-GBSA) Calculations

MM-GBSA method is used to calculate ligand binding free energy in complex with biological macromolecules. It also enables to check whether energy of the ligand in docking studies will be preserved during simulations and gives more reliable results compared to docking score that calculates free energy in a different way. MM-GBSA module of Prime was used to calculate for the pose with the top- docking score [801. OPLS2005 force field and VSGB solvation model was used for calculations [82]. Flexible residue distances from ligand are defined as 3Å.

3.6. Molecular Dynamics (MD) Simulations

Chemical and biological processes happen in progress of time in nature; that is why, they should be observed dynamically in their native-like environment to develop an idea about how they behave in nature. Molecular dynamics is a computational method that simulates the behavior of complex systems in time including the macromolecules embedded in membrane. Hence, it provides real-like results compared to static prediction methods as mentioned above such as homology modeling and molecular docking [83].

The initial protein-ligand complex structure generated in docking is used for MD simulations. Simulations were performed via Desmond module of Schrodinger [83]. CamAKHR homology model with its bound ligand pose were embedded into a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane bilayer and the system was solvated in an orthothrombic box with layers of explicit TIP3P water molecules, and 0.15M NaCl was added (as shown in Figure 3.1.). Receptor orientation in lipid bilayer was specified according to the orientation of PDB ID: 3sn6 template that is found in the orientations of proteins in membranes (OPM) database [84, 85]. The simulations were performed in NPT ensemble, Nose-Hoover thermostat and Martyna-Tobias-Klein barostat protocols were utilized to provide 310 K temperature and 1.01325 bar pressure of the systems [86, 87]. Time step of simulation was assigned to 2.0 fs. Proteins and water molecules were relaxed using stepwise methods. Finally, 100 ns of production simulations were carried out for each system.

3.7. Conservation of Important Amino Acids in Insects

In order to understand the conservation pattern of important amino acids which are important for ligand binding, AKHR sequences of 6 different insects (*Carausius morosus, Bombyx mori, Drosophila melanogaster, Zootermopsis nevadensis, Apis mellifera, Tribolium castaneum*) species were aligned via Clustal Omega [88].

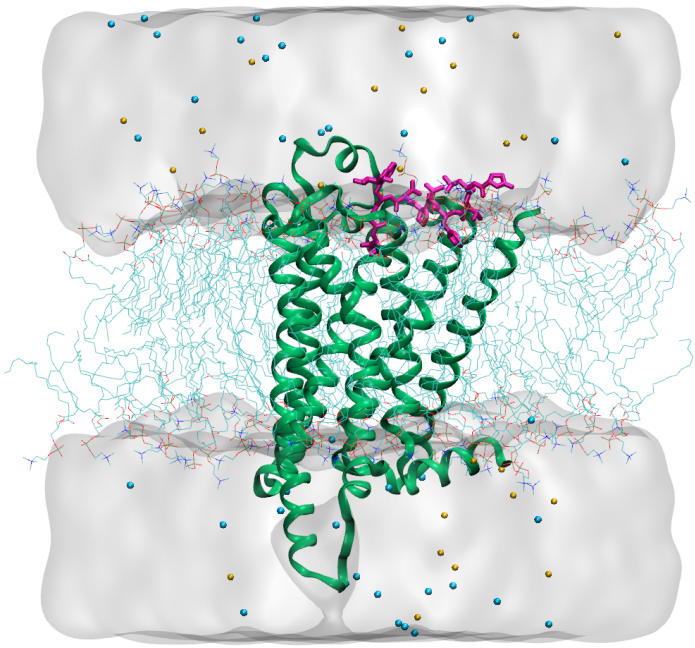


Figure 3.1. CamAKHR- CamAKH complex in native-like cell environment

3.8. Search of Known Insecticides

Known insecticides were searched in PubChem Compounds and their structures were exported [89]. Then, Protein Preparation Wizard module of Schrodinger was used to prepare these molecules for further steps. After constraint docking applied as in previous steps, MM-GBSA of obtained insecticide with the top-docking score was calculated and its MD simulation for 100ns was conducted. Then, MM-GBSA of the last 50ns of simulation was calculated again to compare with the one of CamAKH.

3.9. Mutation Validation Studies

Each of four residues (Asn356, Arg269, Tyr424 and Phe449) which were used as constraints in docking studies were mutated to Alanine one by one. Then, docking and MM-GBSA calculations were performed again for each of them.

4. RESULTS

4.1. Homology Modeling and Protein Preparation

When target sequence (CamAKHR) was given to Swiss-model web server, it gave 50 possible templates with different coverage and identity percentage. When these templates were ranked in terms of coverage, beta-2 adrenergic receptors with PDB IDs 4qkx and 3sn6 were chosen as the best templates ignoring the chimera proteins in the list. Then, initial models were built based on these two templates and model quality validation was obtained via Swiss-model software.

Model quality evaluation of Swiss-model software depends on two estimations named as GMQE and QMEAN. GMQE (Global Model Quality Estimation) is a quality estimation method which demonstrates the expected accuracy of a model built with that alignment, the template and the coverage of the target. Ultimate GMQE scores are expressed as a number in the range of 0-1 and higher numbers indicate higher reliability of model quality.

QMEAN [90] stands for Quantitative Model Energy Analysis. It is a quality estimation method which gives global and local absolute quality estimates of a model. QMEAN Z-score is an estimation of the “degree of nativeness” of structural features in the model on a global scale. It shows whether the QMEAN score of model is comparable to the QMEAN score of experimental structures of similar size. Whereas positive numbers mean that model scores higher than the experimental one on average, negative numbers mean the opposite. Scores of -4.0 or below indicates low-quality model and also shown as thumbs-down symbol.

Besides global QMEAN scoring, there is QMEAN local scoring providing “Local Quality” plot. This plot demonstrates the expected similarity of each residue of model (x-

axis) to the native structure (y-axis). Residues with a score lower than 0.6 are considered as low quality, whereas the peaks indicate higher quality [91].

Furthermore, “Comparison plot” compares the model quality score of constructed model with the scores belong to experimental structures with similar sizes. Every dot indicates an individual experimental protein structure. The darkest black dots represent experimental structures with a normalized QMEAN score within 1 standard deviation of the mean that means absolute value of Z-score is between 0 and 1. Experimental structures with an absolute Z score between 1 and 2 are shown in grey. Other experimental structures that are even further from the mean are light grey. Red star represents the actual model.

The initial model based on 3sn6 beta-adrenergic receptor lacks some part of TM helix7 and all the intracellular short helix8 and has 27.39% identity with the target sequence. For its quality validation, Swiss-model provided Ramachandran plot, global and local quality estimations as structure assessment for this initial model.

In Figure 4.2., residues in the initial model corresponding to the scores below 0.6 indicates low quality. Scores of seven peaks shown by arrows are higher than 0.6 indicating that residues corresponding to these peaks have high quality.

In Figure 4.3., CamAKHR initial model represented as a red star has a normalized QMEAN score higher than 0.5. Its absolute Z score is bigger than 2 indicating the low-quality model.

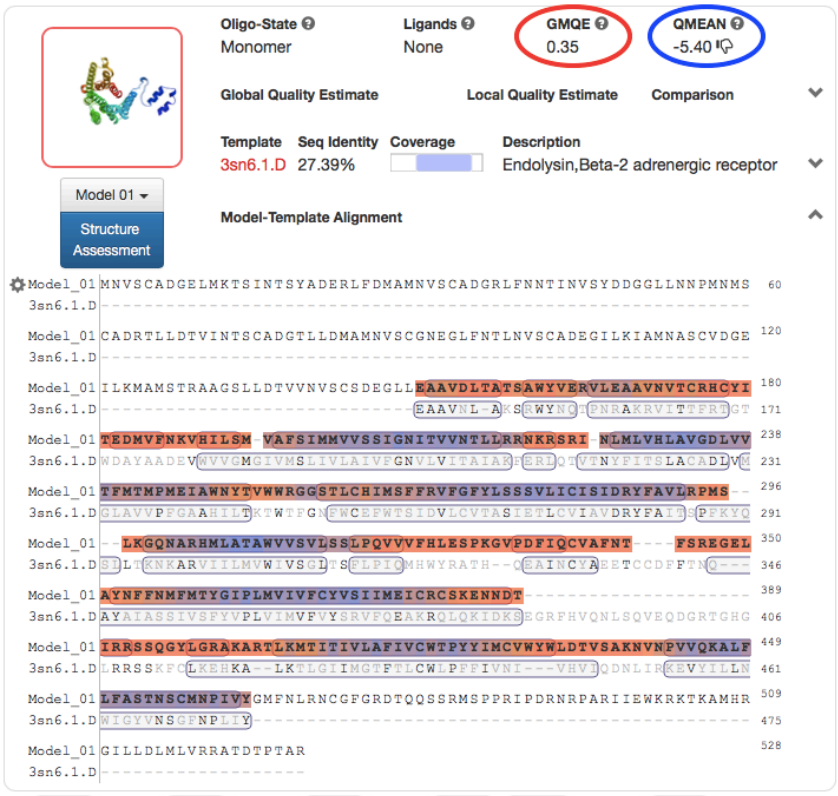


Figure 4.1. Quality validation features of initial CamAKHR model built. GMQE score is shown in red circle and QMEAN score is shown in blue circle.

