FUNCTIONAL AND STRUCTURAL INSIGHTS INTO A NOVEL INSECT G PROTEIN-COUPLED RECEPTOR, ALLATOSTATIN RECEPTOR TYPE C OF PINE PROCESSIONARY MOTH

by

Aida Shahraki

B.S., Atomic and Molecular Physics, Kharazmi University, 2011

M.S., Biophysics, University of Tehran, 2013

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

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2020

to my beloved partner ...

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Dr. Necla Birgül İyison for her constant support and guidance throughout the project. I would be always grateful for the opportunity of working on this exciting project and indebted for her enormous help with any kind of problems I encountered during my stay in Turkey.

Cancer Signaling Laboratory (CSL), as my home lab, was where I met extraordinary people from whom I learned all molecular biology techniques, and beyond that, where I found my best friends. I would like to especially thank Burçin Duan-Şahbaz as my dearest friend and colleague in CSL. Tolga Aslan, Izzet Akiva, and Tuncay Şeker were definitely among my role-models in my Ph.D. life. I have to thank Gökhan Gün, Barbaros Düzgün and Kübra Kahvecı for being such great friends. I will always be grateful to all the former and current CSL members for having their precious friendship.

I would like to thank my thesis defense committee Prof. Serdar Durdağı, Assoc. Prof. İbrahim Yaman, Assist. Prof. Berna Doğan and Assist. Prof. Igor Kryvoruchko for sparing their precious time to evaluate this work and providing me with their invaluable feedback and thoughts.

The present work would have not been possible without constructive collaborations with other laboratories. The WGS part, as the starting point, became possible thanks to the invaluable help of Dr. Chunguang Liang at Würzburg University, Germany. The functional experiments were all performed in Dr. Martin J. Lohse's laboratory, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Germany. I would like to give special thanks to Ali Işbilir, without whom the FRET and BRET studies could have not been performed.

I would like to express my sincere gratitude to the main collaborator of the project, Dr. Serdar Durdağı, the head of the DRG laboratory, Bahçeşehir University, Turkey, where all the computational studies were performed. I will never be able to thank him enough for the great opportunity he provided me with. He absolutely opened my eyes to a new World which I got fascinated with. I would always be indebted to Dr. Berna Doğan who was more than just a teacher to me, and more a mentor and a very great friend. I will never forget İsmail Erol, Busecan Aksoydan, and my other fabulous friends at the DRG laboratory, where I spent the greatest times of my Ph.D. life.

I would like to thank all the members of the Molecular Biology and Genetics department for their enormous help and support. We shared fantastic memories together that I will never forget. My special thanks go to my fantastic Iranian friend, Ulduz Sobhi Afshar with whom I shared moments of joy and sorrow.

Finally, I would like to express my deepest appreciation to my dear family who have been always there for me throughout the toughest moments of my Ph.D. life. Until the last day of my life, I will be grateful to my mother for encouraging me to pursue my dreams. She has been the hero of my life and will undoubtedly stay so forever. I will never be able to thank enough my beloved partner, Yaser Kalblai, for his endless love that enabled me to move forward. Without his emotional support, I could never bear those loneliest days when I had to stay away from home for months to perform a part of this study. I do sincerely appreciate the dedication and devotion he showed at each and every step of our life.

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) Fund 119Z921.

ABSTRACT

FUNCTIONAL AND STRUCTURAL INSIGHTS INTO A NOVEL INSECT G PROTEIN-COUPLED RECEPTOR, ALLATOSTATIN RECEPTOR TYPE C OF PINE PROCESSIONARY MOTH

Insect neuropeptides regulate different aspects of insect physiology. They exert their function via binding to their cognate receptor belonging to Class A G protein coupled receptors (GPCRs). Neuropeptide receptors are potential targets for next-generation pesticides. Allatostatin (AST) neuropeptides regulate the development of insects through the inhibition of juvenile hormone (JH) secretion. Here, Allatostatin receptor type C (AstR-C) of pine processionary moth, a pest in Mediterranean countries, was extracted from the whole genome sequencing (WGS) data. The receptor was cloned and characterized combining via different approaches. Using resonance energy transfer (RET)-based techniques, kinetics of G protein coupling and β-arrestin recruitment were investigated. Homology modeling, docking and molecular dynamics (MD) simulation approaches were conducted to predict the orthosteric pocket of the receptor which was validated by *in silico* and *in vitro* methods. The binding pocket was subjected to virtual screening studies to find agonists. As a result, it was found that binding of the native ligand at sab-nanomolar and nanomolar ranges to the receptor induces Gi/o protein coupling and β-arrestin recruitment, respectively. Kinetics studies revealed that a brief stimulation of the receptor at nanomolar range is enough to obtain a long-lasting response. The accuracy of the predicted orthosteric pocket was validated via G protein activation assay. $Q271^{6.55}$ (Ballesteros-Weinstein numbering) was found to be critical for G protein-dependent activation of AstR-C. Virtual screening studies resulted in obtaining a small molecule capable of activating the receptor.

ÖZET

Yeni Bir Böcek G Proteinine-Bağlı Reseptörü Olan Çam Kese Kurdu C Tipi Allatostatin Reseptörünün İşlevi ve Yapısına Dair

Böcek nöropeptitleri böcek fizyolojisinde çok farklı görevler üstlenir. Bu peptitler G proteinine-bağlı reseptörlerin (GPCR) A Sınıfı'na dahil olan ilgili reseptörlere bağlanarak işlevlerini yerine getirir. Nöropeptit reseptörleri yeni jenerasyon pestisit geliştirme çalışmalarının hedef molekülüdür. Allatostatin (AST) nöropeptitleri gençlik hormonu (GH) salgılanmasını engellemek yoluyla böcek gelişimini düzenler. Bu çalışmada, tüm genom dizileme (WGS) analizi yapılarak Akdeniz ülkelerinde yaygın bir zararlı türü olan çam kese kurdunun C tipi Allatostatin reseptörü (AstR-C) ortaya çıkarılmıştır. Reseptör klonlanmış, sonra *in silico* ve *in vitro* yöntemler birleştirilerek tanımlanmıştır. Rezonans enerji transferine (RET) dayanan tekniklerle, G protein yaklaşma kinetiği, aktivasyon ve β-arrestin yaklaşması değerlendirilmiştir. Reseptörün ortosterik cebini tahmin etmek için homoloji modelleme, kenetlenme ve moleküler dinamik (MD) simülasyonları gerçekleştirilmiş, *in silico* ve *in vitro* yöntemlerle desteklenmiştir. Agonist adayları bulmak amacıyla bağlanma cebinde sanal tarama çalışmaları gerçekleştirilmiştir. Sonuç itibariyle, doğal ligandın reseptöre nanomolar-altı ve nanomolar seviyelerde bağlanmasının aynı sırayla Gi/o bağlanmasını ve β-arrestin yaklaşmasını indüklediği görülmüştür. Kinetik çalışmalara göre, reseptörün nanomolar seviyede kısa bir süre uyarılması uzun süreli bir tepki almak için yeterlidir. Tahmini ortosterik cebin doğruluğu G protein aktifleşme deneyi ile değerlendirilmiştir. $Q271^{6.55}$ (Ballesteros-Weinstein numaralandırma sistemi) amino asidinin AstR-C'nin G proteinine-bağlı aktifleşmesi için kritik olduğu ortaya çıkmıştır. Sanal tarama çalışmaları ise reseptörü aktifleştirme yeteneğine sahip bir adet küçük molekül elde edilmesi ile sonuçlanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

g	Gram
l	Liter
M	Molar
m	Meter
n	Nano
p	Pico
V	Volt
U	Unit

α	Alpha
$\mathring{A}$	Angstrom
β	Beta
Δ	Delta
δ	Delta
γ	Gamma
κ	Kappa
μ	Micro
τ	Tau

®	registered trademark

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
7TM	7-Transmembrane
AST	Allatostatin
AstR-C	Allatostatin Receptor Type C
BRET	Biloluminescence Resonance Energy Transfer
BPB	Bromophenol Blue
BSA	Bovine Serum Albumin
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
Cα	Carbon Alpha
CA	Corpus Allatum
cAMP	Cyclic Adenosine Monophosphate
CC	Corpus Cardiacum
cDNA	Complementary Deoxyribonucleic Acid
CFP	Cyan Fluoroscent Protein
CO_2	Carbon Dioxide
cryo-EM	Cryo-Electron Microscopy
C-terminus	C-terminal Loop
ddH_2O	Double Distilled Water
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide
ECL	Extracellular Loop
EDTA	Ethylenediaminetetraacetic Acid
EtBr	Ethidium Bromide
EtOH	Ethanol

FAO	Food and Agriculture Organization
FBS	Fetal Bovine Serum
FRET	Förster Resonance Energy Transfer
GDP	Guanosine Diphosphate
GIRK	G Protein-Coupled Inwardly Rectifying Potassium Channel
GPCR	G Protein-Coupled Receptor
GRK	G Protein Receptor Kinase
GTP	Guanosine 5'-triphosphate
HCl	Hydrochloric Acid
hr	hours
ICL	Intracellular Loop
IM	Intermediate
JH	Juvenile Hormone
kb	Kilobase
LB	Luria-Bertani Broth
$MgCl_2$	Magnesium Chloride
MaSurCA	Maryland Super Read Cabog Assembler
MD	Molecular Dynamics
min	Minutes
MOPS	3-(N-morpholino) Propanesulfonic Acid
mRNA	Messenger Ribonucleic Acid
NaCl	Sodium Chloride
NaOAc	Sodium Acetate
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
ns	Nanosecond
N-terminus	N-terminal Loop
OD	Optical Density
OPM	Orientation of Protein in Membrane
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline

PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PTTH	Prothoracicotropic Hormone
RET	Resonance Energy Transfer
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
rpm	Revolutions per Minute
RT	Room Temperature
RT-PCR	Reverse Transcription PCR
SD	Standard Deviation
SDM	Site-Directed Mutagenesis
sec	Seconds
SGA	String Graph Assembler
TAE	Tris-Acetic Acid EDTA
Taq	*Thermnus aquaticus*
TpitAstRC	Allatostatin Receptor Type C of Pine Processionary Moth
TM	Trans-membrane
UV	Ultraviolet
WGS	Whole Genome Sequencing
WT	Wild Type
v/v	Volume to Volume
w/v	Weight to Volume
YFP	Yellow Fluorescent Protein

1. INTRODUCTION

1.1. Insects

Insects belong to phylum Arthropoda and class Hexapoda. They have a characteristics body divided into three parts; head, thorax, and abdomen. The thorax itself is further segmented into legs and wings [1].

Insects are the most abundant and diverse organisms living on Earth. Over a million species of insects are already identified, and entomologists estimate that the number can be escalated five to seven times by new discoveries [2].

Considering the vast population of these organisms, it can be deduced that insects are very well adaptable. They almost live in any habitable place on the planet and play a key role in ecological interactions between all living organisms. Insects are the main consumers of plants, and predators of the plant eater organisms at the same time. They are also food source for other animals [3].

1.2. Insects as Pests

Two mega Taxa, insects and plants, have been subjected to coevolution for millions of years and benefited from each other. At the expense of providing insects with shelter, plants have benefited from these organisms for pollination and seed dispersal. However, some herbivorous insect species can harm plants seriously that has resulted in the development of defense mechanisms in plants [4]. Taken together, from the evolutionary point of view, it can be inferred that insects and plants, having mutual benefits, can coexist in an ecological balance. The balance, however, was interrupted after *Homo sapiens* entered the scene. These newcomers basically relied on agriculture for survival and obviously tried to change the balance between insects and plants toward the latter.

According to a definition suggested by Food and Agriculture Organization (FAO), any species that affects plants or their products adversely is a pest. In addition, vectors of pathogens and parasites that can cause diseases in humans or other animals are defined as pests. Entomologists, however, believe that the term pest is an arbitrary label that can be changed constantly. One insect can be considered a pest at certain time, while beneficial at another point of time [4].

In Mediterranean forests, outbreak insects are the most serious biotic disturbances that cause severe defoliations and adversely affect the growth and vitality of forests [5].

Frequent outbreaks of defoliators negatively influence the trees and result in the persistence of trees that are already damaged [6]. This is in particular very destructive in conifers as these trees unlike broad-leaved trees such as larch do not display any re-flush growth following foliage consumption [7].

1.3. Pine Processionary Moth

In Mediterranean countries, conifer forests are mainly defoliated by pine or cedar processionary moths (Lepidoptera, Thaumetopoeidae): *Thaumetopoea pityocampa* (Dennis and Schiff.) (T. pityocampa) [8]. Among other pine defoliators that live in Mediterranean regions, *T. pityocampa* is the most serious and harmful species [9].

Pine processionary moth is univoltine, meaning it produces one brood or generation per year. Adults lay eggs in the middle of summer, between July and August, but the timing can changed slightly in different geographical regions [8]. They lay their eggs in the canopy of the pine trees. Eggs hatch in 45 days which means that larvae hatches in early September. During the autumn and winter, larvae of the insect mainly lives in silk nests on the most insulated parts of trees. The nests protect them from the cold. Being on top of the trees, larvae benefit from sunlight as well. During the night, if the temperature is not very low (not lower than -7 °C), they leave the nest to feed [10].

Pine processionary moth causes the most serious defoliation to pine forests when it is in the fourth and fifth instar developmental stages, that happens in between January to March [11]. At this stage, the average weight of larvae is 0.80 g, and the average length is 30mm [12]. Trees attacked by larvae at this stage, will have all their leaves consumed by the pest and have a distinct appearance as if they had been burned [13].

Mature needles of conifers are mainly preferred by larvae, however, if in need, they also feed on young needles which is potentially the main contributing factor for the defoliation of trees, especially in the crown [6]. Forming processions, fully developed larvae abandon their nests and move toward the ground, where they will enter pupal stage. They often burrow very close to the tree they were living on [11].

Defoliation influences the growth and the reproduction capacity of pine trees negatively. Based on the measurements done on some parameters such as the cone production of male and female, mature cone size, seed production and seed weight, it is proven that the reproductivity of these trees is also affected following the defoliation [14]. It is estimated that the growth of pine trees is reduced by 20% the damage caused by *T. pityocampa*. However, the damage can be increased from 30 to 95%, at severe defoliation [15].

In certain conditions, severe and repeated attacks by defoliators can result in the death of trees. Age of the trees is a factor. Young trees are more susceptible in general. Poor soil condition is another contributing factor. Trees weakened by defoliators, are more sensitive to be attacked by secondary pests such as bark Beetles, which can deliver the killer blow [16].

Climate change and the associated temperature rise during winters are the other threat to pine forests. In fact, pine processionary moth is sensitive to temperatures lower than -7 °C, and as the temperature drops, more insects die in their nests, directly or due to starvation [17]. It has been one of the natural controls over the outbreaks of *T. pityocampa*, but with the recent trend of increased temperature during winters,

outbreaks are more expected [18].

Besides detrimental effects on pine forests, *T. pityocampa* is harmful to mammalian species as well. They can release urticating hair that can cause allergic responses in humans, dogs, and cats [19] and the direct contact of airborne material release by the insect can initiate the allergy responses [20, 21].The symptoms mainly appear at the area of the contact. In some severe cases, anaphylactic shocks are also reported [22, 23].

1.4. Pine Processionary Moth in Turkey

In Turkey, the pest insect, *T. pityocampa* mostly damages *Pinus brutia* trees which is the prevalent pine tree in this region. Majority of forests areas of the country are covered by this type, contributing to more than 3 million hectares. The forests are mainly stretched along the coastal line, namely, Mediterranean, Marmara and Aegean regions [3]. According to several records, *T. pityocampa* damages to about 1.5 million hectares of pine forests each year [24]. Figure 1.1 is adapted from Ipekdal *et al.* [25], and it illustrates the distribution of *T. pityocampa* in pine forests of Turkey and North Cyprus.

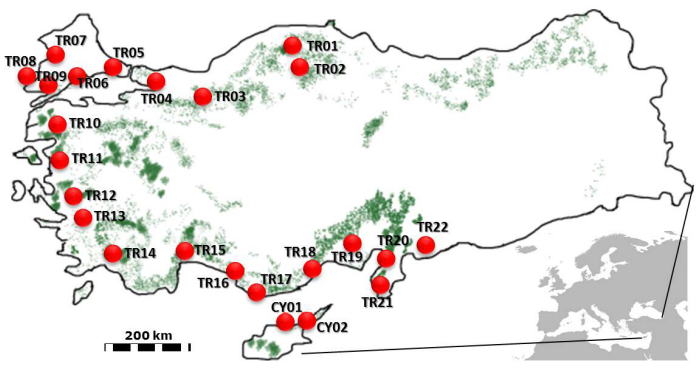


Figure 1.1: Geographical distribution of *T. pityocampa* in coniferous forests in Turkey and North Cyprus [25].

1.5. Available Approaches to Combat the Outbreak of Pine Processionary Moth

In countries such as Turkey that are mostly covered by pine forests, regular protection of coniferous forests against the outbreaks of pine processionary moth is of great importance. Protection methods include chemical, mechanical and physical, biological measures. One of the first lines of the defense is to destroy egg-batches and silk nests physically or chemically. Biological measures include pheromone traps, vegetal oils and entomopathogenic fungi, parasitoids, and microbial pathogens. All the mentioned methods have some limitations.

Physical and mechanical approaches are economically very demanding. Additionally, direct contact to larva puts the health of the labors at risk through causing allergic responses. Using chemical insecticides is not an environmentally friendly approach and results in adverse effects on the ecosystem of the region. This method, as well, is related with health issues. The direct correlation of chemical pesticides with chronic diseases is universally known [26]. Biological approaches such as pheromone traps are proved not to be efficient when applied solely [27]. Some notable results are acquired from biological defenses such as fungi and parasitoids. One of the approaches that has been used in Europe and Mediterranean countries for so long is the aerial application of *Bacillus thuringiensis subsp. kurstaki* [28, 29]. But the main concern regarding biological interventions is disrupting the balance in the ecosystem. The introduced fungi or parasitoid pathogens can affect other beneficial organisms and disrupt the ecosystem balance [30].

1.6. Two Main Hormones Regulate the Post-Embryonic Development of Insects

Hormones play a critical role in the life cycle of insects. The significance of hormones in the regulation of the physiology of insects was first recognized by Vincent Wigglesworth (1899–1994) in 1934.

Post-embryonic development of insects is mainly regulated by two hormones, Juvenile Hormone (JH) and 20-hydroxy-ecdysone (20E) [31]. 20E synthesizes in prothoracic glands, and JH is secreted from Corpora allatum (CA) of insect positioned in the brain. JH takes the responsibility of regulating the physiology of insects throughout late embryonic and larval stages, while 20E is critical for instar larval molting steps and metamorphosis. Before each molting, JH secretion is stopped and available JH in tissues start to be degraded [32]. When JH is at its lowest level, prothoracicotropic hormone (PTTH) starts to be released which, in turn, initiates the secretion of 20E. At this point, larvae of insect molts to the pupal stage. Later release of 20E in the absence of JH causes the adult molting step to proceed (Figure 1.2) [33].

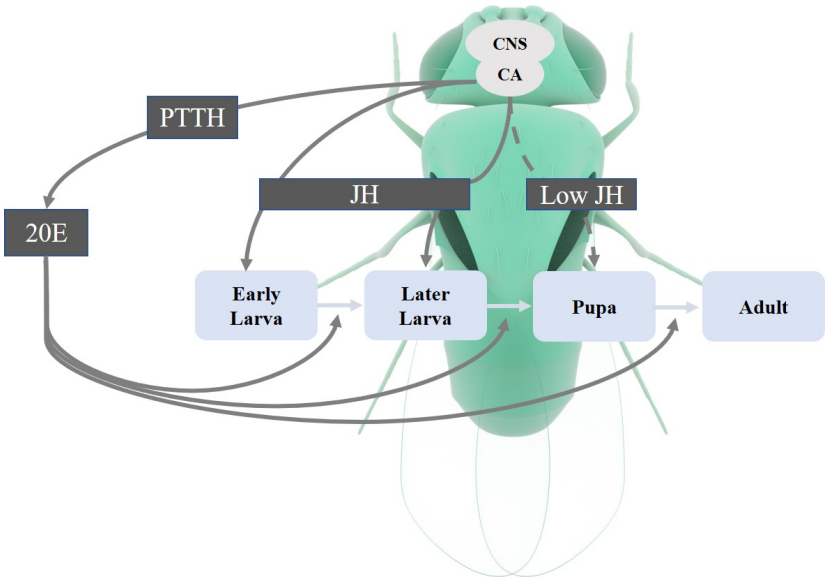


Figure 1.2: Hormones and developmental stages of insects.

1.7. JH and its Effect on the Development of Insects

JH, is a lipid-like molecule released from the endocrine glands positioned behind the brain called CA [34, 35]. In 1967, the chemical structure of JH was identified by Röller and colleagues [36]. At least eight JH-like compounds are found to be produced by insects so far that includes 0, I, II, III, JH3 bisepoxide [JHB3], 4-Methyl-JH, 80-OH-JHIII, 120OH-JHIII. JHIII is the most common type produced in different insect

species [34, 37, 38].

This endocrine regulating hormone influences various physiological and developmental processes such as organ looping, ovarian development, pheromone production, locomotion, migration, diapause regulation memory, learning, food intake and lifespan [39–42].

Due to the profound and versatile effect of the hormone on almost every aspect of insect life, during the recent years manipulation of JH has been proposed as one possible strategy to control the outbreaks of pest species [43]. Induction of the excess amounts of the hormone in larval stage is shown to delay the metamorphosis. Whereas its withdrawal causes precocious metamorphosis [39]. Disruption of the developmental time results in imbalances in the nervous and muscular system [44] and the abdominal differentiation [45].

1.8. Control Mechanisms over JH Release

Secretion of JH is regulated by neurosecretory cells. In the central nervous system (CNS), in response to insulin signaling, insulin-like peptides are produced by median neurosecretory cells called insulin-producing cells (IPC). Insulin production, in turn, directly or indirectly results in the release of JH from CA [46].

Another level of control over JH secretion is exerted through two peptides, named allatotropins (stimulatory peptides) and allatostatins (inhibitory peptides), secreted by neurosecretory cells. These cells are located in the brain and secreted peptides are sent to CA via axonal projections [38].

Besides, physiological controls, environmental factors such as temperature, nutrition and photoperiod have found to influence the release of JH [35]. The general control mechanisms over JH is summarized in Figure 1.3.

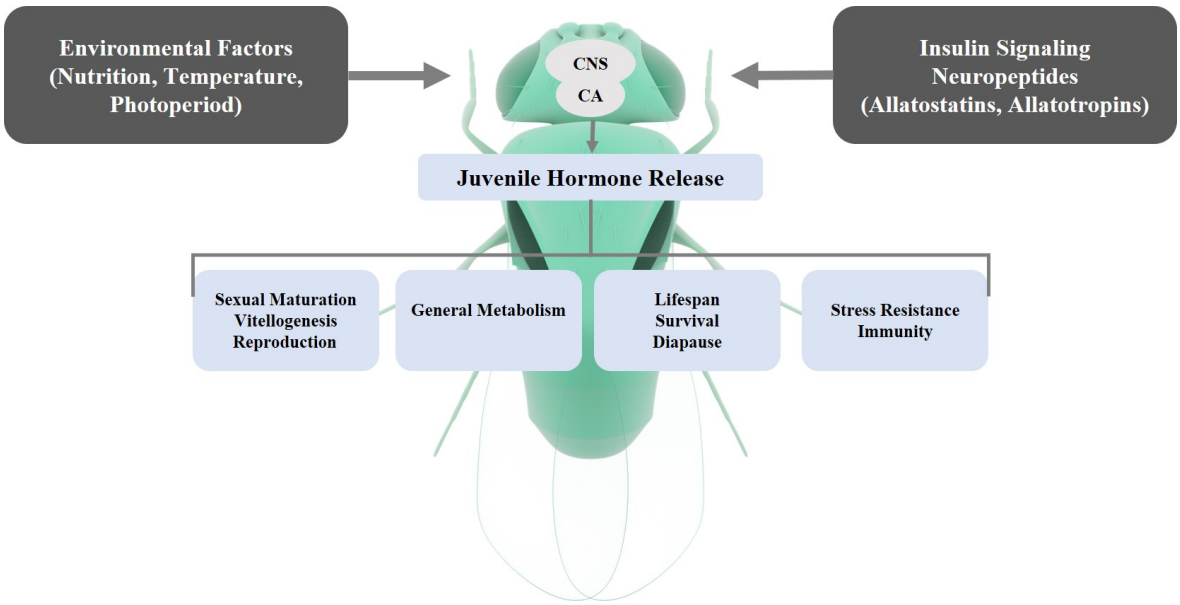


Figure 1.3: JH control mechanisms and downstream effects on the physiology of insects.

1.9. Allatostatin/Juvenile Hormone Pathway

Allatostatins (ASTs) are insect neuropeptides that control many physiological behaviours [34]. These neuropeptides have been proven to inhibit the JH biosynthesis and accordingly regulate insect metamorphosis, reproduction, food intake, growth, and development [47]. Although the inhibitory function of ASTs on JH secretion is very well known and accepted, the exact mechanism of the inhibition is still poorly understood [48, 49].

There are three different AST types in insects categorized according to their conserved motif in their C-terminal core. Type A has FGLamide (FGLa) and type B possesses W(X)6Wamide stretch of amino acids at this region. PISCF stretch of amino acids is conserved in the C type [50].

In a study performed by Nouzova *et al.*in mosquitoes, an inhibitory mechanism of AST on JH production was proposed which is illustrated in Figure 1.4.

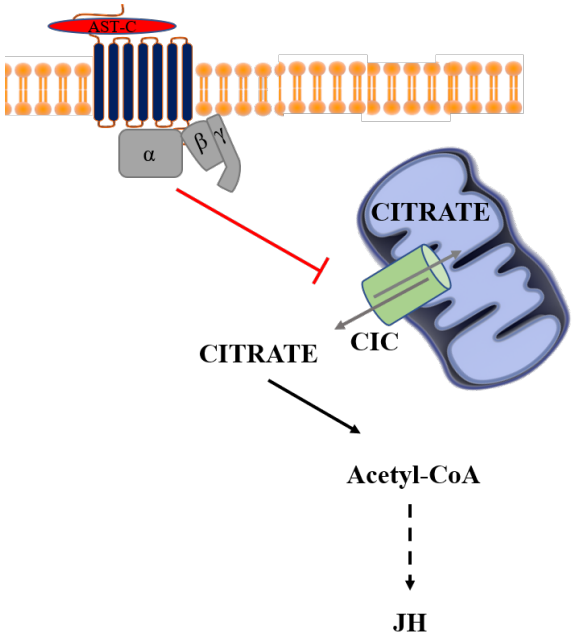


Figure 1.4: The proposed mechanism for the inhibition of JH production after activation of the cognate receptor by AST-C.

The group has shown that AST-C inactivate the mitochondrial transporter protein (CIC) [51], a transporter that has the role of transporting citrate from mitochondria to the cytosol. Citrate is known as an intermediate that is broken down to Acetyl-CoA, that itself is the main building block of the synthesis of JH.

In fact, the sustainable synthesis of JHIII in the cytosol is directly correlated with Acetyl-CoA availability [52]. In summary, upon the secretion of AST-C, CIC is blocked, and citrate is not transported to the cytosol, consequently, Acetyl-CoA is not produced, and JH III as the final product of the cycle will not be synthesized (Figure 1.4).

1.10. ASTs and Their Receptors

Allatoregulatory peptides such as ASTs exert their role through binding to G protein-coupled receptors (GPCRs) of rhodopsin family [53]. The receptors of AST-A and AST-C have been proven to be orthologs of galanin receptor [54] and somato-

statin/opioid receptors [55] in human, respectively .

GPCRs are the largest class of drug targets. One third of all the available drugs target these receptors [56]. GPCRs are seven transmembrane receptors having extracellular and intracellular loops, ECLs and ICLs, respectively. Although all the GPCRs have similar structural architecture, they show low identity in their sequences, and, accordingly they are categorized into different classes or families.

The classification of human GPCRs is done differently. Some have denoted classes A-F based on the overlapping classification [57], while others have categorized GPCRs into glutamate, rhodopsin, adhesion, frizzled and secretin (GRAFS) according to the prototypical members [58]. A separate family is allocated to taste receptors recently [59]. Within each class, GPCRs are further grouped based on various ligands including ions, neurotransmitters, lipids, carbohydrates, nucleotides, amino acids, peptides and proteins [60].

Insect neuropeptide receptors mostly belong to the superfamily of GPCRs. Based on the versatile and profound effect of neuropeptides on the physiology of insects, their cognate receptors can be considered as potential targets for the development of next-generation pesticides. Targeting AST receptor, for instance, can inhibit the release of JH from CA which in turn influences diverse developmental and physiological behaviors of insects.

1.11. Characteristic of GPCRs

A common architecture is shared among all GPCRs. All the members of this receptor family have seven transmembrane (7TM) helices passing through the lipid bilayer and forming three extracellular and three ECLs and ICLs [61]. The ligand binding pocket is in the extracellular face, whereas, G protein coupling site is located at the intracellular part. Although, both regions show high diversity in the shape and sequence, the G protein binding site is more conserved [62].

Due to the inherent flexibility of GPCRs, these proteins can exist in different conformational states, namingly active, intermediate, and inactive. Upon binding of the cognate ligand to the receptor, the equilibrium is disrupted and shifts toward active-like conformation. However, small degree of basal activity observed in some GPCRs indicates that Apo receptors (receptor without a bound ligand) can acquire active-like conformation as well [63].

Some significantly conserved key residues and motifs are identified in class A GPCRs that connect the extracellular pocket to the intracellular G protein coupling site, and enable the transition between different states [64]. CWxP [65], PIF [66], Na+ pocket [67], NPxxY [68], and DRY [69] are the evolutionarily conserved motifs present in the 7TM domain and the intracellular region.

1.12. Downstream Pathways Initiated via GPCR Activation

GPCR activation following the binding of agonist molecule in the extracellular region induces G protein recruitment in an allosteric manner [70]. The 3-dimensional (3D) conformational changes happened in the structure of the protein upon stimulation lead to the exchange of guanosine diphosphate (GDP) to guanosine-5'-triphosphate (GTP) in heterotrimeric G protein complex. G protein is composed of three subunits Gα, Gβ and Gγ. Subsequent to the exchange of GDP to GTP, conformational changes occur in Gα subunit that lead to the dissociation of Gα and G$\beta\gamma$ subunits. After this step, Gα and G$\beta\gamma$ modulate different downstream effector proteins. Adenylyl cyclases, cGMP phosphodiesterases, phospholipase C and RhoGEFs are among the effector proteins regulated by Gα subunit [71]. G$\beta\gamma$, on the other hand, recruits GPCR kinases (GRKs) to the membrane and so regulate G protein coupled inwardly rectifying potassium channels (GIRKs). Voltage dependent calcium channels, adenylyl cyclases, phosphoinosite 3 kinase and mitogen-activated protein kinases are also regulated by this subunit [72]. Upon the hydrolysis of GTP to GDP, the response is terminated, and subunits re-associate.

The attenuation and desensitization of G protein dependent signaling are initiated following the phosphorylation of the intracellular region by GRKs that further recruits arrestins to this the activated receptor. Arrestins for so long have been known for their role as scaffold proteins recruiting clathrin-coated pit machinery and initiating endocytosis of the GPCRs [73].

Four isoforms of these adapter proteins are identified, numbered from 1 to 4, from which two isoforms, arrestin 1 and 4, specifically bind to visual GPCRs. So, only two arrestins, arrestin 2 and 3 bind to nonvisual GPCRs, that are called β-arretsin1 and β-arretsin2, respectively [74]. Besides, their mentioned role in desensitization of GPCRs, signaling role of arrestin is also identified in recent years. These scaffold proteins can recruit various signaling molecules to the GPCRs and activate distinct downstream signaling pathways, that are known as G protein independent pathways. So, arrestins can serve as multifunctional proteins playing roles as adaptor/scaffold or signaling transducers [75]. Two pathways initiated by activated GPCRs are summarized in Figure 1.5.

1.13. Orthosteric Binding Pocket of Class A GPCR

Despite the versatility of ligands that bind to class A GPCRs, the binding pocket of these receptors have a distinct architecture commonly shared within all members. This domain which is referred to as orthosteric binding site is located in the extracellular face of the 7TM and surrounded by ECLs. Different ligands with very diverse chemistry are capable of binding to the orthosteric site. The ligand in class A can be as diverse as monoamines, lipids, acetylcholine, nucleosides and peptides [76–79]. The flexibility of the receptor in binding to different ligands and discriminating them arises from the varied size, shape, and residue composition of orthosteric pocket Within each subclass, the degree of similarity between binding pocket increases. For instance, ECLs in opioid receptors form a structural fold that is similar with all other members of the subclass, however, the sequence composition and similarity is very low [80].

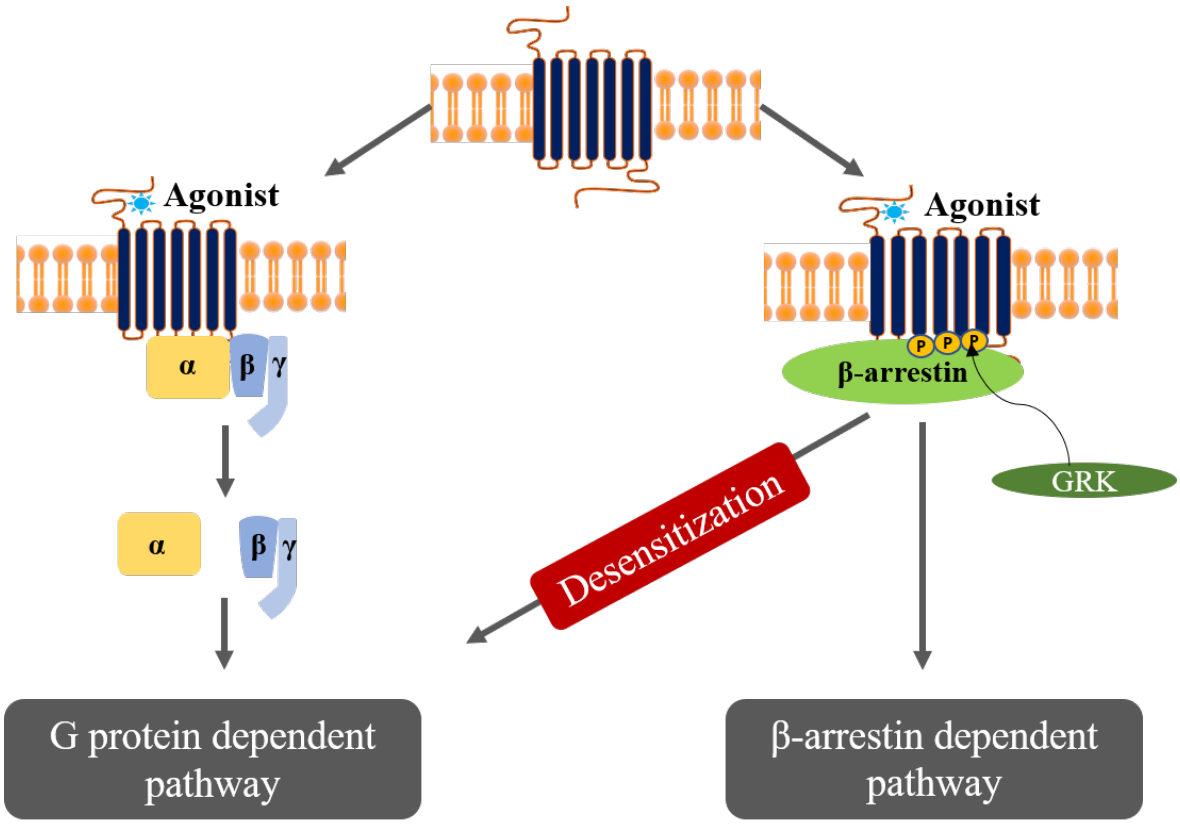


Figure 1.5: Summary of signal transduction by GPCRs.

Some positions at the orthosteric pockets were also shown to be very important for binding of a specific type of ligand to GPCRs. For example, in receptors that have peptide ligand such as angiotensin, endothelin, bradkynin or neurotensin, alignment of the binding site residues have highlighted the presence of positively charged residues at position 4.64, 5.42, and 6.55 (Ballesteros-Weinstein numbering [81]).These residues showed high conservation rate among the subfamily, and importance for receptor activity. In other group of peptides with amidated C-terminal including cholecystokinin, gastrin, orexin, neuropeptide Y, prolactin-releasing peptide, pyroglutamylated RFamide peptide, neuropeptide FF, kisspeptin, vasopressin, oxytocin, bombesin, alignment studies showed that asparagine or glutamine are highly conserved at positions 3.32, 4.60 and 6.55 [82]. Site-directed mutagenesis studies in A2A adenosine receptors revealed the importance of residues positioned at 5.29, 5.30, 6.51 and 6.55 in coordinating the agonist and antagonist molecules bearing bicyclic core [83].

1.14. Agonists, Antagonists, and Inverse Agonists

GPCRs can possess different conformations that in turn determines their state; active, inactive, and intermediate. These receptors do not act like on/off switches and instead there is an equilibrium between different states. A population of receptors at each time are coupled to the downstream effectors that results in the basal activity. Binding of different ligands shifts the equilibrium toward a specific conformation that further determines the downstream signaling pathway [84].

Agonist molecules bind to the orthosteric site and promotes the exchange activity to a level higher than of the basal level and stabilize an active conformation of receptors. Efficacy of a ligand is defined by its ability in activating a specific signaling pathway.

Inverse agonists, on the other hand, favor the inactive conformation and following the binding push the equilibrium toward lower activity compared to the unliganded state. Antagonists, however, have no effect on the basal activity of the receptor and instead compete with agonist and inverse agonists in binding to the orthosteric pocket.

These molecules are further classified into full and partial according to their efficacy which is determined based on the maximal activity obtained. For instance, partial agonists have weaker maximal activity compared to full agonists when applied at the saturation level. The diverse effects of ligands on GPCRs based on the downstream biological response they produce is depicted in Figure 1.6.

1.15. Methods to Study GPCRs

Designing drugs to activate/inactivate GPCRs, as one of the most important pharmacological targets in drug discovery researches, is the final goal of investigations done on these receptors. To this aim, identification and understanding the activation mechanisms are of great importance. Many diverse approaches have been applied so far that includes structural, biophysical, biochemical, and computational methods. X-ray,

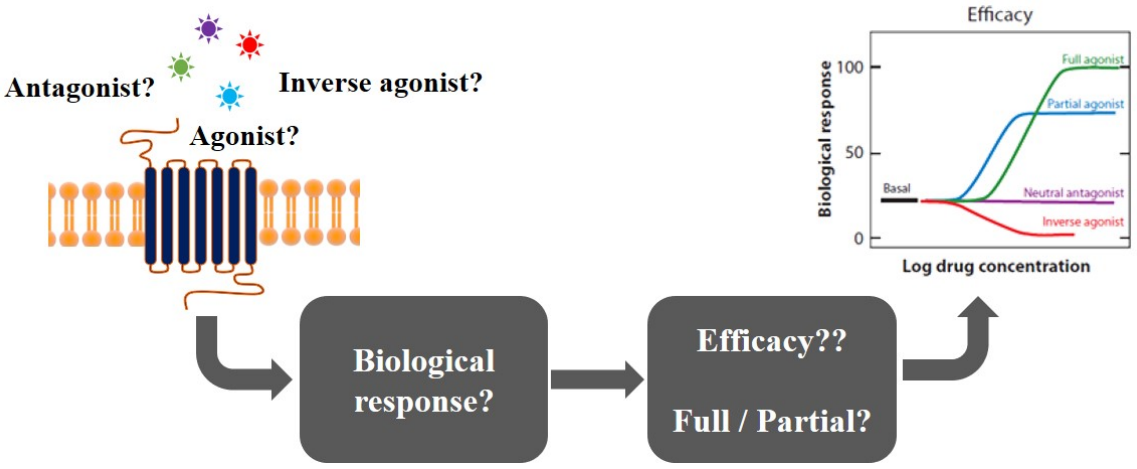


Figure 1.6: How different ligands cause different downstream responses with different efficacy.

NMR , and cryo-electron microscopy (cryo-EM) at the frontline of structural methods were very successful to obtain information about the topology of GPCRs in high resolutions, however, challenges in samples preparation and high quality structure data have slowed down the outcome rate. Methods developed based on resonance energy transfer (RET) such as Förster resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) were also very helpful in providing functional and pharmacological data, and many biosensors have been developed to investigate different mechanisms involved in GPCR activation, including ligand binding, G protein coupling or arrestin recruitment.

Above all, computational approaches such as homology modeling, molecular dynamics (MD) simulations and molecular docking simulations have boosted the speed of GPCR studies significantly. The value of these methods is more appreciated knowing that from about 700 class A GPCRs, to date structure of 54 unique receptor is resolved (data taken from GPCRdb [85]), and others could have been studied owing to the power of computational approaches available. Computational methods can be applied at different stages of drug discovery researches, from model building to binding pocket identification and virtual screenings to hunt for new chemistries for GPCRs.

Taken together, although not sufficient on their own, these methods when combined to other *in vitro* and *in vivo* studies can increase the speed and accuracy of GPCR drug design researches significantly.

2. PURPOSE

Insect neuropeptide receptors are ideal targets for developing next-generation pesticides. Targeting these receptors by activators or inhibitors regulates the downstream signaling pathways that consequently affects the physiology of insects. Despite the great opportunity offered by these receptors in controlling the outbreak of pest species, they have been underexploited, and very low number of these receptors have been identified so far.

The purpose of this study was to characterize AstR-C of pine processionary moth, *Thaumetopoea pityocampa*, a well-know pest in Turkey. Activation of allatostatin receptors in insects result in the inhibition of JH secretion, a hormone that regulates the physiology of insects at the larval stage. According to the promising mechanism offered by the activation of these receptors, we aimed to characterize the structure and function of AstR-C in this specific species. We believed that outcomes of this study will provide precious insights that can be utilized in the future studies aiming to develop small molecules with potential usage as pest control agents.

3. MATERIALS

3.1. General Kits, Enzymes and Reagents

Kits, enzymes and reagents used at different experiments are listed in Table 3.1.

Table 3.1: List of kits, enzymes, and reagents.

Name	**Supplier**
CellMask™	Thermo Fisher Scientific, USA
Dulbecco's Modified Eagle Medium (DMEM)	Gibco, UK
DNA ladder 1kb	GeneOn, Germany
Effectene Transfection Reagent Kit	Qiagen, Germany
Fetal Bovine Serum (FBS)	Gibco, UK
Hoechst 33342 Solution	Thermo Fisher Scientific, USA
ImProm-II Reverse Transcription System	Promega, USA
MinElute Gel Extraction Kit	Qiagen, Germany
NucleoSpin® Plasmid (NoLid), Mini kit for DNA plasmid DNA purification	MACHERY – NAGEL, USA
Penicillin/Streptomycin (10X)	Hyclone, USA
Proteinase K	Thermo Fisher Scientific, USA
Q5® High-Fidelity DNA Polymerase	NEB, USA
Restriction Endonucleases	Thermo Fisher Scientific, USA
NucleoSpin® RNA kit for RNA purification	MACHERY – NAGEL, USA
RNase A Enzyme	Thermo Fisher Scientific, USA
RNase-ZAP	Sigma-Aldrich, Germany
T4 DNA Ligase	NEB, USA
Taq DNA polymerase	Thermo Fisher Scientific, USA

Table 3.1. List of kits, enzymes, and reagents (cont.)

Trypsin-EDTA (0.5 mM EDTA,0.025% Trypsin)	Hyclone, USA
VectaShield Antifade Mounting Medium	Vector Laboratories, USA
ZymoPURE^{TM}II Plasmid Midiprep kit	Zymo Research, USA

3.2. Chemicals

Different chemicals utilized in different experiments are listed in Table 3.2.

Table 3.2: Chemicals used in this study.

Name	**Supplier**
Acetic Acid	Merck, USA
Agar	Conda, Spain
Ammonium Perslfate (APS)	Sigma-Aldrich, USA
Ampicillin	BIOCHROM, Germany
β-mercaptoethanol	Merck, USA
Bovine Serum Albumin (BSA)	AppliChem, Germany
Bromophenol Blue	Fluka, USA
Calcium chloride dehydrate	AppliChem, Germany
Cetyltrimethyl Ammonium Bromide (CTAB)	Sigma-Aldrich, USA
Chloroform	Emsure, Germany
Chloroform: Isoamylalchohol	Emsure, Germany
D-Glucose	Sigma-Aldrich, USA
DMSO	Sigma-Aldrich, USA
EDTA	AppliChem, Germany
Ethanol	Emsure, Germany

Table 3.2. Chemicals used in this study (cont.)

Ethidium Bromide	Sigma-Aldrich, USA
EZMix Tryptone	Sigma-Aldrich, USA
Formamide	Emsure, Germany
Glycerol	Sigma-Aldrich, USA
HEPES	Sigma-Aldrich, USA
Isopropanol	Emsure, Germany
Kanamycin	Fluka, USA
Magnesium Chloride ($MgCl_2$)	Sigma-Aldrich, USA
Methanol	Emsure, Germany
Paraformaldehyde	Sigma-Aldrich, USA
Phenol: Chloroform: Isoamylalchohol	Emsure, Germany
Phosphate Saline Buffer (PBS)	Gibco, UK
Potassium Chloride (KCl)	Sigma-Aldrich, USA
Sodium Acetate (NaOAc)	Sigma-Aldrich, USA
Sodium Chloride (NaCl)	Thermo Fisher Scientific, USA
Sodium Dodecyl Sulfate (SDS)	AppliChem, Germany
Sodium Hydroxide (NaOH)	Sigma-Aldrich, USA
Tris-Base	AppliChem, Germany
Triton X-100	AppliChem, Germany
Tryptone	Sigma-Aldrich, USA
Yeast Extract	Conda, Spain

3.3. Biological Materials

3.3.1. Bacterial Strains

Bacterial strain used in this study was *E. coli* DH5α (genotype: F-Ψ80d lacZΔ-M15- Δ(lacZYA-argF) U169 end A1 recA1 hsdR17 (rk-, mk+) supE44Λ- thi-1 gyrA96

relA1 phoA).

3.3.2. Mammalian Cell Lines

Human embryonic kidney cell line HEK-TSA was used. Cell culture were done in Dr. M.J. Lohse lab, MDC, Berlin, Germany. This cell line was grown in this lab routinely.

3.4. Buffers and Solutions

3.4.1. Genomic DNA Extraction Solution

2X CTAB extraction buffer was used to extract the genomic DNA from the tissue of pine processionary moth. Description of 2X CTAB extraction solution is given in Table 3.3.

Table 3.3: Genomic DNA extraction solutions.

Solution/Buffer	**Content**
2X CTAB extraction buffer	100 mM Tris-HCl 1.4 M NaCl 30 mM EDTA 2% (w/v) CTAB

3.4.2. DNA Gel Electrophoresis

To separate the DNA samples according to their length, these samples were loaded to electrophoresis gel for which the details are provided in Table 3.4.

Table 3.4: DNA Gel Electrophoresis solutions.

Solution/Buffer	**Content**
50X Tris-acetic acid EDTA (TAE)	2 M Tris-acetate 50mM EDTA pH 8.5
TE Buffer	10 mM Tris–HCl 1 mM EDTA, pH 8.0
Ethidium bromide (EtBr)	10 mg/ml (stock solution) 30 ng/ml (working solution)
10X Tris Borate EDTA (TBE)	108 g Tris base 55 g Boric acid 9.3 g EDTA Double ddH_2O up to 1 L
6X Loading Buffer	0.1 ml 1 M Tris-HCl, pH 7.6 0.3 ml 1% Bromophenol Blue (BPB) 6 ml 100% glycerol 1.2 ml 0.5M EDTA up to 10 ml ddH_2O

3.4.3. RNA Gel Electrophoresis

To check the quality of RNA samples and evaluate their integrity, the extracted RNA samples were loaded to a gel and electrophoresis was performed. The details of the gel preparation are given in Table 3.5.

Table 3.5: RNA Gel Electrophoresis solutions.

Solution/Buffer	**Content**
Diethylpyrocarbonate (DEPC) treated water	1% (v/v) Diethylpyrocarbonate
10X Morpholino Propane Sulfonic Acid (MOPS)	41.8 g MOPS 20ml 0.5M EDTA 16.8ml 3M NaOAc DEPC treated water up to 1 L ph 7.0
EtBr RNA Loading Buffer	0.72 ml formamide 0.16 ml 10X MOPS 0.26 ml formaldehyde 0.18 ml DEPC treated water
	0.1 ml 80% Glycerol 0.08 ml Bromophenol blue 50 μg EtBr

3.4.4. Bacterial Culture Solutions and Antibiotics

Bacterial culture solutions were used at different steps, each culture was supplemented with different antibiotics according to the plasmid that was used. The details of bacterial solutions and antibiotics stock preparation is given in Table 3.6.

Table 3.6: Solutions and antibiotics used in bacterial cultures.

Solution/ Buffer	**Content**
Luria-Bertani medium (LB)	10 g tryptophan 5 g yeast extract 10 g NaCl ddH_2O up to 1 L, autoclaved.

Table 3.6. Solutions and antibiotics used in bacterial cultures. (cont.)

Luria-Bertani Agar	10 g tryptophan 5 g yeast extract 10 g NaCl 15 g Agar ddH_2O up to 1 L, autoclaved.
Ampicillin stock	100 mg/ml in 50 % Ethanol Filter-sterilized and stored at -20°C 100 μg/ml (working concentration)
Kanamycin stock	50 mg/ml in ddH_2O Filter-sterilized and stored at -20°C 50 μg/ml (working concentration)

3.4.5. Culture Media

Dulbecco's modified Eagle's medium (DMEM) and fetal bovine serum (FBS) were purchased from Gibco (Paisley, UK). Penicillin/Streptomycin was commercially obtained from BIOCHROM AG (Berlin, Germany).

Cell culture media was prepared by supplementing DMEM with FBS (10%) and Penicillin/Streptomycin (1%).

10X phosphate buffered saline (PBS) solution was used in cell culture at washing step. The details for this solution is provided in Table 3.7.

Table 3.7: Solutions used in cell culture.

Solution/ Buffer	**Content**
10X PBS	80 g NaCl 2 g KCl 14.4 g $Na_2HPO_4.2\ H_2O$ 2.4 g KH_2PO_4 ddH_2O up to 800 ml adjust the pH to 7.4 by HCl ddH_2O up to 1 L.

3.4.6. Staining Solutions

The membrane and nucleus of cells were stained with CellMask^{TM} and Hoechst 33342 Solution, respectively. The detail of these solutions are given in Table 3.8.

Table 3.8: Solutions used in cell staining.

Solution/Buffer	**Content**
Hoechst 33342 Solution	Solubilized in DMSO at 10 mg/ml
1X CellMask^{TM}	10 μL of the stain to 10 mL of PBS

3.4.7. FRET Buffers

In FRET studies, different solutions including 10X FRET buffer and experimental buffers were used. The description of these buffers are provided in Table 3.9.

Table 3.9: Solution used in FRET studies.

Solution/Buffer	**Content**
10X FRET Buffer	10 mM HEPES 140 mM NaCl 5.4 mM KCl 1 mM $MgCl_2$ 2 mM $CaCl_2$ Adjust pH to 7.3 ddH_2O up to 1 L.
Experimental Buffer	PBS supplemented with 0.1% BSA (v/w)

3.5. Nucleic Acids

3.5.1. Plasmids

Two empty plasmids were used in this study. pEYFP-N1(Clontech, CA, USA) and pcDNA3.1(+) (Invitrogen, CA, USA) were commercially obtained.

3.5.2. FRET and BRET Biosensors

Different FRET biosensors were used including Gαi_1, Gαi_2, Gαi_3 [86], Gαq [87], Gαs [88] and Gα_{13} [89]. In these biosensors, Gα subunit is tagged with mTurqoise2 and Gγ is tagged with mVenus. Human Gβ_1-wt and bovine Gγ_2-wt were also used. For BRET studies, GRK_2 (G protein Receptor Kinase 2) and β-arrestin2-Nluc [90] were utilized.

3.5.3. Primers

Primers used in polymerase chain reactions (PCRs), sequencing and cloning were purchased from Macrogen Europe (AZ, Netherlands). Primers used in cloning of the WT receptor in pcDNA3.1(+) and p-EYFP-N1 plasmids are given in Table 3.10. The bold characters represent restriction enzyme cutting sites (RE).

Table 3.10: Primers used in cloning and sequencing.

Primer Code	Sequence (5'-3)	Application	RE
WT-F	5'- **CTCGAG**ATGGAGCTCGAA -3'	Cloning	HindIII
WT-R	5'- **AAGCTT**GAGTCGCGAATG -3'	Cloning	BamHI
sYFP-F	5'-**AAGCTT**ATGGAGCTCGAAGAC- 3'	Cloning	XhoI
sYFP-R	5'- **GGATCC**TCAGAGTCGCGAAT-3'	Cloning	HindIII
T7	5′- TAATACGACTCACTATAGGG-3′	Sequencing	-
SP6	5′-ATTTAGGTGACACTATAG-3′	Sequencing	-
CMV-F	5'-CGCAAATGGGCGGTAGGCGTG-3'	Sequencing	-
GFP-2R	5'- GGACCTGTGTATTGTGTGGC-3'	Sequencing	-

The sequence of the primers used in site-directed mutagenesis (SDM) and their codes are given in Table 3.11.

Table 3.11: Primers used in SDM studies.

Primer Code	Sequence (5'-3')
D181A-F	5'-CGCTTATAGTGGCTAACAATGGAC-3'
D181A-R	5'-GTCCATTGTTAGCCACTATAAGCG-3'
N182A-F	5'-CGCTTATAGTGGATGCCAATGGACCTTCG-3'
N182A-R	5'-CGAAGGTCCATTGGCATCCACTATAAGCG- 3'
N188A-F	5'-GGACCTTCGTGTGCCATTGTGTGGC- 3'

Table 3.11.List of primers used in SDM studies (cont.)

N188A-R	5'-GCCACACAATGGCACACGAAGGTCC- 3'
Q200A-F	5'-GATTTAAACAAAGGTGCGACTACCTTTACTC-3'
Q200A-R	5'-GAGTAAAGGTAGTCGCACCTTTGTTTAAATC-3'
Q271A-F	5'-GTACTGGGCGTGCGCAGTCGCTCTGATC- 3'
Q271A-R	5'-GATCAGAGCGACTGCGCACGCCCAGTAC- 3'
Q278A-F	5'-CATTCCTCATTATCGCAATGTCAGTCCG- 3'
Q278A-R	5'-CGGACTGACATTGCGATAATGAGGAATG- 3'

3.6. Peptide

AST-C of *T. pityocampa* was ordered from GenScript Biotech, New Jersey, USA.

3.7. Molecules

Small molecules selected from OTAVA peptidomimetics library were purchased from OTAVA chemicals.

3.8. Online Tools

AUGUSTUS, CCTOP, ClusPro, Clustal Omega, Cutadopt 1.6, I-TASSER, NEB Tm Calculator, Pfam, PrimerX, Geneious, SignalP 5.1, SWISS-MODEL, tblastn, tblastx and TMHMM were utilized in various procedures.

3.9. Disposable Labware

The list of disposable labwares utilized at different experiments performed in this study is given in Table 3.12.

Table 3.12: List of disposable labware used in this study.

Name	**Supplier**
Centrifuge Tubes (15 ml, 50 ml)	TPP, Switzerland
Microfuge Tubes (1.5 ml, 2 ml)	Axygen, USA
PCR Tubes (0.2 ml)	Axygen, USA
Cryovial Tubes	Grenier Bio One, UK
Cell Culture Plates (10 cm, 6-well, 96-well)	TPP, Switzerland
Cell Scraper	TPP, Switzerland
Serological Pipettes (5 ml, 10 ml, 25 ml)	Capp, Denmark
Syringes (10 ml, 50 ml)	Set Medikal, Turkey
Insulin syringes (1 ml)	Set Medikal, Turkey
Pipette Tips (Filtered)	Capp, Denmark
Pipette Tips (Bulk)	Axygen, USA
96-well plates, Black Flat Bottom	GmbH, Germany
96-well plates, White Flat Bottom	GmbH, Germany

3.10. Equipment

Different equipment utilized in this study are listed in Table 3.13.

Table 3.13: Equipment used in this study.

Name	**Supplier**
Agarose Gel Electrophoresis	Mini-sub Cell GT, BioRad, USA
Autoclave	Midas 55, Prior Clave, UK

Table 3.13. Equipment used in this study (cont.)

Analogue-digital converter	Digidata 1440A, Axon Instruments, USA
Balances	DTBH 210, Sartorius, Germany
Carbon dioxide tank	Genc Karbon, Turkey
Cell culture incubator	Hepa Class 100, Thermo, USA
Centrifuges	ultracentrifuge J2MC, Beckman Coulter, USA
	Mini Centrifuge 17307-05, Cole Parmer, USA Centrifuge 5415R, Eppendorf, USA Centrifuge, Allegra X-22, Beckman Coulter,USA
Cold room	Birikim Elektrik Soğutma, Turkey
Deepfreezers	-20 °C, Arçelik, Turkey -70 °C Freezer, Harris, UK -80 °C ULT Freezer, Thermo Fisher Scientific, USA -150 °C, MDF-1156, Sanyo,Japan
Documentation System	Gel Doc XR System, Bio-Doc, Italy G-BOX Chemi XX6, Syngene, UK
Freezing Container	Mr.Frosty, Nalgene, Thermo Fisher Scientific, USA
Heat blocks	DRI-Block DB-2A, Techne, UK
Hemocytometer	Improved Neubauer, Weber Scientific International Ltd, UK
Ice Machine	Scotsman Inc., AF20, Italy
Laboratory Bottles	Isolab, Germany
Laminal Flow Cabinet	Labcaire BH18, UK
Magnetic Stirrers	M221 Elektro-mag, Turkey
Micropipettes	Finnpipette, Thermo Fisher Scientific, USA

Table 3.13. Equipment used in this study (cont.)

Microscopes	Axiovert 200 inverted microscope, Zeiss, GERMANY, with an oil immersion objective (plan-NEOFLUAR 63x/1.25). iXon Ultra EMCCD camera (Andor, Belfast, UK)
	Leica DMi8 inverted microscope, Leica Microsystems, Germany) Dichroic beamsplitter T505lpxr Visitron Systems Puchheim, Germany Emission filter 470/24 nm and YFP emission filter 535/30 nm (Chroma technology corp.) Leica TCS SP5 or TCS SP8 with an oil immersion objective (HC PL APO 63x/1,40-0,60 oil), Leica Microsystems, Germany.
Microwave oven	M1733N, Samsung, Malaysia
Plate Reader	Synergy Neo2 plate reader BioTek Instruments, equipped with a monochromator.
pH meter	WTW pH330i, Germany
Pipettor	Pipetus-akku,Hirscmann Labogerate, Germany
Power Supply	Biorad, USA
Refrigerators	2082C, Arçelik, Turkey 4030T, Arçelik, Turkey
Shakers	VIB Orbital Shaker, InterMed, Denmark Lab-Line Universal Oscillating Shaker, USA
Software	Bio3D package 2.2.0 Clampex 10.3 Excel (Microsoft office 365)

Table 3.13. Equipment used in this study (cont.)

	FastQC, version 0.11.7 FASTX-toolkit program GraphPad 6.07, California, USA ImageJ, Image Analysis Software, NIH, USA Maestro molecular modeling package MaSurCA assembler Quantity One, Bio-Rad, Italy R studio SGA assembler Visiview 4.0 imaging software (Visitron Systems) VMD XStella 1.0, Stella, Germany
Spectrophotometer	NanoDrop 1000, USA
Thermocyclers	Applied Biosystems 2720 Thermal Cycler BIORAD DNAEngine Peltier Thermal Cycler
Vacuum pump	KNF Neuberger, USA
Vortex	Vortexmixer VM20, Chiltern Scientific, UK
Water baths	TE-10A, Techne, UK
Water purification system	UTES, TURKEY

4. METHODS

4.1. Whole Genome Sequencing of Pine Processionary Moth

4.1.1. Sampling of Pine Processionary Moth

Larvae of pine processionary moth or *T. pityocampa* was gathered from Kemerburgaz Forest, Istanbul, Turkey. Insects were anesthetized using CO_2 and decapitated. Immediately after the decapitation, head tissue was transferred to liquid nitrogen and with a pestle and mortar it was smashed. Head tissue samples were stored at -80 °C for later usage.

4.1.2. Genomic DNA Isolation

Tissue samples prepared at section 4.1.1 was used. genomicDNA (gDNA) of the insect was extracted from head tissue using CTAB method [91]. Cetyltrimethyl ammonium bromide (CTAB) (Sigma-Aldrich, USA) is a cationic detergent. 2X CTAB extraction buffer was prepared. Before adding the tissue sample to the prepared CTAB buffer, 2 U of proteinase K (Thermo Fisher Scientific, USA) was added. Then, 40 mg of the frozen head tissue was mixed in the prepared mixture of CTAB buffer and proteinase K. The mixture was vortexed gently and was incubated for 20 to 30 min at 50 °C in water bath. During the incubation, the mixture was inverted 3 to 5 times each 5 min. The mixture was then transferred to another water bath at 65 °C and incubated for 15 min. Again, inversions were done each 5 min. Then, it was cooled down slightly and 500 μl of cold chloroform: isoamylalcohol (Emsure, Germany) was added on top, then, inverted 2 to 3 times. The mixture was put at shaker and shook gently for 15 min at room temperature. After the incubation, the sample was centrifuged for 10 min at 12,500 rpm and the clear supernatant (400 μl) was transferred to a fresh tube. To precipitate the DNA, ethanol participation was done in which 2X volume of 95% ethanol and 0.1X volume of 3 M sodium acetate was added to the supernatant and the

tube was inverted gently. The mixture was incubated at -80 °C overnight and the next day, it was taken out of the deep freezer, thawed briefly, and spun at maximum speed at 4 °C. The supernatant was discarded carefully, and the pellet was washed with 1 ml of cold 70% ethanol and spun 5 min at 4 °C. Again, the supernatant was discarded, and the wash step repeated. The pellet was dried and then eluted.

4.1.3. RNA Elimination and gDNA Purification

Isolated gDNA was incubated with 1 μl RNase A, for 30 min at 37 °C. Equal volume of Phenol: Chloroform: Isoamylalchohol (25:24:1) (Emsure, Germany) was added and the sample was centrifuged at 14000g for 5 min. The upper phase was taken into a new tube. Then, ethanol precipitation was done. 2X volume of 95% ethanol along with 0.1X volume of ammonium acetate (7.5M) were added to the taken upper phase. The sample was incubated overnight at -80 °C. In the next morning, the sample was thawed briefly and was pelleted by spinning at maximum speed for 10 min at 4 °C. Then the pellet was washed by cold 70% ethanol and spun at full speed for 5 min at full speed. The supernatant was taken carefully, and the wash step was repeated. The pellet was left to dry and then re-suspended in 100 μl of nuclease free water.

4.1.4. DNA Quality and Integrity Control

The quality of the isolated gDNA was checked using nanodrop spectrophotometer. The gDNA sample was loaded on 0.7% agarose gel and run at 70 V.

4.1.5. Library Construction and Whole Genome Sequencing

This part is done by Macrogen Korea company. The non-human Hiseq X platform, Truseq PCR-free with 350 bp inserts, 150 bp paired-end sequencing method was used. 1 g intact gDNA was required for this method. To construct the sequencing library, the gDNA sample is fragmented randomly and the adapters are ligated to the

5' and 3' end of fragments. For cluster generation, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing.

4.1.6. Raw Data Control

Raw data were used for *de novo* assembly. FastQC, version 0.11.7 program was used to evaluate the quality of the sequenced data [92]. FASTX-toolkit program (Hannon Lab) was used to trim the reads and clip the primer and adapters, if present.

4.1.7. *De novo* Assembly

Sequence reads were preprocessed with Cutadapt 1.6 [93]. String Graph Assembly (SGA) [94] and Maryland Super Read Cabog Assemble (MaSuRCA) programs [95] were used for the De novo assembly. *Kmer* which is the number of the overlapping nucleotides at the edge of reads that results in the larger DNA fargments was determined during the procedure. The assembly statistics including total base identified by each program, maximum scaffold span, *N50* scaffold span of assemblers was compared. *N50* is the sequence length of the shortest contig at 50% of the total genome length.

4.1.8. Genome Comparison

The genome data of eight species belonging to Lepidoptera family deposited in i5K database, including *Agrotis ipsilon* (*A. ipsilon*), *Helicoverpa armigera H. armigera*, *Helicoverpa zea* (*H. zea*), *Heliothis virescens* (*H. virescens*), *Mamestra configurata* (*M. configurata*), *Spodoptera frugiperda* (*S. frugiperda*), *Spodoptera litura* (*S. litura*) and *Trichoplusia ni* (*T. ni*) were compared with *T. pityocampa* in terms of GC content and genome size.

4.1.9. Phylogenetic Tree Construction

The phylogenetic tree was constructed using Geneious alignment, Version 7.1 [96]. The alignment of protein sequences was conducted using the Neighbour-Joining method with 1000 reiterations. The input file was prepared by aligning the protein sequence of cytochrome oxidase subunit I (partial sequences) of the above mentioned species and *Drosophila melanogaster* (*D. melanogaster*) in the Clustal Omega online tool [97].

4.1.10. Sequence of AstR-C and AST-C in *T. pityocampa*

The nucleotide sequence of AstR-C and AST-C of *T. pityocampa* were derived from the whole genome sequencing data. The sequence of AstR-C of *T. pityocampa* will be denoted as "TpitAstRC" in the rest of the text. The sequence of AstR-C and AST-C of *D. melanogaster* and *H. armigera* were extracted from databases and queried against the new assembly contigs. NCBI-tblastn was applied with default parameters, except the E-value was adjusted to 10^{-3}. Smith-Waterman optimal alignments were achieved. The collected hits were then filtered using a Perl script to find out non-redundant orthologs, with an additional filter of identity higher than 50% and coverage higher than 50%. Afterwards the candidate genes and gene fragments were manually extended and curated using raw sequence reads. AUGUSTUS online tool [98] was used to identify 5'UTR and 3'UTR including introns. Nucleotide sequences were translated to protein and aligned with ortholog proteins of *D. melanogaster* and *H. armigera* in Clustal omega online tool [99]. To determine the protein family of the receptor, the sequence of the receptor, TPitAstRC was checked in pfam [100]. Conserved residues and motifs were investigated in the sequence. The topology of TpitAstRC was also evaluated in CCTOP [101]. The preprohormone sequence of the native ligand, AST-C was subjected to SignalP version 5.1 to obtain the mature peptide sequence [102]. Dibasic cleavage was performed between two Arginines accroding to the rules suggested by Veenstra *et al.* [103].

4.1.11. Data Deposition in NCBI

The Whole Genome Shotgun project and the coding sequence of TpitAstRC and AST-C were deposited at DDBJ/ENA/GenBank of NCBI.

4.2. Molecular Cloning

4.2.1. RNA Isolation

Tissue samples prepared at Section 4.1.1 was used. 0.4-0.5 mg of the head tissue sample was weighted and transferred to a clean tube. 1 mL Trizol was added to the sample and the mixture was smashed into finer particles and homogenized in Magna-Lyzer. Without touching the pellet, homogenates were taken into another tube. Then, it was incubated at room temperature for 5 min. Chloroform (Emsure, Germany) in a 1:5 Trizol:Chloroform ratio was added to the mixture and it was vortexed for 20 sec. It was followed with an incubation step at room temperature for 2 min. Centrifugation was applied at 10000 g for 18 min. The supernatant was taken into a microcentrifuge tube and 1 volume of 100% ethanol was added. The sample was inverted 6 times. 700 μl of it was transferred to a NucleoSpin® RNA Column, and the procedure was followed as recommended in the protocol of NucleoSpin® RNA (Macherey Nagel, Germany).

4.2.2. MOPS Gel Electrophoresis

All RNA analyses were exerted under RNase-inhibiting conditions. Surfaces and instruments were cleaned using RNase-ZAP (Sigma-Aldrich, Germany). At all steps, DEPC-treated water was used. MOPS gel was prepared in 0.8% ratio. Before loading samples to gel, 400 ng of it was denatured in denaturing mixture. Denaturation was done at 55 °C for 15 min. 4X loading dye was added to the samples. After mixing, samples were loaded to MOPS gel. Electrophoresis was done at 60 V for 40 min. Samples with high integrity were checked for quantity in nanodrop spectrometer.

4.2.3. Reverse Transcription of TpitAstRC

complementary DNA (cDNA) synthesis was done using ImProm-II Reverse Transcription System, Promega (Madison, USA) according to manufacturer's recommendations. This obtained cDNA library was used in reverse transcription poly chain reaction (RT-PCR) method to amplify the open reading frame of TpitAstRC.

4.2.4. RT-PCR

Using the cDNA obtained in Section 4.2.3, PCR reaction was prepared as it is provided in Table 4.1 and run according to Table 4.2. For cloning the WT receptor in pcDNA3.1(+) plasmid and pEYFP-N1, WT-F and WT-R pair and sYFP-F and sYFP-R pair of primers for which the details are given in Table 3.10 were used, respectively.

Table 4.1: PCR reaction reagents and amounts.

Component	**Final concentration**	**Volume for 20 μl Reaction**
5X Q5 Buffer	1X	4 μl
10 μlM Forward Primer	0.2 μM	0.4 μl
10 μlM Reverse Primer	0.2 μM	0.4 μl
dNTPs (2.5 mM each)	50 μM each	0.4 μl
Q5 DNA Polymerase (2U/μl)	0.4 U	0.4 μl
100% DMSO	20%	0.4 μl
ddH_2O		Up to 20 μl

The melting temperature of each calculated via online tool of NEB TM calculator, version 1.9.10. The fragment with the desired length was excised from the gel.

Table 4.2: PCR reaction conditions.

Step	Temperature	Time	Cycle
Initial Denaturation	98 °C	30 sec	X1
Denaturation	98 °C	10 sec	X30
Annealing	Calculated for each pair of primers	15 sec	X30
Extension	72 °C	20 sec/1 kb	X30
Final Extension	72 °C	7 min	X1

4.2.5. Extraction of DNA Samples from Agarose Gel

Samples amplified by RT-PCR loaded into in 1% agarose gel and run in 90 V for 30 min. The desired product size observed in gel was extracted. Agarose Gel DNA Extraction Kit (Sigma-Aldrich, Germany) was used according to the protocol and the DNA was extracted.

4.2.6. Digestion of DNA and Ligation

For each cloning, PCR product and the relevant plasmid were digested with restriction enzymes according to the Table 3.10. Ligation of inserts and linearized plasmid was performed using T4 DNA Ligase (NEB, USA) according to manufacturer's instruction. In brief, 50-100 ng digested plasmid and inserts were mixed with 1:1 or 1:3 molar ratio for each 10 μl of reaction volume. The mixture was incubated at room temperature for 30 min. Then, enzyme inactivation was done at 65 °C for 10 min.

4.2.7. Preparation of Chemically Competent *E. coli* Cells

Glycerol stock of *E. coli* DH5α strain aliquoted previously were taken and added to 5 ml LB medium (inoculation). Culture was applied to overnight growth at 37 °C

while shaking at 200 rpm. Next morning, 250 μl of the culture grown overnight was taken and 25 ml LB was added to it. The inoculated medium was applied to further growth in shaker until optical density (OD) reached to value between 0.4 to 0.6 when measured at 590 nm. At this point, the culture was centrifuged for 10 min at 4000 g, 4 °C. Supernatant was discarded and 2.5 ml ice-cold sterile 50 mM $CaCl_2$ was added to the pellet, and it was resuspended. 10% glycerol was added to the resuspended cells, and aliquoted in 100 μl. Aliquots were immediately transferred to liquid nitrogen and stored at -80 °C.

4.2.8. Transformation of the Chemically Competent DH5α

Previously prepared competent cells were taken out of freezer and for each cloning one vial was used. Before adding ligation mixture, vial was kept for 15 min on ice. After the addition of the 10-50 ng plasmid of 15-20% of ligation mixture to the thawed competent cell, the mixture was incubated on ice for 20-30 min. Heat shock step was done by placing tubes at 42 °C for 45-60 sec following with 2 min of ice incubation. Then, 500 ml LB medium was added on top and the mixture was incubated at 37 °C for 1 hr while shaking at 200 rpm. After the incubation, 100-200 μl of cell was taken and spread on relevant antibiotic containing LB agar plates. The upside-down plates were then incubated overnight at 37 °C.

4.2.9. Colony PCR

To screen the positive colonies, those bearing the desired DNA fragment, Colony PCR was performed. Colonies were selected from LB agar plates and mixed in PCR solution prepared beforehand. The PCR mixture was prepared according to Table 4.3 For colony PCRs, Taq polymerase (Thermo FIsher Scientific, USA) enzyme was used, and other components were added. The mixture was applied to amplification process done in a Thermal cycler.