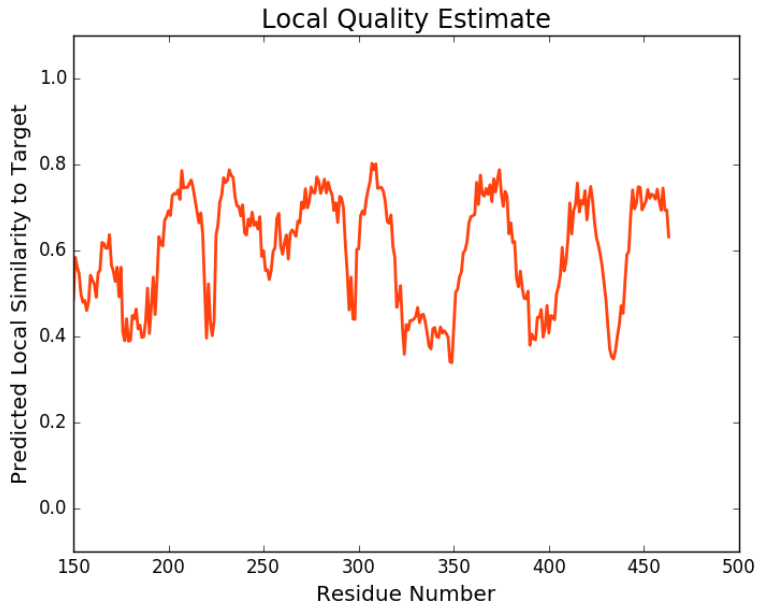


Figure 4.2. Local quality estimate plot of initial CamAKHR model

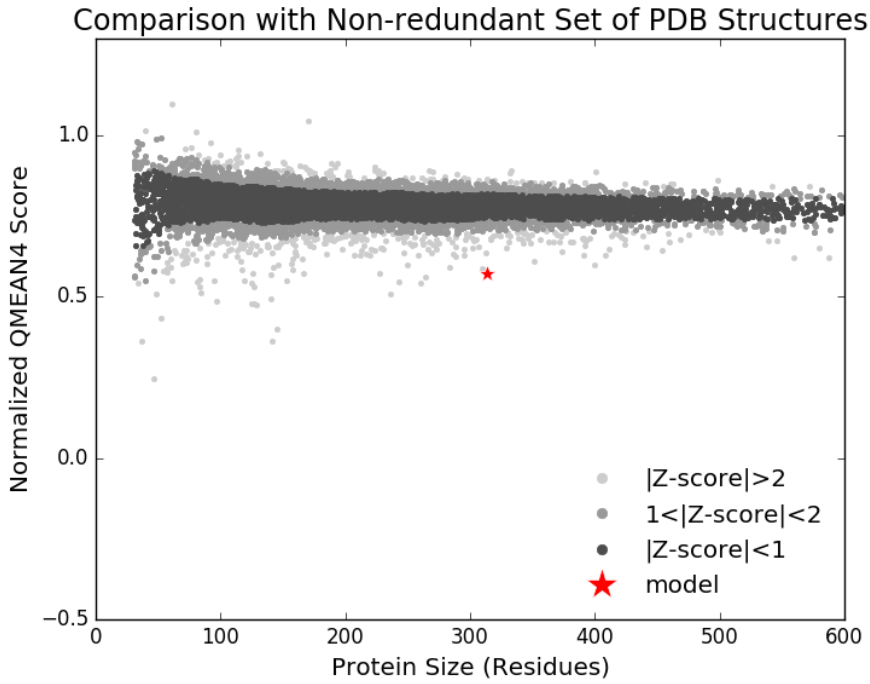


Figure 4.3. Comparison plot of initial CamAKHR model

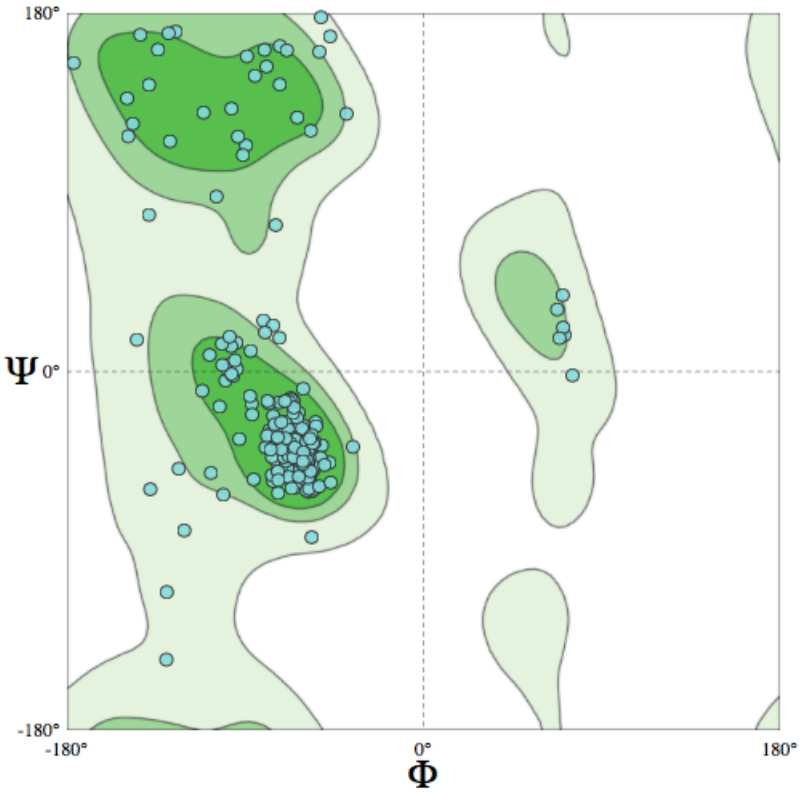


Figure 4.4. Ramachandran plot of initial CamAKHR model

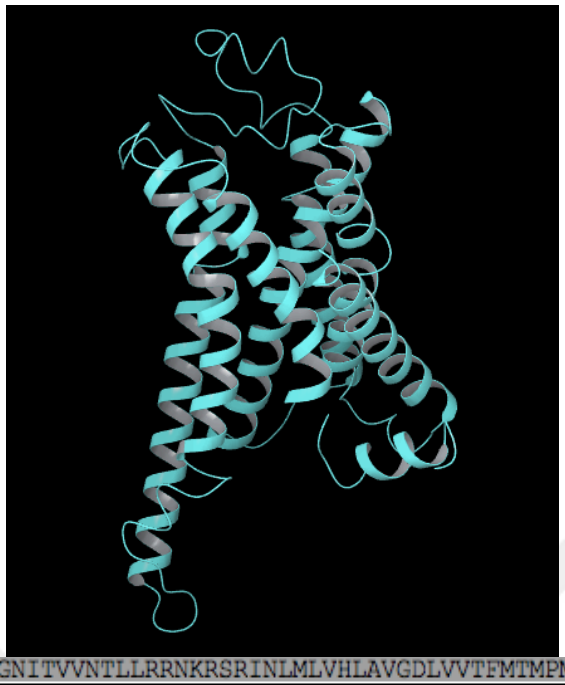
Ramachandran plot enables to visualize energetically favored regions for backbone dihedral angles (Ψ) against of amino acid residues (Φ) in a protein. For this, Φ (Phi; N-Cα) and Ψ (Psi; Cα-C) angles are utilized. The number of Φ/ Ψ pairs determine the border (contour) lines for all amino acids except Gly, Pro, pre-Pro. The darkest green area shows the most favored region in which atoms do not come closer than the sum of their Van der Waals radii. The medium green area is called allowed area in which atoms can get a little bit closer to each other. Other areas are considered as disallowed area in which atoms get closer much more than the sum of their Van der Waals radii and clashes are possible. Ramachandran plot also provides secondary structure properties of a protein such as helices and sheets as shown in Figure 4.4. [92, 93].

Table 4.1. MolProbity results of the initial CamAKHR model based on 3sn6 template [94, 95]

Parameters	Scores	Residues
MolProbity Score	2.33	
Clash Score	16.21	(D267 PHE-D321 VAL), (D230 HIS-D230 HIS), (D273 PHE-D420 TRP), (D239 THR-D243 MET), (D350 LEU-D432 LEU)
Ramachandran Favoured	91.67%	
Ramachandran Outliers	2.24%	D435 VAL, D224 ILE, D253 VAL, D402 LYS, D223 ARG, D428 VAL, D189 VAL
Rotamer Outliers	1.42%	D274 TYR, D323 VAL, D230 HIS, D457 CYS
C-Beta Deviations	5	D432 LEU, D428 VAL, D204 SER, D326 LEU, D189 VAL
Bad Bonds	2 / 2572	D432 LEU, D428 VAL
Bad Angles	52 / 3499	D208 ASN, (D346 ARG-D347 GLU), (D318 LEU-D319 PRO), (D328 SER-D329 PRO), D327 GLU, (D243 MET-D244 PRO), D281 ILE, D239 THR, D346 ARG, D273 PHE, (D459 ASN-D460 PRO), (D293 ARG-D294 PRO), D431 TRP, D440 VAL, D263 HIS, (D421 THR-D422 PRO), (D332 VAL-D333 PRO), D326 LEU, D183 ASP, D325 HIS, (D363 ILE-D364 PRO), D436 SER
Cis Non-Proline	1 / 304	(D348 GLY-D349 GLU)

In Table 4.1., MolProbity score of CamAKHR initial model was observed as 2.33. Clashes between Phe267-Val321, His230-His230, Phe273-Trp420, Thr239-Met243 and Leu350-Leu432 were observed with 16.21 clash score. 91.67% of the atoms in the model were found in the most favored region in Ramachandran plot (Figure 4.4.), whereas 2.24% of them were found in the disallowed region as outliers. These outliers were Val435, Ile224, ValL253, Lys402, Arg223, Val428 and Val189. Rotamer outliers can be considered as outliers which show differences with the rotamers (different orientations of side chains) recorded in known X-ray crystals. Rotamer outliers are found in 1.42% of atoms as Tyr274, Val323, His230 and Cys457. Five C-β deviations are found in Leu432, Val428, Ser204, Leu326 and Val189. Two bad bonds were observed in Leu432 and Val428 out of 2572 bonds. 52 bad angles were found out of 3499 as Asn208, (Arg346-Glu347), (Leu318-Pro319), (Ser328-Pro329), Glu327, (Met243-Pro244), Ile281, Thr239, Arg346, Phe 273, (Asn459-Pro460), (Arg293-Pro294), Trp431, Val440, His263, (Thr421-Pro422), (Val332-Pro333), Leu326, Asp183, His325, (Ile363-Pro364), Ser436, (Asn225-Leu226), Glu182, Ser204, Phe240, (Asn441-Pro442), His304, (Glu182-Asp183), His230, Val229, Trp310, Val185, Trp429, Asp433, Tyr430, Asn387, His177, Val435, (Ile224-Asn225), Thr421, Asn342, Phe335, Asn459, Val321, (Leu432-Asp433). There was one cis non-proline in Gly348-Glu349 observed.

The other model based on 4qkx has a long N-term with helices and 23.97% identical to target sequence. Because both templates have lower similarity than 30%, both templates were chosen to build a model with a better quality. Lacking parts of the constructed model based on 3sn6 was completed manually by adding a part of TM7 and all the intracellular short helix8 of the model based on 4qkx template and renumbered in Maestro. After renumbering, model was refined to get better quality model removing clashes and bad contacts. At the end, 293 amino acid-length model with 2350 heavy atoms out of 4759 and +13 charge was obtained as in Figure 4.5.



NKVHILSMVAFSIMMVVSSIGNITVVNTLLRRNKRSRINLMLVHLAVGDLVVTFMTMPMEIAWNYTVWWRGGST
LCHIMSFFRVFGFYLSSSVLICISIDRYFAVLRPMSLKGQNARHMLATAWVVSVLSSLPQVVVFHLESPKGVPD
FIQCVAFNTFSREGELAYNFFNMFMTYGIPLMVIVFCYVSIIMEICRCSKENNDTIRRSSQGYLGRAKARTLKM
TITIVLAFIVCWTPYYIMCVWYWLDTVSAKNVNPVVQKALFLFASTNSCMNPIVYCRSPDFRIAFQELLCL

Figure 4.5. Constructed model of CamAKHR with its sequence

When 3sn6 template (after deletion of unnecessary parts) and the constructed model structures are aligned, RMSD value was found as 1.151 Å (In Figure 4.6. and Figure 4.7.).

Figure 4.6. Protein structure alignment of CamAKHR model and 3sn6 template

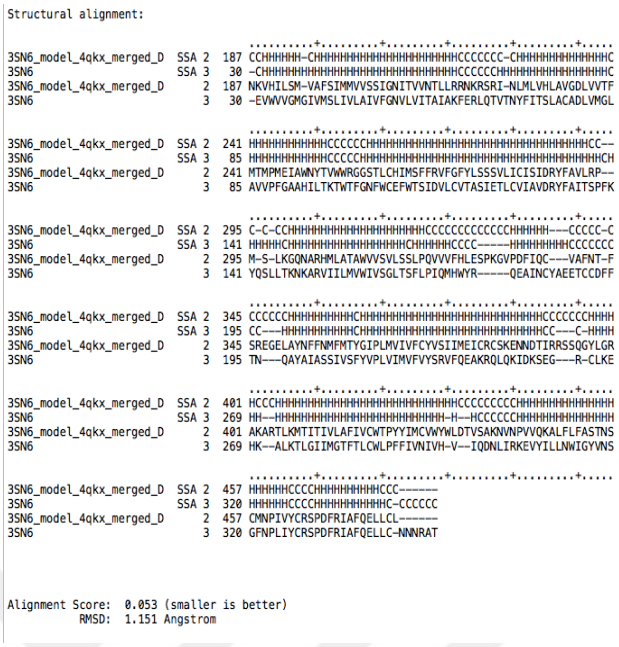


Figure 4.7. Multiple sequence alignment of CamAKHR model and 3sn6 template

4.2. Ligand Preparation

After common modifications of AKH/RCHP family were applied to CamAKH peptide as explained in methods part (Figure 4.8 and Figure 4.9.), its energy was minimized (to value -1640.133 kcal/mol) with OPLS2005 force field via Macromodel module of Maestro. After energy minimization, 10 amino acid-length modified ligand was obtained.