Table 4.3: Colony PCR reaction reagents and amounts.

Component	**Final concentration**	**Volume for 25 μl Reaction**
10X Taq Buffer	1X	2.5 μl
10 μM Forward Primer	0.4 μM	1 μl
10 μlM Reverse Primer	0.4 μM	1 μl
dNTPs (2.5 mM each)	50 μM each	0.4 μl
Taq DNA Polymerase	0.4 U	0.4 μl
100% DMSO	5%	1.25 μl
$MgCl_2$ (20 mM)	2 mM	2 μl
ddH_2O		Up to 25 μl

The condition of PCR is given in Table 4.4. Then, the products were loaded into 1% agarose gel and electrophoresis was run at 90 V for 30 min.

Table 4.4: Standard PCR protocols for the Taq DNA Polymerases.

Step	**Temperature**	**Time**	**Cycle**
Initial Denaturation	95 °C	30 sec	X1
Denaturation	95 °C	20 sec	X30
Annealing	Calculated for each pair of primers	30 sec	X30
Extension	72 °C	1 min/1 kb	X30
Final Extension	72 °C	7 min	X1

4.2.10. Plasmid Purification

Positive colonies identified by colony PCR were grown overnight in LB to which suitable antibiotic is added. Next morning, plasmid purifications was performed using NucleoSpin Plasmid Miniprep kit (Macherey-Nagel, UK) and ZymoPURE^{TM}II Plasmid Midiprep kit (Zymo Research, USA) for mini and midi/maxi isolation scales, respectively, according to the manufacturer's protocols. Plasmids intended to be used in transfections were purified with midi kit. Spectrophotometric measurements were done to check the quality of the isolated plasmids. OD was measured and those with OD260/280 between 1.8-2.00 were considered as good quality.

4.2.11. Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was used to substitute some residues with Alanine. Using the plasmid in which WT sequence of TpitAstRC is cloned (pcDNA3-WT) as the template, explained in Section 4.2.4, and primers listed in Table 3.11, different mutations were introduced in the binding pocket of TpitAstRC. A modified PCR reaction was performed in which the forward and reverse reaction mixture were prepared in two different tubes, each having one of the two primers (forward or reverse), and then each was subjected to 10 cycles of PCR. Then, after 10 cycles, contents of two tubes were mixed, making the final volume 50 μl, and PCR reaction was continued for 14 cycles. The details of PCR mixtures are given in Table 4.5.

Table 4.5: PCR reaction mix for SDM.

Component	**Forward Reaction**	**Reverse Reaction**
Q5 reaction buffer	5 μl	5 μl
10 μlM Forward Primer	1.25 μl	-
10 μlM Reverse Primer	-	1.25 μl
dNTPs (2.5 mM each)	1 μl	1 μl
Q5 DNA polymerase (2U/ μl)	0.25 μl	0.25 μl

Table 4.5. PCR reaction mix for SDM. (cont.)

100% DMSO	2 μl	2 μl
$MgCl_2$ (20 mM)	2 mM	2 μl
Template pcDNA3-WT	50 ng	50 ng
ddH_2O	Up to 25 μl	Up to 25 μl

PCR product was a mixture of parental plasmid and the product having the point mutation. To remove the parental copy, PCR product was digested overnight using 7 μl of dpnI enzyme (Thermo Fischer Scientific, USA) which specifically cuts at the methylated sites and consequently would digest the template plasmid (Figure 4.1). In the next morning, the digestion reaction was deactivated by incubating the digested product at 37 °C for 20 min. Transformation and Plasmid Isolation were performed as explained in Section 4.2.8 and 4.2.10. Samples were sent for sanger sequencing to Macrogen Korea company.

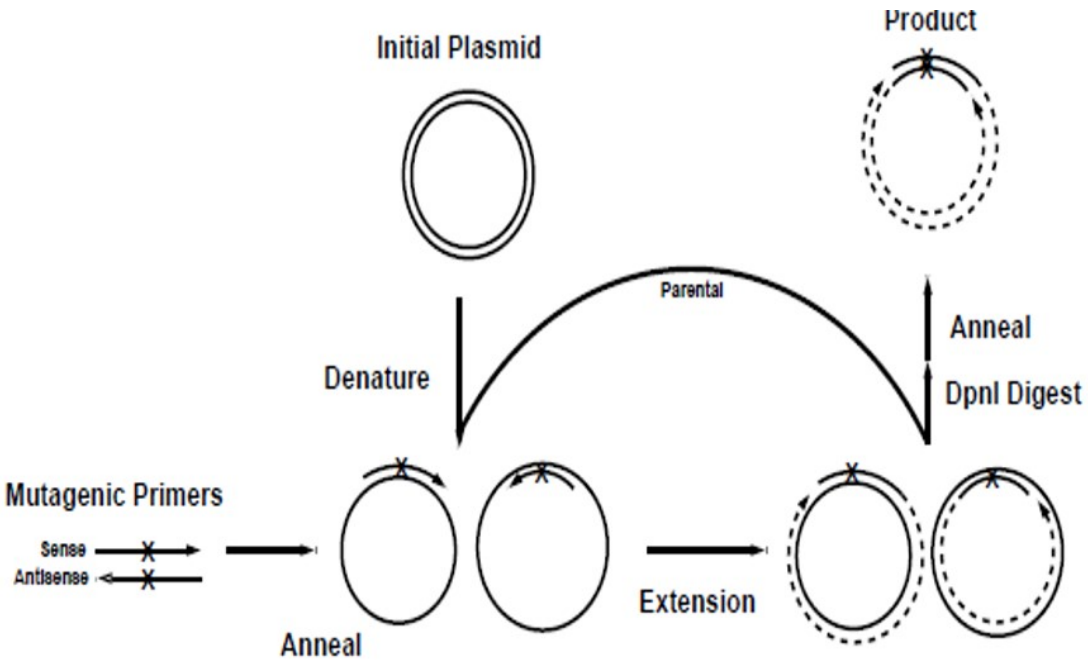


Figure 4.1: Schematic representation of PCR product and digestion.

4.3. Studies on Mammalian Cell Culture

4.3.1. Growth and Maintenance of HEK-TSA Cells

HEK-TSA cells were grown on 10 cm cell culture dish. Cell culture medium was prepared using Dulbecco's modified eagle medium (DMEM) (Gibco, UK) to which 10% FBS (Gibco, UK), 1% Penicillin/Streptomycin (Hyclone, USA) and 4.5 g/l glucose (high glucose) were added. Complete media was stored at 4 °C, and before each use, it was incubated at sterile water bath at 37 °C to reach to this temperature. Cells were cultivated in incubator at temperature of 37 °C supplemented with 5% CO_2.

4.3.2. Passaging

Cells were routinely grown, and before reaching to 90% confluency were passaged to new cell culture dishes. The passaging is started with aspirating the old medium. Then, cells are washed with 1X calcium and phosphate free PBS. To detach the adherent cells from the dish surface, trypsin (0.025%, ready to use) was added to cells and petri dish was incubated at 37 °C for 3-5 min. Then, 2-3 volumes of the added trypsin, complete DMEM was added to the cells to inactivate the enzymatic activity of trypsin. Cells were resuspended, dispersed, and transferred to 50 ml falcon. Then, cells were centrifuged, and the supernatant was discarded. They were resuspended in fresh complete DMEM and dispersed in 1:5 ratio.

4.3.3. Cryopreservation

To preserve the grown cells, when at 90% confluency, they were harvested and trypsinized. Then, trypsin was neutralized by adding 10 volumes of growth medium. In heamocytometer, cells were counted. Then, precipitated by centrifugation for 3 min at 500 g. Meanwhile, the freezing medium (5% DMSO, 10% FBS and 85% DMEM) was prepared, and 1 ml of it was transferred into 2 ml screw capped-cryotubes. 1-2 million cells were transferred to each vial. Freezing-box was filled with isopropanol and

cryo-tubes were placed in it. The box was kept at -80 °C overnight. The next day, cell stocks were translocated to the storage tank filled with liquid nitrogen.

4.3.4. Thawing

When new stock of cells was needed, one vial was taken from the nitrogen tank. Before use, it was incubated at 37 °C water bath and thawed immediately. In a falcon to which 4 ml fresh medium was put, cells were added, and they were precipitated at 500 g for 3 min. The supernatant was discarded, and resuspension was done in an adequate amount of fresh complete medium. Then, cells were transferred to a fresh petri dish and incubated at incubator.

4.3.5. Transient Transfection

Transient transfected cells were used in RET-based functional assays. For G protein activation assay, dose response curve (DRC) and temporal kinetics of the activation, pcDNA3-WT and different Gα biosensors were used. However, for G protein recruitment assay sYFP-WT plasmid was substituted by pcDNA3-WT plasmid. Modified versiosn of G biosensors were used. In β-arrestin kinetics studies, sYFP-WT and β-arrestin plasmid were used for the transfection. The relevant amounts and engineering of plasmids will be explained in the due sections, thoroughly. In general, Effectene Transfection Reagent Kit (Qiagen, Germany) was used for as transfection reagent following the manufacturer's instructions. The amount of plasmid and the volume of were determined according to the manufacturer's protocol. Before each transfection, cells were seeded one day before, and at 70-80% confluency, they were transfected by the transfection mixture.

4.3.6. Staining Cells

HEK-TSA cells were seeded in 6-well plates on Poly-L-Lysine (PLL) (Sigma-Aldrich, Germany)-coated cover slips and transfected as described above. Cells were

washed once with pre-warmed 1X PBS 24 hr after transfection and fixed with 1 ml ice-cold 4% paraformaldehye (PFA) for 20 min at room temperature, and then washed 3 times with warm 1X PBS. 1X CellMask^{TM} (Thermo Fisher Scientific, USA) and Hoechst 33342 Solution (Thermo Fisher Scientific, USA) were used according to the manufacturer's protocols, in order to label the cell membrane and nucleus, respectively.

4.3.7. Confocal Microscopy

Labeled cells on cover slips were mounted on glass slides using VectaShield Antifade Mounting Medium (Vector Laboratories, USA). Samples were imaged using a Leica TSC SP8 confocal microscopy setup equipped with an HC PL APO 40x/1.30 Oil CS2 objective. Localization of TpitAstRC was imaged via illumination of EYFP (λex/λem: 514/518-580 nm), cell membrane was imaged via CellMask^{TM} (λex/λem: 649/655-700 nm) and the nuclei were imaged via Hoechst 33342 stain (UV laser, λex/λem: 405/460-490 nm). Images were obtained with the LAS X software in a 1024 x 1024 pixels format, consisting of 4 averaged line scans. The scan speed was set to 400 Hz and pinhole was set to Airy 1.

4.3.8. Peptide and Small Molecules Dissolvation

The native peptide and small molecules were dissolved in PBS supplemented with 0.1% BSA. The final concentration was 1 mM.

4.3.9. G Protein Activation Assay

At different 10 cm petri dishes, pcDNA3-WT plasmid (Section 4.2.4) with each FRET biosensor were transfected to HEK-TSA cells transiently. 0.5 μg of pcDNA3-WT and 1.5 μg of G protein biosensors from Section 3.5.2 using Effectene Transfection Reagent Kit (Qiagen, Germnay) according to the manufacturer's instructions were added. For Gs activation, since this sensor was not available in one plasmid, cells were transfected with 0.5 μg receptor, 0.8 μg ECFP-Gs, 0.5 g G1, 0.2 μg EYFP-

Gγ2. Twenty-four hours later, cells were reseeded to poly-D-lysine pre-coated black-wall, black-bottomed. Twenty-four hours after reseeding, cells expressing the FRET sensors were subjected to FRET measurement. Before the measurement, cells were washed to substitute the DMEM with the experimental buffer, and then basal FRET ratio was measured in 90 μL buffer. Subsequently, 10 μL of 10-fold ligand solution or buffer (negative control) was applied to each well and the stimulated FRET ratio was recorded. All FRET experiments were conducted at 37 °C with a Synergy Neo2 plate reader (BioTEK) equipped with 420/50 nm excitation and 485/20 nm emission filters for CFP. Acceptor emission of YFP were detected with a 540/25 nm (FRET) filter.

4.3.10. Temporal Kinetics of G protein activation

The same constructs and cell culture procedure as G protein activation assay was used. But the kinetics measurements were performed on a Zeiss Axiovert 200 inverted microscope equipped with an oil immersion 63x objective lens and a dual-emission photometric system. Another difference with DRC was that here, during the reseeding step, cells were transferred to Poly-L-Lysine-coated coverslips in 6-well cell culture dishes. The kinetics were studies in 16 hours after re-seeding. Coverslips were placed in a metal chamber, washed with PBS and supplemented with HBSS. Ligand application during live FRET measurement was performed using a high-speed perfusion system (ValveLink 8.2, Automate Scientific). Cells were excited with light from a polychrome IV. Illumination was set to 40ms out of a total integration time of 100 ms. CFP (480 $\pm$ 20 nm), YFP (535 $\pm$ 15 nm), and FRET ratio (YFP/ CFP) signals were recorded simultaneously (beam splitter DCLP 505 nm) upon excitation at 436 $\pm$ 10 nm (beam splitter DCLP 460 nm). Fluorescence signals were detected by photodiodes. The signal was then digitalized using an analogue-digital converter (Digidata 1440A, Axon Instruments). All data were recorded on a PC running Clampex 10.3 software (Axon Instruments). Individual traces were fit to a one component exponential decay function to extract the exponential time constant, tau. The half-time of activation (t1/2) is defined as τ*ln2. In dynamic experiments, cells were stimulated with AST-C ligand. Further processing of the data was done in Excel (Microsoft Office) and graphs

and statistics were conducted using GraphPad version 6.07 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

4.3.11. Temporal Kinetics of G Protein Recruitment

Plasmid named sYFP-WT from Section 4.2.4 was used which is fluorescently tagged in its C-terminus. FRET biosensor Gαi, Gβ and Gγ-CFP tagged were also utilized. For transient expression of the SYFP-WT and the G protein subunits, 1.5 $\times$ 106 HEK-TSA cells were seeded onto a 55mm dish and transfected the day after with 2 g of plasmids using Effectene Transfection Reagent (Qiagen, Germany) according to the manufacturer's protocol. Kinetics measurements were performed at the same measurement setup as explained at Section 4.3.10.

4.3.12. β-arrestin Recruitment Assay

sYFP-WT plasmid was used (explained at Section 4.2.4). SYFP-WT, GRK_2 (G protein Receptor Kinase 2) and arrestin2-Nluc were transfected to HEK-TSA cells. Cells were washed to substitute DMEM with the experimental buffer and incubated with furimazin (1:1000 of 90 μl HBSS) for arrestin2-Nluc for 2–5 min at 37 °C to allow for substrate diffusion and the basal BRET ratio was measured. Following this, 10 μl of 10-fold ligand solution or buffer was applied to each well and the stimulated BRET ratio was recorded for 20 minutes. BRET experiments were performed at 37 °C with Synergy Neo2 (BioTEK) plate reader equipped with a 460/40 nm filter to select the NanoLuc emission.

4.4. Computational Studies

4.4.1. Homology modeling

SWISS-MODEL online tool [104] was used to build tertiary structure of TpitAstR-C. Different templates available in Protein Data Bank (PDB) with PDB IDs of 4DJH,

4N6H, 5C1M and 6DDE were taken and various models were generated. All these templates showed high sequence identity and similarity to TpitAstRC. Primarily, the built models were evaluated using the QMEAN [105] and Ramachandran plot. Those models with acceptable values were selected and subjected to 25 ns molecular dynamics (MD) simulations. N-terminus of the receptor was modeled utilizing I-TASSER webserver [106], and merged with the constructed models. The structure of the native ligand, AST-C, was also built using I-TASSER webserver either. The distance between the two Sulfur atoms of Cysteine residues were set to be kept at 2.05 Å to have the disulfide bond between these two residues. Five models were generated by the program and ranked according C-score was values that is a metric the program assigns at the range of [-5,2] to the constructed models. Bigger numbers are related to models with higher quality.

4.4.2. Protein Preparation

Before proceeding with the MD simulation, protein structures of the system needed to be prepared. This was done in "Protein Preparation Wizard" module of Maestro molecular modeling package [107]. Both the receptor and the ligand were prepared here. First, the structure of these proteins was refined and then they subjected to minimization step. Problems present in the structures including missing hydrogen atoms, side chains, loops or flipped residue were resolved in "Prime" module of Maestro. The protonation states at pH 7.4 was assigned using Epik [108]. Optimized Potential for Liquid Simulation (OPLS) force field [109] with version OPLS2005 was used for the minimization and optimization processes.

4.4.3. System Preparation

Gα part of Gi complex available at PDB (PDB ID: 6DDE) was placed at the intracellular interface of TpitAstRC. The structure of Gα acquired some modification including some unresolved gaps that needed to be filled. To fill the gaps, "Crosslink Proteins" module of Schrödinger2017 was used. The conformation of the linker was

predicted utilizing Simple *de novo* loop creation. The energy calculation of this linker was performed via Prime module, selecting implicit solvent energy calculation for the energy calculation. Five residues were inserted at each gap. For the longer part, to link the 55^{th} residue to the 182^{nd} one, 10 residues taken from the sequence of Gα positioned in between to link these two residues. The other gap was filled with seven residues. The orientation of the models built for TpitAstRC in the membrane was obtained using the Orientations of Proteins in Membrane (OPM) database [110]. The "Desmond System Builder" module of Meastro was used to assemble the system in a way that it best resembled the biological environment consisting of solvent, membrane, counter ions and water. The protein was embedded in POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid bilayer and TIP3P explicit water [111] was selected. 0.15 M NaCl was added to the system to neutralize.

4.4.4. Running MD Simulations

Desmond package was used for the MD simulations [112]. OPLS 2005 force filed was used for the simulations [109]. The default algorithm of desmond was utilized at the equilibration step. In brief, the system was subjected to a short MD simulation of 24 ps at the temperature of 300 K and the pressure of 1.01325 bar using Berendsen thermostat and Berendsen barostat method with constant fast temperature relaxation constant and normal pressure relaxation constant. The production step of MD simulation was run at 300 K which is the recommended temperature when using POPC lipid bilayer. Nose-Hoover thermostat [113] and Martyna-Tobias-Klein method [114] were applied to the system to keep the temperature at 300 K and the pressure at 1.01325 bar. Long-range electrostatic interactions were calculated applying the particle mesh Ewald method [115]. A cut-off radius of 9.0 Å was used for both van der Waals and Coulombic interactions. The time-step was assigned as 2.0 fs. NPγT ensemble was used during production simulations applying surface tension of 4000 bar/Å. It is the recommended surface tension for NPγT ensemble. At different steps, according to the purpose of the study, MD simulations with varied time lengths were run. For instance, MD simulations performed for the binding pocket verification of TpitAstRC and AST-C run for 500 ns

and it was repeated three times applying random seed numbers.

4.4.5. Analyzing MD Simulation Trajectories

MD simulation analysis was performed using "Simulation Interaction Diagram" or "SID" module of the Schrödinger package. The trajectory files collected during MD simulations were used for the analysis. The report prepared by the program includes different parts.

4.4.6. Root Mean Square Deviation

The average change in displacement of a selection of atoms at each frame with respect to a reference frame is calculated. The metric is called root mean square deviation (RMSD) and during each MD simulation it is calculated for all frames. The RMSD for frame x is calculated as shown in Eq. 4.1.

$$\mathbf{RMSD}_x = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(r'_i(t_x) - r_i(t_{ref})^2} \tag{4.1}$$

Where N represents the number of atoms, t_{ref} is the reference time that typically the first frame is used as the reference and it is regarded as time t=0; and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t_x.

For every frame in the simulation trajectory, the procedure is repeated. All the obtained frames are first aligned on the reference frame backbone, and then the RMSD is calculated based on the atom selection. The atom selection can be chosen as carbon alpha (Cα), backbone or side chains. Here, all RMSDs are calculated using Cα atom selection. The RMSD changes during time is plotted. The plot gives insight into the conformational changes that happens during the MD simulation time in the structure. If the structure is equilibrated during the simulation time, the plot is expected to

reach to a plateau. Decreasing or increasing RMSD changes toward the end of the MD simulation time are indicator of non-equilibrated structure.

4.4.7. Root Mean Square Fluctuation

Unlike RMSD, root mean square fluctuation (RMSF) characterizes local changes that happens in the protein structure. RMSF for residue i is calculates as provided in Eq. 4.2.

$$\mathbf{RMSF}_i = \sqrt{\frac{1}{T}\Sigma_{i=1}^{T}(< r'_i(t_x) - r_i(t_{ref} >)^2} \tag{4.2}$$

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i, r' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance is taken over the selection of atoms in the residue.

RMSF-reside plot depicts the fluctuations of each residue of the structure. Looking at this plot, the regions with higher fluctuations can be identified immediately. The flexibility and rigidity of the different areas can be inferred. For some parts such as N-terminus and C-terminus or loop regions high fluctuations are expected.

4.4.8. Molecular Mechanics/Generalized Born Surface Area Analysis

If a ligand is bound to the protein, free energy of binding can be calculated by Molecular mechanics combined with generalized born surface area (MM/GBSA) analysis [116]. The approach is utilized for estimating the free energy of the binding of ligands, which are usually small, to biological macromolecules. "Prime MM-GBSA" module reports energies for the complex structure and, also for the ligand and protein, separately, as well as the energy differences relating to strain and non-strain (NS) mode of ligand and receptor. In each MD simulation, 100 frames were written and for each

frame, MM/GBSA calculations was done. The average of non-strain ΔG (ΔG-NS) is calculated by the script. Dividing ΔG-NS to the number of the heavy atoms of each ligand, binding efficiency of the binding was calculated. ΔG-NS will be denoted as "ΔG" in the rest of the text.

4.4.9. Molecular Docking Studies

ClusPro web server [117] was used for the docking studies which is an online tool for protein-protein docking. A mask file was provided to the prgram to assign the regions that are not allowed for the docking runs. The contributing residues of the receptor in the ligand-protein interaction were evaluated by using the "Ligand Interaction Diagram" of the Maestro package. To investigate the stability of protein-ligand complex, it was applied to MD simulation runs.

4.5. Verification of the Orthosteric Pocket of TpitAstRC

4.5.1. *In silico* Verification

In silico Alanine substitution was applied to the residues residing in the binding pocket that were shown to be contributing to the Receptor-Ligand interaction. Alanine substitution was done in Apo form of the structure without coupled Gα in the intracellular face, and Holo form both with and without the bound Gα.

New docking poses were generated for the mutant receptors by running docking simulations in ClusPro. All systems including Apo WT and mutants, Holo without G WT and mutants and Holo with Gα protein WT and mutants were subjected to 200 ns MD simulations.

The effect of the mutations on the state (active, intermediate, and inactive) of the receptor was investigated. This was done by measuring the Δd that is calculated

as given in Equation 4.3, 4.4 and 4.5.

$$\mathbf{d}_2 = (2 \times 41) - (6 \times 38) \tag{4.3}$$

$$\mathbf{d}_1 = (3 \times 44) - (7 \times 52) \tag{4.4}$$

$$\mathbf{\Delta d} = \mathrm{d}_2 - \mathrm{d}_1 \tag{4.5}$$

d_1 and d_2 values were obtained by measuring the distance between specific residues. In the given formula, residues are specified by generic numbering offered by GPCRdb [118]. In TpitAstRC, d_2 is the distance between M75$^{2\times41}$ and L254$^{6\times38}$ and d_1 is the distance between C129$^{3\times49}$ and L308$^{7\times52}$. Δd was measured along the MD simulation time using the "Simulation Event Analysis" module of Schrödinger. The free energy of binding was calculated at Holo form.

Principal component analysis (PCA) were performed in which large-amplitude motions along MD simulations could be extracted using a dimensional reduction method called PCA. In this study Bio3D package of Grant *et al.* [119] was utilized using R program. All the frames of trajectories were aligned with respect to an initial reference state before performing PCA to eliminate translational and rotational motions of protein and just to focus on internal fluctuations. Only Cα atoms of proteins were used for PCA to focus on backbone movements. Here, we have applied PCA for both TpitAstRC and Gα protein separately. We have performed PCA for both WT and Q271A mutated systems, for which MD simulations was extended to 500 ns, to elucidate the effect of mutation in addition to determine the overall combined motions of proteins. Both Holo and Apo forms systems were considered to elucidate the effect of ligand binding on receptor.

The cross-correlation between atomic fluctuations/displacements are useful to provide information about the effect of mutations, binding of ligand, etc. on the protein structure [120]. Here, Bio3D package in R program was used and Cα atoms of proteins

were utilized to focus on backbone of proteins. For both receptor TpitAstRC and Gα proteins cross correlation analysis were performed in four different systems for which PCA also applied. Dynamic cross-correlation maps (DCCM) were plotted to visualize the correlation between residues in Bio3D package.

4.5.2. *In vitro* Verification

The identified binding pocket-residing residues of TpitAstRC were substituted to alanine using SDM as described in Section 4.2.11. The mutant receptors were tested in G protein activation assay to investigate the effect of point and combination of point mutations on the G protein-dependent activation of the receptor. Different mutant receptors were transfected to HEK-TSA cell line and the measurements were done in different doses of AST-C.

4.6. High-Throughput Virtual Screening in OTAVA Peptidomimetics Library

4.6.1. Library Preparation

Peptidomimetics β-turn library was downloaded from OTAVA chemicals. Ligands needed to be prepared before proceeding with docking. The preparation was performed using the "LigPrep" of maestro package 2017. OPLS2005 [109] was used as the force field in which the structure would be minimized. PROPKA was used for the prediction of the protonation states at pH 7.4 [108]. The preparation of the ligands was done, and all the prepared ligands were concatenated and saved as ".maegz" file.

4.6.2. Grid Generation

"SiteMap" module of Schrödinger was used for the binding site identification. In order to generate the grid, the site with the highest probability was chosen. The "Receptor Grid Generation" of Schrödinger package was used to create a grid at the

binding pocket site. Before, proceeding with grid generation, the protein needed to be prepared by "Protein Preparation Wizard" of Schrödinger. A grid suitable for peptide docking was generated. The program puts the box automatically around the map. The coordinates and the box size can be modified. The box size was increased to 20 Å in our system.

4.6.3. High-Throughput Virtual Screening Workflow

The "virtual screening workflow" module of Schrödinger was used. The workflow can be adjusted in a way to apply different docking algorithms in consecutive steps to a given library of molecules. In this study, ligands were subjected to high-throughput virtual screening (HTVS) and Glide/SP (standard- precision) [121]. In the HTVS step, 100 poses of each ligand were generated, and the program was set to keep 10% of the best compounds with good scoring state. The molecules fulfilling this condition, were applied to docking Glide/SP , in which 10 poses were generated for each ligand and 10% of good poses were retained. Ligands were kept flexible in the grid that results in a acquisition of different conformations of ligands. Beside the conformation, the position and the orientation of ligand relative to the receptor changes during the docking are considered which results in obtaining different ligand poses. After the minimization of each pose in the grid, the final scoring was carried out in Glide [122]. GlideScore uses a composite Emodel score which is a combination of GlideScore, the nonbonded interaction energy, and, in case of flexible docking, the excess internal energy of the ligand conformation. Glide uses Emodel to rank among different poses of a ligand, while Glidescore is used by the program to rank chemically distinct species. The finally selected molecules were applied to stepwise simulation studies.

Again, before the MD simulations, system preparation step was done and the prepared systems were subjected to consecutive MD simulations including 1 ns, 5 ns and 50 ns.

4.6.4. Data Analysis and Visualization

ImageJ (National Institute of Health) was used to analyze the raw microscopy images. Further processing of the data were done in Excel (Microsoft Office). All concentration−response data were fitted using nonlinear regression models with Prism 6 (GraphPad Software, San Diego, CA, USA). Visual Molecular Dynamics (VMD) software [123] program was used for visualization and image generation.

5. RESULTS

5.1. Whole Genome Sequencing Analysis

5.1.1. Sample Preparation

gDNA samples were collected from larvae and sent for sequencing. Quality and quantity of the samples were tested by the company before the sequencing (Table 5.1).

Table 5.1: Genomic DNA samples sent for sequencing.

Sample Name	**Concentration (ng/μl)**	**Result**
TpitHeadS1	57.62	Pass
TpitHeadS2	85.11	Fail/smear
TpitHeadS3	62.04	Fail/smear
TpitHeadS4	12.57	Fail/smear
TpitHeadS5	31.49	Fail/smear

The integrity assessment of the gDNA samples were performed via agarose gel electrophoresis by Macrogen Company. 1 out of 5 samples passed the quality control step (Figure 5.1). The concentration of this sample was 57.62 ng / μl and a sharp band around 10 kb was observed which was an indication of a high-quality gDNA.

5.1.2. *De novo* Assembly

A total of 164 GB of raw sequence reads was obtained via sequencing the gDNA sample by HiSeq 150bp paired-end method. Before proceeding with assembly, the raw data needed to be investigated for adapters and primers used for the sequencing. The quality of the reads in the raw data was evaluated using FastQC method, and according

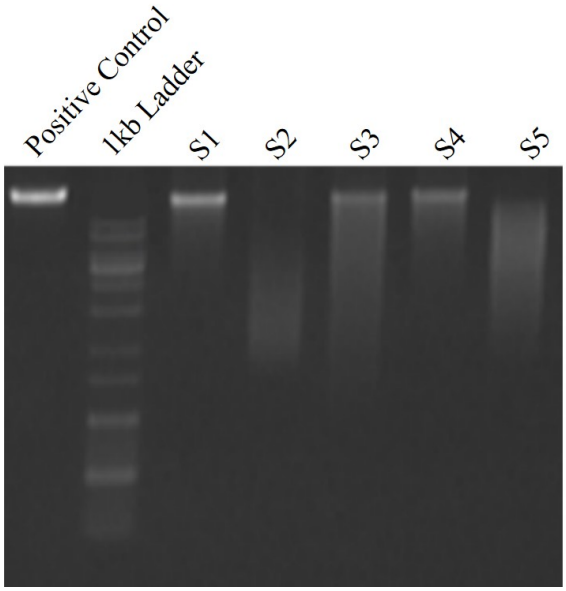


Figure 5.1: Quality control of the DNA samples.

to the quality tests, it was decided to trim 10 bp from the end of each read (Figure 5.2).

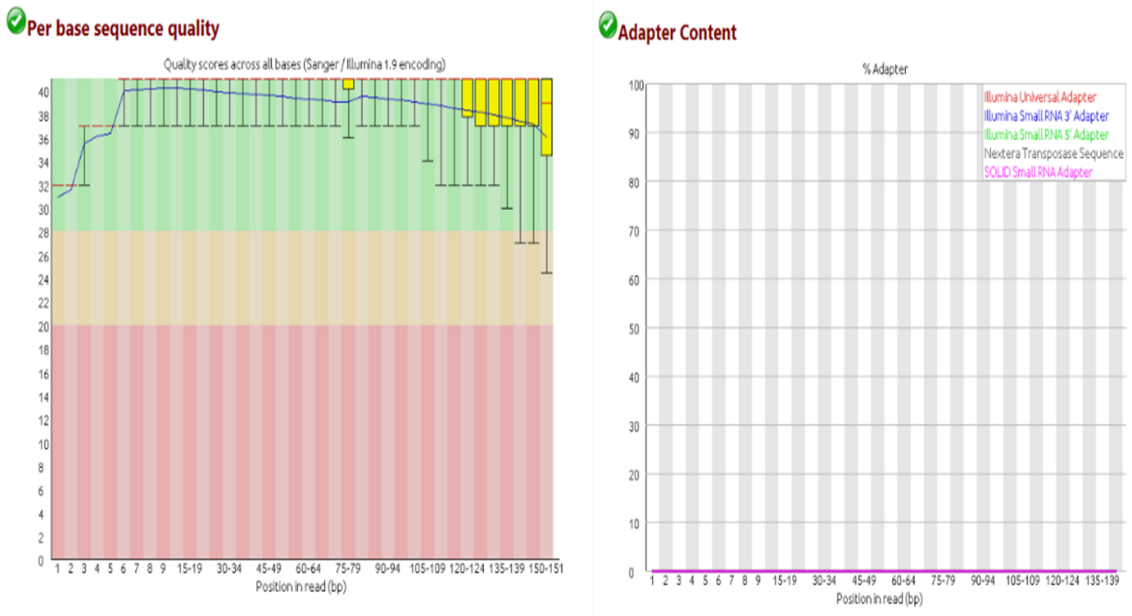


Figure 5.2: Quality control of raw data obtained from FastQC program, a) Per base sequence quality, b) Adapter content. The plot shows that no adapter is present in final reads.

It is very common to obtain low quality sequencing results toward the end of the sequencing reads. The adaptors and primers were investigated. Two illumina adapters R1: AGATCGGAAGAGCACACGTCTGAAC, R2: AGATCGGAAGAGC GTCGTGTAGGGA were found in the raw data and clipped.

The trimmed and clipped raw data were then subjected to assembly process using two different assemblers. More than one assembler was used at this point since different assembler programs have varied performance in assembling the data based on the intrinsic properties of the raw data. SGA assembler and MaSuRCA assembler were utilized, and the resulted data by each program were investigated comparing N50 value that is a metric generally used to evaluate the performance of different assemblers and is defined as a weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value. Selecting K-mer 71bp was proven to result in the best N50 values in both programs. Altogether, it was found that MaSuRCA assembler did a better performance in assembling the data with with N50 value of 28,901 bp. In total, 38,646 contigs (593,440,828 bp) were collected and the longest contig was 219,425 bp (Table 5.2).

Table 5.2: Assembly statistics of two assemblers.

Assembler	**Total Bases (Mbp)**	**Max Scaffold (bp)**	**N50 (bp)**
SGA	644.67	173,781	20,602
MaSuRCA	593.44	219,425	28,901

5.1.3. Genome Comparison

To have a better understanding regarding the assembled data, genome size and GC content of the of *T. pityocampa* were compared within the other species of Lepidoptera order belonging to Noctuidae family (see Table 5.3 and Figure 5.3).

Table 5.3: Comparisons of the genome size and GC content of *T. pityocampa* with other species of Lepidoptera.

Scientific Name	**Common Name**	**Genome Size (Mbp)**	**GC%**
T. pityocampa	Pine processionary moth	593.44	37.0
A. ipsilon	Black cutworm moth	486.92	37.6
H. armigera	Cotton	337.08	37.4
H. zea	Corn earworm	341.14	37.4
H. virescens	Tobacco budworm	403.15	33.2
M. configurata	Bertha armyworm	571.32	36.9
S. frugiperda	Fall armyworm	413.95	36.0
S. litura	Not assigned	438.90	36.8
T. ni	Cabbage looper	350.50	35.2

Our assembly results showed that having the genomic size of 588.04 Mbp, *T. pityocampa* has the largest genome among other species in Table 5.3. The size was comparable to *Mamestra configurata*. The GC content of this species was 37%. This value was varied from 35.3% to 37.6% in the other chosen species.

5.1.4. Data Deposition in NCBI

The assembled data were deposited in NCBI (*Thaumetopoea pityocampa* ASH-NBICSL-2017, Accession number: WUAW00000000). The coding sequence of AstR-C and AST-C were deposited in GenBank under MN871948 and MT254058 accession numbers, respectively. The receptor will be referred to as TpitAstRC hereafter.

5.2. Characteristic Features of TpitAstRC and AST-C

5.2.1. TpitAstRC

The sequence of the receptor was submitted to pfam to investigate its GPCR class. The results showed that TpitAstRC belongs to 7TM Rhodopsin class (see Table 5.4.).

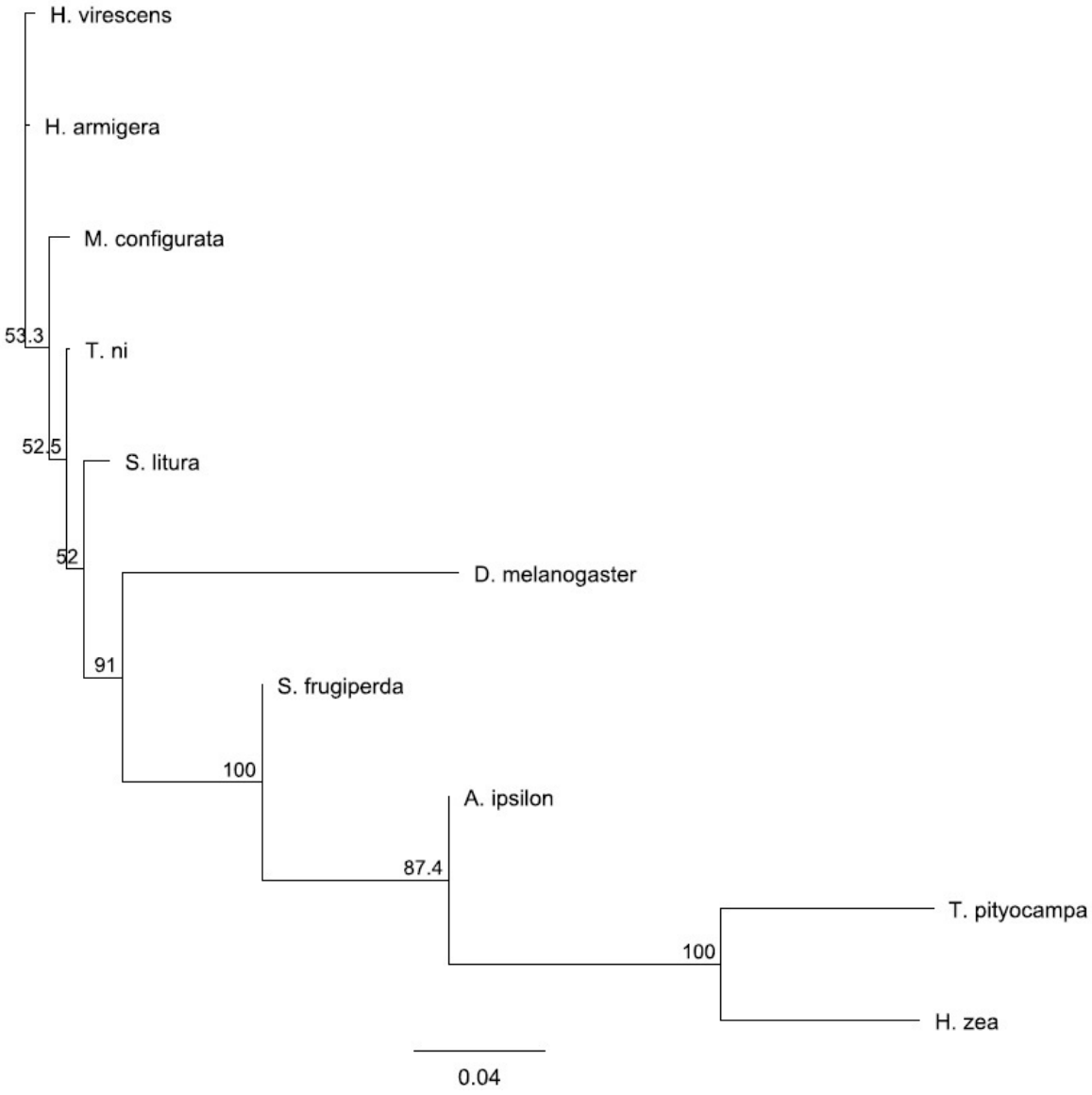


Figure 5.3: Evolutionary relationships of genomic data.

Table 5.4: Sequence analysis of TpitAstRC in pfam using Hmmscan.

Family ID	**Description**	**Start**	**End**	**E-value**
7tm-1	7 transmembrane receptors (Rhodopsin family)	55	309	1.7E-58
7TM-GPCR-Srsx	Serpentine type 7TM GPCR chemoreceptor Srx	49	324	2.8E-13

This class is also known as class A GPCRs. As a member of classA GPCRs, the sequence of the receptor was evaluated to find the conserved residues and motifs available in this family (Figure 5.4). The sequence of other GPCRs including $\beta 2$ adrenergic receptor (UniprotKB: P07550), μ opioid receptor (UniprotKB: P35372) and adenosine A2A receptor (UniprotKB: P29274) were aligned with TpitAstRC protein sequence in clustal omega. The receptor was found to possess all the conserved residues and motif. Numbering in Figure 5.4 shows Ballesteros-Weinstein numbering. All the conserved residues in each TM is shown. Critical motifs in GPCR activation such as DRY and NPxxY are represented. An exception at position 6.30 (Ballesteros-Weinstein numbering) [64]. At this position, a Glutamic acid was expected, however, it was substituted with a Histidine in TpitAstRC.

The topology of the receptor was also checked in CCTOP to check the different structural parts of TpitAstRC including N-terminus, transmembrane regions, intracellular and extracellular loops and C-terminus (Figure 5.5).

The prediction are given both as 1D and 2D representation. Blue color shows the extracellular parts and red represents intracellular regions. The transmembrane segments are shown in orange. According to the result, TpitAstRC with 92.109 reliability has seven transmembrane segments . The second extracellular loop and the third intacellular loops are the longest when compared to the others, and C-terminus with 103 amino acids is much longer than N-terminus that has 38 residues.

Figure 5.4: Alignment of the protein sequence of TpitAstRC and other well-known class A GPCRs.

a)

MELEDVEMFGSDNISYDYNGTVNGTYERCPNVNLPVVYVVAQVLYAIVCV
VGLLGNTLVIYVVLRYSKMQTVTNMYIVNLAIADECFLIGIPFLIITMSV
RSWPFGSFFCKAYMISTGINQFTSSIFLCIMSADRYIAVCHPIAAPRLRT
PFVSRIVSAAAWTASALVMTPIFMYTTLIVDNNGPSCNIVWPEKDLNKGQ
TTFTLYSFALGFAAPLTLIFVFYCLVIRKLKTVGPKNKSKEKKRSHRKVT
KLVLTVIAVYVLCWLPYWACQVALIYSQATECASHVTITVFLVAACFSYS
NSAMNPILYAFLSDNFKKSFLKACTCAAGKDVNATLHVENSMMPRKRGVG
GARAQARAAESKGGLSVPCAVGSRSEASTALTSKSVAGSDLMPMETRPAT
LVPLIATNGLTHSRL

b)

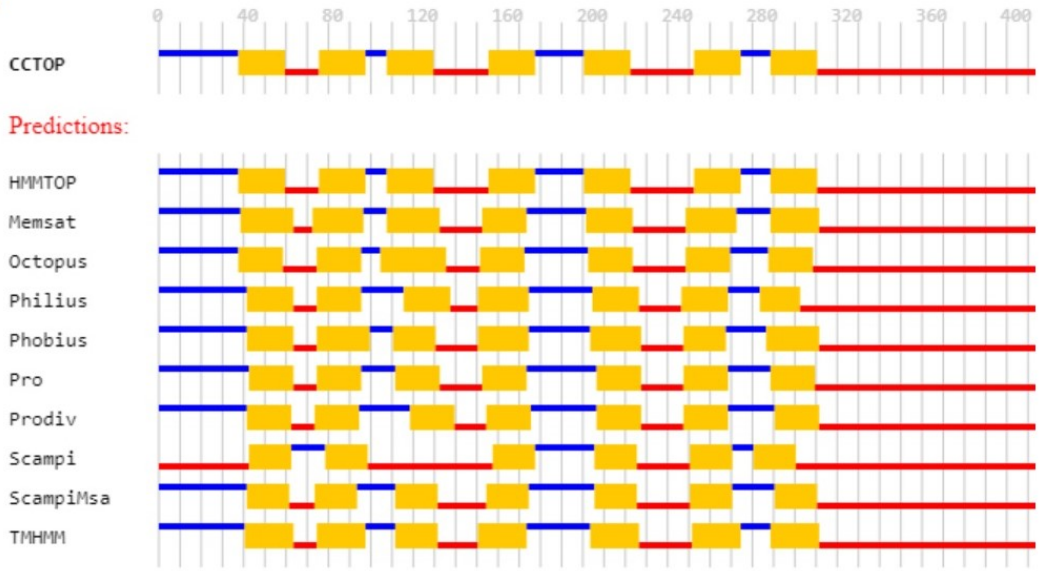


Figure 5.5: Topology of TpitAstRC according to CCTOP. a) 1D representation. b) 2D representation.

5.2.2. AST-C

Sequence of AST-C of *T. pityocampa* was extracted from the WGS data and were deposited in NCBI under the above-mentioned accession number . The gene was interrupted with two introns. Introns were determined in AUGUSTUS and removed. Exons were translated into the longest open reading frame. The obtained protein was the precursor of AST-C neuropeptide. In insects, bioactive peptides are produced from long precursors that go through the process of maturation in the endoplasmic reticulum [124]. In case of AST-C, only one type of neuropeptide is generated from the precursor protein.

To find this mature version, SignalP-5.0 was used to predict the signal peptide, and dibasic cleavage sites were predicted according to the rules stated by Veenstra, 2000 [103] (Figure 5.6). SignalP results showed that the cleavge site for the signal peptide is positioned between position 26 and 27; AHA-AP, with probability value of 0.8541. The signal sequence and dibasic cleavage site are highlighted in yellow and green, respectively. The sequence obtained after dibasic cleavage, shown in red, is further modified post-translationally in N-terminus and glutamine is changed to PyroGlutamate (Pyro[E]), highlighted in blue in Figure 5.5. The sequence of the mature neuropeptide was compared with AST-C peptide of other lepidopteran peptides available in the literature and it was found to be identical with other Lepidopteran AST-Cs [125–127].

5.3. Molecular Cloning

5.3.1. RNA Extraction

After extracting RNA sample from the tissue, the quantity and quality of the extracted sample were evaluated. The concentration was 585 ng/μl. The integrity of the RNA sample was controlled. It was shown to be intact and not degraded (Figure 5.7). In insects, except for aphids, single band pattern is common and 18S rRNA and

```
CLUSTAL O(1.2.4) multiple sequence alignment

Dmel(AST-C,isoformA,NP_523542.1)        MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Dmel(AST-C,isoformA,AAF53062.1)         MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Dmel(AST-C,isoformB,NP_001162948.1)     MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Dmel(AST-C,isoformB,ACZ94235.1)         MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Dmel(AST-C,AAK40100.1)                  MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Dmel(AST-C,AAK40081.1)                  MMKFVQILLC---YGLLLTLFFALSEARPSGAETGPDSDGLDGQDAEDVRGAYGGGYDMP   57
Harm(AST-C,AGH25547.1)                  -MKSSAYNVFLAIVAAALALLFVTVNAAPMEADDEAAE--------NTLVAHPDGDMEMS   51
Tpit(AST-C)                             -MKSTACSLSLTILVATQALLLVAAHAAPMAD-DEQSE--------NMVMGRPGSEVD-L   49
                                         **     :         :*::.  .* *        .        : : .  ..  :

Dmel(AST-C,isoformA,NP_523542.1)        AQ------A--------IYPNIPMDRLQMLFAQYRPTSYSAYLRSPTYGNVNELYRLPES   103
Dmel(AST-C,isoformA,AAF53062.1)         AQ------A--------IYPNIPMDRLQMLFAQYRPTSYSAYLRSPTYGNVNELYRLPES   103
Dmel(AST-C,isoformB,NP_001162948.1)     AQ------A--------IYPNIPMDRLQMLFAQYRPT-YSAYLRSPTYGNVNELYRLPES   102
Dmel(AST-C,isoformB,ACZ94235.1)         AQ------A--------IYPNIPMDRLQMLFAQYRPT-YSAYLRSPTYGNVNELYRLPES   102
Dmel(AST-C,AAK40100.1)                  AQ------A--------IYPNIPMDRLQMLFAQYRPT-YSAYLRSPTYGNVNELYRLPES   102
Dmel(AST-C,AAK40081.1)                  AQ------A--------IYPNIPMDRLQMLFAQYRPT-YSAYLRSPTYGNVNELYRLPES   102
Harm(AST-C,AGH25547.1)                  APWDAINTAALRKLLLQLDAEERMGRVTRSWPQAEPRGWGLR---ALDGRLAR--QWRAD   106
Tpit(AST-C)                             GAWDPINAAALRKFLLKLDAEERMDRVTRSWPQDAPRGWAGR---SMDGRLAR--QWRTD   104
                                        .      *       :  :  *.*:   : *  *  :.        *.: .  :   .

Dmel(AST-C,isoformA,NP_523542.1)        KRQVRYRQCYFNPISCFRK    122
Dmel(AST-C,isoformA,AAF53062.1)         KRQVRYRQCYFNPISCFRK    122
Dmel(AST-C,isoformB,NP_001162948.1)     KRQVRYRQCYFNPISCFRK    121
Dmel(AST-C,isoformB,ACZ94235.1)         KRQVRYRQCYFNPISCFRK    121
Dmel(AST-C,AAK40100.1)                  KRQVRYRQCYFNPISCFRK    121
Dmel(AST-C,AAK40081.1)                  KRQVRYRQCYFNPISCFRK    121
Harm(AST-C,AGH25547.1)                  KRQVRFRQCYFNPISCFRK    125
Tpit(AST-C)                             KRQVRFRQCYFNPISCFRK    123
                                        *****.**************
```

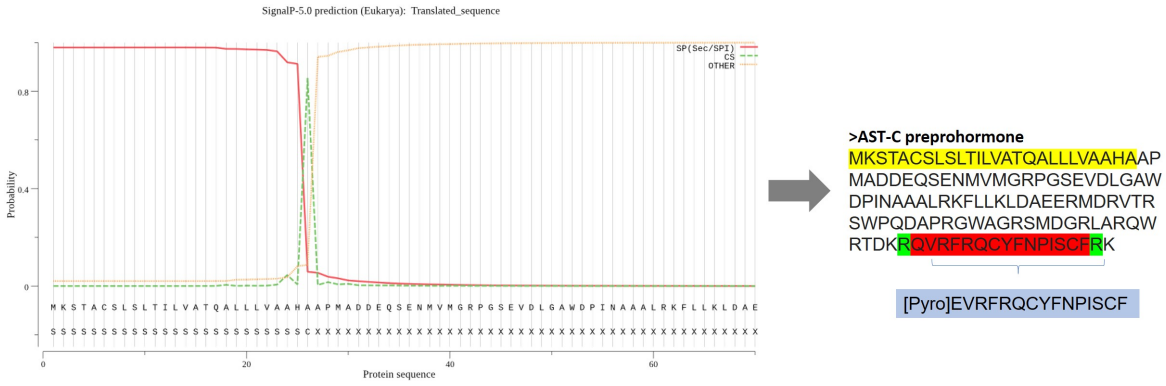


Figure 5.6: AST-C preprohormone. a) Alignment of the AST-C preprohormone of *D. melanogaster* (shown as Dmel) and *H. armigera* (shown as Harm). b) SignalP-5.0 was used to predict the signal peptide.

28S rRNA comigrate together. Human RNA was used as positive control.

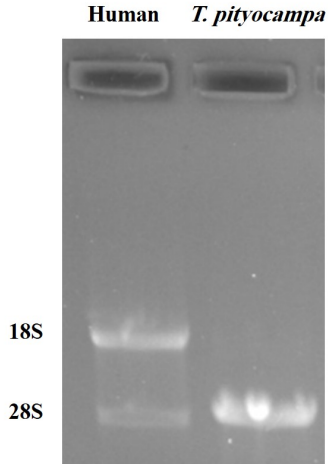


Figure 5.7: Quality control of the RNA sample in MOPS gel electrophoresis in denaturing condition.

5.3.2. Cloning in pcDNA3.1(+) plasmid

The extracted RNA sample was used to synthesize cDNA library, and this cDNA pool was used in PCR reactions as template. To clone TpitAstRC in pcDNA3.1 plasmid, WT-F and WT-R from Table 3.10 were used. PCR reaction was prepared according to Table 4.1 and run as detailed in Table 4.2. PCR product was loaded to agarose gel. The expected length of the product was 1245bp (Figure 5.8).

The PCR product was extracted from gel and purified. Then, the plasmid and PCR products were digested by HindIII and BamHI enzymes and ligation was performed. The accuracy of the cloned fragment was validated by sequencing using T7 and SP6 primers.

5.3.3. Cloning in SYFP plasmid

To clone TpitAstRC in SYFP-N1 plasmid, sYFP-F and sYFP-R from Table 3.10 were used. PCR reaction was prepared according to Table 4.1 and run as detailed in Table 4.2. PCR product was loaded to agarose gel (Figure 5.9).

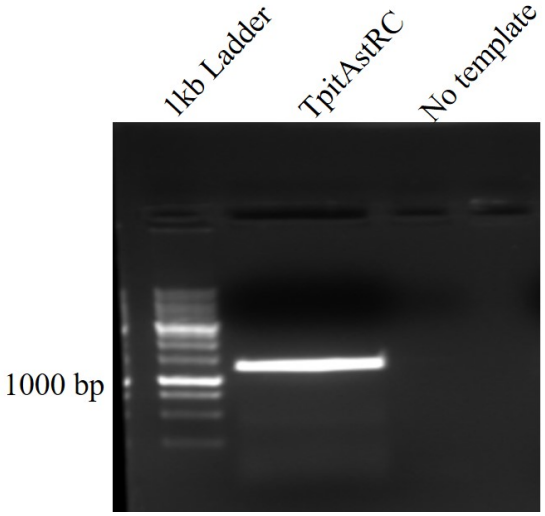


Figure 5.8: RT-PCR product for TpitAstRC using WT-F and WT-R primers.

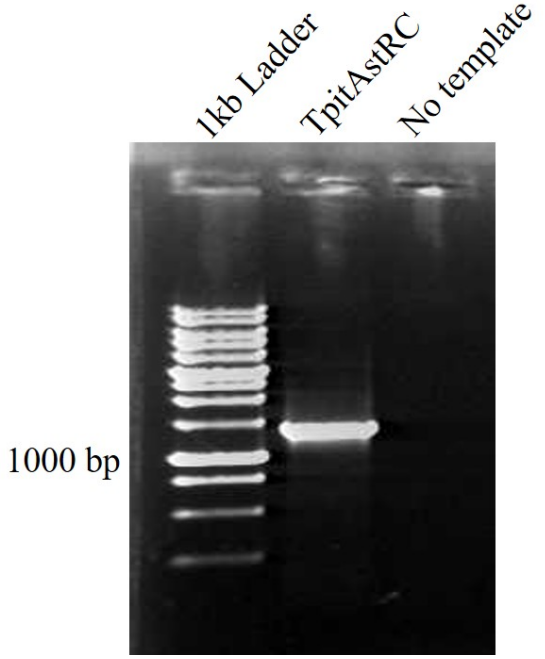


Figure 5.9: RT-PCR product for TpitAstRC using sYFP-F and sYFP-R primers.

The PCR product was extracted from gel and purified. Then, the plasmid and PCR products were digested by XhoI and HindIII enzymes and ligation was performed. The accuracy of the cloned fragment was validated by sequencing using CMV-F and GFP-2R primer.

5.4. Functional Studies

5.4.1. Cell Localization of TpitAstRC

As a proof for translation of a transmembrane protein from this sequence, cellular localization of TpitAstRC was investigated. The receptor bearing YFP in its C-terminus (sYFP-WT) was overexpressed in HEK-TSA cells. It was shown that TpitAstRC mainly expresses in the plasma membrane but not in nuclear or cytoplasmic regions (Figure 5.10).

5.4.2. G Protein Coupling of TpitAstRC

As a novel GPCR, among many other unknown features, no data were available about the Gα subtype that is coupled to TpitAstRC receptor.

In this part, using FRET biosensors, G protein activation assay was utilized to find the G protein subtype that couples to the receptor. In these sensors, alpha subunit was tagged by CFP and the gamma subunit was tagged by YFP (Figure 5.11a).

Gs, Gi1, Gi2, Gi3, Gq and G13 were the FRET biosensors used here. A decrease in FRET signal is expected following the binding of the stimulated receptor to the relevant Gα subtype. Dose-response curve (DRC) shows the changes of FRET signal upon the application of different doses of AST-C. The presented data are representative of at least three biological replicates. G protein activation was observed when Gαi sensors were used. Hence, it was deduced that TpitAstRC favored coupling to the Gαi subtype (Figure 5.11b).

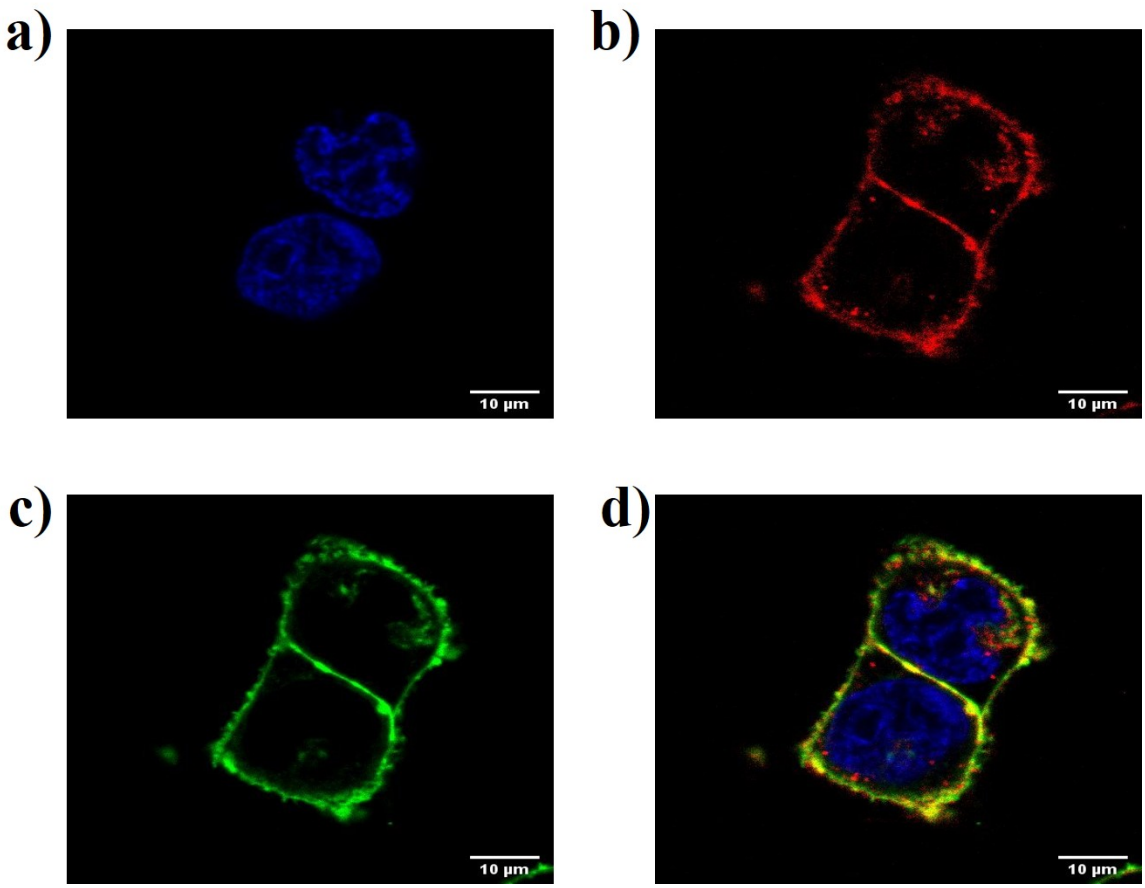


Figure 5.10: Cell localization of TpitAstRC. a) Nucleus staining by Hoechst 33258. b) Plasma membrane staining by CellMask^{TM} Deep red. c) TpitAstR-C fluorescently tagged (YFP) in the C-terminus. d) Merged image.

It is of note that three different available Gαi sensors (Gαi_1, Gαi_2 and Gαi_3) were tested (Figure 5.11c). Results indicated that all used Gαi sensors couple to TpitAstRC with an EC_{50} at sub-nanomolar range. However, Gαi_2, yielded the best ΔFRET. Thus, in the following experiments Gαi_2 was used.

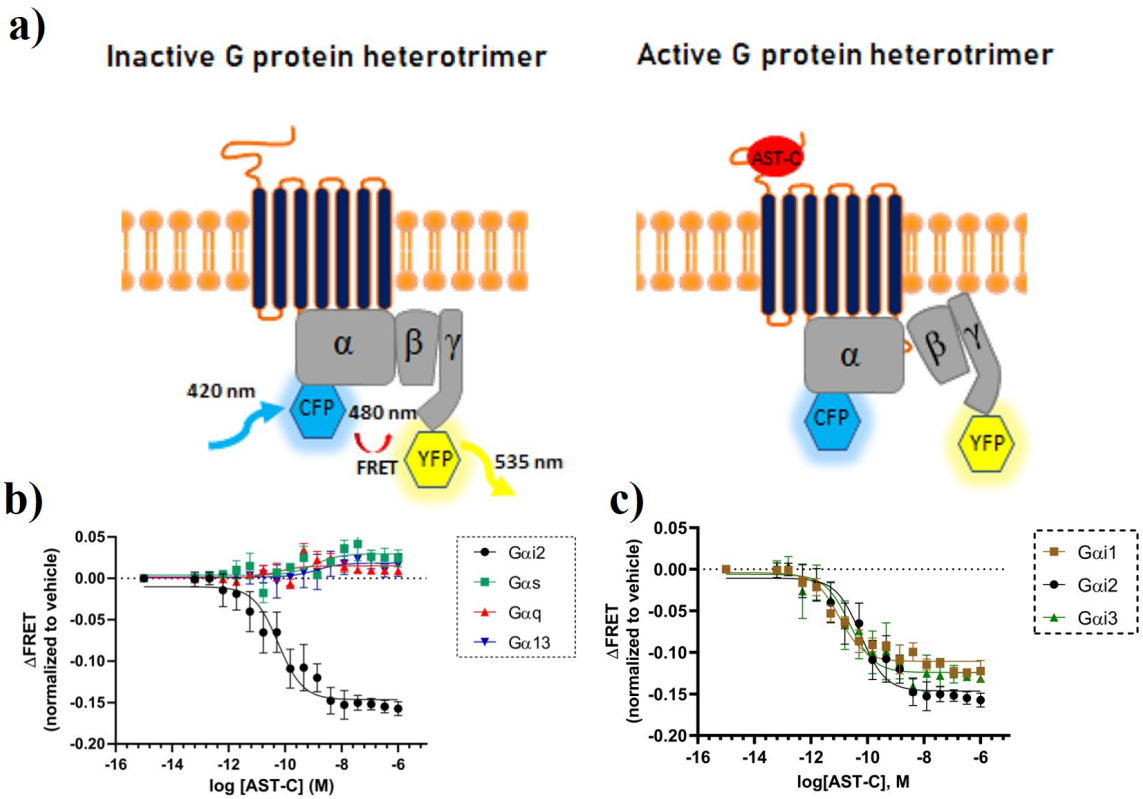


Figure 5.11: G protein activation assay. a) Schematic representation of G protein activation assay. ΔFRET measurements using. b) different G protein biosensors and c) three different Gαi sensors.

5.4.3. Temporal Kinetics of G Protein Recruitment and Activation

Temporal kinetics were performed to understand the time required for the recruitment and activation of G protein complex to the receptor. So, we conducted the kinetics studies at two events, G protein recruitment, and G protein activation. The measurements were done in single cell stimulated with AST-C ligand. The ligand was applied with a perfusion pipette for 10 seconds, and then it was washed off. In the recruitment experiment, C-terminus of the receptor and the gamma subunit of G pro-

tein complex were tagged with YFP and CFP, respectively. At 1 nM concentration of the native peptide, AST-C, on average, a tau value of 4.7 and 74.1 second were yielded for the association and dissociation of G protein complex, respectively (Figure 5.12a). So, the G protein complex is recruited to TpitAstRC within seconds and and its dissociation from the receptor requires much longer time following the substitution of the ligand with FRET buffer.

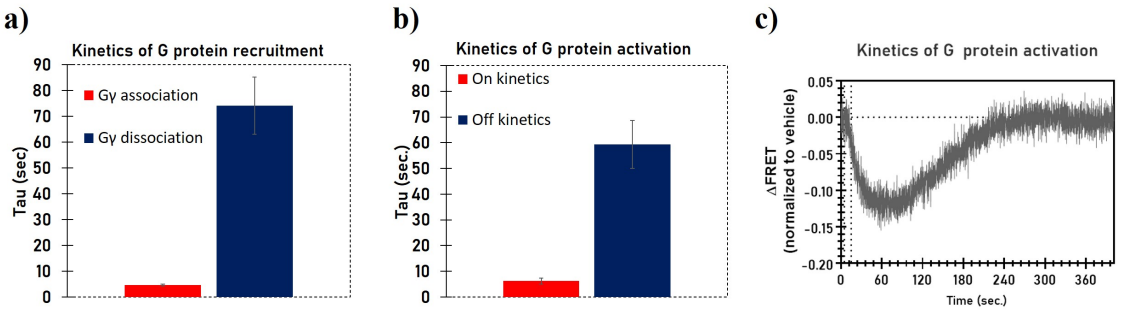


Figure 5.12: Temporal kinetics studies of TpitAstRC. a) G protein recruitment. b) G protein activation. c) A representative trace showing the temporal kinetics of the receptor. The error bars represent standard deviation.

The kinetics of G protein activation were investigated as well to evaluate the *on/off* kinetics of the G protein after it becomes activated by the administration of the native ligand. Fluorescent tags used in this experiment were identical to the ones used in G protein activation assay (Figure 5.11a). The experiment here shows the time that G protein remains active following a brief application of the ligand (*on* kinetics, G protein dissociation) and the time required for it to return to its basal level (*off* kinetics, G protein re-association) when the ligand is being washed *off* by perfusing buffer (instead of ligand) to the cell. The inactivation of G protein was observable at 1 nm concentrations with the τ value of 6.2 sec for *on* kinetics and 59.3 sec for *off* kinetics (Figure 5.12b). Figure 5.12c depicts a representative trace acquired in the G protein activation kinetics experiments.

5.4.4. TpitAstRC/ β-arrestin Recruitment Assay

Arrestins are one of the most important cytosolic effectors of GPCRs. β-arrestin1 (arrestin-2) and β-arrestin2 (arrestin-3) that belong to non-visual arrestin family, bind to the receptors stimulated by agonist molecules [128]. Studying β-arrestin-dependent downstream signaling of GPCRs and desensitization happening in the turn-over process of GPCRs are very crucial in understanding the pharmacology of these receptors.

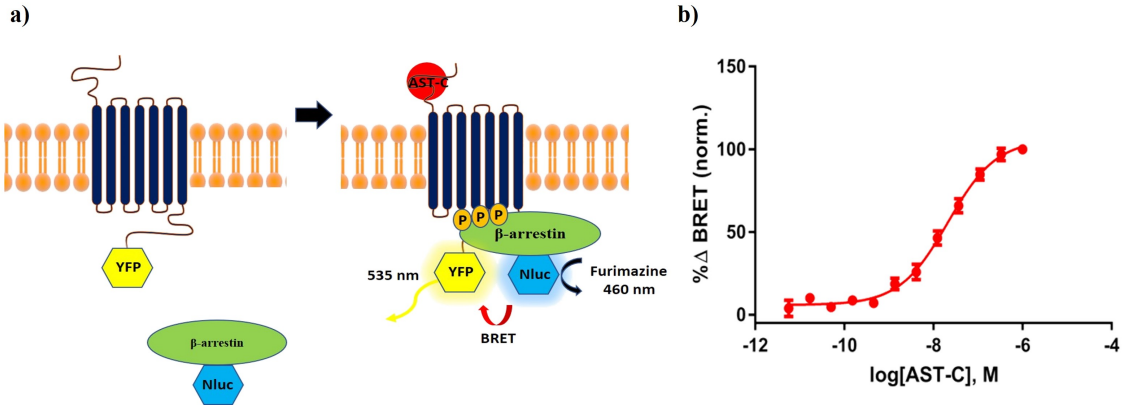


Figure 5.13: β-arrestin recruitment assay. a) Schematic representation of β-arrestin recruitment assay. b) %ΔBRET measurements using different concentrations of the ligand.

In this study, the recruitment of β-arrestin2 to TpitAstR-C upon application of different concentrations of AST-C was investigated using a BRET-based method. Biosensor used here was β-arrestin tagged with NanoLuc luciferase (19 kDa; Nluc) [129] that played role as the donor. The acceptor was YFP at the C-terminus of the receptor. To activate luciferase, Furimazine, the substrate of the enzyme, was applied (Figure 5.13a).

The BRET signal was measured before and after the administration of AST-C, and it was shown that β-arrestin is recruited to TpitAstRC at EC_{50} of nanomolar range (Figure 5.13b). Results from each 96-well plate experiment were normalized to max-min values from the same plate. Data were fit to Hill equation, using the four-parameter dose-response fit function of GraphPad Prism6. The presented data are

representative for at least three different transfections performed on three biological experimental days.

5.5. Structural Studies

5.5.1. Predicted Structure of TpitAstRC

Homology modeling of TpitAstRC was done in SWISS-MODEL webserver [104]. To build a reliable model, different templates were chosen, and models were constructed based on those templates.

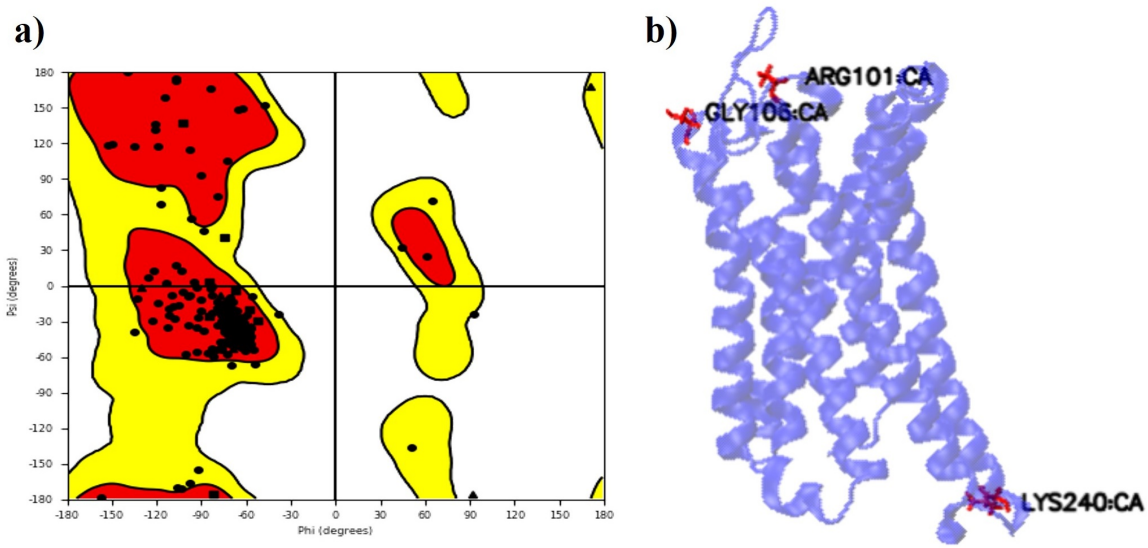


Figure 5.14: Model analysis. a) Ramachandran plot of the (Red: favored region, Yellow: allowed region). b) Receptor localization of outlier residues.

A model was constructed based on *Mus musculus* μ opioid receptor (PDB code: 6DDE) showed good quality. This template is resolved with an agonist (DAMGO) and Homo sapiens nucleotide-free Gi [130] and is in the active state. Thus, this template was an ideal starting point to build the model based on it. In addition, 6DDE template showed 37.15% sequence identity with TpitAstRC, that is in the acceptable range in homology modeling approaches ($\geq$30% sequence identity is required). Ramachandran plot was investigated for this model and outlier residues were identified (Figure 5.14a

and b).

To relax the structure, the constructed model was subjected to short MD simulation (25 ns). Analyzing the MD simulation trajectories, RMSD and RMSF changes for Cα were evaluated during the MD simulation time, 25 ns (Figure 5.15a and b).