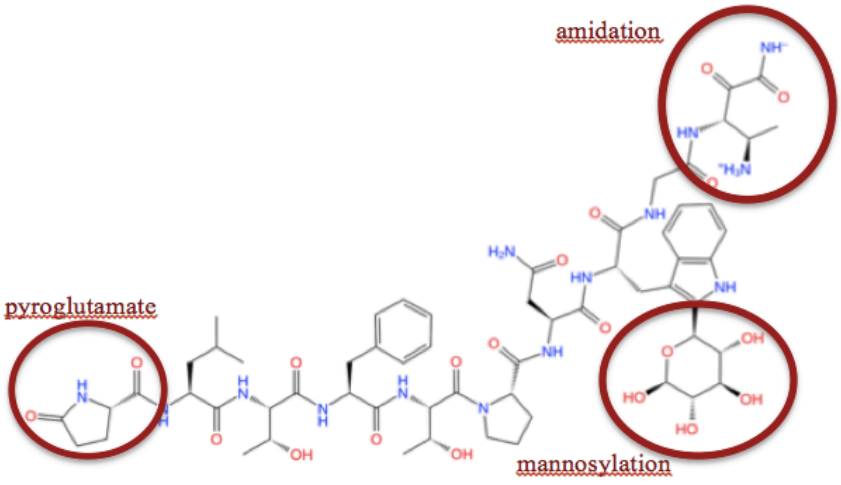


Figure 4.8. Ligand 2D structure with modified residues

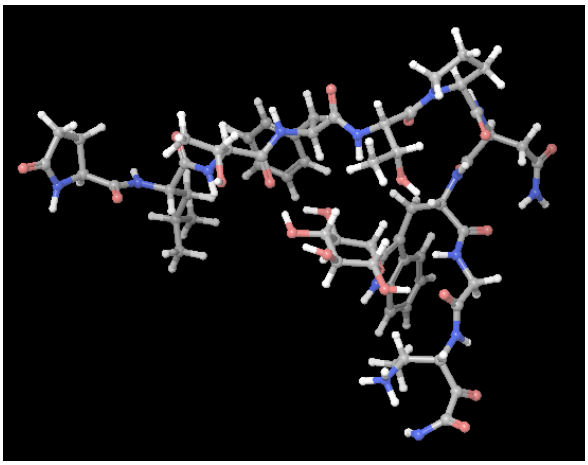


Figure 4.9. Ligand 3D structure representation

In order to obtain the lowest energy conformation of the ligand, conformational search was performed and 101 conformers were produced.

4.3. Molecular Docking and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) Calculations

Grid box was intended to be produced depending on the coordinates of active sites in the core region of the receptor. The grid coordinates were specified as 6.35, -20.93, -57.46 in x, y, and z axes, respectively.

After 101 conformers of ligand were generated, positional constraints were applied to four residues as mentioned in methods part and docking was performed again with them. Only two poses out of 101 (one pose for each conformer) were found to be docked. They were very similar to each other especially in the parts buried in transmembrane core region, next steps were performed with the pose having the top-docking score.

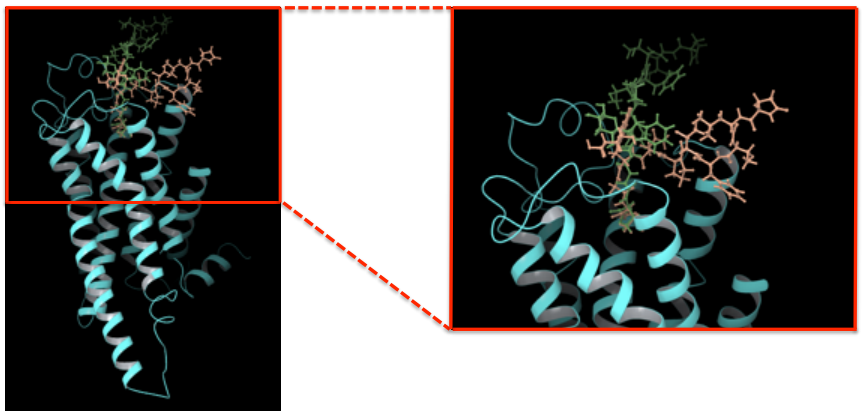


Figure 4.10. Representative binding of two poses out of 101. Orange colored ligand represents the pose with top-docking score and the green one represents the other.

MM-GBSA that gives more accurate energy values was calculated for the pose that has the top-docking score. MM-GBSA for this pose was calculated as -38.179 kcal/mol which can be evaluated as moderate, it was chosen to be used in the following simulation.

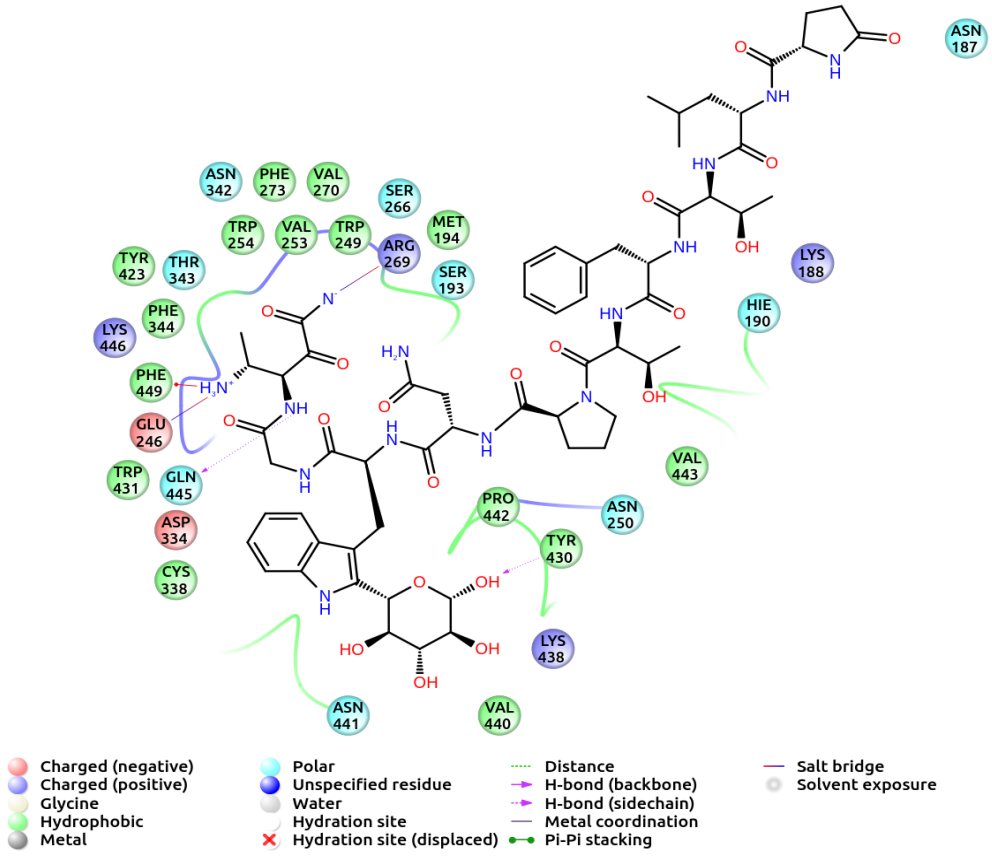


Figure 4.11. 2D protein-ligand interaction diagram of CamAKH within CamAKHR

In Figure 4.11., 2D protein-ligand interaction diagram of CamAKH (with top-docking score) within CamAKHR model binding pocket was shown. Arg269 and Glu246 forms salt bridges with the last residue of the ligand. Gln445 forms hydrogen bond and Phe449 stacks with the same residue as well. Tyr430 establishes hydrogen bond with mannose group of tryptophan residue of the ligand.

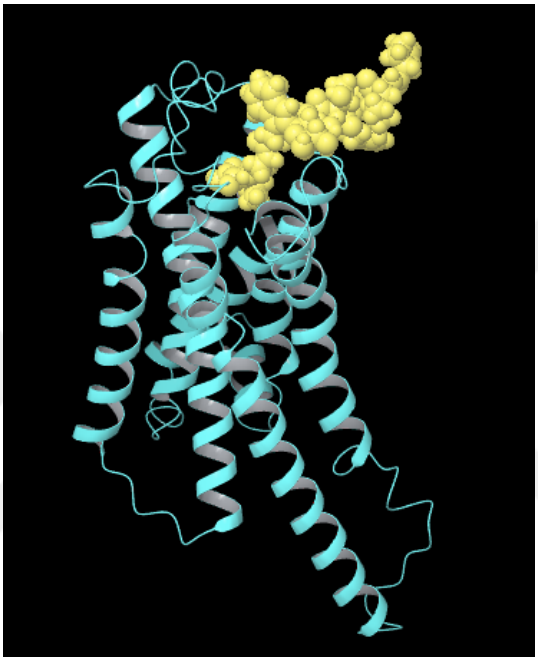


Figure 4.12. Representative binding of CamAKH within CamAKHR model

In Figure 4.12., the ligand within the binding cavity of the CamAKHR model was shown. After docking step, this pose of CamAKH was used for MD simulations.

4.4. Molecular Dynamics

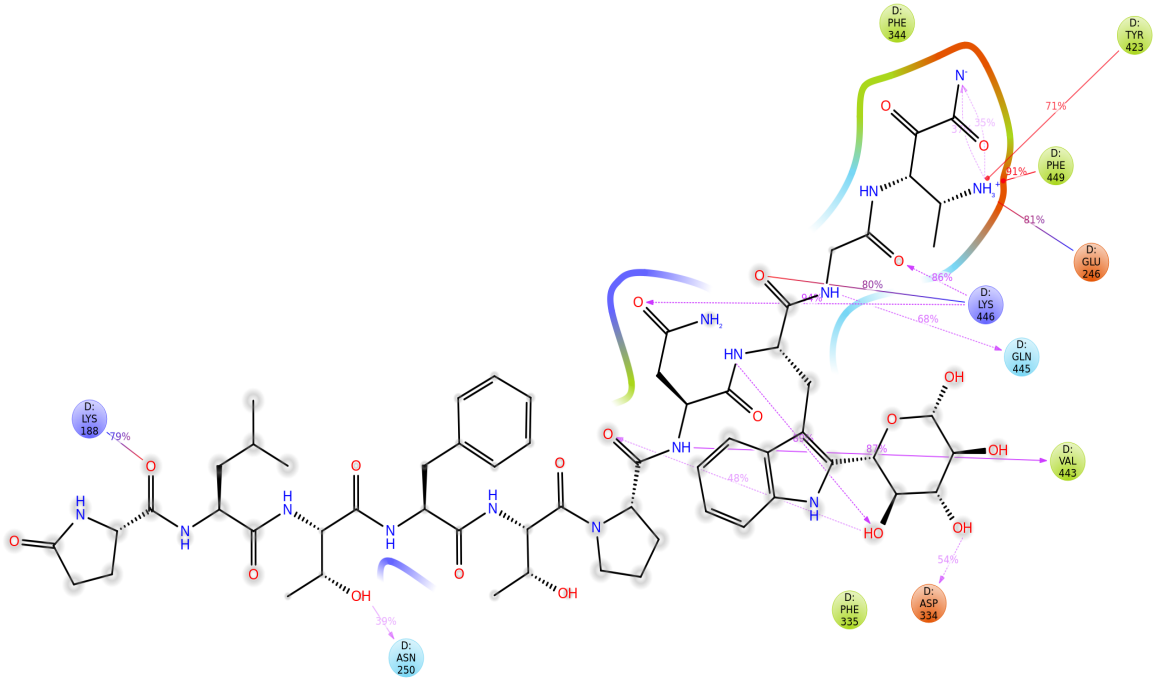


Figure 4.13. 2D protein-ligand interaction diagram of CamAKH within CamAKHR model binding pocket during MD simulations