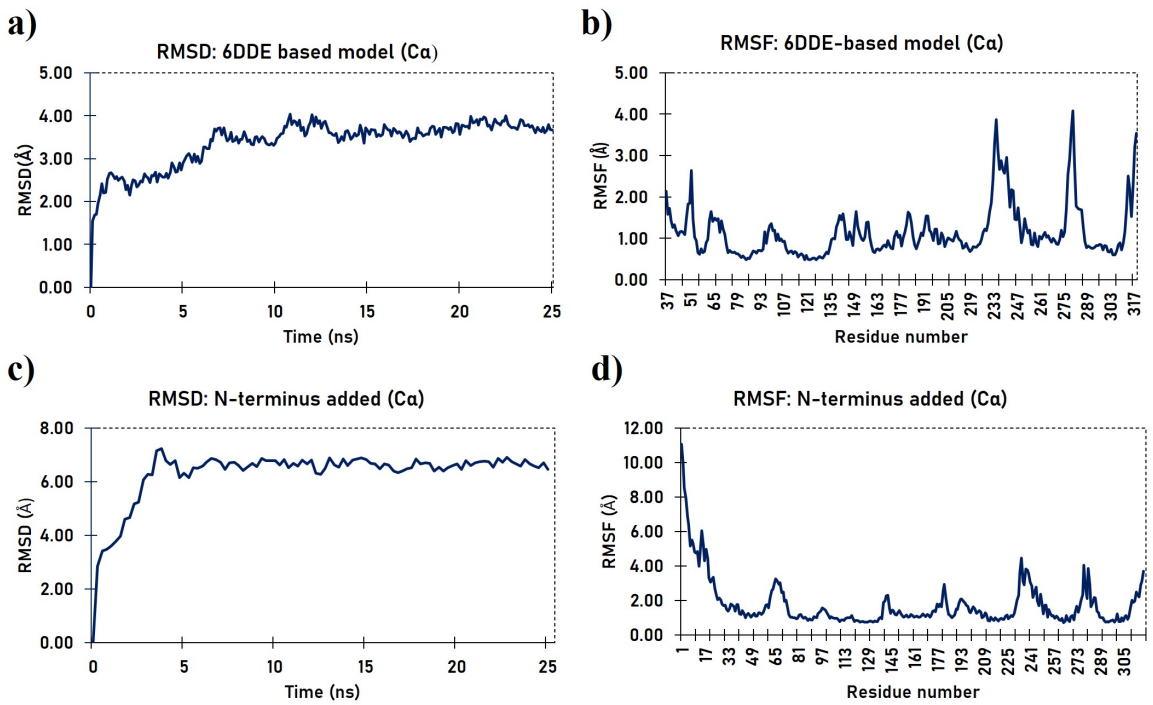


Figure 5.15: RMSD and RMSF plots of the model. a) RMSD, and b) RMSF changes of the model based on 6DDE.pdb. c) RMSD, and d) RMSF changes of the model with N-terminus. MD simulation time was 25 ns. The analysis is given for Cα.

N-terminus of the receptor could not be modeled in the modeling process due to the high flexibility of this region in GPCRs. Thus, we used I-TASSER webserver to build this part with *ab initio* approaches [106]. Then, it was merged with the model. The model to which N-terminus was appended subjected to MD simulation. Figure 5.14c and 5.14d show the RMSD and RMSF changes of this model. Comparing with the RMSD and RMSF values before and after appending the N-terminus to the model, it is obvious that it causes an increase in the fluctuations of the system. N-terminus residues are not embedded in the membrane bilayer and as a result have high flexibility, thus, the observed higher values are expected. Besides N-terminus, other fluctuations are relevant to ECL and ICLs.

5.5.2. Inserting Gα in the Intracellular Interface

To have a better understanding on the binding mode of the ligand to the receptor and obtaining the most active-like structure of AstR-C, Gα part of G protein complex resolved in *Mus musculus* μ opioid receptor was inserted at the intracellular part of the final model. Comparing the sequence of the resolved protein with the full sequence of Gα deposited in uniport (UniprotKB: P63096), two gaps were identified; one between the 55^{th} and 182^{nd} residues, and a another between 233^{rd} and 241^{st} residues. To crosslink the gaps, "crosslink proteins module" of Schrödinger molecular modelling package was used. Residues were taken from the original sequence of Gα and inserted in between the gaps and the gaps were crosslinked. The crosslinked Gα was inserted in the intracellular interface of TpitAstRC (Figure 5.16).

5.5.3. Running MD Simulation for TpitAstRC-Gα Complex

Setting the system up by putting the receptor and Gα complex embedded in POPC membrane, solvents, and ions, 500 ns MD simulation was run. Then, the stability of the receptor and Gα were evaluated separately using RMSD and RMSF (Figure 5.16a, b, c, d, e and d). The analysis for TpitAstRC were conducted with/without N-terminus, and RMSD and RMSF were calculated (Figure 5.17a, b, c, and d).

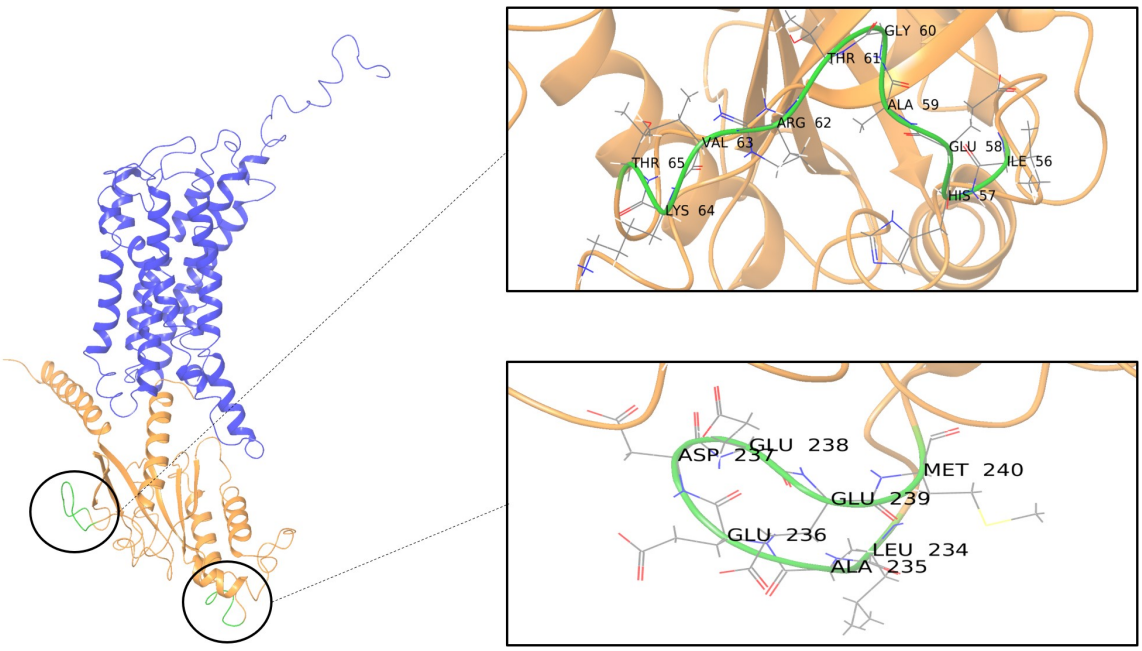


Figure 5.16: 3D representation of TpitAstRC with the crosslinked Gα in the intracellular interface. The crosslinked gaps are illustrated in the right side.

As discussed before, high flexibility of the N-terminus region contributes to the higher RMSD values in the system, and it is obvious when comparing these values in Figure 5.13a and 5.13b. The first 36 residues of TpitAstRC show RMSF values as high as ≈14 Å (Figure 5.16d). Residues residing in ECL and ICL, expectedly had high fluctuation values. C-terminus also showed high fluctuations. Overall, the receptor showed a good stability and the RMSD plot reached to a plateau that lasted during the MD simulation time.

The structure of Gα was investigated as well (Figure 5.16e and f). Although RMSD values were higher compared to TpitAstRC, the structure reached stability during the MD time. The higher values are accepted for this protein since it is a cytosolic protein with higher levels of flexibility compared to TpitAstRC which is the membrane embedded protein.

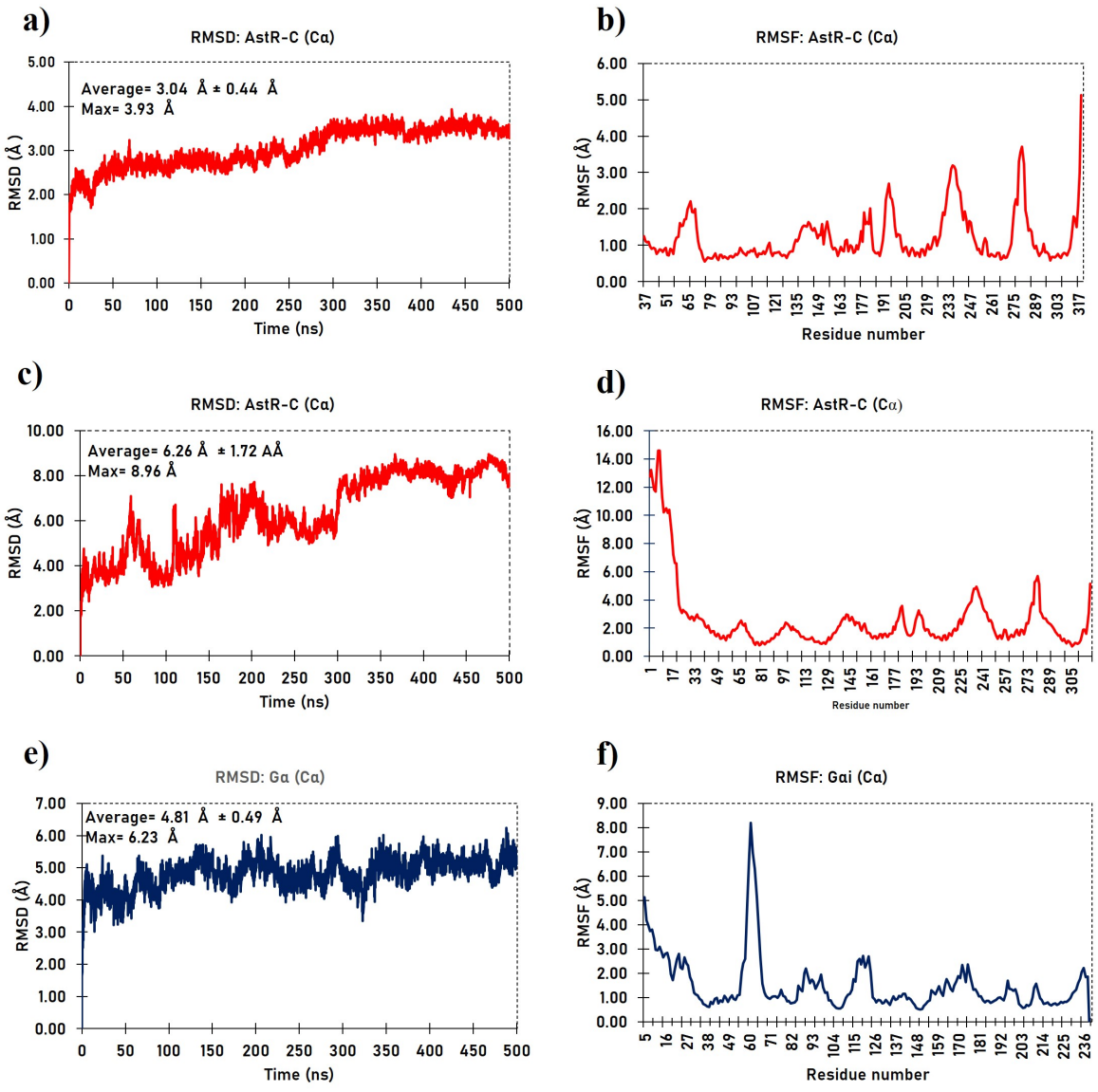


Figure 5.17: System stability evaluation. a) RMSD, and b) RMSF changes without N-terminus. c) RMSD, and d) RMSF changes with N-terminus. e) RMSD, and f) RMSF changes of Gα. MD simulation time was 500 ns. The analysis is given for Cα.

5.5.4. Structure of AST-C

AST-C has 15 amino acid which makes it relatively larger peptide. So, it can fold into various conformations adopting secondary conformations. Predicting the structure of such a large peptide *ab initio* can be very challenging and most likely inconclusive for many molecular modeling programs. Hence, again homology modeling was chosen to be utilized for the model construction. I-TASSER program was used to this aim. After obtaining different models, considering the literature regarding the structure of neuropeptide AST-C, the best model showing the best C-score was chosen.

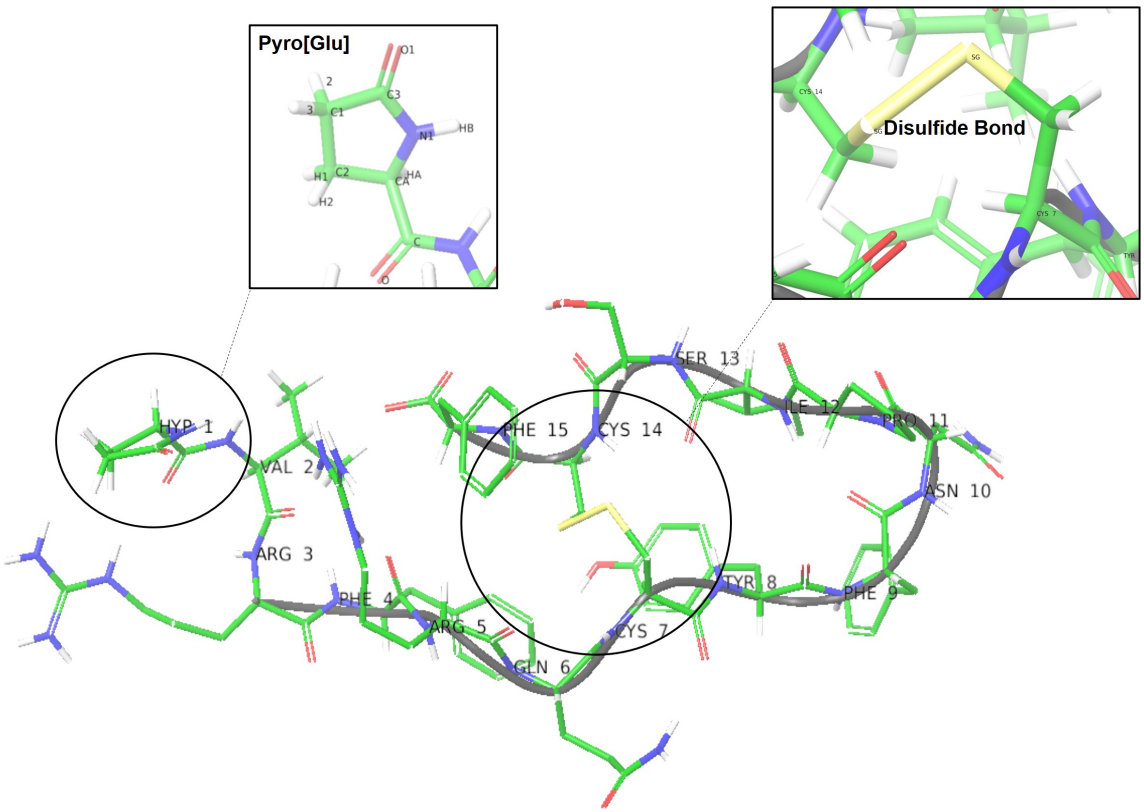


Figure 5.18: 3D structure of AST-C.

C-score is a metric used by I-TASSER which shows the accuracy of the constructed model. Accordingly, models showing C-scores higher than -1.5 are expected

to manifest a structure with the most probable folding to the physiological environment [106]. The C-score of the built model was -1.16. In the modeling procedure, a distance restraint of 2.05 Å (i.e. required for disulfide bond formation) was introduced between two sulfur atoms of Cys7 and Cys14. This was done based on the available literature regarding a disulfide bond between two Cysteins in the secondary structure of Allatostatin peptides. This bond is critical for the structure and function of these peptides. After having the model, the first residue needed to be modified from Glutamine to Pyroglutamate. The 3D structure of the final model constructed for AST-C is depicted in Figure 5.18.

5.5.5. Orthosteric Binding Pocket of TpitAstRC

The orthosteric binding pocket of TpitAstRC was predicted combining docking and MD simulation approaches. 3D models of the receptor and the native ligand were utilized. ClusPro web server was used to perform protein-protein docking. A mask file containing the not-allowed regions for the ligand to bind to the receptor was provided to the program. In GPCRs, the orthosteric pocket is localized in the upper half of the receptor, so the bottom-half was assigned as the prohibited site. The ligand, AST-C, was kept flexible, while the receptor was rigid. 1000 rotamers of ligand were generated, and it was shown that 946 of the rotamers cluster together in an identical pose at the expected orthosteric pocket. The energy of binding for the best pose was -1621.8 kcal/mol. This pose was taken and prepared for MD simulation studies. The simulation system was assembled and subjected to 500 ns MD simulation. The RMSD and RMSF changes of the different components of the system including the receptor, the ligand and Gα were evaluated (Figure 5.19a, b, c, d, e).

The stability of the ligand, AST-C, was assessed using two different RMSD fitting modes. In the first one, the translational motion of the ligand at the binding pocket was considered. Here, RMSD changes of AST-C were calculated with respect to the first frame of the receptor. This was denoted as “Lig-fit-Protein”. In the other fitting mode, assigned as “Lig-fit-Ligand”, however, the rotational motion or in other words

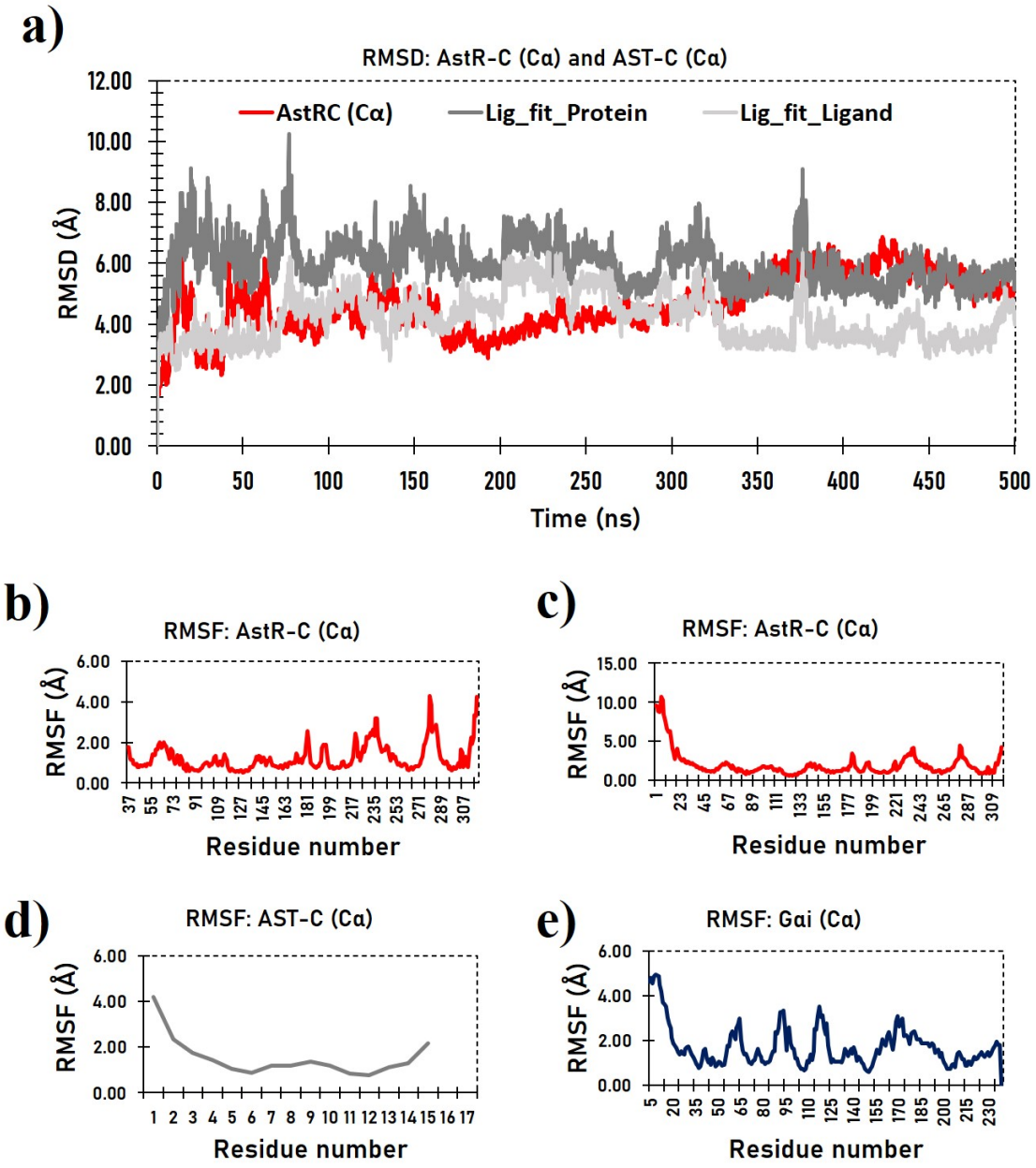


Figure 5.19: Stability of Receptor-Ligand during 500 ns MD simulation. a) RMSD of TpitAstRC, AST-C (Lig-fit-Protein) and (Lig-fit-Ligand). RMSF changes of the receptor c) with, d) without N-terminus. RMSF changes of d) AST-C and e) Gα.

internal fluctuations of the ligand at the binding pocket were investigated. The ligand showed to be deviating from the first frame in both fitting modes during the first 100 ns of the MD simulation, and then reached to a stable mode. (Figure 5.18 a).

Figure 5.19b and 5.19c show the fluctuations of the residues of the receptor, with and without N-terminus. Again, as expected, N-terminus and C-terminus of TpitAstRC manifested higher RMSF values. Among ECL and ICLs, ECL3 had the highest fluctuation. For the ligand, except for the first 4 residues, no considerable fluctuations were observed and overall, it remained stable (Figure 5.18d). Fluctuations of Gα was also investigated, the location of highly fluctuating residues was checked and confirmed to be in the regions such as N-terminus, loops and C-terminus (Figure 5.18e).

The positioning of the ligand in the binding cavity of the receptor is depicted in Figure 5.20a. The interaction between TpitAstRC and AST-C was tracked by investigating the trajectories of the MD simulations, and it was observed that AST-C was in close contact with the ECLs of TpitAstRC and not with the transmembrane residues (Figure 5.20b). ECL2 was found to be mainly involved in binding pocket formation. Five residues including D181, N182, N188, E193 and Q200 were in contact with the ligand. This loop is the longest extracellular loop of TpitAstRC. Residues of ECL3 was also involved in the interaction with the ligand as well. Q278 was one of the residues in this loop that showed long-lasting contact with the ligand. No interaction was observed between the ECL1 and AST-C. In ECL3, shown with green, Q278 forms a relatively stable contact with AST-C.

Different types of interactions including hydrogen and ionic bonds and hydrophobic interactions were spotted between the receptor and ligand. 2D diagram of receptor-ligand interaction is represented in Figure 5.20c. Investigating the interaction fraction of all residue of the binding pocket during the MD simulation time (500 ns), it was found that hydrogen binding was exploited by different residues of TpitAstRC. Hydrophobic interactions, salt and water bridges were also involved (Figure 5.20d).

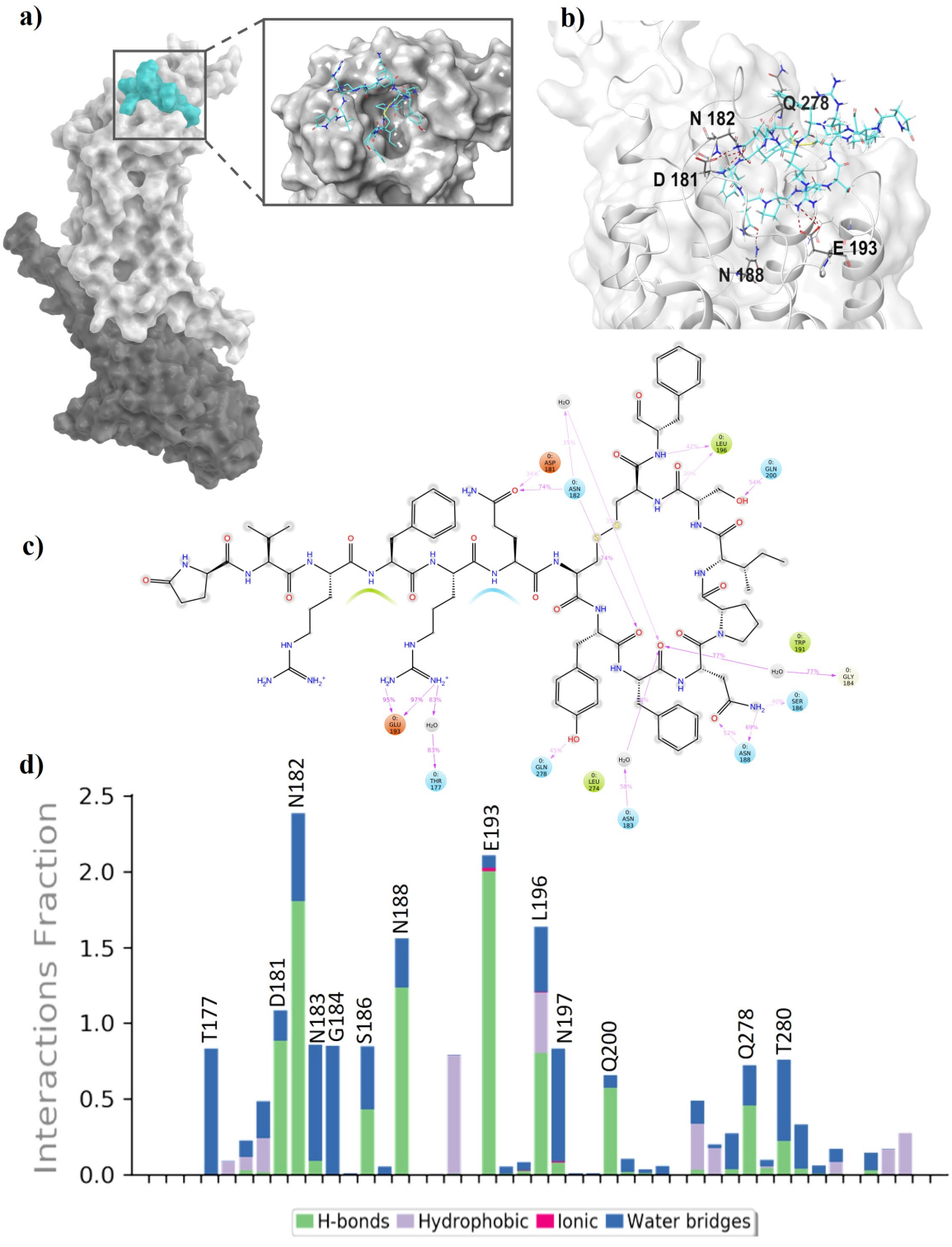


Figure 5.20: Receptor-Ligand interaction. a) Localization of AST-C in the binding pocket. b) Residues of AST-C in contact with TpitAstRC. c) 2D representation of Protein-Ligand interaction. d) Protein-Ligand interaction fraction diagram.

5.6. Verification of the Binding Pocket of TpitAstRC

5.6.1. *In silico* Verification

The importance of the residues of TpitAstRC that were shown to be actively involved in the receptor-ligand contacts were investigated. This was critical in testing the accuracy of the predicted 3D structure of the receptor.

To this aim, some of the spotted residues of the identified binding pocket were selected and substituted to Alanine *In silico*. In addition, $Q271^{6.55}$ (Ballesteros-Weinstein numbering) was chosen, due to the well-known importance of this position for the ligand binding in Class A GPCRs.

The 200 ns MD simulation trajectories of the WT receptor and other mutant receptors were investigated at Apo and Holo forms. It is of note that the Holo form was considered for the ligand-bound receptor that either have or have not coupled to Gα, and denoted as Holo and Holo-No Gα.The state of the structure, naming active, inactive, or intermediate, was evaluated calculating Δd, a metric suggested by GPCRdb that obtains via measuring the distance of two pairs of the residues and then subtracting. According to GPCRdb, for class A GPCRs, Δd below 2.0 Å shows a structure at inactive state, between 2 to 7.15 Å is related to intermediate states of the structure. Values higher than 7.15 Å are attributed to structures at active state.

During the 200 ns of the MD simulation time, at Apo form, WT receptor was found to be in an intermediate state (Table 5.5). Intermediate state is shown as "IM" in the table. Alanine substitution at D181 and N182 positions shifted the state toward more inactive states. E193A and Q278A did completely the opposite and resulted in structures at active state. The structure remained at an intermediate state for N188A, Q200A and Q271A mutants.

Binding of AST-C to the receptor increased the Δd values in general, and expectedly, shifted the structures toward more active states (Table 5.5).

Table 5.5: Δd values and state of WT and mutant receptors at Apo and Holo forms (with/without Gα).

Apo	**WT**	**D181A**	**N182A**	**N188A**	**E193A**	**Q200A**	**Q271A**	**Q278A**
d_2(Å)	21.7 ± 0.9	19.3 ± 0.6	19.5 ± 0.7	22.5 ± 0.8	22.7 ± 0.5	18.7 ± 1.2	21.9 ± 0.7	23.2 ± 0.5
d_1(Å)	18.3 ± 0.9	20.4 ± 0.7	21.0 ± 0.9	15.4 ± 0.5	14.5 ± 0.5	15.6 ± 0.6	15.3 ± 0.5	14.2 ± 0.5
Δd(Å)	3.4	-1.04	-1.44	7.09	8.2	3.08	6.6	9.38
State	IM	Inactive	Inactive	IM	Active	IM	IM	Active
Holo	**WT**	**D181A**	**N182A**	**N188A**	**E193A**	**Q200A**	**Q271A**	**Q278A**
d_2(Å)	21.5 ± 0.9	24.6 ± 1.1	24.5 ± 1.5	25.5 ± 1.5	22.8 ± 0.9	27.6 ± 1.4	21.3 ± 0.6	22.6 ± 0.8
d_1(Å)	14.6 ± 0.6	14.1 ± 0.4	14.5 ± 0.9	14.6 ± 0.9	14.3 ± 0.5	18.2 ± 1.0	19.5 ± 0.7	17.6 ± 0.6
Δd(Å)	6.9	10.15	5.0	10.8	8.4	9.4	1.8	4.9
State	IM	Active	IM	Active	Active	Active	Inactive	IM
Holo (Gα)	**WT**	**D181A**	**N182A**	**N188A**	**E193A**	**Q200A**	**Q271A**	**Q278A**
d_2(Å)	23.5 ± 0.9	21.4 ± 0.7	22.4 ± 0.5	25.1 ± 1.0	22.1 ± 0.4	22.7 ± 0.8	21.5 ± 0.5	21.5 ± 0.6
d_1(Å)	12.0 ± 0.5	15.5 ± 0.5	14.2 ± 0.3	14.9 ± 0.5	13.7 ± 0.3	19.1 ± 1.0	18.2 ± 0.6	17.5 ± 0.6
Δd(Å)	11.5	5.8	8.2	10.5	8.4	3.6	3.3	4.0
State	Active	IM	Active	Active	Active	IM	IM	IM

This was significant for WT receptor, in particular, for which binding of the ligand transitioned the state from intermediate to active. Despite the general trend observed for WT and mutant constructs, point mutation at Q271 position shifted the state of the structure to inactive. WT receptor and mutants were superimposed at Apo form, and TM6 and ECL3 in specific were investigated. It was observed that Q271A mutant possesses a distinct conformation at this region compared to the other constructs, with more inwardly positioned TM6 and ICL3 (Figure 5.21.). At Holo form, it was shown that mutations introduced in the binding pocket change the ligand binding pose compared to the WT receptor (Figure 5.22).

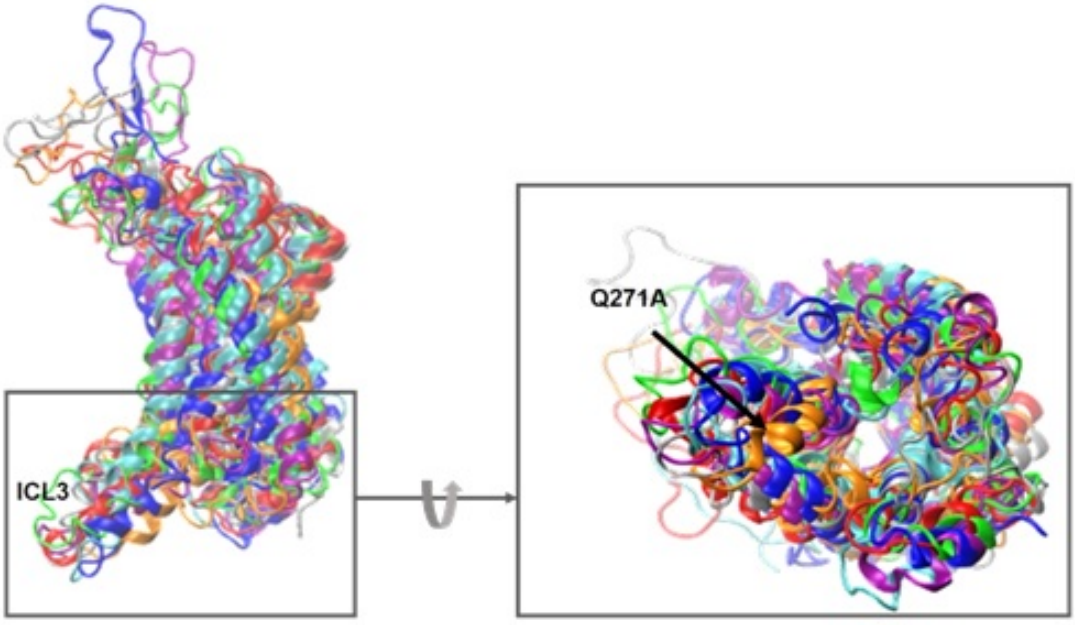


Figure 5.21: Superimposition of the WT and other mutant receptors at Apo form. Rotated and zoomed depiction is represented in the right side.

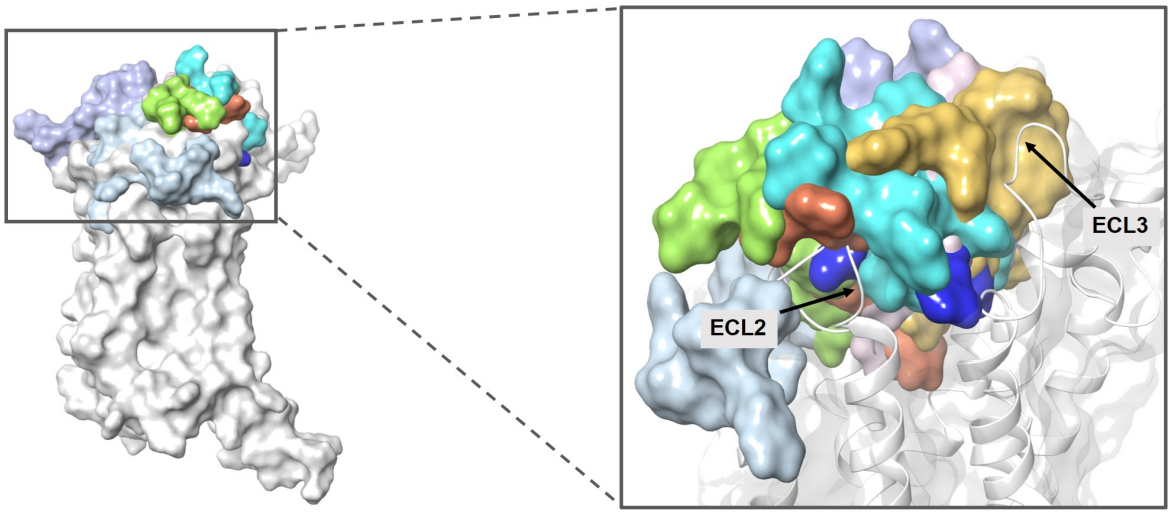


Figure 5.22: Effect of point mutations on the binding pose of AST-C in different mutant receptors. Turquoise shows AST-C in WT receptor.

While ATS-C is mainly positioned between ECL2 and ECL3 at WT receptor (shown in turquoise), in the mutant receptors it either moves toward one of the ECLs or the funnel of the receptor.

5.6.2. *In vitro* Verification

In the next step, some of the residues already investigated via *in silico* studies were selected for *in vitro* site-directed mutagenesis experiments. WT and Alanine substituted receptors were transfected to HEK-TSA cells, and FRET-based G protein activation assay was conducted to test the effect of point mutations on the G protein-dependent activation of TpitAstRC. In addition, a combination of some of these point mutations were generated in order to investigate the collective effect of these residues in forming stable contact with the ligand. The changes in FRET signal of mutant receptors in different concentrations of AST-C were measured and compared with WT (Figure 5.23 and Table 5.6).

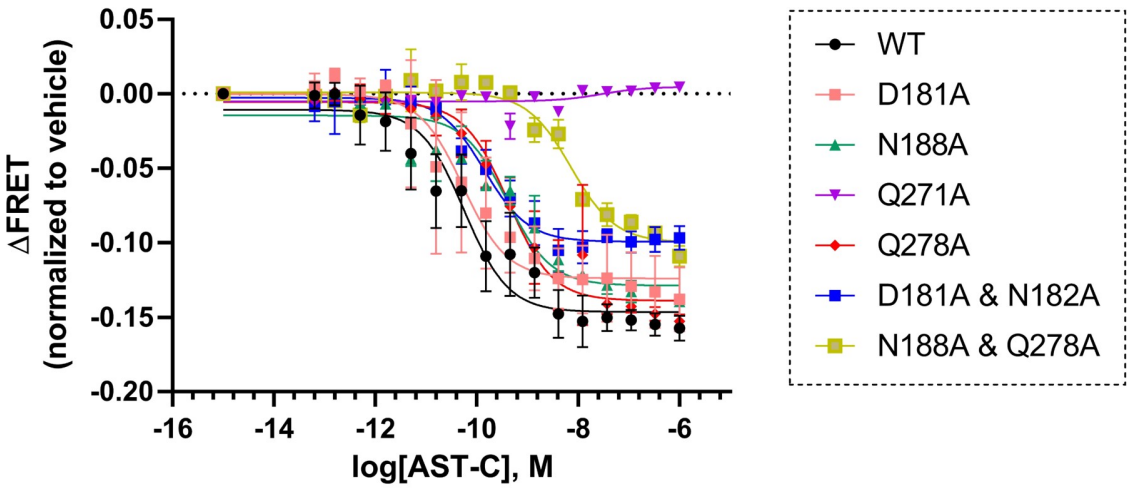


Figure 5.23: ΔFRET changes of different receptors at increasing concentrations of AST-C. The data were fit to Hill equation, using the four-parameter dose-response fit function. The data are representative for at least three biological replicates.

In general, Alanine substitution of residues of the binding site resulted in a right shift in EC_{50} values. The observed effect implied the importance of these residues in the G protein-dependent activation of the receptor. Having more than one mutation in the binding pocket caused more pronounced effects in the increase of EC_{50} values. In addition, the maximum response, ΔFRET, was decreased. Compared to WT TpitAstRC, the decrease in double mutant receptors was almost 30%. The results here showed the significance of the selected residues in forming the orthosteric pocket of TpitAstRC and ligand binding interactions. Hence, it was deduced that the selected residues are truly involved in protein-ligand interaction.

Table 5.6: EC_{50} values of WT and mutant TpitAstRC.

Receptor Constructs	**EC_{50} (M)**	**R^2 (Goodness of fit)**
WT	0.057	0.89
D181A	0.053	0.77
N188A	0.39	0.95
Q271A	-	0.25
Q278A	0.4	0.91
D181A + Q278A	0.13	0.93
N188A + Q278A	7.20	0.94

5.6.3. Q271A Mutant Receptor

In silico and *in vitro* experiments showed that Alanine mutation at position 6.55 results in a distinct behavior of the receptor in G protein dependent signaling. Before any further analysis, we confirmed the plasma membrane localization of this point mutation (Figure 5.24).

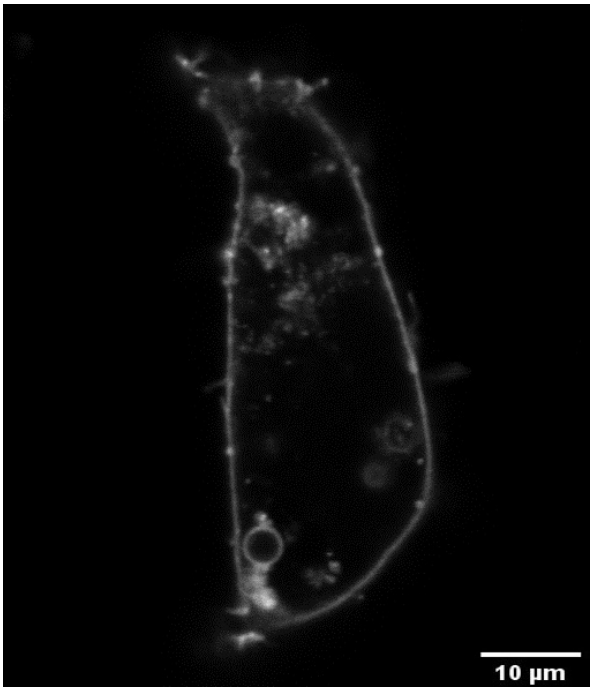


Figure 5.24: Cell localization of Q271A mutant receptor.

Eliminating the probability of not being expressed in the cell membrane, we hypothesized that this point mutation can either exerts its effect at Apo or Holo forms. Hence, we decided to extend the already performed MD simulations to 500 ns for this point mutation and then analyze the internal.

PCA analysis was performed for Apo and forms of WT and mutant receptor. The total mean square displacement (or variance) of atom positional fluctuations were projected onto the subspace defined by the largest three principal components. First three PCs of the system covered more than 60% of the movements in the receptor (Figure 5.25).

In Figure 5.24 the coloring, from blue to red, shows the changes in the displacement of residues along the MD simulation time. Comparing these displacements in WT and mutant receptors, the pattern of the movements is obviously different between the two structures.

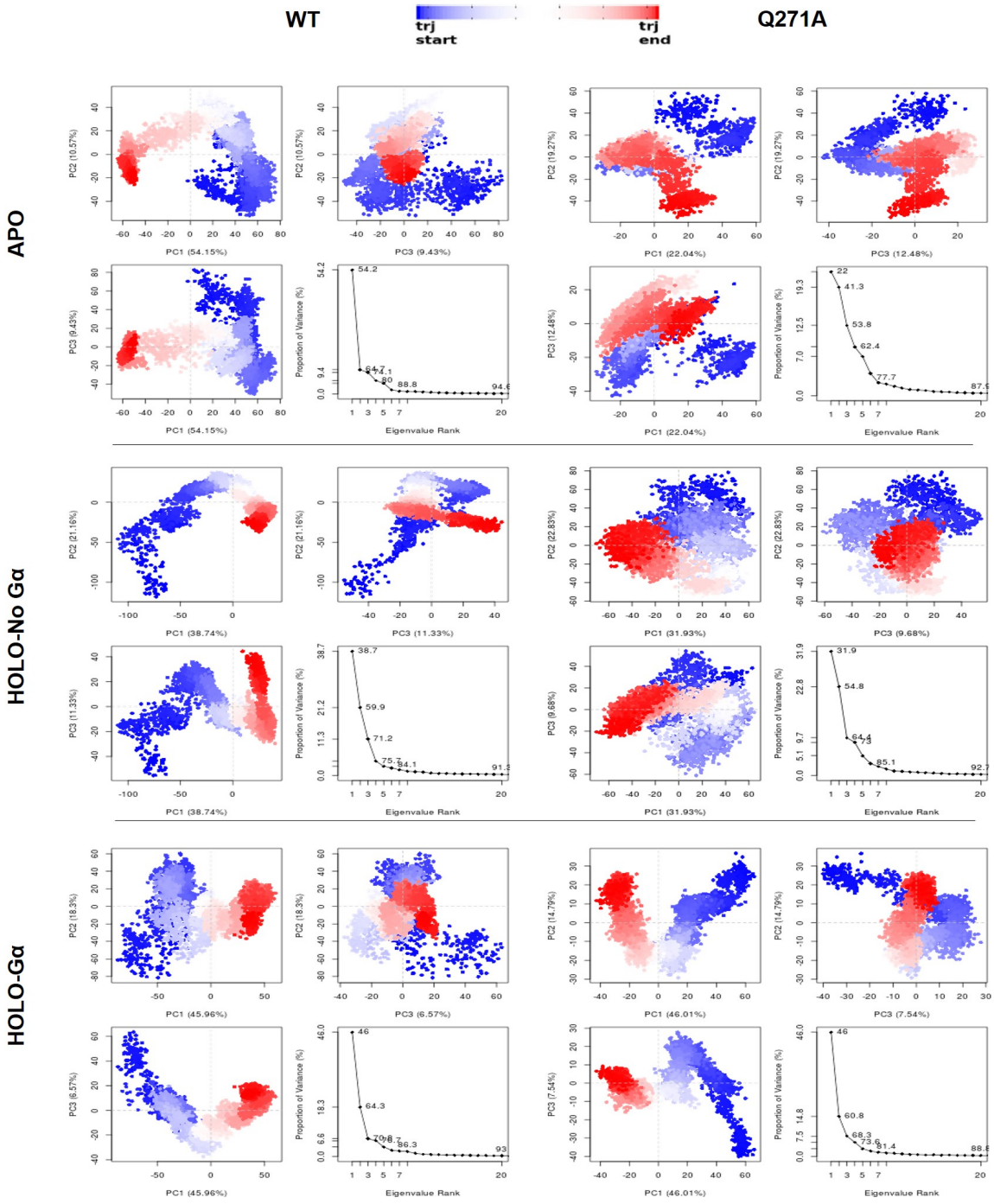


Figure 5.25: PCA results for trajectories with instantaneous conformations.

Calculating the eigenvalue magnitudes for each PC, it was observed that WT receptor has higher magnitudes compared to Q271A mutant receptor (Figure 5.26). The similar trend was detected at both Apo and Holo forms (with/without Gα).

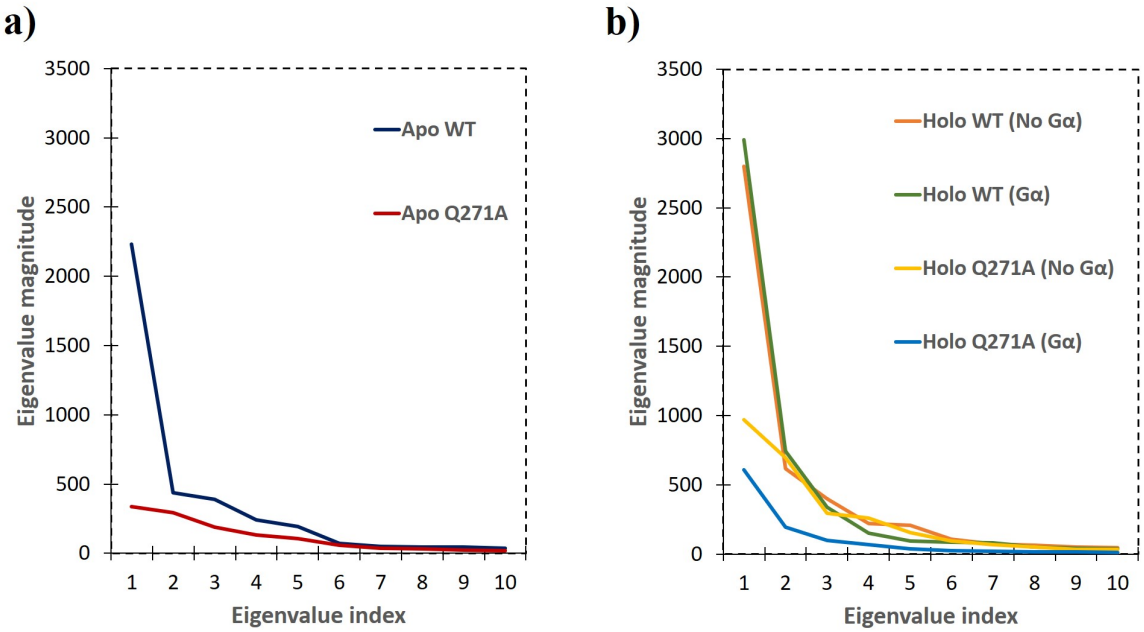


Figure 5.26: Eigenvalue magnitudes. Analysis of Eigenvalues corresponding to eigenvalue indexes for of the first 10 modes of action of (a) WT and, (b) Q271A receptors.

Scrutinizing MD simulation trajectories, we checked the correlated and uncorrelated internal motions of the structures. Dynamical cross-correlation map was obtained for each structure and compared (Figure 5.27). In the maps, X and Y axis represents the residues of the structure and the movements of each residue compared to other residues of the protein is shown with a color range of red to blue. Red color shows highly correlated motions, and blue color, on the other hand, depicts those in opposite direction. A very distinct pattern of movements was observed for Q271A receptor. The population of both the correlated and uncorrelated displacements were drastically reduced in this structure.

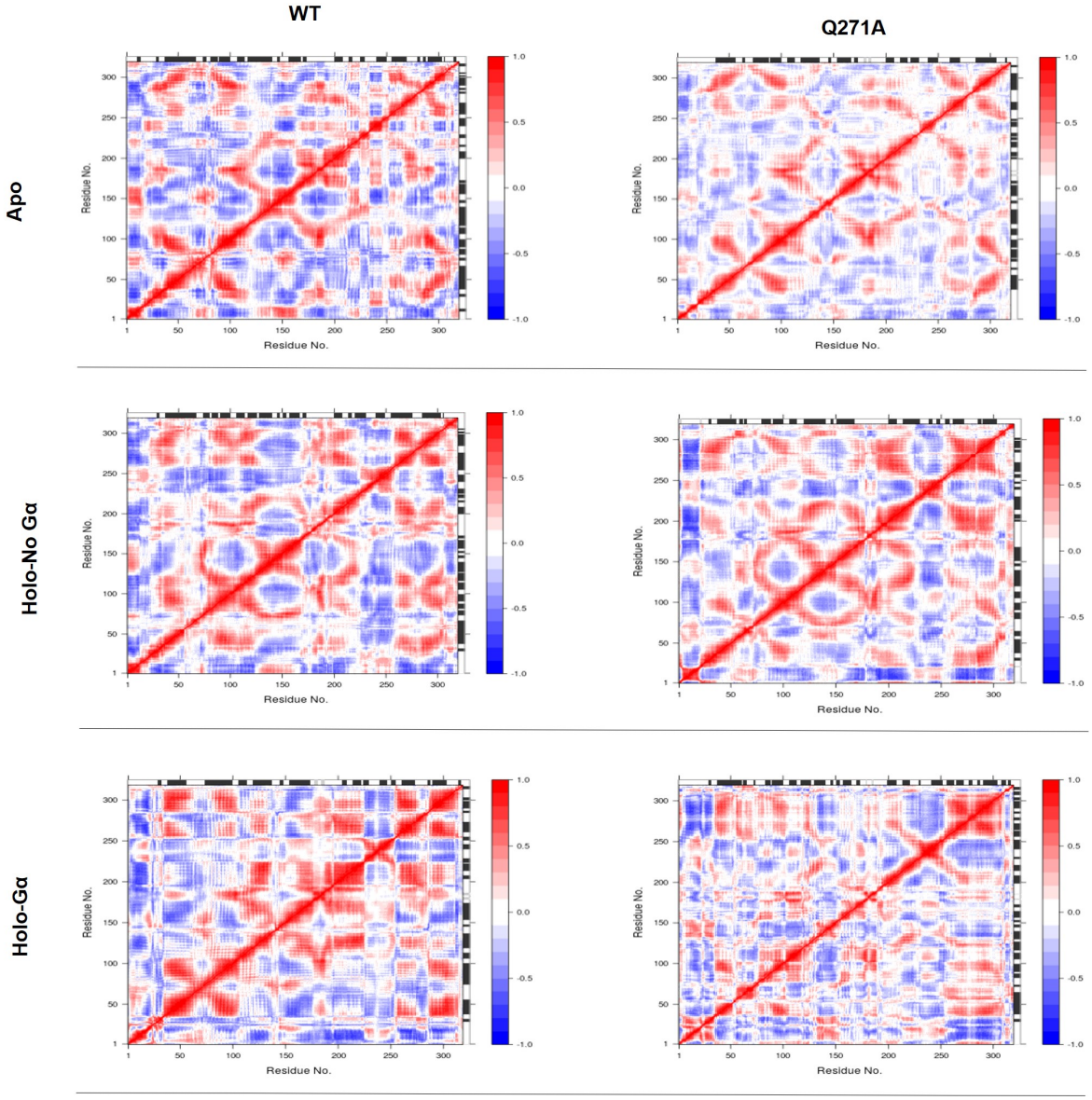


Figure 5.27: Dynamical cross-correlation map. Correlated (red) and un-correlated (blue) displacements were compared between WT and Q271A receptor at different states.

5.7. Virtual Screening Studies to Find Agonist Molecules

5.7.1. Grid Generation

A grid needed to be generated before proceeding with the virtual screening studies. Different sites were identified by "Sitemap module" of Schrödinger. These pockets were ranked according the likelihood of the cavity to be a ligand-binding pocket. In TpitAstRC two sites were identified in the upper half of the receptor, and SiteI with the highest probability was taken (Figure 5.28). In fact, this pocket showed the highest probability value among all other sites, and it was positioned in the orthosteric site of the receptor. The grid was inserted at this site.

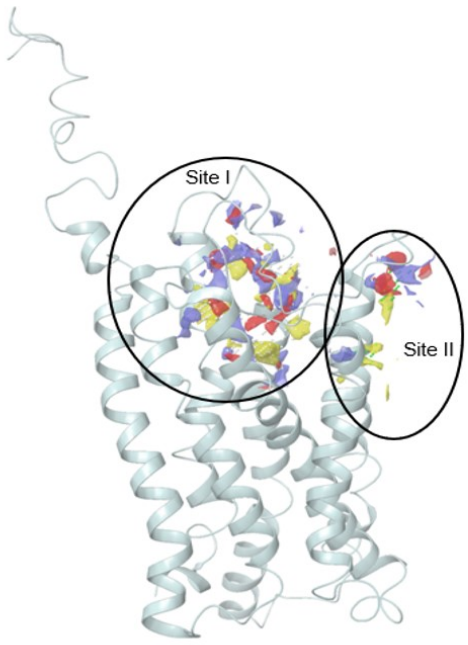


Figure 5.28: Two top-ranked binding sites of TpitAstRC identified by SiteMap.

5.7.2. Screening OTAVA β-turn Peptidomimetics Library

β-turn peptidomimetics library contains 1058 compounds. After ligand preparation step, the number of compounds reached to $\approx$ 3000. This is because of the conversion from 2D to 3D structure in that each molecule have different chiral centers

which results in various stereoisomers and tautomers of a compound. The screening was performed targeting the binding pocket of TpitAstRC, thus structure-based virtual screening approach was exerted here. The aim was to search for small molecules that can activate TpitAstRC.

In the generated grid, small compounds were subjected to high-throughput virtual screening (HTVS) workflow. In the first step of the workflow, 28277 poses were generated, and all were used in the screening. The highest and lowest (absolute number) docking score were -8.59 and -1.62 kcal/mol, respectively. For the finally obtained docking poses, binding efficiency (BE) values were also calculated dividing the binding energy to the number of the heavy atoms of each compound. Thus, two ranking methods were combined to narrow down the list, one based on the free energy of binding and another according to the BEs. The reason for considering the BE values as an additional filter was because of the varied number of heavy atoms of the compounds. In fact, bigger molecules are expected to rank higher due to the higher number of heavy atoms, however, these compounds might not bind to the target very efficiently. Therefore, we took dual-ranking approach in all steps. Docking score of -7.0 kcal/mol and BE of -2.0 kcal/mol were set as cutoff. 209 compounds that fulfilled this cutoff. were subjected to 1 ns MD simulation.

In order to calculate the free energy of binding during the MD simulation time, MM/GBSA analysis was performed taking 100 frames of the generated trajectories for each molecule. The best free energy of binding obtained at this step was -90.98 kcal/mol (Figure 5.29).

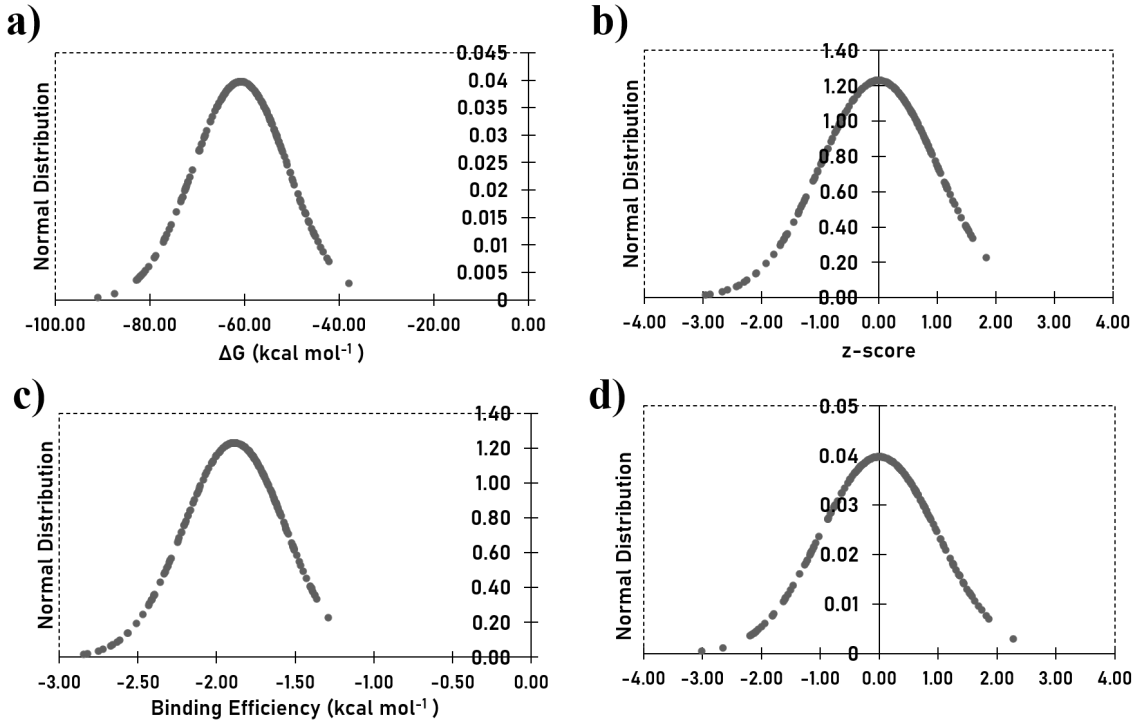


Figure 5.29: Statistical analysis performed for 1 ns MD simulation results. Normal distribution of a) ΔG, and b) z-score normal distribution. Normal distribution of c) BE, and d) z-score.

Hits were ranked, and normal distribution of ΔG and BE were plotted. Another plot for the same set of data were generated calculating z-score. The first row of the Figure 5.28 shows analysis of free energy values, ΔG, (normal distribution and z-score), while the second row depicts the analysis result of BE values (normal distribution and z-score). In both lists, z-score of -1.00 was set as the cutoff and compounds with z-score lower than -1.00 were taken. Then, two lists were combined, and unique compounds were selected. At this step, 44 unique hits were chosen to be subjected to 5 ns MD simulation runs. The same procedure was pursued for the analysis of 5 ns simulation results (Figure 5.30). The molecules of the both list with z-score better than -1.0 were chosen, and prepared for the next 50 ns simulation run (11 poses).

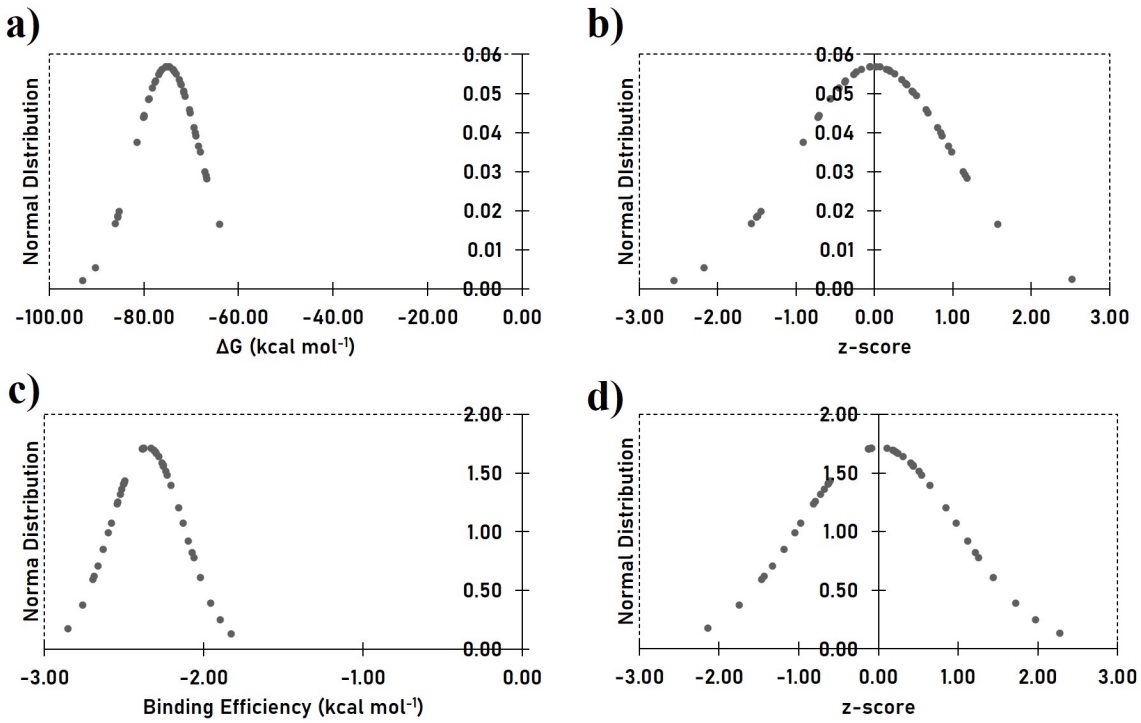


Figure 5.30: Statistical analysis performed for 5 ns MD simulation results. Normal distribution of (a) free energy of binding (ΔG), and (b) z-score normal distribution. Normal distribution of (c) binding efficiency (BE) and (d) represents and z-score.

For the finally selected 11 poses, in addition to MM/GBSA analysis, RMSD, RMSF and the interaction between the hit molecules and the receptors were evaluated during the MD simulation time. Binding energy and BE of the final hits is given

in Table 5.7. The highest ΔG and BE were -95.19 and -2.66 kcal/mol, respectively, obtained for the compound with code P7119988823.

Table 5.7: ΔG and BE values of final hits.

Compound Code	**ΔG (kcal/mol)**	**BE (kcal/mol)**
P7119988823-1	-95.91	-2.66
P7112410434-2-2	-80.65	-2.06
P7119812493-2	-78.67	-2.45
P7119812967-1	-78.42	-2.37
P7717730050-1-4	-77.17	-2.57
P7112410434-2-1	-75.57	-1.93
P7717730057-1-2	-72.49	-2.49
P7717730050-1-1	-72.08	-2.40
P0118650037-2	-68.55	-2.28
P7717730057-1-3	-67.37	-2.32
P7717730057-1-6	-67.13	-2.31

Seven unique molecules were found. The 2D structures of these compounds is illustrated in Figure 5.31. Investigating the interactions between the small molecules and the receptor, at least one residue of the receptor previously detected to be in contact with the native ligand, AST-C, was detected (Protein-Ligand interaction fraction diagrams are given in APPENDIX C).

Seven unique molecules found from the screening in peptidomimetics library were tested by *in vitro* experiments. G protein activation assay was performed to evaluate the activation of TpitAstRC following the administration of small compounds.P7112410434 compound was the only candidate that behaved as agonist and resulted in the G protein-dependent activation of the receptor at EC_{50} of 0.2 μM (Figure 5.32).

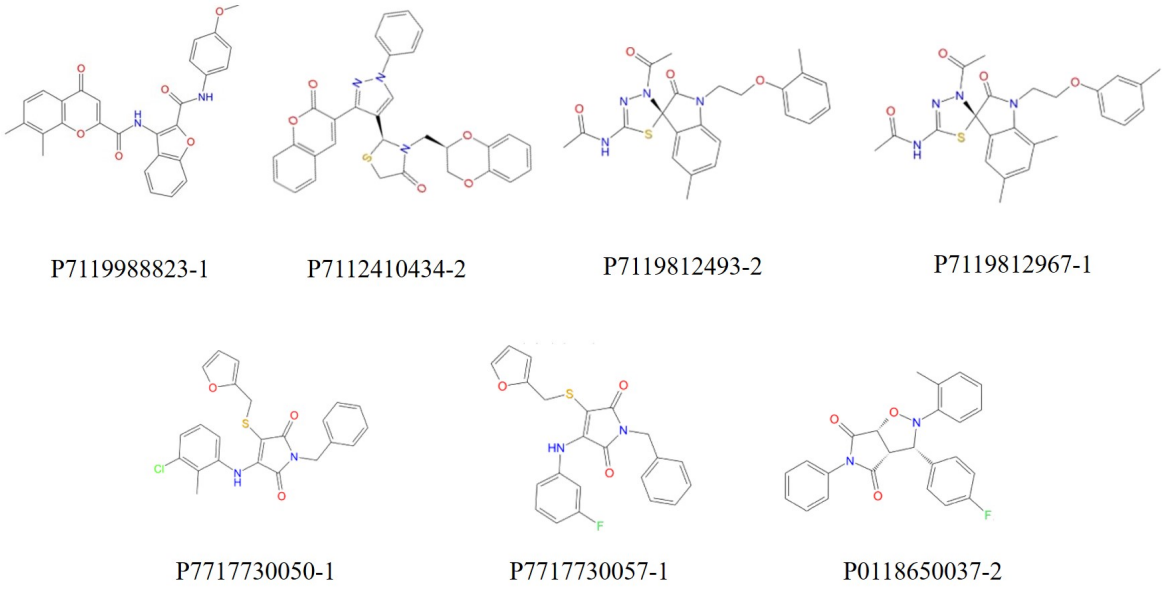


Figure 5.31: 2D structure of compounds.

Two different poses of this compound were top-ranked in 50 ns MD simulations, each showing a different mode of binding in the binding pocket. According to the MM/GBSA results, one of these two poses ranked as the second best hit in terms of the free energy of binding with ΔG and BE of -80.65 and -2.06 kcal/mol, respectively. In this pose, two hydrogen bonds were inspected between Q200 and Q271 residues of the receptor and the ligand. Here, hydrophobic interactions were more prevalently involved in the protein-ligand interaction.

P7717730057 and P7119988823 were found to be inverse agonists of the receptor and behave in apposite direction of agonists. Interestingly, P7119988823 was the first hit in our virtual screening list having the highest free energy of binding, -95.91 and BE of -2.66 kcal/mol. Although this compound binds tighter to the receptor compared to the agonist compound, in the downstream it results in a completely different effect.

Other four compounds including P118650037, P7119812493, P7119812967 and P7717730050 were behaved like antagonists. However, it is not possible to decide whether they have bound to the receptor and then inhibited the activation or they

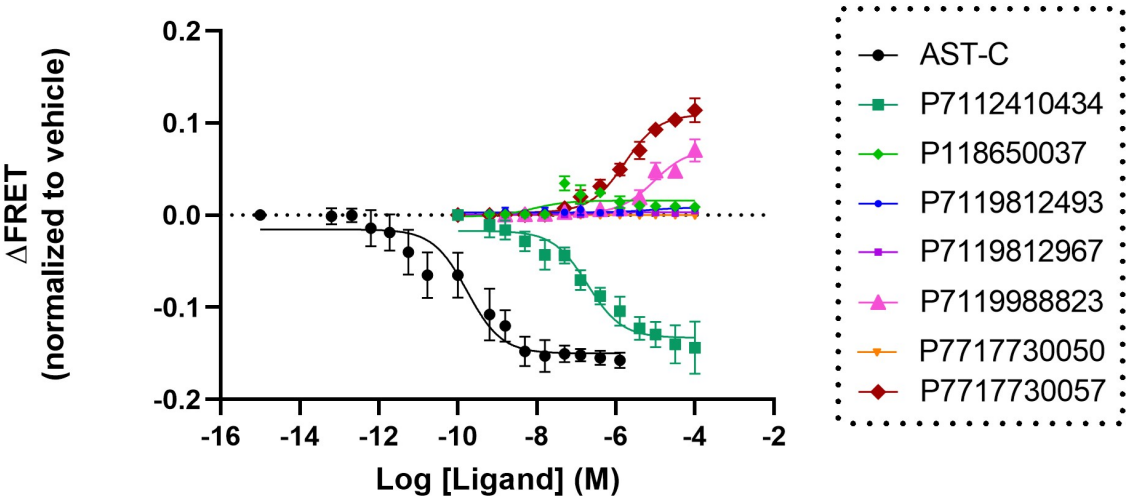


Figure 5.32: G protein activation assay performed for the finally selected seven compounds.

could not bind to the receptor.

6. DISCUSSION AND CONCLUSION

Insect neuropeptide receptors are responsible for the regulation of many biological processes including growth, development, reproduction, and behavior. These receptors have been proposed as potential targets for pest control agents since manipulation of their activity can disrupt the normal fitness of the organisms in a direction to control their population. Different arthropod GPCRs can be targeted by agonists or antagonists to be activated or blocked, respectively, according to their function.

Despite their importance, low number of insect receptors have been identified so far. In addition, even for the known receptors their pharmacology is not well studied. Like vertebrate GPCRs, arthropod GPCRs are expected to have high level of complexity. The complexity arises from possible modes for the binding of ligands to receptor and various downstream signaling pathways that can be initiated upon the stimulation of receptors with different ligands.

In this work, a very comprehensive study was performed on Allatostatin receptor type C of *T. pityocampa* to acquire insights into the structure and function of the receptor with a future perspective of designing small molecules capable of manipulating the physiology of the insect to control its outbreaks.

6.1. WGS Data Revealed the Sequence of TpitAstRC and AST-C

Similar to human genome, genome of arthropods encodes a significant number of GPCRs each of which can be considered as potential targets and can be evaluated. The genome studies performed in fruit fly, *D. melanogaster*, resulted in the identification of 200 GPCR [131]. The same approach was used for investigating GPCRs of *Anopheles gambiae* (malaria mosquito), *Tribolium castaneum* (red flour beetle), and *Bombyx mori* (silkworm), and as a result 276, 84, and 111 GPCRs were annotated, respectively [132].

Considering the importance of the identification of arthropod genomes in agriculture, energy production, medicine and security of food, an initiative called i5k was launched in 2011 with the aim of providing the genome data of 5000 insects.

Previously no data were available on the genome of the insect and only *de novo* transcriptome analysis was available in the databases [133] in which no information could be obtained for the target receptor of this study, TpitAstRC.Here, the WGS analysis of pine processionary moth was performed for the first time and the genome was assembled *de novo*. The prepared gDNA samples were sent to be sequenced in Macrogen Company. One out of five gDNA samples showed acceptable quality which could have been improved by modifying the first step of the tissue preparation in a way to keep the tube of the sample in contact with liquid nitrogen during the smashing step and decreasing the temperature as much as possible.

The assembled genome was compared in terms of the genome size and GC content with other evolutionarily close species for which the genomic data were available in i5K database. It is of note that no genome data could be found for other species of Notodontidae family that *T. pityocampa* belongs, and species selected for the comparison were chosen from the closest family that is Noctuidae.

The sequence of the target receptor, TpitAstRC and its endogenous ligand, AST-C, were derived from the assembled data and used at the next steps of the study. This thesis work supplies a huge data for future studies on the same pest. For instance, in an attempt to find more potential targets for next-generation pesticide development, other GPCRs with important physiological functions including Octopamine/Tyramine receptor, Short neuropeptide F receptor, FMRFamide receptor and Allatostatin receptor type A can be extracted from the data obtained here. We could find some signatures of these GPCRs within the genome but did not include since they were not the focus of this thesis work.

6.2. RET-Based Studies Revealed Pharmacological and Functional Insights on TpitAstRC/AST-C System

In accordance with the low number of insects GPCRs identified, not many functional studies have been performed either. Among the completed studies with the aim of characterizing these receptors, biogenic amine-binding receptors that belong to Class A GPCRs were extensively studies [134].These receptors include dopamine receptors, octopamine/tyramine receptor, and serotonin receptors.

Here, RET-based methods were utilized to evaluate the activation of TpitAstRC, upon treatment with the endogenous ligand, AST-C. GPCRs after being stimulated with their cognate ligand, following the conformational changes, couple to cytoplasmic G protein heterotrimers [135]. G protein heterotrimers constitute of α, β and γ subunits.Gα subunits are classified into four subtypes, Gαs, Gαi/o, Gαq and Gα12/13. In general, each GPCR recognizes one specific type and accordingly initiate a downstream cascade [136]. Following the binding of the native ligand to the receptor dissociation of Gα and G$\beta\gamma$ occurs. In G protein activation assay, by tagging Gα and G$\beta\gamma$ with different fluorescent proteins, and measuring the changes in fluorescent intensity that happens following the dissociation of these partners, G protein coupling of GPCRs can be studied. Provided that the stimulated receptor binds to the relevant subtype of Gα protein, due to the dissociation of G protein subunits and the increase in the distance between them, a reduction in FRET signal happens.