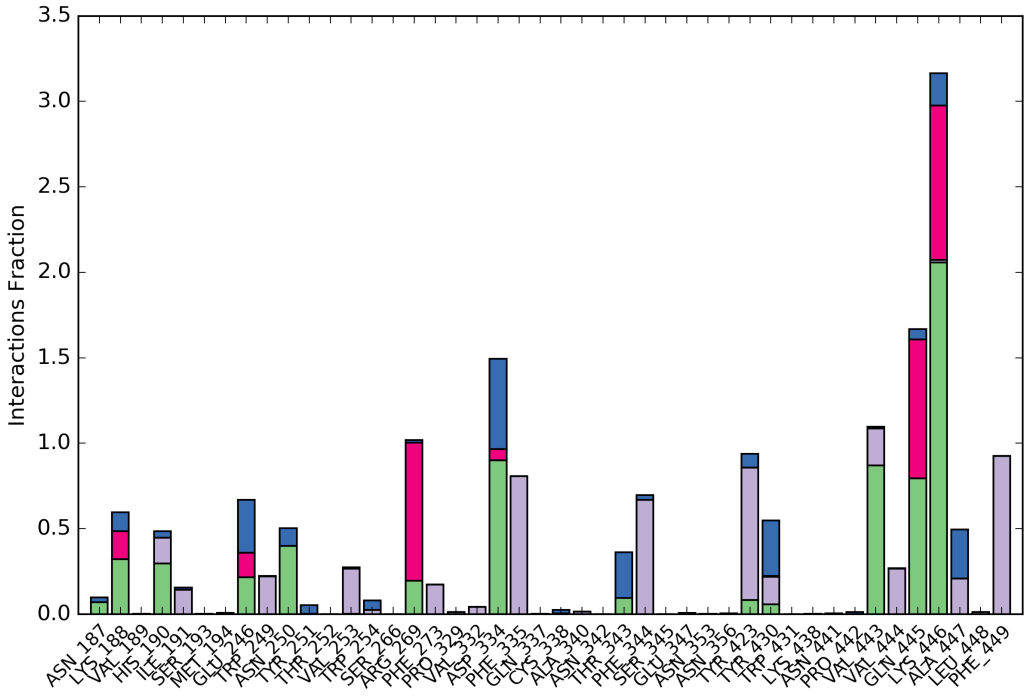


Figure 4.14. The occupation levels of important amino acids in close proximity to the CamAKH ligand during MD simulations

In Figure 4.13. and Figure 4.14., protein- ligand interaction of CamAKH within CamAKHR model binding pocket in which the crucial amino acid residues with their occupation levels was analyzed.

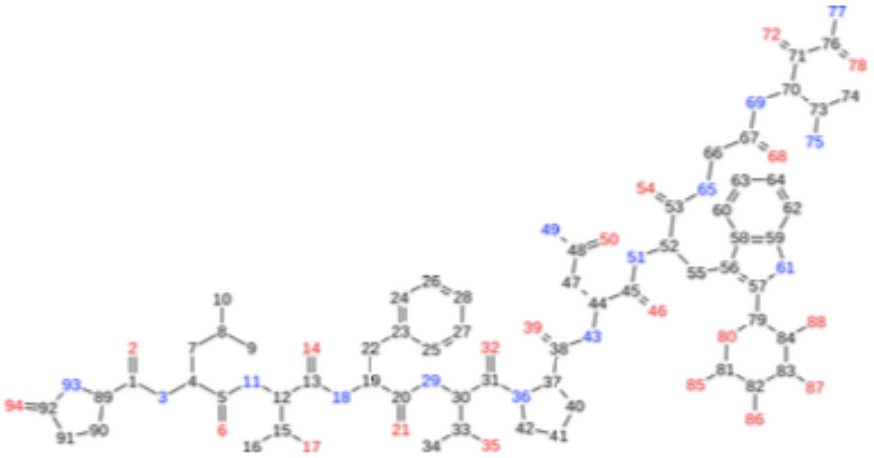
Lys446 in the active site of CamAKHR model was highly pronounced via the formation of electrostatic interactions including salt bridges (80%) and hydrogen (86%, 94%) bonds. Gln445 (68%) and Asp334 (54%) also included the same interactions with the ligand. Arg269 also demonstrated high level of ionic bonds with the ligand. Val443 (87%) made hydrogen bonds using its backbone. Using their side chains, Asn250 (39%) and Gln445 formed hydrogen bonds with the ligand. Lys188 (79%) and Glu246 (81%) showed salt bridges with the first and the last residue of the ligand, respectively. Tyr423 (71%), Phe335, Phe344 and Phe449 (91%) established hydrophobic interactions. Also, Phe449 stacking with the last residue of ligand showed the highest occupancy level along the simulations and this largest interaction originating from Phe449 is hydrophobic connection with NH_3^+.

Ligand Root Mean Square Fluctuation (L-RMSF) is used to understand local changes in ligand atoms; thus, it gives an idea about the interaction of ligand with receptor and their entropic role in binding.

RMSF for atom i is calculated as following formula:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r'_i(t)) - r_i(t_{ref}))^2}$$

In this formula, t is trajectory time for RMSF calculation, t_{ref} is the reference time (commonly first frame t=0 used), r is atom i position in t_{ref} time and r' is the position of atom i after aligned on the reference frame.



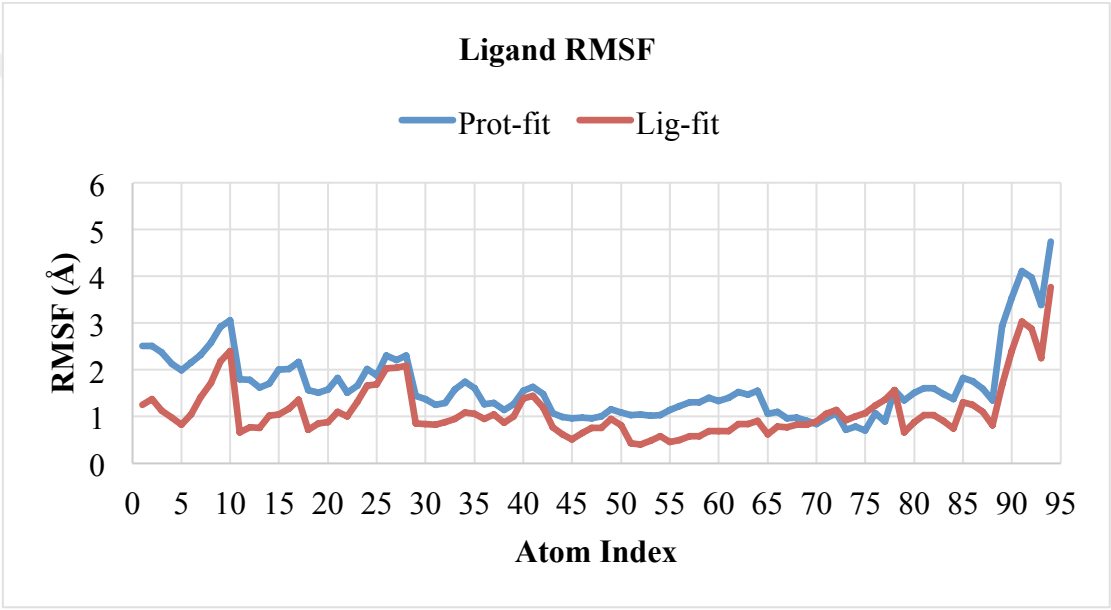


Figure 4.15. L-RMSF of CamAKH Ligand

"Fit on Protein" (abbreviated as Prot-fit) indicates the fluctuations of ligand with respect to receptor in the protein-ligand complex. It is measured firstly by aligning CamAKHR-CamAKH complex on CamAKHR backbone and then by calculating CamAKH RMSF on heavy atoms of the CamAKH.

"Fit on Ligand" (abbreviated as Lig-fit) indicates the fluctuations of atoms in the ligand regardless of the protein.

According to ligand RMSF in Figure 4.15, there was not any significant difference between Prot-fit and Lig-fit curves during MD. There were not so much fluctuations in the

curves, either. The most flexible part of the ligand was found in atoms of N-terminus pyroglutamate residue with the highest RMSF values.

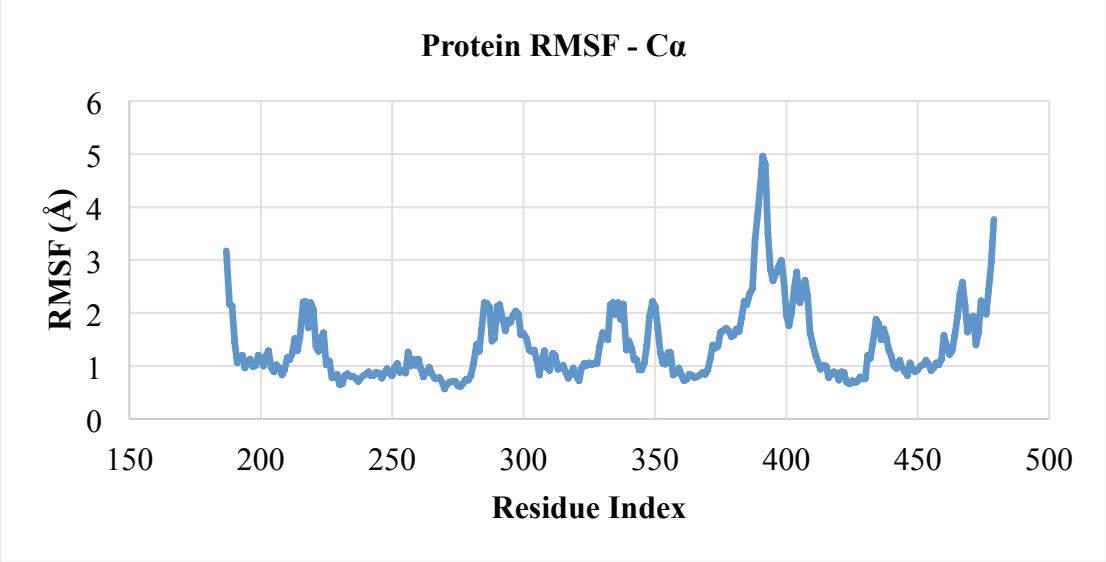


Figure 4.16. RMSF of CamAKHR model

Root Mean Square Fluctuation (RMSF) is used to understand local changes in protein residues. RMSF for residue i is calculated as following formula:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r'_i(t)) - r_i(t_{ref}))^2 >}$$

In this formula, t is trajectory time for RMSF calculation, t_{ref} is reference time, r_i is residue position and r' is the position of atoms after aligned on the reference.

Peaks in the RMSF plot indicate the most flexible regions of protein throughout the simulation. Terminal and loops demonstrated more flexible structure corresponding to peaks, whereas transmembrane alpha helices and beta strands show more rigid structures. Seven transmembrane helices with the lowest RMSF values were observed as in Figure

4.16. RMSF values of crucial residues which are constrained in docking and of transmembrane helices were shown in Table 4.2. below.