The results of G protein activation assay revealed that Gαi is the relevant G protein subtype coupled to the receptor at sub-nanomolar ranges of the native ligand. The result obtained for G protein coupling of TpitAstRC was expected due to the evolutionary relationship between human Somatostatin receptors and Allatostatin receptors in insects. Human somatostatin receptors are proven to couple with Gαi/o subtype as well [137]. The Gαi coupling of TpitAstRC can be validates by testing cAMP levels following the ligand administration. A decrease in cytoplasmic cAMP is expected, as Gαi coupling results in the reduction of intracellular cAMP level via inactivation of

adenyl cyclase, an enzyme that catalyzes ATP to cAMP conversion.

Besides the determination of G protein subtype preference of TpitAstRC, the kinetics of G protein coupling upon stimulation with AST-C was investigated. The kinetics studies were evaluated at two events including G protein recruitment and G protein activation. Here, the time required for the recruitment and activation of the receptor following a brief application of native ligand was investigated. In the recruitment event,results showed that the G protein complex associates to the receptor in ≈4 seconds after the application of 1 nm concentration of AST-C, but the dissociation lasts longer (≈70 seconds). In G protein activation event, *on* and *off* kinetics of G protein were investigated and it was observed that at the same concentration (1 nM) of AST-C, ≈6 seconds after the application of the ligand, G protein subunits are dissociated (dissociation step) and ≈60 seconds later reunite (re-association step). At both kinetics events, a brief administration of the ligand was found to be enough to initiate G protein coupling and the following activation within ≈4-6 seconds. The termination of the initiated effect, however, lasted much longer (60 to 70 seconds). This suggested the potency of AST-C ligand in activating the G protein dependent signaling pathway. At higher concentrations of the endogenous ligand, the re-association and *off* kinetics could not be observed in G protein recruitment and activation events, respectively, for 400 seconds measurement time.

Comparing the acquired G protein activation time to other GPCRs activated by small molecules at ms range, the time obtained here was longer which can be attributed to the large size of the peptide ligand, AST-C, and its binding modes [138–141].

The recruitment of β-arrestin as another important intracellular effector of GPCRs was evaluated as well and it was shown that this scaffold protein, which is important in initiating desensitization and G protein-independent pathways, was recruited to the receptor at micromolar ranges of AST-C. Thus, in TpitAstRC, compared to G protein-dependent pathway, β-arrestin is being recruited at almost thousand times higher concentrations of the native ligand. In G protein activation assay, an EC_{50} value

of sub-nanomolar ranges was observed. We showed the recruitment of β-arrestin here, however, nothing can be deduced regarding the downstream effect of the β-arrestin recruitment. It can either initiate the desensitization of the TPitAstRC or other β-arrestin dependent signaling pathways.

To our knowledge, the present work and another similar study done in our laboratory (by Ali Işbilir, Molecular Biology and Genetics, Boğazici University) on Allatostatin receptor type C of *Carausius morosus*, stick insect, are the only insect neuropeptide receptors investigated by cutting-edge RET-based methods with the aim of the identification of downstream effectors. These pharmacological evaluations are of great importance for future pesticide design studies. More specific drugs with less side effects can be designed as a result of such studies.

6.3. *In silico* and *In Vitro* Studies Provided Significant Structural Insights on TpitAstRC and AST-C

As a novel neuropeptide receptor, no structural information was available for this protein. Here, the goal was to acquire structural insights on the structure of TpitAstRC. In addition, the potential structure of the native ligand, AST-C was investigated. Having reliable models for the receptor and the ligand, the binding mode of receptor-ligand was evaluated and orthosteric pocket was identified.

Before proceeding with structural studies, the sequence of AstR-C was investigated for characteristics features and motifs of ClassA GPCRs. All conserved motifs were found in the sequence with the only exception observed at position 6.30 (Ballesteros-Weinstein generic numbering). At this position, a Glutamic acid is expected, however, it was substituted with a Histidine in TpitAstR-C. In Class A GPCR, E6.30 plays a crucial role in the activation of the receptor since it is involved in the formation of an ionic lock with two more key residues naming $R^{3.50}$ and $T^{6.34}$. Looking for similarities among other GPCRs, it was found that opioid receptors have Histidine at this position as well. It is shown that this substitution does not affect the activation

process of these proteins since Histidine is still capable of forming hydrogen bond network with $R^{3.50}$ and $T^{6.34}$ [142]. Thus, TpitAstRC possesses all the substantial motifs required for the receptor activation Class A GPCRs.

In *silico* homology modeling approaches offers a reasonable alternative for X-ray crystallography and cryo-EM, to investigate the structure of proteins. The method uses the structural information of other proteins with solved structure deposited in PDB and construct a model based on the similarity. Homology modeling reduces the required time and money significantly. However, the validation of the models generated using this approach can be challenging. The problem become more serious when the sequence identity between the target protein and those available in PDB is low. In recent years, the number of resolved structure for GPCRs is proliferated that has contributed positively to the application of homology modeling and the accuracy of the constructed models.

Utilizing homology modeling approaches, we constructed 3D models for both the receptor and endogenous ligand. Opioid receptor of *Mus musculus* with PDB ID 6DDE was proven to be the best template based on the sequence identity. Bound to an agonist, the structure of this template was resolved in active state. Hence, it was an ideal template according to the aim of our project which was to find agonist molecules using the 3D structure of TpitAstRC. In fact, in drug discovery studies, the initial state of the template should be considered before starting to construct a model. According to the final aim of the project being to design agonists or antagonists, one should try to choose templates in active or inactive states, respectively. Expectedly, the conformation of the protein in general, and binding pocket residues in specific at each of the mentioned state is so that it allows a specific binding mode. Thus, using an active template to build a model the binding way that best suits for the binding of agonist or antagonist binding, that in turn, increases the chance of finding small molecules that behave as agonists. In addition, to even more increase the probability of finding agonist molecules, we decided to insert Gα protein in the intracellular interface of TpitAstRC to keep the receptor at the active state during the MD simulation runs.

For the endogenous ligand, very little details regarding the 3D structure of these ligands were available in the literature and we tried to look for any piece of data that helps obtaining a more reliable structure. We found that there is a structurally very important disulfide bond between Cys7 and Cys14 of the ligand that we considered in modeling studies. The 3D model of AST-C had a turn-like structure.

After acquiring reliable models for the receptor and the native ligand, protein-protein docking simulation was performed to find the residues of the receptor involved in the ligand binding. The obtained pose was subjected to long MD simulations. The results of three independent replicates of MD simulations (500 ns) were scrutinized to identify the residues of TpitAstRC that are involved in interaction with the ligand. It was shown that the ligand locates between ECL2 and ECL3. But, in terms of the number of residues in contact with the ligand, ECL2 and not ECL3 was the main loop in establishing protein-ligand interaction. The outcome was in line with the literature over class A GPCRs which shows the importance of ECL2 in forming the binding cavity and ligand binding [143]. Our results illustrated that only one residue of ECL3(Q278) is involved in long-lasting interaction with AST-C. On the other hand, none of the residues of ECL1 were detected as ligand binder. The significance of ECL2 in formation of the binding pocket and interaction with the ligand can be also deduced considering the length of this region. Having 32 residues, it is the longest of ECLs. ECL1 and ECL3 each have only 13 residues. Thus, ECL2 as a long loop that is not embedded in the membrane bilayer can move more freely and interact with the ligand.

The significance of the identified residues in forming the binding cavity of the receptor was investigated both *in silico* and *in vitro*. These residues were substituted with Alanine virtually and then docking studies were performed for the mutant receptors and the ligand binding site was compared. The results showed that point mutations in the identified orthosteric pocket changes the binding mode of the native ligand. In the mutant receptors, the ligand moves from its position in WT receptor (between ECL2 and ECL3) and possess a distinct location. In most of the mutants the ligand moved toward the funnel of the receptor. It can be explained in part by the reduced steric

clash in the binding pocket following the Alanine substitution, due to the smaller side chain of Alanine compared to the substituted ones that allow the ligand to move deeper in the cavity. In general, the changes in the positioning of the ligand following Alanine substitution is expected as each substitution changes the conformation of the receptor that in turn can results in distinct binding modes.

Alanine substitution was also performed *in vitro* by conducting site-directed mutagenesis experiments in which some residues previously tested *in silico* were substituted with Alanine. The effect of mutations on G protein activation of the receptor was evaluated comparing the EC_{50} values of the mutant receptors with WT receptor. Results showed that almost all substitutions results in a shift in the EC_{50} values toward higher concentrations of the native ligand, meaning that higher amounts of AST-C are required for G protein-dependent activation of the receptor. In double mutant receptors, in addition to the EC_{50} shift, Alanine substitution reduced the FRET maximum response as well. Maximum response shows the efficacy of the ligand in activating the G protein signaling, and following the mutation in the binding site of TpitAstRC, at same AST-C doses, the capacity of the receptor in G protein activation was reduced when compared to WT receptor. It implies the substantial effect of the mutations in the binding site on G protein-dependent activation of the receptor via changing the conformation in the structure that allosterically conveyed to the G protein binding site located in the intacellular interface. Hence, sit-directed mutagenesis experiments followed with G protein activation assay illustrated the accuracy of the constructed 3D model of the receptor and it showed that the identified residues were truly involved in establishing interaction with AST-C.

In silico and *in vitro* results revealed the drastic effect of residue positioned at 6.55 on the G protein dependent activation of TpitAstRC. When substituted by Alanine, no FRET signal was detected following the administration of the native ligand that shows incapability of this mutant receptor in initiating G protein activation. Different replicates of 500 ns MD simulation runs did not show any direct interaction between this residue and AST-C. However, we decided to perform analysis for this

residue as well, due to its well-known role in the ligand binding in most of Class A GPCRs [144, 145]. Running MD simulation for the Apo form of this mutant receptor and comparing the 3D structure with WT by superimposing the structures, a distinct positioning of ICL3 of the receptor was observed for Q271A mutatant. ICL3 movement is of great importance in G protein activation of GPCRs, and this specific positioning of ICL3 at Apo form might inhibit the further movements required for the activation of the receptor following the binding of the native ligand. In addition, Δd calculations performed for state determination in WT and mutant receptors at Apo form and two holo forms, showed that ligand binding to Q271A mutant receptors shifts the state of the structure to inactive. The result here was in accordance with *in vitro* experiments in which AST-C treatment did not result in G protein activation. Considering results from both methods, we hypothesized that Q271A substitution may change the structure of the receptor so that it inhibits the propagation of the conformational changes required for the receptor activation following the binding of the ligand. Consequently, we evaluated this hypothesis via atomic structural studies. Internal fluctuations of the receptor at different forms, Apo and Holo, were investigated via PCA and cross-correlation analysis and results were compared between the WT and Q271A mutant receptors. Comparing the PC movements of WT and Q271A, a different pattern of PCs was observed. Additionally, eigenvalue magnitudes of this mutant receptor was significantly lower than those of WT. GPCRs are highly flexible allosteric proteins that their activation requires many conformational changes in the structure following the ligand binding but as PC analysis revealed, Q271A mutant receptor has considerably less internal motions that we speculated might be the underlying factor for the loss of activation. It should be noted that similar effect of point mutation at position 6.55 was reported by Change *et al.* [146], in kappa opioid receptor (κOR), where their results showed that Alanine substitution disrupts the TM6 and ICL3 outward movements.

The effect of this point mutation on β-arrestin recruitment of TpitAstRC can be evaluated in BRET studies to test whether the Alanine substitution is only fatal for G protein-dependent signal pathway or it influences the β-arrestin dependent pathway as well.

6.4. Virtual Screening Studies Resulted in an Agonist for TpitAstRC

There is an urgent need for finding new insecticides with novel modes of action to control the pest population in pine forests. Besides the serious environmental and public health issues inherent in using the conventional insecticides, resistance to these chemicals have reduced their effect. According to the classification of all the available insecticides based on their physiological target made by the international association of crop protection called insecticide resistance action committee (IRAC) [147], these chemicals are divided into five groups (Figure 6.1). One of these five groups accommo-

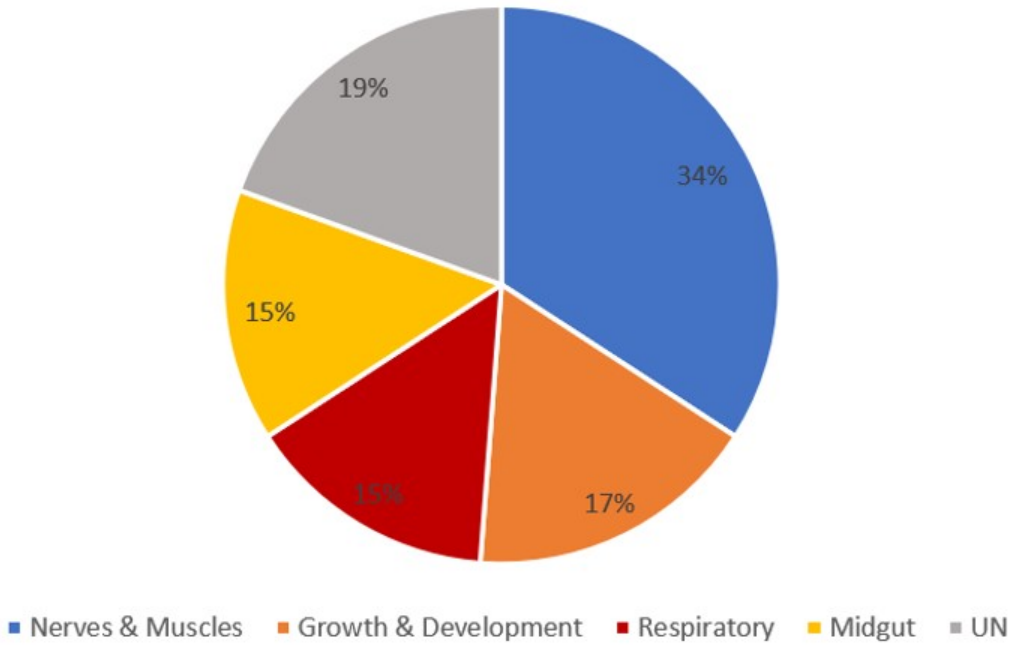


Figure 6.1: Classification of the available insecticides based on their physiological targets.

date chemicals that do not have an identified target, called undefined (UN). The other four regulate substantial physiological processes of insects including nerves and muscles, growth and development, respiration, and midgut. Together these four categories contain 32 members, each having many subgroups.

Looking at this classification, the importance of targeting insect GPCRs in developing pest control agents is more understood. Since as mentioned before, many of these receptors are involved in the regulation of different physiological functions in insects.

However, to the best of our knowledge to date, only two small molecules targeting invertebrate GPCRs are registered as insecticides by IRAC. Formamide class of insecticides, amitraz and chlordimeform, are the only class of insecticide that acts as agonist of invertebrate Octopamine receptor, with very promising results [148]. Octopamine receptor does not have any ortholog protein in vertebrates, hence no off-target effect is anticipated in vertebrates. However, recent studies have shown that amitraz can cross-activate α-adrenergic receptors and consequently can cause side effects in vertebrate species [149]. This further emphasizes the importance of the characterization of more neuropeptide receptors in invertebrates in order to be able to design more specific molecules.

In this study, obtaining a reliable 3D structure of TpitAstRC by applying *in silico* approaches, we exerted virtual screening methods to hunt for agonist molecules. Agonist molecules were desired since the the activation of the Allatostatin receptors results in the inhibition of JH secretion through a not very clear mechanism that in turn regulates the the physiological processes in insects. Thus, molecules capable of activating these receptors offer good alternatives for conventional insecticides.

High throughput screening approaches are routinely utilized in the studies aiming the discovery of small molecules used as GPCR drugs. Again, in case of arthropod GPCRs very few studies are conducted using high throughput screening methods [150].

The main reason for not using the peptide itself in pest control approaches is the sensitivity of these biomolecules to proteases that reduce their bioavailability. In addition, using peptides as pest control agents is economically very demanding. Peptidomimetics, on the other hand, offer a very suitable alternative in that they resemble

the peptide structures and in general have similar pharmacokinetics as native peptides and can diffuse into cells.

The screening studies were performed in β-turn peptidomimetics library OTAVA chemicals with the objective of finding small molecules capable of activating TpitAstRC. This library was selected due to the similarity they show with the 3D structure of the native ligand. Our model showed that AST-C folds into a turn-like structure, and accordingly we hypothesized that performing the screening studies using molecules that resemble the 3D structure of the native ligand increases the chance of finding more hits. β-turn peptidomimetics compounds are generated using two selection procedures, pharmacophore screening and similarity search. In the first approach, real β-turn structures are used to create pharmacophore models. While in the latter, known scaffolds for β-turn peptidomimetics including pentameric and hexameric cycle scaffolds, bicyclic scaffolds are utilized.

1058 small peptidomimetics molecules were screened. Docking and MD simulations were combined and by careful investigation of the free energy of binding, seven hits were chosen. These molecules were then ordered and tested in functional studies to evaluate their potency in activating G protein-dependent signaling pathway of TpitAstRC. Only one molecule was found to act as agonist. It activated the G protein-dependent signaling at micromolar ranges. Others were either antagonists or inverse agonists of the receptor. It is of note that the top-ranked molecule in our list behaved as antagonist. This result emphasizes the importance of evaluating the hits obtained from *in silico* studies in biological assays. As affinity of binding used by docking methods for sorting the molecules does not provide information regarding the efficacy of a given compound.

One out of seven molecules showed our desired effect. Considering the very careful methodology utilized in our study, we expected higher success rate. But in general, the success rate in drug discoveries based on virtual screening methods is not very high. In recent years, however, the number of hits found by this method tend to increase. This is

mainly due to the improvements in docking algorithms and the involvement of artificial intelligence in virtual screening workflows. In our work there are various underlying factors that contribute to the low success rate. First of all, the screening studies here were performed utilizing the 3D model of the receptor and not the structure obtained from X-ray crystallography or cryo-EM. Additionally, in this built model, only on representative structure of the receptor was used in the screening studies. This means that in docking studies only one fixed repertoire of the binding pocket with fixed side chain positioning of binding pocket-residing residues was used. Alternatively, multiple representative structures of TpitAstRC could have been selected and parallel screening workflows could have been run. This would provide different side chain sampling of residues at the binding pocket that in turn could increase the chance of finding more hits. Secondly, more than one library could have been screened. Increasing the number of the molecules in the screening procedure could have elevated the chance of finding agonist molecules. In future studies, the screening studies can be repeated applying the two above-mentioned approaches.

The effect of the found agonist on the physiology of pine processionary moth should be evaluated in bioassays.

6.5. Remarks

To sum up, the present research, undoubtedly, is one of the most extensive and comprehensive studies performed on an insect neuropeptide receptor, and we believe that considering the new mode-of-action proposed by these promising targets, structural and functional insights obtained here will positively contribute to the development of next-generation pesticides in the future.

REFERENCES

1. Gurr, G. M., S. D. Wratten and W. E. Snyder, "Biodiversity and insect pests", *Biodivers Insect Pests Key Issues Sustain Manag*, pp. 1–20, 2012.

2. Pedigo, L. P. and M. E. Rice, *Entomology and pest management*, Waveland Press, 2014.

3. Canakcioglu, H. and T. Mol, "Forest entomology, Istanbul University", , 1998.

4. Luckmann, W. H. and R. L. Metcalf, "The pest management concept", *Introduction to insect pest management. Wiley, New York*, pp. 1–31, 1994.

5. Sangüesa-Barreda, G., J. J. Camarero, A. García-Martín, R. Hernández and J. de la Riva, "Remote-sensing and tree-ring based characterization of forest defoliation and growth loss due to the Mediterranean pine processionary moth", *Forest Ecology and Management*, Vol. 320, pp. 171–181, 2014.

6. Hódar, J. A., J. Castro and R. Zamora, "Pine processionary caterpillar Thaumetopoea pityocampa as a new threat for relict Mediterranean Scots pine forests under climatic warming", *Biological conservation*, Vol. 110, No. 1, pp. 123–129, 2003.

7. Jacquet, J.-S., C. Orazio and H. Jactel, "Defoliation by processionary moth significantly reduces tree growth: a quantitative review", *Annals of forest science*, Vol. 69, No. 8, pp. 857–866, 2012.

8. Tsankov, G., "Egg parasitoids of the pine processionary moth, Thaumetopoea pityocampa (Den. & Schiff.)(Lep., Thaumetopeidae) in Bulgaria: Species, importance, biology and behaviour 1", *Journal of Applied Entomology*, Vol. 110, No. 1-5, pp. 7–13, 1990.

9. Tiberi, R., A. Niccoli, M. Curini, F. Epifano, M. Marcotullio and O. Rosati, "The role of the monoterpene composition inPinus spp. needles, in host selection by the pine processionary caterpillar, Thaumetopoea pityocampa", *Phytoparasitica*, Vol. 27, No. 4, pp. 263–272, 1999.

10. Démolin, G., "Bioecologia de la procesionaria del pino Thaumetopoea pityocampa Schiff. Incidencia de los factores climaticos", *Boletin del Servicio de Plagas Forestales*, Vol. 12, pp. 9–24, 1969.

11. Hódar, J. A., R. Zamora and J. Castro, "Host utilisation by moth and larval survival of pine processionary caterpillar Thaumetopoea pityocampa in relation to food quality in three Pinus species", *Ecological Entomology*, Vol. 27, No. 3, pp. 292–301, 2002.

12. Kanat, M. and M. H. Alma, "Insecticidal effects of essential oils from various plants against larvae of pine processionary moth (Thaumetopoea pityocampa Schiff)(Lepidoptera: Thaumetopoeidae)", *Pest Management Science: formerly Pesticide Science*, Vol. 60, No. 2, pp. 173–177, 2004.

13. Hedde, M., M. Aubert, F. Bureau, P. Margerie and T. Decaëns, "Soil detritivore macro-invertebrate assemblages throughout a managed beech rotation", *Annals of Forest Science*, Vol. 64, No. 2, pp. 219–228, 2007.

14. Jacquet, J.-S., A. Bosc, A. P. O'Grady and H. Jactel, "Pine growth response to processionary moth defoliation across a 40-year chronosequence", *Forest ecology and management*, Vol. 293, pp. 29–38, 2013.

15. Kulman, H., "Effects of insect defoliation on growth and mortality of trees", *Annual review of entomology*, Vol. 16, No. 1, pp. 289–324, 1971.

16. Roques, A., J. Rousselet, M. Avcı, D. N. Avtzis, A. Basso, A. Battisti, M. L. B. Jamaa, A. Bensidi, L. Berardi, W. Berretima *et al.*, "Climate warming and past

and present distribution of the processionary moths (Thaumetopoea spp.) in Europe, Asia Minor and North Africa", pp. 81–161, 2015.

17. Robinet, C., J. Rousselet and A. Roques, "Potential spread of the pine processionary moth in France: preliminary results from a simulation model and future challenges", *Annals of forest science*, Vol. 71, No. 2, pp. 149–160, 2014.

18. Tiberi, R., M. Bracalini, F. Croci, G. Tellini Florenzano and T. Panzavolta, "Effects of climate on pine processionary moth fecundity and on its egg parasitoids", *Ecology and evolution*, Vol. 5, No. 22, pp. 5372–5382, 2015.

19. Rodriguez-Mahillo, A. I., M. Gonzalez-Muñoz, J. M. Vega, J. A. López, A. Yart, C. Kerdelhué, E. Camafeita, J. C. Garcia Ortiz, H. Vogel, E. Petrucco Toffolo *et al.*, "Setae from the pine processionary moth (Thaumetopoea pityocampa) contain several relevant allergens", *Contact Dermatitis*, Vol. 67, No. 6, pp. 367–374, 2012.

20. Maier, H., W. Spiegel, T. Kinaciyan, H. Krehan, A. Cabaj, A. Schopf and H. Hönigsmann, "The oak processionary caterpillar as the cause of an epidemic airborne disease: survey and analysis", *British Journal of Dermatology*, Vol. 149, No. 5, pp. 990–997, 2003.

21. Werno, J., M. Lamy *et al.*, "Animal atmospheric pollution: the urticating hairs of the pine processionary caterpillar (Thaumetopoea pityocampa Schiff.)(Insecta, Lepidoptera).", *Comptes Rendus de l'Académie des Sciences. Series 3, Sciences de la Vie*, Vol. 310, No. 8, pp. 325–331, 1990.

22. Bosma, A. and H. Jans, "A severe anaphylactic shock caused by spraying the oak processionary caterpillar (Thaumetopoea processionea) in North Brabant", *Nederlands tijdschrift voor geneeskunde*, Vol. 142, No. 27, pp. 1567–1569, 1998.

23. Vega, J., I. Moneo, A. Armentia, J. Vega, R. De la Fuente and A. Fernandez,

"Pine processionary caterpillar as a new cause of immunologic contact urticaria", *Contact Dermatitis*, Vol. 43, No. 3, pp. 129–132, 2000.

24. Atmiş, E., S. Özden and W. Lise, "Urbanization pressures on the natural forests in Turkey: An overview", *Urban Forestry & Urban Greening*, Vol. 6, No. 2, pp. 83–92, 2007.

25. Ipekdal, K., C. Burban, C. Kerdelhue and S. S. Çağlar, "Distribution of two pine processionary moth species in Turkey evidences a contact zone", *Turkish Journal of Zoology*, Vol. 39, No. 5, pp. 868–876, 2015.

26. Nicolopoulou-Stamati, P., S. Maipas, C. Kotampasi, P. Stamatis and L. Hens, "Chemical pesticides and human health: the urgent need for a new concept in agriculture", *Frontiers in public health*, Vol. 4, p. 148, 2016.

27. Jactel, H., P. Menassieu, F. Vétillard, B. Barthélémy, D. Piou, B. Frérot, J. Rousselet, F. Goussard, M. Branco and A. Battisti, "Population monitoring of the pine processionary moth (Lepidoptera: Thaumetopoeidae) with pheromone-baited traps", *Forest Ecology and Management*, Vol. 235, No. 1-3, pp. 96–106, 2006.

28. Panzavolta, T., "Use of Bacillus thuringiensis var. kurstaki formulations for the control of harmful defoliators in the order Lepidoptera", *papers presented bacterial toxins for insect control, COST*, Vol. 862, 2005.

29. Gindin, G., A. Navon, N. Saphir, A. Protasov and Z. Mendel, "Differing susceptibility toBacillus thuringiensis formulations ofThaumetopoea wilkinsoni populations between forests with differentBt management in Israel", *Phytoparasitica*, Vol. 35, No. 2, p. 179, 2007.

30. Goettel, M. and A. Hajek, "Evaluation of non-target effects of pathogens used for management of arthropods", *Evaluating indirect ecological effects of biological control*, pp. 81–97, 2001.

31. Boulahbel, B., N. Aribi, S. Kilani-Morakchi and N. Soltani, "Insecticidal activity of azadirachtin on Drosophila melanogaster and recovery of normal status by exogenous 20-hydroxyecdysone", *African Entomology*, Vol. 23, No. 1, pp. 224–233, 2015.

32. Flatt, T., M.-P. Tu and M. Tatar, "Hormonal pleiotropy and the juvenile hormone regulation of Drosophila development and life history", *Bioessays*, Vol. 27, No. 10, pp. 999–1010, 2005.

33. Riddiford, L. M., "Cellular and molecular actions of juvenile hormone I. General considerations and premetamorphic actions", *Advances in insect physiology*, Vol. 24, pp. 213–274, Elsevier, 1994.

34. Gade, G., K.-H. Hoffmann and J. H. Spring, "Hormonal regulation in insects: facts, gaps, and future directions", *Physiological reviews*, Vol. 77, No. 4, pp. 963–1032, 1997.

35. Tobe, S. S. and B. Stay, "Structure and regulation of the corpus allatum", *Advances in Insect Physiology*, Vol. 18, pp. 305–432, Elsevier, 1985.

36. King, L. and S. Tobe, "The structure of a juvenile hormone-binding lipophorin from the hemolymph of Diploptera punctata", *Insect biochemistry and molecular biology*, Vol. 22, No. 8, pp. 817–827, 1992.

37. Gilbert, L. I., "Insect hormones", *Endocrinology*, pp. 127–136, Springer, 2005.

38. Richard, D. S., S. W. Applebaum, T. J. Sliter, F. C. Baker, D. A. Schooley, C. C. Reuter, V. C. Henrich and L. I. Gilbert, "Juvenile hormone bisepoxide biosynthesis in vitro by the ring gland of Drosophila melanogaster: a putative juvenile hormone in the higher Diptera", *Proceedings of the National Academy of Sciences*, Vol. 86, No. 4, pp. 1421–1425, 1989.

39. Nijhout, H., "Insect Hormones Princeton Univ", *Press, Princeton, NJ*, 1994.

40. Zera, A. J. and L. G. Harshman, "The physiology of life history trade-offs in animals", *Annual review of Ecology and Systematics*, Vol. 32, No. 1, pp. 95–126, 2001.

41. Zera, A. J., "The endocrine regulation of wing polymorphism in insects: state of the art, recent surprises, and future directions", *Integrative and comparative biology*, Vol. 43, No. 5, pp. 607–616, 2003.

42. Dingle, H. and R. Winchell, "Juvenile hormone as a mediator of plasticity in insect life histories", *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, Vol. 35, No. 4, pp. 359–373, 1997.

43. Dhadialla, T. S., G. R. Carlson and D. P. Le, "New insecticides with ecdysteroidal and juvenile hormone activity", *Annual review of entomology*, Vol. 43, No. 1, pp. 545–569, 1998.

44. Riddiford, L. M. and M. Ashburner, "Effects of juvenile hormone mimics on larval development and metamorphosis of Drosophila melanogaster", *General and comparative endocrinology*, Vol. 82, No. 2, pp. 172–183, 1991.

45. Restifo, L. L. and T. G. Wilson, "A juvenile hormone agonist reveals distinct developmental pathways mediated by Ecdysone-inducible broad complex transcription factors", *Developmental genetics*, Vol. 22, No. 2, pp. 141–159, 1998.

46. Riddiford, L. M., "How does juvenile hormone control insect metamorphosis and reproduction?", *General and comparative endocrinology*, Vol. 179, No. 3, pp. 477–484, 2012.

47. Audsley, N. and R. Weaver, "Neuropeptides associated with the regulation of feeding in insects", *General and comparative endocrinology*, Vol. 162, No. 1, pp. 93–104, 2009.

48. Weaver, R. and N. Audsley, "Neuropeptide regulators of juvenile hormone synthesis: structures, functions, distribution, and unanswered questions", *Annals of the New York Academy of Sciences*, Vol. 1163, No. 1, pp. 316–329, 2009.

49. Stay, B. and S. S. Tobe, "The role of allatostatins in juvenile hormone synthesis in insects and crustaceans", *Annu. Rev. Entomol.*, Vol. 52, pp. 277–299, 2007.

50. Wang, M., L. Zhang, X. Wang, Y. Ling, X. Wu, X. Li, Y. Mi and X. Yang, "Exploring the N-terminus region: Synthesis, bioactivity and 3D-QSAR of allatostatin analogs as novel insect growth regulators", *Chinese Chemical Letters*, Vol. 29, No. 9, pp. 1375–1378, 2018.

51. Kaplan, R. S. and J. A. Mayor, "Structure, function and regulation of the tricarboxylate transport protein from rat liver mitochondria", *Journal of bioenergetics and biomembranes*, Vol. 25, No. 5, pp. 503–514, 1993.

52. Nouzova, M., C. Rivera-Perez and F. G. Noriega, "Allatostatin-C reversibly blocks the transport of citrate out of the mitochondria and inhibits juvenile hormone synthesis in mosquitoes", *Insect biochemistry and molecular biology*, Vol. 57, pp. 20–26, 2015.

53. Verlinden, H., M. Gijbels, E. Lismont, C. Lenaerts, J. V. Broeck and E. Marchal, "The pleiotropic allatoregulatory neuropeptides and their receptors: A mini-review", *Journal of Insect Physiology*, Vol. 80, pp. 2–14, 2015.

54. Birgül, N., C. Weise, H.-J. Kreienkamp and D. Richter, "Reverse physiology in Drosophila: identification of a novel allatostatin-like neuropeptide and its cognate receptor structurally related to the mammalian somatostatin/galanin/opioid receptor family", *The EMBO Journal*, Vol. 18, No. 21, pp. 5892–5900, 1999.

55. Kreienkamp, H.-J., H. J. Larusson, I. Witte, T. Roeder, N. Birgül, H.-H. Hönck, S. Harder, G. Ellinghausen, F. Buck and D. Richter, "Functional annotation

of two orphan G-protein-coupled receptors, Drostar1 and-2, from Drosophila melanogaster and their ligands by reverse pharmacology", *Journal of Biological Chemistry*, Vol. 277, No. 42, pp. 39937–39943, 2002.

56. Santos, R., O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea *et al.*, "A comprehensive map of molecular drug targets", *Nature reviews Drug discovery*, Vol. 16, No. 1, pp. 19–34, 2017.

57. Kolakowski Jr, L. F., "GCRDb: a G-protein-coupled receptor database.", *Receptors & channels*, Vol. 2, No. 1, p. 1, 1994.

58. 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.

59. Nordström, K. J., M. Sällman Almén, M. M. Edstam, R. Fredriksson and H. B. Schiöth, "Independent HHsearch, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families", *Molecular biology and evolution*, Vol. 28, No. 9, pp. 2471–2480, 2011.

60. Southan, C., J. L. Sharman, H. E. Benson, E. Faccenda, A. J. Pawson, S. P. Alexander, O. P. Buneman, A. P. Davenport, J. C. McGrath, J. A. Peters *et al.*, "The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands", *Nucleic acids research*, Vol. 44, No. D1, pp. D1054–D1068, 2016.

61. Zhou, Q., D. Yang, M. Wu, Y. Guo, W. Guo, L. Zhong, X. Cai, A. Dai, W. Jang, E. I. Shakhnovich *et al.*, "Common activation mechanism of class A GPCRs", *Elife*, Vol. 8, p. e50279, 2019.

62. Vass, M., A. J. Kooistra, D. Yang, R. C. Stevens, M.-W. Wang and C. de Graaf, "Chemical diversity in the G protein-coupled receptor superfamily", *Trends in Pharmacological Sciences*, Vol. 39, No. 5, pp. 494–512, 2018.

63. Fleetwood, O., P. Matricon, J. Carlsson and L. Delemotte, "Energy Landscapes Reveal Agonist Control of G Protein-Coupled Receptor Activation via Microswitches", *Biochemistry*, Vol. 59, No. 7, pp. 880–891, 2020.

64. Bylund, D. B., "International Union of Pharmacology nomenclature of adrenoceptors", *Pharmacol. Rev.*, Vol. 46, pp. 121–136, 1994.

65. Filipek, S., "Molecular switches in GPCRs", *Current opinion in structural biology*, Vol. 55, pp. 114–120, 2019.

66. Kato, H. E., Y. Zhang, H. Hu, C.-M. Suomivuori, F. M. N. Kadji, J. Aoki, K. K. Kumar, R. Fonseca, D. Hilger, W. Huang *et al.*, "Conformational transitions of a neurotensin receptor 1–G i1 complex", *Nature*, Vol. 572, No. 7767, pp. 80–85, 2019.

67. Eddy, M. T., M.-Y. Lee, Z.-G. Gao, K. L. White, T. Didenko, R. Horst, M. Audet, P. Stanczak, K. M. McClary, G. W. Han *et al.*, "Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor", *Cell*, Vol. 172, No. 1-2, pp. 68–80, 2018.

68. Rasmussen, S. G., B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski *et al.*, "Crystal structure of the β 2 adrenergic receptor–Gs protein complex", *Nature*, Vol. 477, No. 7366, pp. 549–555, 2011.

69. Alhadeff, R., I. Vorobyov, H. W. Yoon and A. Warshel, "Exploring the free-energy landscape of GPCR activation", *Proceedings of the National Academy of Sciences*, Vol. 115, No. 41, pp. 10327–10332, 2018.

70. Latorraca, N. R., J. K. Wang, B. Bauer, R. J. Townshend, S. A. Hollingsworth, J. E. Olivieri, H. E. Xu, M. E. Sommer and R. O. Dror, "Molecular mechanism of GPCR-mediated arrestin activation", *Nature*, Vol. 557, No. 7705, pp. 452–456, 2018.

71. Kristiansen, K., "Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function", *Pharmacology & therapeutics*, Vol. 103, No. 1, pp. 21–80, 2004.

72. Khan, S. M., R. Sleno, S. Gora, P. Zylbergold, J.-P. Laverdure, J.-C. Labbé, G. J. Miller and T. E. Hébert, "The expanding roles of G$\beta\gamma$ subunits in G protein–coupled receptor signaling and drug action", *Pharmacological reviews*, Vol. 65, No. 2, pp. 545–577, 2013.