Table 4.2. RMSF values corresponding to transmembrane helices and constrained residues in CamAKHR model

Residues	**RMSF values (Å)**
Asn356	1.256
Arg269	0.697
Tyr424	0.664
Phe449	0.883
TM1 (190-217)	1.203
TM2 (226-252)	0.835
TM3 (259-292)	1.121
TM4 (299-321)	1.094
TM5 (352-388)	1.363
TM6 (406-431)	1.092
TM7 (444-462)	1.046

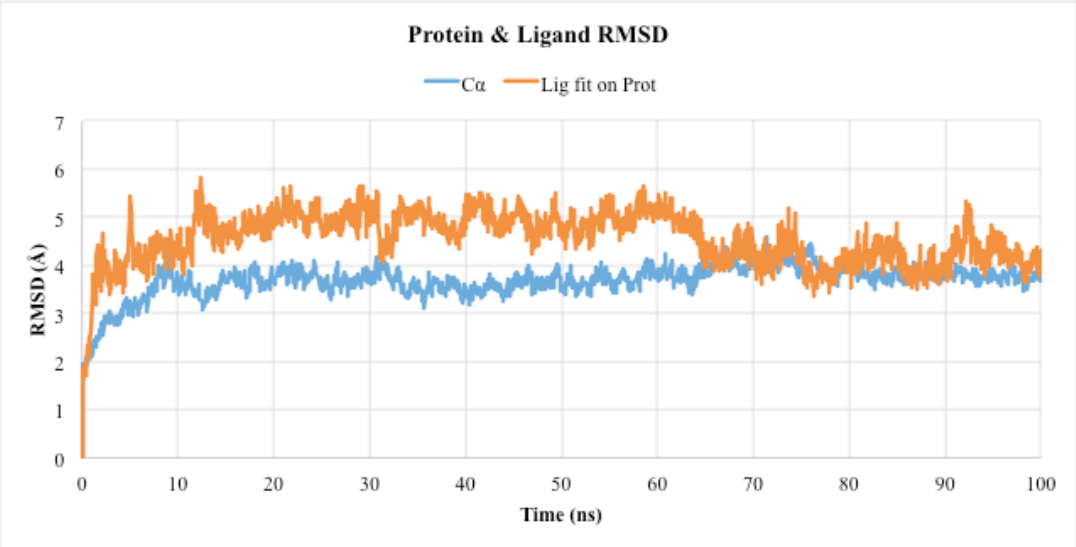


Figure 4.17. RMSD-time graph of CamAKHR model complexed with CamAKH

Root Mean Square Deviation (RMSD) calculation is used to measure the average change in displacement of a selection of atoms which belongs to a specific frame with respect to a reference frame. This calculation is performed for every frame in simulation trajectory. RMSD for frame x is calculated as follows:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r'_i(t_x)) - r_i(t_{ref}))^2}$$

In this formula, N is the number of atoms in selected atoms, t_{ref} is reference time (commonly first frame t=0 used), r' is the position of the selected atoms in frame x after aligned on the reference frame and t_x is time at where frame x is recorded.

Protein RMSD calculation gives insights into the structural conformation of the receptor throughout the simulation and indicates whether the receptor reached equilibration.

Ligand RMSD gives an idea about the stability of the ligand with respect to the protein and its binding pocket. “Lig Fit Prot” demonstrates the RMSD of a ligand and measured firstly by aligning the protein-ligand complex on the protein backbone of the reference and then calculating RMSD of the heavy atoms of the ligand.

According to these RMSD values shown in Figure 4.17, the system has equilibrated because the simulation converges meaning protein RMSD value stabilizes around a fixed value toward the end of the simulation. When the average for RMSD values of the protein and ligand were calculated for each frame, they were found as 3.68 Å and 4.55 Å, respectively. Also, significant fluctuations in both ligand and protein were not observed.

4.5. Conservation of Important Amino Acids in Insects

Among thirteen important amino acids, six of them were found as the most conserved residues in six different insect species. These residues are Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 as can be seen in Figure 4.18.

Figure 4.18. Multiple sequence alignments of AKHR in six different insect species

4.6. Search of Known Insecticides

351 results were found in known insecticides in PubChem Compounds. After constrained docking, MM-GBSA of 104753 coded compound (called neosaxitoxin) with the top-docking score was calculated as -36.137 kcal/mol. Its MM-GBSA was calculated again for the last 50ns of MD simulation as -30.03 kcal/mol in average, whereas MM-GBSA of peptide for the last 50ns of MD was calculated as -71.73 kcal/mol as in Figure 4.19.

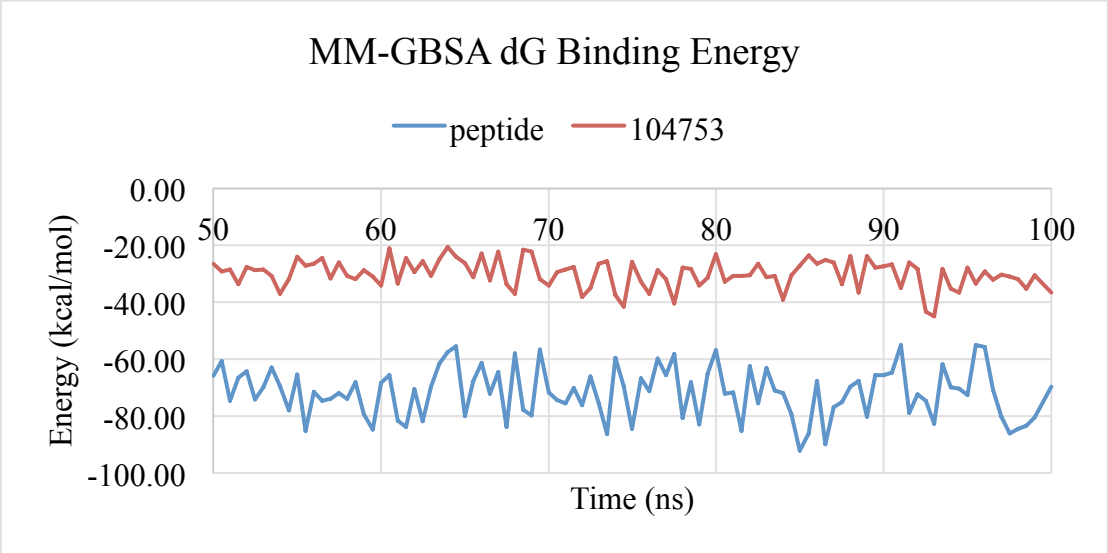


Figure 4.19. MM-GBSA dG binding energy graph of peptide and 104753 during last 50ns of MD simulation

4.7. Mutation Validation Studies

In silico mutations in crucial residues were generated to verify the importance of these residues. When Asn356, Arg269 and Tyr424 were mutated to Ala one by one, no poses were obtained in constrained docking which was applied with the same procedure in previous steps.

When Phe449 was mutated to Ala (in F449A mutant), there were many poses obtained. MM-GBSA of the pose with the top-docking score was calculated as -3.507 kcal/mol.

5. DISCUSSION

The purpose of this thesis was to characterize a gonadotropin-like receptor called as adipokinetic hormone receptor in stick insect *Carausius morosus* and to find significant residues in its ligand binding pocket bioinformatically. In order to achieve this goal, the sequence of CamAKHR was obtained from transcriptome data of this insect and its 3D structure was built via homology modeling based on two beta-adrenergic templates. After model quality validation and model refinement steps, docking was performed with its native CamAKH ligand. Thus, its ligand binding cavity with important residues was identified. These were verified via MD simulations and the stability of the system in native-like cell environment was observed.

The most important step in structural studies can be considered building the most proper homology model by choosing the right template. To ensure this, model quality validation methods are used. In this study presented in this thesis, Swiss-model software provided a detailed model report including Ramachandran plot, global and local quality estimations and comparison plot with non-redundant set of PDB structures. These were provided for the initial model built based on only 3sn6 beta-adrenergic receptor template with 27.39% sequence identity.

GMQE (Global Model Quality Estimation) score which is a number between 0 and 1 gives the expected accuracy of a model combining the template, the alignment and the coverage of the target. If the score is closer to 1, the model has more reliable quality. For the initial CamAKHR model, this score is 0.35 very close to zero; therefore, this initial model has low-quality.

QMEAN (Qualitative Model Energy Analysis) that give an estimation of the “degree of nativeness” of structural features in this initial model on a global scale by comparing QMEAN score of this initial model with QMEAN score of experimental structures of similar protein size. In Figure 4.1., QMEAN was -5.40 in this initial model. Because values of -4.0 or below mean low-quality model, it can be considered as low-quality model.

Besides global quality estimation, QMEAN scores also provide local estimations in "Local Quality" plot. In this plot, scores of all residues in the model are predicted based on their similarities to the native structure. Whereas the residues corresponding to the peaks indicate highly similar residues to the template, the residues with a score lower than 0.6 indicate poorly similar residues to the template. Thus, it enables to understand which residues should be carefully considered due to not possessing high similarity to the target. In Figure 4.2., when highly similar residues were counted, seven peaks shown by blue arrows were found. Because residues comprising these seven peaks correspond to the residues in transmembrane region of the initial model, residues of seven transmembrane helices have high-quality compared to the rest of the model. Among these residues, especially the residues in the middle parts of the transmembrane helices have higher quality compared to the rest. This result was expected because it is known that one of the main characteristics of GPCRs is that they are consisted of seven transmembrane helices. N-terminus, C-terminus, extracellular and intracellular loops are weakly resolved structures even via X-ray crystallography due to several reasons mentioned previously in Introduction. These may be the reasons why seven peaks corresponding to transmembrane helices show high similarity to the target while other parts of the model show lower similarity.

In "Comparison plot", the score of this initial model is compared with the scores of experimental structures with the similar protein size. If the normalized QMEAN score of the model shows a high standard deviation from the mean, absolute Z score of the model increases indicating low-quality model as in this initial CamAKHR model. In Figure 4.3., the initial model is shown by a red star in the lightest grey area of the plot and has a normalized QMEAN score higher than 0.5 which is good; however, its absolute Z value is higher than 2 which indicates low-quality.

Other than these estimations and comparisons with the known crystals of similar sizes, "Ramachandran plot" which provides energetically allowed regions for amino acid residues of a protein can be considered as a model quality validation method. According to Table 4.1 and Figure 4.4., most of the residues (91.67%) were located in favoured regions in the darkest green and medium green regions, while a small part of residues (2.24%) were located in the disallowed regions in the plot. The latter is called as outlier residues. If

the model has a high-quality, there should be no outliers because clashes should not be present in a high-quality model. In ideal case, around 98% of residues should be in Ramachandran favoured regions to be a good model; therefore, initial CamAKHR model cannot be evaluated as high-quality model.

This plot also gives an insight on the secondary structure of the model. Allowed region in the left-upper, left-bottom and middle, and right-upper parts of the plot indicates β-sheets, righthanded α-helices and lefthanded α-helices.

Furthermore, this initial model has a high clash score (16.21-ideal case:0), rotamer outliers (1.42%- ideal case: <1%), some C-β deviations (5- ideal case:0), bad bonds (2/2572-ideal case:0), a lot of bad angles (52/3499- ideal case:0) and cis non-Proline (1/304, ideal case:0) as shown in Table 5.1.

When the low sequence similarity with the template, bad quality validation scores and plots were evaluated, it was clear that this initial model was a low-quality model. Furthermore, it lacked the half of TM7 and all of intracellular helix 8. These are the reasons why this initial model was not used. Lacking parts in the protein structure were completed using 4qkx beta-adrenoreceptor as template (second best template with 23.97% sequence identity) and then model refinement was performed. Thus, all lacking parts, structural distortions, clashes, unfavourable interactions deviations, outliers and bad angles were avoided.

After the ultimate homology model of CamAKHR was built, some analyses were performed to verify its quality. It was aligned with 3sn6 template again and RMSD value was found 1.151 Å. Because it was smaller than 2 Å, this model can be considered as a proper model. Also, this ultimate model was evaluated in terms of characteristic features of Family A GPCRs such as seven transmembrane helices, conserved residues and motifs (Table 5.1.). If this model is successful, it should contain these features. As expected, it has seven transmembrane helices and an intracellular helix 8 and all other features were shown in Table 5.1. below.

Table 5.1. Characteristics of Family A GPCRs in the ultimate CamAKHR model

TM helices	Residues of TM	Conserved Residues and Motifs
TM1	Hie190 - Arg217	Gly207, Asn208
TM2	Leu226 - Thr252	Asp235, Leu236
TM3	Ser259 - Leu292	Cys282, Asp286, Arg282, Tyr288 (DRY)
TM4	Gly299 - Val321	Trp310, Pro319
TM5	Tyr352 - Asp388	Tyr361, Pro364
TM6	Leu406 - Trp431	Phe416, Trp420, Pro422 (CWXP)
TM7	Val444 - Val462	Asn459, Pro460, Tyr462 (NPXXY)

Taking all these analyses into consideration, it can be clearly said that this ultimate CamAKHR model is reasonable to proceed with the next steps.

After model validation, ligand was prepared adding common modifications. MM-GBSA for the pose with the-top docking score was calculated as -38.179 kcal/mol and docking was performed with this pose (green colored one in Figure 4.10.). Because ligands bind to the upper part of GPCRs for receptor activation, the template used for modeling this part should be considered. Active 3sn6 beta-adrenergic receptor was used to model till TM7; therefore, the binding site of the ligand should have similar coordinates of the ligand in this template. For this purpose, the coordinates of the ligand in 3sn6 template and the residues interacting with this ligand were recorded. These coordinates were used in grid generation. The crucial residues interacting with the ligand which are Asn356, Arg269, Tyr424 and Phe449 were used for constrained docking. Ligand interacting residues of CamAKHR model were found as Arg269, Phe449, Glu246, Gln445 and Tyr430 (Figure 4.11.). Only two of the crucial constraint residues Arg269 and Phe449 were observed during this docking.

When this CamAKHR-CamAKH complex was embedded into a membrane, solvated and ionized, MD simulation was performed. At the end of simulation, ligand-interacting residues were observed as Tyr423, Phe449, Glu246, Lys446, Gln445, Val443, Asp334, Asn250 and Lys188 (Figure 4.13.). Only one of the crucial constrained residues Phe449

was present during MD simulation. Because docking provides a static binding of ligand to the receptor, it gives interactions at an instant. When the native environment of cells were mimicked in MD simulations, real-like interactions between the ligand and the receptor were provided in a period of time (100ns in this simulation). Hence, ligand interacting residues coming from MD simulations are much more reliable than docking meaning that Tyr423 (71%), Phe449 (91%), Glu246 (81%), Lys446 (80%), Gln445 (68%), Val443 (87%), Asp334 (54%), Asn250 (39%) and Lys188 (79%) are very important with given percentages in ligand binding. As in Figure 4.14., there are other important amino acids in close proximity to the binding pocket of CamAKHR during MD simulation in addition to the previously mentioned residues. These residues are Arg269 (one of the crucial amino acids), Phe335, Tyr430 and Phe344 with interaction fraction higher than 0.5 were maintained. Furthermore, Glu246, Gln445, Phe449, Tyr430 and Arg269 residues seem to be important in the ligand interaction during both MD and docking.

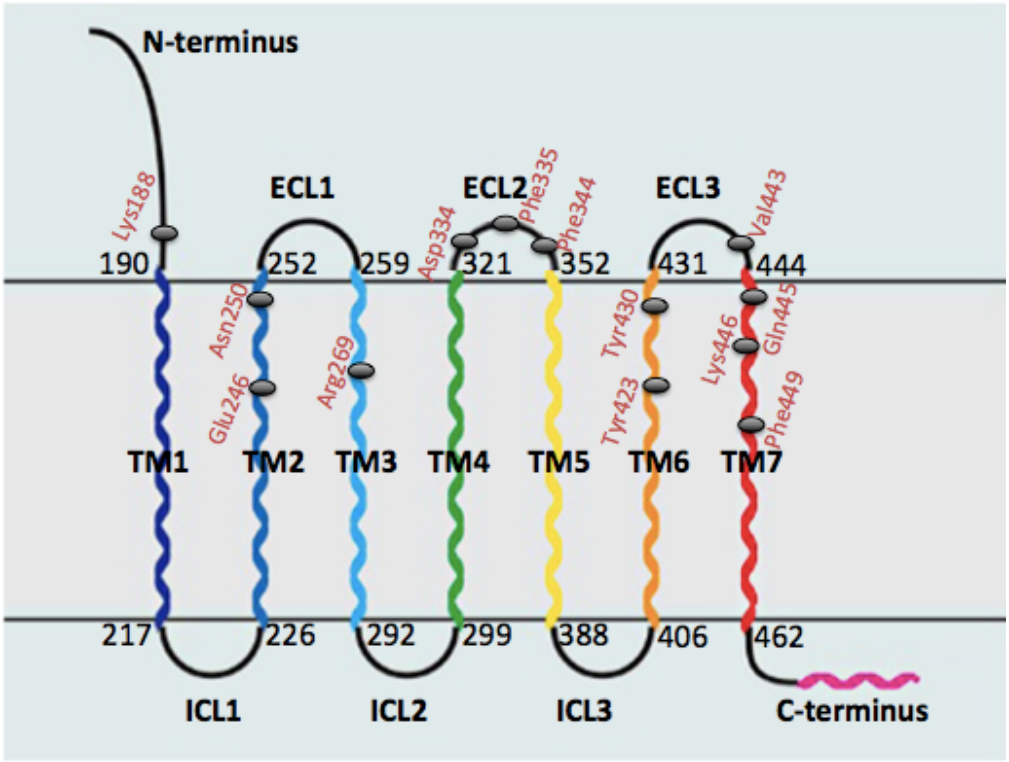


Figure 5.1. Locations of important residues in/around binding pocket of CamAKHR model

As can be seen in Figure 5.1., there are thirteen important residues contributing to ligand binding in CamAKHR- CamAKH complex. These residues are mostly located in TM6, TM7 and ECL2 regions of CamAKHR model.

At the end of MD simulations, three graphs were obtained: protein RMSF, ligand RMSF and protein and ligand RMSD. In ligand RMSF plot (Figure 4.15), both ligand fit on protein (Prot-fit) and ligand's own (Lig-fit) RMSF values are present. Prot-fit shows ligand RMSF in protein-ligand complex, while Lig-fit shows ligand RMSF regardless of the protein at the level of atoms. In both, residues numbered 5-10, 25-28 and 90-95 have higher RMSF values than the rest. These residues correspond to the parts of first two residues of the ligand and cyclic part of the fourth residue; thus, pGlu, Leu and Phe residues can be considered as the most flexible part of the ligand. There is not so much difference between Prot-fit and Lig-fit plots meaning that ligand RMSF is not affected by the presence of protein. Also the fluctuations between atoms are not so much in both plots. These may stem from that the ligand is a big molecule with 94 atoms; therefore, it does not have so much freedom to be flexible. Hence, it seems very stable during MD.

Protein RMSF plot shows RMSF values of Cα atoms of each residue in the model (Figure 4.16). The residues corresponding to the peaks are considered as the most flexible regions of the model, while the ones corresponding to the lowest RMSF values are the most rigid/ stable regions of the model. Because seven transmembrane helices are the most conserved parts of GCPRs, the lowest RMSF containing regions are considered as TMs. As expected, there are seven regions with the lowest RMSF values indicating TMs. The rest of the model (terminal and loops) has higher RMSF values and the ICL3 the longest loop has the highest RMSF implying the highest flexibility.

In Table 4.2., RMSF values corresponding to the crucial constraint amino acid residues and seven transmembrane residues which should have limited flexibility were shown. These are also consistent with the explanation above that stable residues have lower RMSF values.

In Figure 4.17., RMSD values of Cα atoms of the model and of ligand fit on protein during 100ns MD simulation were shown. The system reached equilibration because the

simulation converges indicating that protein RMSD value stabilizes around a fixed value between 4Å and 5 Å toward the end of the simulation. Actually, it was observed that CamAKHR model was stabilized after 10ns of the simulation as in Figure 4.17. There were not so much fluctuations in both ligand and protein indicating that ligand is a big molecule with limited flexibility. Furthermore, because the observed values Lig Fit on Prot are not significantly larger than the RMSD of the protein and both protein and ligand RMSD curves are close to each other during simulation time, it can be likely considered that ligand did not diffuse away from its initial binding site.

Identified thirteen amino acids (in Figure 5.1.) that are important in ligand binding were investigated for their conservation among different insect species. To find this, five different insect AKHR sequences in addition to CamAKHR were aligned and Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 were found out as the most conserved residues. Therefore, proposed binding pocket includes Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 residues. Table 5.2. below supports their conservation because of the fact that conservation requires lower RMSF values to remain more stable.

Table 5.2. RMSF values of six conserved residues

Residue	Glu246	Arg269	Tyr423	Tyr430	Lys446	Phe449
RMSF	0.768	0.697	0.696	0.75	0.819	0.883

Known insecticides were investigated whether they bind to CamAKHR or not, and their binding affinity if they bind. Approximately 350 compounds were docked to CamAKHR and the one coded as 104753 with the top-docking score was used for the following steps. Its MM-GBSA was measured as -36.137 kcal/mol before MD. MM-GBSA for CamAKH was measured as -38.178 kcal/mol previously. Because MM-GBSA scores are close to each other, this compound might be a pesticide candidate for CamAKHR. After MD simulations, MM-GBSA calculations were performed for last 50ns of both CamAKH and 104753. Whereas average MM-GBSA score of CamAKH was -71.73 kcal/mol, score of 104753 was -30.03 kcal/mol. Even though they seemed to be close to each other regarding MM-GBSA scores before MD simulations, the difference

between scores increased after MD simulation. CamAKH peptide showed better score with lower MM-GBSA value compared to its initial value and also to 104753. This means CamAKH peptide can bind more efficiently to CamAKHR than 104753 as a known pesticide. This result was not unexpected. Future work of this study is designing AKH-based antagonist as a novel next generation pesticide and antagonist should be developed to bind to inactive state of a receptor rather than active one. Because our CamAKHR model was constructed based on active conformations of two templates, this model can be considered in active conformation. Because of the conformational changes in CamAKHR after activation, antagonists can show less binding compared to agonists. These known insecticides recorded in database probably act binding to inactive conformations of receptors, therefore binding efficiency of 104753 to active CamAKHR is lower than binding efficiency of agonist CamAKH to CamAKHR. This result also demonstrates that development of AKH-based novel pesticide also requires CamAKHR model based on templates in inactive conformations. Binding efficiency of AKH-based antagonist to CamAKHR models that are built based on both active and inactive conformations should be investigated.

Mutation validation studies were performed in order to verify the importance of residues used as constraints in docking studies *in silico*. When three of these residues (Asn356, Arg269 and Tyr424) were mutated to Alanine (each in different trials), no poses were obtained meaning that they are crucial amino acids for ligand binding and there is no ligand binding when they are mutated. When Phe449 was mutated to Ala (in F449A mutant), there were poses. However, even the pose with the top-docking score gave a high -3.507 kcal/mol MM-GBSA score meaning that this residue is also crucial because the ligand does not tend to stay at its initial site in the model during simulation when this residue is mutated.

6. CONCLUSION

It is essential to develop new methods to avoid pests for sustainable agriculture because old traditional ways are not environmentally- friendly. Therefore, they result in several detrimental impacts not only on pests but also on soil, plants, animals and water; thus, indirectly on humans. It is inevitable to develop new methods that prevent harmful effects of pests without affecting the rest for the sake of nature. One of the latest trends in pesticide studies is the production of neuropeptide- based pesticide. Regarding this issue, adipokinetic hormone receptor essential in several physiological functions in insects including *C.morosus* pest was studied. In order to understand the action mechanism of AKH based antagonist that can be designed as a novel next generation pesticide, binding mechanism of AKH to AKHR that subsequently activates related physiological processes in stick insect *C.morosus* was investigated. There was not any X-ray crystal structure of adipokinetic hormone receptor in the literature, so 3D structure of AKHR in *C.morosus* was built via homology modeling firstly. Its model validation analyses were performed and the model was refined. Then, its native CamAKH ligand binding pocket and important residues within this pocket were identified via docking and MD simulations. During MD simulations, CamAKHR reached stabilization and CamAKH did not show significant displacement from its initial binding site. There were 13 amino acid residues observed in/around ligand binding cavity of CamAKHR. These residues are Tyr423, Phe449, Glu246, Lys446, Gln445, Val443, Asp334, Asn250, Lys188, Arg269, Tyr430, Phe335 and Phe344. These residues are mostly located in TM6, TM7 and ECL2 regions of CamAKHR model. Also, six of them that are Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 were found out as the most conserved residues in different insect species. Therefore, proposed binding pocket includes Glu246, Arg269, Tyr423, Tyr430, Lys446 and Phe449 residues which are corresponding to TM2, TM3, TM6 and TM7 of CamAKHR. Mutation studies have shown that Asn356, Arg269, Tyr424 and Phe449 that are constrained in docking studies are crucial amino acids in ligand binding.

The findings of this work can be used in the generation of neuropeptide-based pesticide with a higher binding capacity than its native CamAKH peptide. There may not

be any known crystals of AKHR yet; however, the model built in this work can provide a 3D structure with crucial residues in ligand binding. Further verification of crucial residues within binding pocket of CamAKHR can be conducted via wet-lab techniques such as site directed mutagenesis after cloning CamAKHR. As mentioned before in discussion part, it can be necessary to build a CamAKHR model based on templates in inactive conformations for binding of AKH-based antagonist efficiently contributing to the production of a new environmentally- friendly pesticide.

REFERENCES

1. International Union for Conservation of Nature and Natural Resources, *The IUCN Red List of Threatened Species*, IUCN Global Species Programme Red List Unit, 2014, 2018.

2. Foottit, R. G. and P. H. Adler, *Insect Biodiversity: Science and Society*, John Wiley & Sons, Chichester, 2009.

3. Capinera, J. L., *Insects and Wildlife: Arthropods and their Relationships with Wild Vertebrate Animals*, John Wiley & Sons, Chichester, 2010.

4. Gullan, P. J. and P. S. Cranston, *The Insects: An Outline of Entomology*, 5th edition, John Wiley & Sons, Oxford, 2014.

5. Resh, V. H. and R. T. Carde (editors), *Encyclopedia of Insects*, 2nd edition, pp. 220–227, Academic Press, London, 2009.

6. Burn, A. J., T. H. Coaker and P. C. Jepson (editors), *Integrated Pest Management*, p.474, Academic Press, London, 1987.

7. Brock, P. D., "Three new species of South African stick insects (Phasmida).", *Journal of Orthoptera Research,* Vol.15, No.1, pp. 37-44, 2006.

8. Picker, M., and C. Griffiths (editors), *Alien and Invasive Animals: A South African Perspective*, Struik-Random House Publishers, Cape Town, 2011.

9. Baker, E., "The worldwide status of phasmids (Insecta: Phasmida) as pests of agriculture and forestry, with a generalised theory of phasmid outbreaks.", *Agriculture & Food Security,* Vol. 4, No.1, 2015.

10. Wojciechowska, M., P. Stepnowski, and M. Gołębiowski, "The use of insecticides to control insect pests.", *Information Systems Journal*, Vol: 13, pp. 210–220, 2016.

11. Pimentel, D. and A. Greiner (editors), *Techniques for reducing pesticide use: environmental and economic benefits*, John Wiley and Sons, Chichester, 1997.

12. Tisdell, C. A. and C. Wilson, "Why farmers continue to use pesticides despite environmental, health and sustainability costs.", *Ecological Economics*, Vol. 39, No. 3, pp. 449–462, 2001.

13. Kodrík, D., H. G. Marco, P. Šimek, R. Socha, and G. Gäde, "The adipokinetic hormones of Heteroptera: a comparative study.", *Physiological Entomology*, Vol. 35, pp. 117–127, 2010.

14. Gäde, G. and G. J. Goldsworthy, "Insect peptide hormones: a selective review of their physiology and potential application for pest control.", *Pest Management Science*, Vol. 59, pp. 1063–1075, 2003.

15. Montgomery, J. (editor), *Insecticides: Occurrence, Global Threats and Ecological Impact*, Nova Science Publishers, pp. 77-92, NY, 2015.

16. Gäde, G., K. H.Hoffmann and J. H. Spring, "Hormonal regulation in insects: facts, gaps, and future directions.", *Physiological Reviews*, Vol. 77, pp. 963–1032, 1997.

17. Nijhout, H. F., *Insect Hormones*, Princeton University Press, New Jersey, 1998.

18. Henry, H. L. and A. W. Norman (editors), *Encyclopedia of Hormones*, pp. 38-42, Academic Press, London, 2003.

19. Li, S., F. Hauser, S. K. Skadborg, S. V. Nielsen, N. Kirketerp-Møller and C. J. Grimmelikhuijzen, "Adipokinetic hormones and their G protein-coupled receptors emerged in Lophotrochozoa.", *Scientific reports*, Vol. 6, No. 32789, 2016.

20. Munte, C. E., G. Gäde, B. Domogalla, W. Kremer, R. Kellner and H. R. Kalbitzer, "C-mannosylation in the hypertrehalosaemic hormone from the stick insect *Carausius morosus*.", *The FEBS Journal*, Vol. 275, pp. 1163–1173, 2008.

21. Gäde, G., "Peptides of the adipokinetic hormone/red pigment-concentrating hormone family: a new take on biodiversity.", *Trends in Comparative Endocrinology and*

Neurobiology: Annals of the New York Academy of Sciences, Vol. 1163, pp. 125–136, 2009.

22. Köllisch, G. V., M. W. Lorenz, M, R. Kellner, P. D. Verhaert and K. H. Hoffmann, "Structure elucidation and biological activity of an unusual adipokinetic hormone from corpora cardiaca of the butterfly, Vanessa cardui.", *European Journal of Biochemistry*, Vol. 267, No. 17, pp. 5502-5508, 2000.

23. Kodrik, D., "Adipokinetic hormone functions that are not associated with insect flight.", *Physiological Entomology,* Vol. 33, No. 3, pp. 171-180, 2008.

24. Krishnan, N., and D. Kodrík, *Endocrine control of oxidative stress in insects,* pp. 261-270, Wiley-Blackwell, New Jersey, 2012.

25. Spencer, I. M., and D. J. Candy, "Hormonal control of diacyl glycerol mobilization from fat body of the desert locust, Schistocerca gregaria.", *Insect Biochemistry,* Vol. 6, No. 3, pp. 289-296, 1976.

26. Van Marrewijk, W. J. A., A. Th. M. Broek, and A. M. Beenakkers, "Regulation of glycogenolysis in the locust fat body during flight.", *Insect Biochemistry,* Vol. 10, No. 6, pp. 675-679, 1980.

27. Vroemen, S. F., W. J. Van Marrewijk, J. De Meijer, A. T. M. Van den Broek and D. J. Van der Horst, "Differential induction of inositol phosphate metabolism by three adipokinetic hormones.", *Molecular and cellular endocrinology*, Vol. 130, No. 1-2, pp. 131-139, 1997.

28. Kodrík, D., N. Krishnan and O. Habuštová, "Is the titer of adipokinetic peptides in Leptinotarsa decemlineata fed on genetically modified potatoes increased by oxidative stress?.", *Peptides,* Vol. 28, No. 5, pp. 974-980, 2007.

29. Gokuldas, M., P. A. Hunt and D. J. Candy, "The inhibition of lipid synthesis in vitro in the locust, Schistocerca gregaria, by factors from the corpora cardiaca.", *Physiological entomology,* Vol. 13, No. 1, pp. 43-48, 1988.

30. Carlisle, J. A. and B. G. Loughton, "Adipokinetic hormone inhibits protein synthesis in Locusta.", *Nature,* Vol. 282., No. 5737, p.420, 1979.

31. Kodrík, D. and G. J. Goldsworthy, "Inhibition of RNA synthesis by adipokinetic hormones and brain factor (s) in adult fat body of Locusta migratoria.", *Journal of insect physiology,* Vol. 41, No. 2, pp.127-133, 1995.

32. Lee, G. and J. H. Park, "Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in Drosophila melanogaster.", *Genetics,* Vol. 167, No. 1, pp. 311-323, 2004.

33. Lorenz, M. W., "Adipokinetic hormone inhibits the formation of energy stores and egg production in the cricket Gryllus bimaculatus.", *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology,* Vol. 136, No. 2, pp. 197-206, 2003.

34. Scarborough, R. M., G. C. Jamieson, F. Kalish, S. J. Kramer, G. A. McEnroe, C. A. Miller and D. A. Schooley, "Isolation and primary structure of two peptides with cardioacceleratory and hyperglycemic activity from the corpora cardiaca of Periplaneta americana.", *Proceedings of the National Academy of Sciences*, Vol. 81, No. 17, pp. 5575-5579, 1984.

35. Kodrík, D., K. Vinokurov, A. Tomčala and R. Socha, "The effect of adipokinetic hormone on midgut characteristics in *Pyrrhocoris apterus* L. (Heteroptera).", *Journal of Insect Physiology*, Vol. 58, pp. 194–204, 2012.

36. O'Shea, M., J. L. Witten and M. Schaffer, "Isolation and characterization of 2 myoactive peptides — further evidence of an invertebrate peptide family.", *Journal of Neuroscience*, Vol. 4, pp. 521–529, 1984.

37. Socha, R., D. Kodrík and R. Zemek, "Adipokinetic hormone stimulates insect locomotor activity.", *Naturwissenschaften,* Vol. 86, No. 2, pp. 85-86, 1999.

38. Goldsworthy, G., K. Opoku-Ware and L. Mullen, "Adipokinetic hormone enhances laminarin and bacterial lipopolysaccharide-induced activation of the prophenoloxidase cascade in the African migratory locust, Locusta migratoria.", *Journal of Insect Physiology,* Vol. 48, No. 6, pp. 601-608, 2002.

39. Van der Horst, D. J., "Insect adipokinetic hormones: release and integration of flight energy metabolism.", *Comparative Biochemistry and Physiology Part B*, Vol. 136, pp. 217–226, 2003.

40. Rosenbaum, D. M., S. G. F. Rasmussen and B. K. Kobilka, "The structure and function of G-protein-coupled receptors.", *Nature,* Vol. 459, No.7245, p. 356, 2009.

41. Fredriksson, R., M. C. Lagerström, L. G. Lundin and H. B. Schiöth, "The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.", *Molecular pharmacology*, Vol. 63, No. 6, pp. 1256-1272, 2003.

42. Sealfon, S. C., H. Weinstein and R. P. Millar, "Molecular mechanisms of ligand interaction with the gonadotropin-releasing hormone receptor.", *Endocrine* reviews, Vol. 18, No. 2, pp. 180-205, 1997.

43. Mirzadegan, T., G. Benkö, S. Filipek and K. Palczewski, "Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin.", *Biochemistry*, Vol. 42, No. 10, pp. 2759-2767, 2003.

44. Ballesteros, J. A. and H. Weinstein, "Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors.", *Neuroscience Methods,* Vol. 25, pp. 366-428, 1995.

45. Ranganathan, A., *The impact of GPCR structures on understanding receptor function and ligand binding*., Doctoral Dissertation, Department of Biochemistry and Biophysics, Stockholm University, 2016.

46. Melia, T. J., C. W. Cowan, J. K. Angleson, and T. G. Wensel, "A comparison of the efficiency of G-protein activation by ligand-free and light-activated forms of rhodopsin.", *Biophysics Journal*, Vol. 73, pp. 3182–3191, 1997.

47. Strange, P. G., "Three-state and two-state models.", *Trends in Pharmacological Sciences*, Vol. 19, pp. 85–86, 1998.

48. Schöneberg, T., A. Schulz and T. Gudermann, "The structural basis of G-protein-coupled receptor function and dysfunction in human diseases.", *Reviews of Physiology, Biochemistry and Pharmacology*, Vol. 144, pp. 145–227, 2002.

49. Krebs, A., C. Villa, P. C. Edwards and G. F. X. Schertler, "Characterisation of an improved two-dimensional p22121 crystal from bovine rhodopsin1.", *Journal of Molecular Biology*, Vol. 282, No. 5, pp. 991-1003, 1998.

50. Schertler, G. F. X., C. Villa and R. Henderson, "Projection structure of rhodopsin.", *Nature,* Vol. 362, No. 6422, p. 770, 1993.

51. Kobilka, B. and G. F. X. Schertler, "New G-protein-coupled receptor crystal structures: insights and limitations.", *Trends in pharmacological sciences,* Vol. 29, No. 2, pp. 79-83, 2008.

52. Serrano-Vega, M. J., F. Magnani, Y. Shibata and C. G. Tate, "Conformational thermostabilization of the β_1-adrenergic receptor in a detergent-resistant form.", *Proceedings of National Academy of Sciences of the United States of America,* Vol. 105, pp. 877–882, 2008.

53. Rasmussen, S. G. F., 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 β_2 adrenergic G-protein-coupled receptor.", *Nature,* Vol. 450, pp. 383–387, 2007.

54. Rosenbaum, D. M., 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 β_2-adrenergic receptor function.", *Science,* Vol. 318, pp. 1266–1273, 2007.

55. Cherezov, V., 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 β_2-adrenergic G-protein-coupled receptor.", *Science,* Vol. 318, pp. 1258–1265, 2007.

56. Warne, T., M. J. Serrano-Vega, J. G. Baker, R. Moukhametzianov, P. C. Edwards, R. Henderson, A. G. W. Leslie, C. G. Tate and G. F. X. Schertler, "Structure of a β_1-adrenergic G-protein-coupled receptor.", *Nature*, Vol. 454, pp. 486–491, 2008.

57. Jaakola, V. P., M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. T. Chien, J. R. Lane, A. P. IJzerman and R. C. Stevens, "The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist.", *Science,* Vol. 322, pp. 1211–1217, 2008.

58. Ghanouni, P., Z. Gryczynski, J. J. Steenhuis, T. W. Lee, D. L. Farrens, J. R. Lakowicz and B. K. Kobilka, "Functionally Different Agonists Induce Distinct Conformations in the G Protein Coupling Domain of the β2Adrenergic Receptor.", *Journal of Biological Chemistry*, Vol. 276, No. 27, pp. 24433-24436, 2001.

59. Worth, C. L., G. Kleinau and G. Krause, "Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models.", *PloS one* , Vol. 4, No. 9, e7011, 2009.

60. Ravna, A.W. and I. Sylte, "Homology Modeling of Transporter Proteins (Carriers and Ion Channels).", *Methods in Molecular Biology*, Vol. 857, pp. 281-299, 2011.

61. Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, "Full-length transcriptome assembly from RNA-seq data without a reference genome.", *Nature Biotechnology*, Vol. 29, No. 7, pp. 644-652, 2011.

62. Orry, A. J. W. and R. Abagyan R. (editors), *Homology Modeling*, pp. 55–82, Humana Press, New Jersey, 2012.

63. Jacobson, M. P., D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw and R. A. Friesner, "A hierarchical approach to all-atom protein loop prediction.", *Proteins: Structure, Function, and Bioinformatics*, Vol. 5, No.2, pp. 351-367, 2004.

64. Jacobson, M. P., R. A. Friesner, Z. Xiang, B. Honig, "On the Role of Crystal Packing Forces in Determining Protein Sidechain Conformations.", *Journal of Molecular Biology,* Vol.320, PP. 597-608, 2002.

65. Tramontano, A., D. Cozzetto, A. Giorgetti and D. Raimondo, "The assessment of methods for protein structure prediction.", *Methods in Molecular Biology,* pp. 43–57, 2008.

66. Liu,T., M. Guerquin and R. Samudrala, "Improving the accuracy of template-based predictions by mixing and matching between initial models.", *BMC Structural Biology*, Vol. 8, No. 24, 2008.

67. Orry, A. J. W. and R. Abagyan R. (editors), *Homology Modeling*, pp. 83–106, Humana Press, New Jersey, 2012.

68. Rasmussen, S. G. F., B. T. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. A. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara and B. K. Kobilka, "Crystal structure of the β 2 adrenergic receptor–Gs protein complex.", *Nature*, Vol. 477, No. 7366, 2011.

69. Weichert, D., A. C. Kruse, A. Manglik, C. Hiller, C. Zhang, H. Hübner, B. K. Kobilka and P. Gmeiner, "Covalent agonists for studying G protein-coupled receptor activation.", *Proceedings of the National Academy of Sciences,* Vol. 111, No. 29, pp. 10744-10748, 2014.

70. Sastry, G. M., M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments.", *Journal of Computer-Aided Molecular Design*, Vol. 27, No. 3, pp. 221-234, 2013.

71. Bas, D. C., D. M. Roger and J. H. Jensen, "Very fast prediction and rationalization of pKa values for protein-ligand complexes.", *Proteins: Structure, Function, and Bioinformatics*, Vol. *73*, No. 3, pp. 765–783, 2008.

72. Siu, S., K. Pluhackova and R. Böckmann, "Optimization of the OPLS-AA Force Field for Long Hydrocarbons.", *Journal of Chemical Theory and Computation*, Vol. 8, No. 4, pp. 1459- 1470, 2012.

73. Banks, J. L., H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A.K. Felts, T.A. Halgren, D.T. Mainz, J.R. Maple, R. Murphy, D.M. Philipp, M.P. Repasky, L.Y. Zhang, B.J. Berne, R.A. Friesner, E. Gallicchio, R. M. Levy, "Integrated Modeling Program, Applied Chemical Theory (IMPACT).", *Journal of Computational Chemistry*, Vol. 26, p. 1752, 2005.

74. Greenwood, J. R., D. Calkins, A. P. Sullivan and J. C. Shelley, "Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution.", *Journal of Computer-Aided Molecular Design*, Vol. 24, pp. 591-604, 2010.

75. Shelley, J. C., A. Cholleti, L. Frye, J. R. Greenwood, M. R. Timlin and M. Uchimaya, "Epik: a software program for pK_a prediction and protonation state generation for drug-like molecules.", *Journal of Computer-Aided Molecular Design*, Vol. 21, pp. 681-691, 2007.

76. MacroModel, version 10.8, Schrödinger, LLC, New York, NY, 2015.

77. Christopher, J. A., S. J. Aves, K. A. Bennett, A. S. Doré, J. C. Errey, A. Jazayeri, F. H. Marshall, K. Okrasa, M. J. Serrano-Vega, B. G. Tehan, G. R. Wiggin and M. Congreve, "Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the mGlu5 Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2- yl)pyrimidin-4-yl]benzonitrile).", *Journal of Medicinal Chemistry*, Vol. 58, No. 16, pp. 6653–6664, 2015.

78. Friesner, R. A., R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes.", *Journal of Medicinal Chemistry*, Vol. 49, pp. 6177–6196, 2006.

79. Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, J. L. Banks, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening.", *Journal of Medicinal Chemistry*, Vol. 47, pp. 1750–1759, 2004.

80. Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, D. E. Shaw, M. Shelley, J. K. Perry, P. Francis and P. S. Shenkin, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.", *Journal of Medicinal Chemistry*, Vol. 47, pp. 1739–1749, 2004.

81. Rastelli, G., A. Rio, G. Degliesposti and M. Sgobba, "Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA.", *Journal of Computational Chemistry,* pp.797-810, 2010.

82. Li, J., R. Abel, K. Zhu, Y. Cao, S. Zhao and R. A. Friesner , "The VSGB 2.0 Model: A Next Generation Energy Model for High Resolution Protein Structure Modeling.", *Proteins*, Vol. 79, No.10, pp. 2794-2812, 2011.

83. Bowers, K. J., D. E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, "Scalable algorithms for molecular dynamics simulations on commodity clusters.", In: SC 2006 Conference*, Proceedings of the 2006 ACM/IEEE Conference on Supercomputing,* USA*,* 11-17 November 2006, IEEE, Florida, p. 43, 2006.

84. Lomize, M. A., A. L. Lomize, I. D. Pogozheva and H. I. Mosberg, "OPM: Orientations of Proteins in Membranes database.", *Bioinformatics*, Vol. 22, No. 5, pp. 623–625, 2006.

85. Lomize, A. L., I. D. Pogozheva, M. A. Lomize and H. I. Mosberg, "Positioning of proteins in membranes: A computational approach.", *Protein Science : A Publication of the Protein Society*, Vol. 15, No. 6, pp. 1318–1333, 2006.

86. Hoover, W. G., "Canonical dynamics: equilibrium phase-space distributions.", *Physical review A* , Vol. 31, No. 3, p. 1695, 1985.

87. Martyna, G. J., D. J. Tobias and M. L. Klein, "Constant pressure molecular dynamics algorithms.", *The Journal of Chemical Physics*, Vol. 101, No. 5, pp.4177-4189, 1994.

88. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.", *Molecular Systems Biology*, Vol. 7, No. 539, 2011.

89. Kim, S., P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, S. H. Bryant, "PubChem Substance and Compound databases.", *Nucleic Acids Research*, 44(D1) :D1202-13, 2016.

90. Benkert, P., M. Biasini and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models.", *Bioinformatics,* Vol. 27, pp. 343-350, 2011.

91. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, 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 Research*, Vol. 46, No. W1, pp. W296-W303, 2018.

92. Ramachandran, G.N., C. Ramakrishnan, V. Sasisekharan, "Stereochemistry of polypeptide chain configurations.", *Journal of Molecular Biology*, Vol. 7, pp. 95–99, 1963.

93. Lovell, S. C., I. W. Davis, W. B. Arendall III, P. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson, "Structure validation by Cα geometry: ϕ, ψ and Cβ deviation.", *Proteins: Structure, Function, and Bioinformatics*, Vol. 53, No. 3, pp. 437-450, 2003.

94. Chen, V. B., W. B. Arendall III, 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 Crystallographica Section D: Biological Crystallography,* Vol. 66, No. 1, pp. 12-21, 2010.

95. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson, "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.", *Nucleic Acids Research*, Vol. 35, No. suppl_2, pp. W375-W383, 2007.