73. Kohout, T. A. and R. J. Lefkowitz, "Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization", *Molecular pharmacology*, Vol. 63, No. 1, pp. 9–18, 2003.

74. Hilger, D., M. Masureel and B. K. Kobilka, "Structure and dynamics of GPCR signaling complexes", *Nature structural & molecular biology*, Vol. 25, No. 1, p. 4, 2018.

75. Shenoy, S. K. and R. J. Lefkowitz, "Multifaceted roles of β-arrestins in the regulation of seven-membrane-spanning receptor trafficking and signalling", *Biochemical Journal*, Vol. 375, No. 3, pp. 503–515, 2003.

76. Rosenbaum, D. M., V. Cherezov, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis, R. C. Stevens *et al.*, "GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function", *science*, Vol. 318, No. 5854, pp. 1266–1273, 2007.

77. Jaakola, V.-P., M. T. Griffith, M. A. Hanson, V. Cherezov, E. Y. Chien, J. R. Lane, A. P. Ijzerman and R. C. Stevens, "The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist", *Science*, Vol. 322, No. 5905, pp. 1211–1217, 2008.

78. Hanson, M. A., C. B. Roth, E. Jo, M. T. Griffith, F. L. Scott, G. Reinhart, H. Desale, B. Clemons, S. M. Cahalan, S. C. Schuerer *et al.*, "Crystal structure of a lipid G protein–coupled receptor", *Science*, Vol. 335, No. 6070, pp. 851–855, 2012.

79. Manglik, A., A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka and S. Granier, "Crystal structure of the μ-opioid receptor bound to a morphinan antagonist", *Nature*, Vol. 485, No. 7398, pp. 321–326, 2012.

80. Granier, S. and B. Kobilka, "A new era of GPCR structural and chemical biology", *Nature chemical biology*, Vol. 8, No. 8, pp. 670–673, 2012.

81. Ballesteros, J. A. and H. Weinstein, "[19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors", *Methods in neurosciences*, Vol. 25, pp. 366–428, Elsevier, 1995.

82. Tikhonova, I. G., V. Gigoux and D. Fourmy, "Understanding peptide binding in Class AG protein-coupled receptors", *Molecular pharmacology*, Vol. 96, No. 5, pp. 550–561, 2019.

83. Jaakola, V.-P., J. R. Lane, J. Y. Lin, V. Katritch, A. P. IJzerman and R. C. Stevens, "Ligand Binding and Subtype Selectivity of the Human A2A Adenosine Receptor Identification and Characterization of Essential Amino Acid Residues", *Journal of Biological Chemistry*, Vol. 285, No. 17, pp. 13032–13044, 2010.

84. Weis, W. I. and B. K. Kobilka, "The molecular basis of G protein–coupled receptor activation", *Annual review of biochemistry*, Vol. 87, pp. 897–919, 2018.

85. Pándy-Szekeres, G., C. Munk, T. M. Tsonkov, S. Mordalski, K. Harpsøe, A. S. Hauser, A. J. Bojarski and D. E. Gloriam, "GPCRdb in 2018: adding GPCR structure models and ligands", *Nucleic acids research*, Vol. 46, No. D1, pp. D440–D446, 2018.

86. van Unen, J., A. D. Stumpf, B. Schmid, N. R. Reinhard, P. L. Hordijk, C. Hoffmann, T. W. Gadella Jr and J. Goedhart, "A new generation of FRET sensors for robust measurement of Gαi1, Gαi2 and Gαi3 activation kinetics in single cells", *PLoS One*, Vol. 11, No. 1, p. e0146789, 2016.

87. Adjobo-Hermans, M. J., J. Goedhart, L. van Weeren, S. Nijmeijer, E. M. Manders, S. Offermanns and T. W. Gadella, "Real-time visualization of heterotrimeric G protein Gq activation in living cells", *BMC biology*, Vol. 9, No. 1, p. 32, 2011.

88. Hein, P., F. Rochais, C. Hoffmann, S. Dorsch, V. O. Nikolaev, S. Engelhardt, C. H. Berlot, M. J. Lohse and M. Bünemann, "Gs activation is time-limiting in initiating receptor-mediated signaling", *Journal of Biological Chemistry*, Vol. 281, No. 44, pp. 33345–33351, 2006.

89. Mastop, M., N. R. Reinhard, C. R. Zuconelli, F. Terwey, T. W. Gadella Jr, J. van Unen, M. J. Adjobo-Hermans and J. Goedhart, "A FRET-based biosensor for measuring Gα13 activation in single cells", *PloS one*, Vol. 13, No. 3, p. e0193705, 2018.

90. Janzen, W. P., *High throughput screening: methods and protocols*, 190, Springer Science & Business Media, 2001.

91. Chen, H., M. Rangasamy, S. Y. Tan, H. Wang and B. D. Siegfried, "Evaluation of five methods for total DNA extraction from western corn rootworm beetles",

PLoS one, Vol. 5, No. 8, p. e11963, 2010.

92. Andrews, S. *et al.*, "FastQC: a quality control tool for high throughput sequence data", , 2010.

93. Martin, M., "Cutadapt removes adapter sequences from high-throughput sequencing reads", *EMBnet. journal*, Vol. 17, No. 1, pp. 10–12, 2011.

94. Simpson, J. T. and R. Durbin, "Efficient de novo assembly of large genomes using compressed data structures", *Genome research*, Vol. 22, No. 3, pp. 549–556, 2012.

95. Zimin, A. V., G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg and J. A. Yorke, "The MaSuRCA genome assembler", *Bioinformatics*, Vol. 29, No. 21, pp. 2669–2677, 2013.

96. Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0", *Systematic biology*, Vol. 59, No. 3, pp. 307–321, 2010.

97. Pei, J., "Multiple protein sequence alignment", *Current opinion in structural biology*, Vol. 18, No. 3, pp. 382–386, 2008.

98. Stanke, M. and B. Morgenstern, "AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints", *Nucleic acids research*, Vol. 33, No. suppl_2, pp. W465–W467, 2005.

99. McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez, "Analysis tool web services from the EMBL-EBI", *Nucleic acids research*, Vol. 41, No. W1, pp. W597–W600, 2013.

100. El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart *et al.*, "The Pfam protein

families database in 2019", *Nucleic acids research*, Vol. 47, No. D1, pp. D427–D432, 2019.

101. Dobson, L., I. Reményi and G. E. Tusnády, "CCTOP: a Consensus Constrained TOPology prediction web server", *Nucleic acids research*, Vol. 43, No. W1, pp. W408–W412, 2015.

102. Nielsen, H., "Predicting secretory proteins with SignalP", *Protein function prediction*, pp. 59–73, Springer, 2017.

103. Veenstra, J. A., "Mono-and dibasic proteolytic cleavage sites in insect neuroendocrine peptide precursors", *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, Vol. 43, No. 2, pp. 49–63, 2000.

104. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli *et al.*, "SWISS-MODEL: homology modelling of protein structures and complexes", *Nucleic acids research*, Vol. 46, No. W1, pp. W296–W303, 2018.

105. Benkert, P., S. C. Tosatto and D. Schomburg, "QMEAN: A comprehensive scoring function for model quality assessment", *Proteins: Structure, Function, and Bioinformatics*, Vol. 71, No. 1, pp. 261–277, 2008.

106. Yang, J. and Y. Zhang, "I-TASSER server: new development for protein structure and function predictions", *Nucleic acids research*, Vol. 43, No. W1, pp. W174–W181, 2015.

107. Sastry, G. M., M. Adzhigirey, T. Day, R. Annabhimoju and 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.

108. Bas, D. C., D. M. Rogers 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.

109. Damm, W., A. Frontera, J. Tirado-Rives and W. L. Jorgensen, "OPLS all-atom force field for carbohydrates", *Journal of Computational Chemistry*, Vol. 18, No. 16, pp. 1955–1970, 1997.

110. Lomize, M. A., I. D. Pogozheva, H. Joo, H. I. Mosberg and A. L. Lomize, "OPM database and PPM web server: resources for positioning of proteins in membranes", *Nucleic acids research*, Vol. 40, No. D1, pp. D370–D376, 2012.

111. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of chemical physics*, Vol. 79, No. 2, pp. 926–935, 1983.

112. 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 *et al.*, "Scalable algorithms for molecular dynamics simulations on commodity clusters", *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, pp. 43–43, IEEE, 2006.

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

114. 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.

115. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, "A smooth particle mesh Ewald method", *The Journal of chemical physics*, Vol. 103, No. 19, pp. 8577–8593, 1995.

116. Genheden, S. and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities", *Expert opinion on drug discovery*, Vol. 10, No. 5, pp. 449–461, 2015.

117. Comeau, S. R., D. W. Gatchell, S. Vajda and C. J. Camacho, "ClusPro: a fully automated algorithm for protein–protein docking", *Nucleic acids research*, Vol. 32, No. suppl_2, pp. W96–W99, 2004.

118. Isberg, V., C. de Graaf, A. Bortolato, V. Cherezov, V. Katritch, F. H. Marshall, S. Mordalski, J.-P. Pin, R. C. Stevens, G. Vriend *et al.*, "Generic GPCR residue numbers–aligning topology maps while minding the gaps", *Trends in pharmacological sciences*, Vol. 36, No. 1, pp. 22–31, 2015.

119. Grant, B. J., A. P. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. Caves, "Bio3d: an R package for the comparative analysis of protein structures", *Bioinformatics*, Vol. 22, No. 21, pp. 2695–2696, 2006.

120. Salmas, R. E., M. Yurtsever and S. Durdagi, "Investigation of inhibition mechanism of chemokine receptor CCR5 by micro-second molecular dynamics simulations", *Scientific reports*, Vol. 5, No. 1, pp. 1–12, 2015.

121. 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, No. 21, pp. 6177–6196, 2006.

122. Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and 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, No. 7, pp. 1750–1759, 2004.

123. Humphrey, W., A. Dalke, K. Schulten *et al.*, "VMD: visual molecular dynamics",

Journal of molecular graphics, Vol. 14, No. 1, pp. 33–38, 1996.

124. Clynen, E., J. Huybrechts, P. Verleyen, A. De Loof and L. Schoofs, "Annotation of novel neuropeptide precursors in the migratory locust based on transcript screening of a public EST database and mass spectrometry", *BMC genomics*, Vol. 7, No. 1, pp. 1–15, 2006.

125. Kramer, S. J., A. ToSCHI, C. A. Miller, H. Kataoka, G. B. Quistad, J. P. Li, R. L. Carney and D. A. Schooley, "Identification of an allatostatin from the tobacco hornworm Manduca sexta.", *Proceedings of the national academy of sciences*, Vol. 88, No. 21, pp. 9458–9462, 1991.

126. Xu, G., G.-X. Gu, Z.-W. Teng, S.-F. Wu, J. Huang, Q.-S. Song, G.-Y. Ye and Q. Fang, "Identification and expression profiles of neuropeptides and their G protein-coupled receptors in the rice stem borer Chilo suppressalis", *Scientific reports*, Vol. 6, p. 28976, 2016.

127. Secher, T., C. Lenz, G. Cazzamali, G. Sørensen, M. Williamson, G. N. Hansen, P. Svane and C. J. Grimmelikhuijzen, "Molecular cloning of a functional allatostatin gut/brain receptor and an allatostatin preprohormone from the silkworm Bombyx mori", *Journal of Biological Chemistry*, Vol. 276, No. 50, pp. 47052–47060, 2001.

128. Salahpour, A., S. Espinoza, B. Masri, V. Lam, L. Barak and R. R. Gainetdinov, "BRET biosensors to study GPCR biology, pharmacology, and signal transduction", *Frontiers in endocrinology*, Vol. 3, p. 105, 2012.

129. Hall, M. P., J. Unch, B. F. Binkowski, M. P. Valley, B. L. Butler, M. G. Wood, P. Otto, K. Zimmerman, G. Vidugiris, T. Machleidt *et al.*, "Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate", *ACS chemical biology*, Vol. 7, No. 11, pp. 1848–1857, 2012.

130. Koehl, A., H. Hu, S. Maeda, Y. Zhang, Q. Qu, J. M. Paggi, N. R. Latorraca, D. Hilger, R. Dawson, H. Matile *et al.*, "Structure of the μ-opioid receptor–G i protein complex", *Nature*, Vol. 558, No. 7711, pp. 547–552, 2018.

131. Brody, T. and A. Cravchik, "Drosophila melanogaster G protein–coupled receptors", *The Journal of cell biology*, Vol. 150, No. 2, pp. F83–F88, 2000.

132. Hill, C. A., A. N. Fox, R. J. Pitts, L. B. Kent, P. L. Tan, M. A. Chrystal, A. Cravchik, F. H. Collins, H. M. Robertson and L. J. Zwiebel, "G protein-coupled receptors in Anopheles gambiae", *Science*, Vol. 298, No. 5591, pp. 176–178, 2002.

133. Gschloessl, B., H. Vogel, C. Burban, D. Heckel, R. Streiff and C. Kerdelhué, "Comparative analysis of two phenologically divergent populations of the pine processionary moth (Thaumetopoea pityocampa) by de novo transcriptome sequencing", *Insect biochemistry and molecular biology*, Vol. 46, pp. 31–42, 2014.

134. Gether, U., "Uncovering molecular mechanisms involved in activation of G protein-coupled receptors", *Endocrine reviews*, Vol. 21, No. 1, pp. 90–113, 2000.

135. Hanlon, C. D. and D. J. Andrew, "Outside-in signaling–a brief review of GPCR signaling with a focus on the Drosophila GPCR family", *Journal of cell science*, Vol. 128, No. 19, pp. 3533–3542, 2015.

136. Offermanns, S., "G-proteins as transducers in transmembrane signalling", *Progress in biophysics and molecular biology*, Vol. 83, No. 2, pp. 101–130, 2003.

137. Law, S. F., K. Yasuda, G. I. Bell and T. Reisine, "Gi alpha 3 and G (o) alpha selectively associate with the cloned somatostatin receptor subtype SSTR2.", *Journal of Biological Chemistry*, Vol. 268, No. 15, pp. 10721–10727, 1993.

138. Vilardaga, J.-P., M. Bünemann, C. Krasel, M. Castro and M. J. Lohse, "Measurement of the millisecond activation switch of G protein–coupled receptors in living cells", *Nature biotechnology*, Vol. 21, No. 7, pp. 807–812, 2003.

139. Hoffmann, C., G. Gaietta, M. Bünemann, S. R. Adams, S. Oberdorff-Maass, B. Behr, J.-P. Vilardaga, R. Y. Tsien, M. H. Ellisman and M. J. Lohse, "A FlAsH-based FRET approach to determine G protein–coupled receptor activation in living cells", *Nature methods*, Vol. 2, No. 3, pp. 171–176, 2005.

140. Rochais, F., J. Vilardaga and V. Nikolaev, "Bü nemann, M., Lohse, MJ, and Engelhardt, S.(2007). Real-time optical recording of beta1-adrenergic receptor activation reveals supersensitivity of the Arg389 variant to carvedilol", *J. Clin. Invest*, Vol. 117, pp. 229–235.

141. Reiner, S., M. Ambrosio, C. Hoffmann and M. J. Lohse, "Differential signaling of the endogenous agonists at the β2-adrenergic receptor", *Journal of Biological Chemistry*, Vol. 285, No. 46, pp. 36188–36198, 2010.

142. Trzaskowski, B., D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski and S. Filipek, "Action of molecular switches in GPCRs-theoretical and experimental studies", *Current medicinal chemistry*, Vol. 19, No. 8, pp. 1090–1109, 2012.

143. Peeters, M., G. Van Westen, Q. Li and A. IJzerman, "Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation", *Trends in pharmacological sciences*, Vol. 32, No. 1, pp. 35–42, 2011.

144. Holleran, B. J., I. Domazet, M.-E. Beaulieu, L. P. Yan, G. Guillemette, P. Lavigne, E. Escher and R. Leduc, "Identification of transmembrane domain 6 & 7 residues that contribute to the binding pocket of the urotensin II receptor", *Biochemical pharmacology*, Vol. 77, No. 8, pp. 1374–1382, 2009.

145. Tikhonova, I. G. and E. Poerio, "Free fatty acid receptors: structural models and elucidation of ligand binding interactions", *BMC structural biology*, Vol. 15, No. 1, p. 16, 2015.

146. Cheng, J., X. Sun, W. Li, G. Liu, Y. Tu and Y. Tang, "Molecular switches of the

κ opioid receptor triggered by 6′-GNTI and 5′-GNTI", *Scientific reports*, Vol. 6, p. 18913, 2016.

147. Sparks, T. C. and R. Nauen, "IRAC: Mode of action classification and insecticide resistance management", *Pesticide biochemistry and physiology*, Vol. 121, pp. 122–128, 2015.

148. Meyer, J. M., K. F. Ejendal, L. V. Avramova, E. E. Garland-Kuntz, G. I. Giraldo-Calderón, T. F. Brust, V. J. Watts and C. A. Hill, "A "genome-to-lead" approach for insecticide discovery: pharmacological characterization and screening of Aedes aegypti D 1-like dopamine receptors", *PLoS Negl Trop Dis*, Vol. 6, No. 1, p. e1478, 2012.

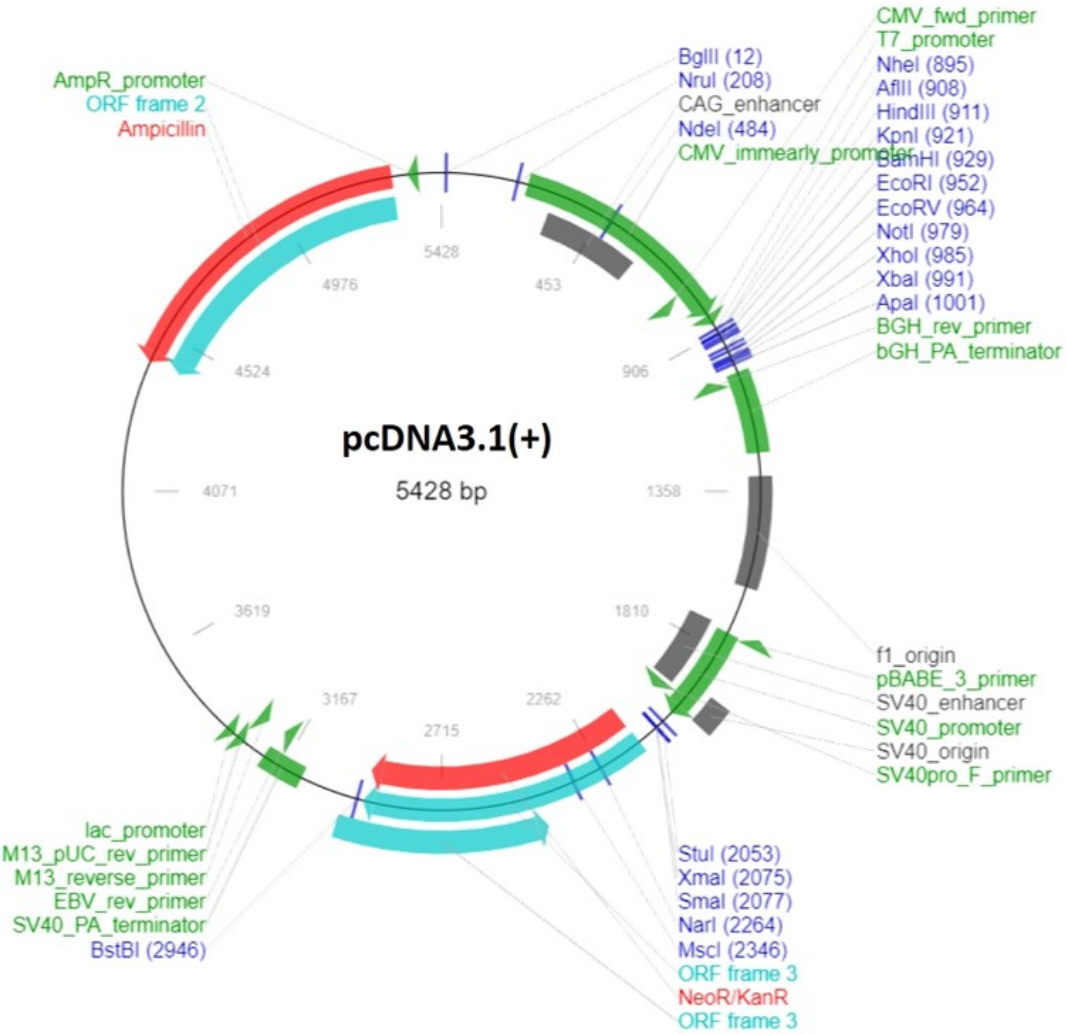
149. Yilmaz, H. L. and D. Yildizdas, "Amitraz poisoning, an emerging problem: epidemiology, clinical features, management, and preventive strategies", *Archives of disease in childhood*, Vol. 88, No. 2, pp. 130–134, 2003.

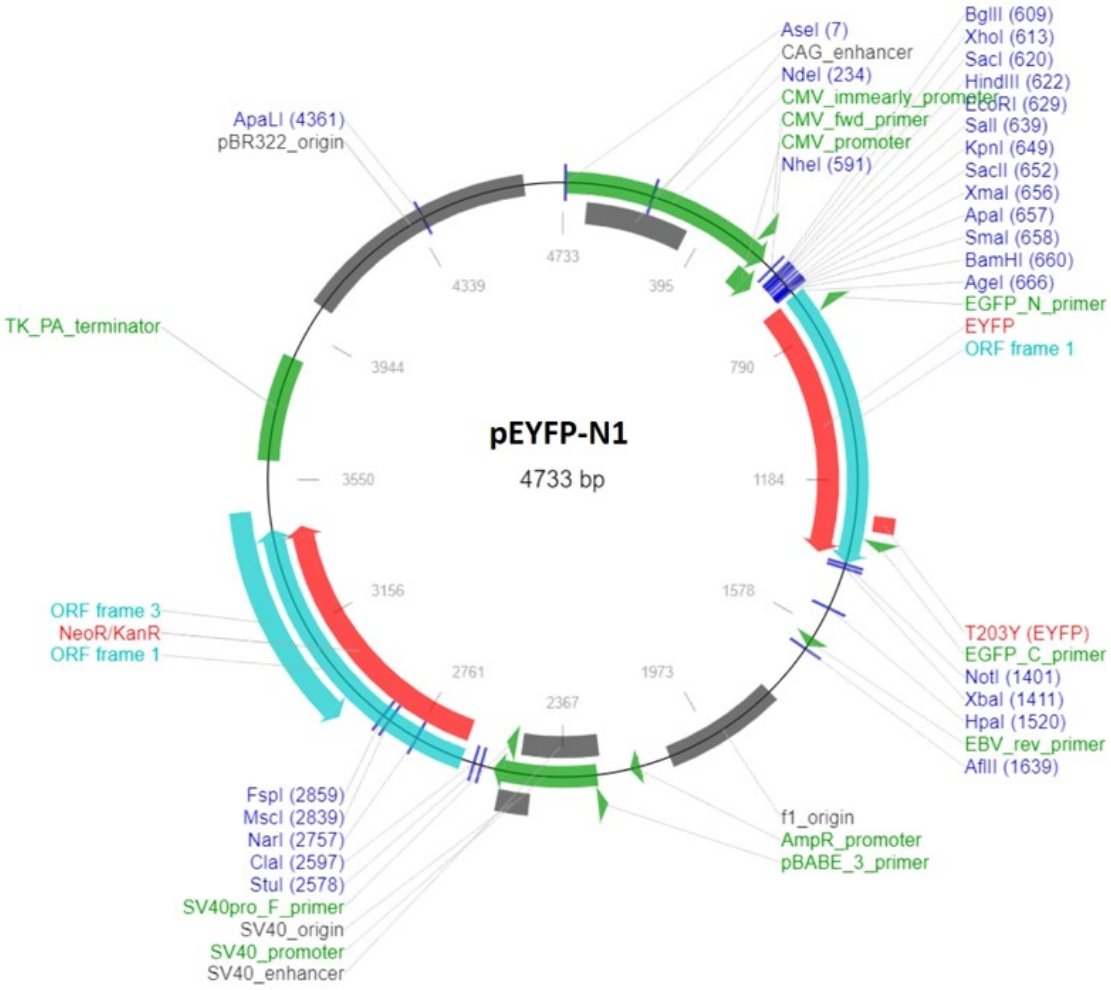
150. Ejendal, K. F., C. W. Dessauer, T. E. Hébert and V. J. Watts, "Dopamine D2 receptor-mediated heterologous sensitization of AC5 requires signalosome assembly", *Journal of signal transduction*, Vol. 2012, 2012.

APPENDIX A: AMINO ACID ABBREVIATIONS

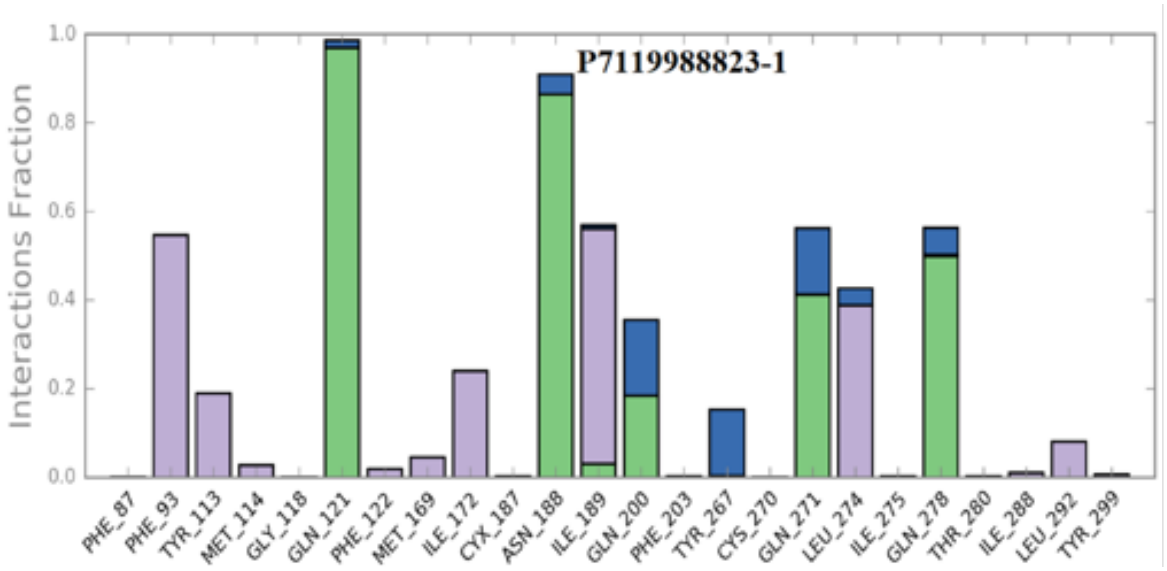
Amino acid	**One-Letter Code**	**Three-Letter Code**
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thre
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

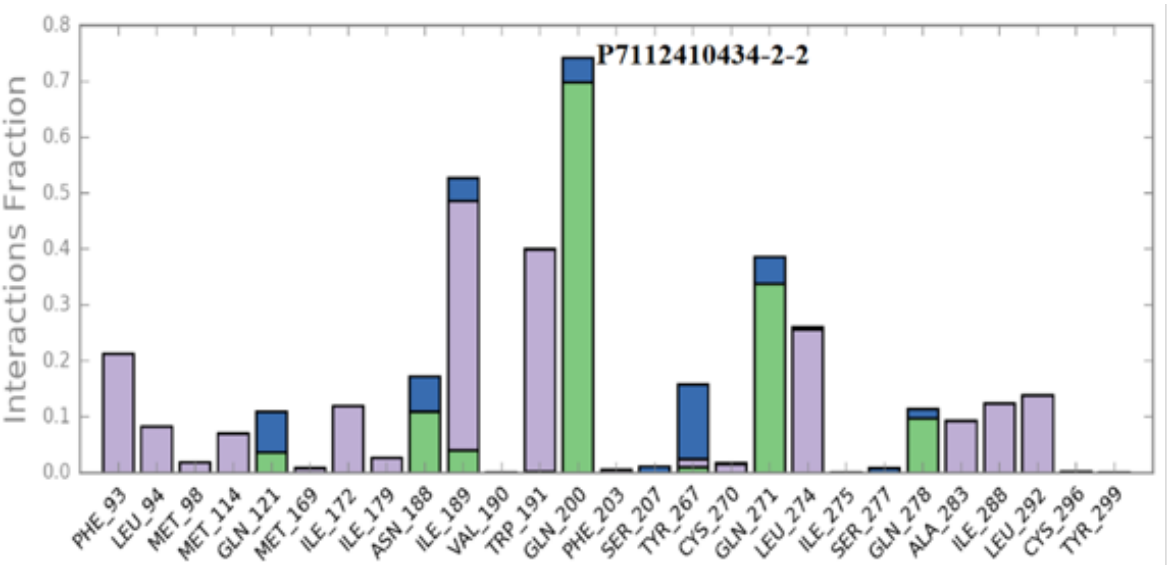
APPENDIX B: PLASMID MAPS

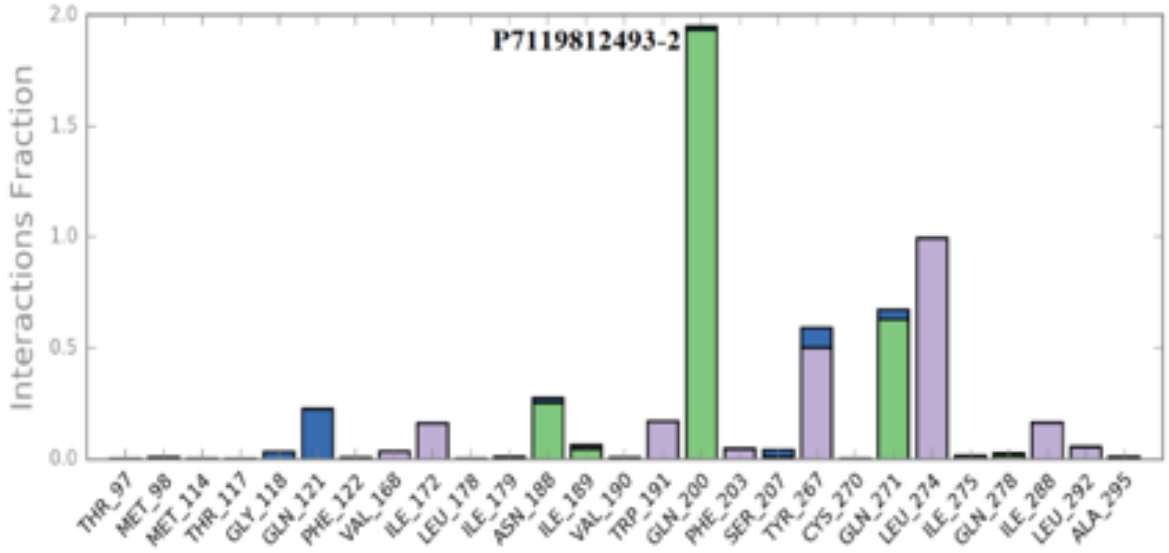


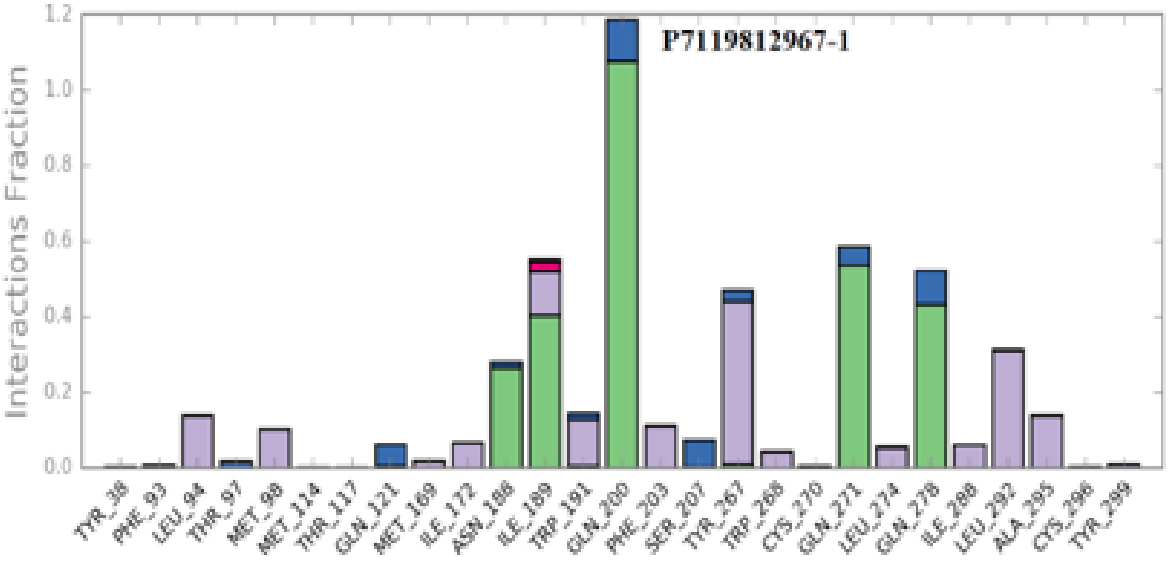

pEYFP-N1
4733 bp
4733
395
790
1184
1578
1973
2367
2761
3156
3550
3944
4339
AseI (7)
CAG_enhancer
NdeI (234)
CMV_immearly_promoter
CMV_fwd_primer
CMV_promoter
NheI (591)
BglII (609)
XhoI (613)
SacI (620)
HindIII (622)
EcoRI (629)
SalI (639)
KpnI (649)
SacII (652)
XmaI (656)
ApaI (657)
SmaI (658)
BamHI (660)
AgeI (666)
EGFP_N_primer
EYFP
ORF frame 1
T203Y (EYFP)
EGFP_C_primer
NotI (1401)
XbaI (1411)
HpaI (1520)
EBV_rev_primer
AflII (1639)
f1_origin
AmpR_promoter
pBABE_3_primer
FspI (2859)
MscI (2839)
NarI (2757)
ClaI (2597)
StuI (2578)
SV40pro_F_primer
SV40_origin
SV40_promoter
SV40_enhancer
ORF frame 3
NeoR/KanR
ORF frame 1
TK_PA_terminator
ApaLI (4361)
pBR322_origin

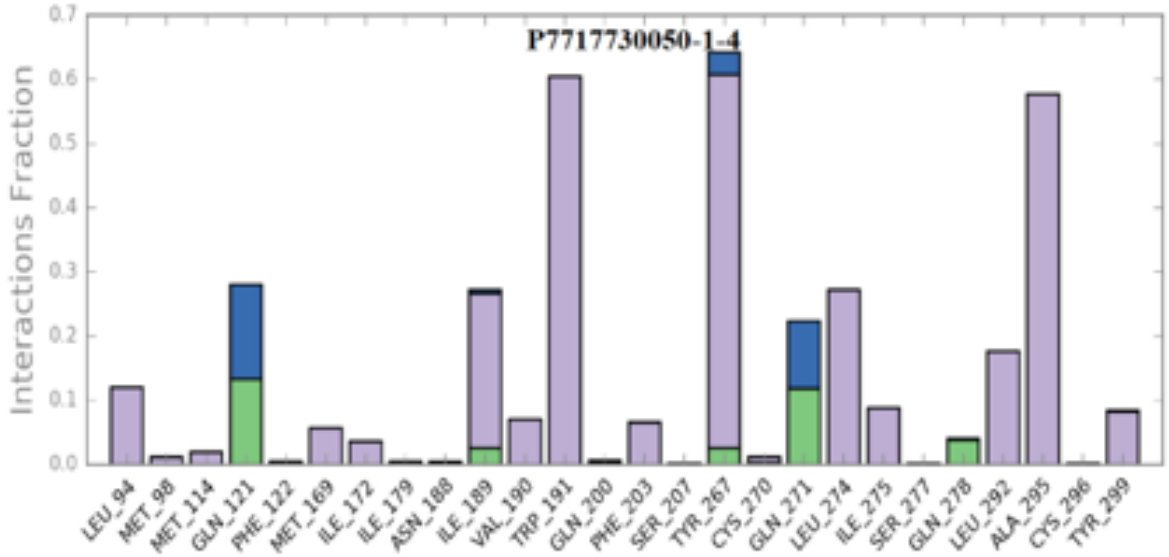
APPENDIX C: RECEPTOR-LIGAND INTERACTION FRACTION DIAGRAM

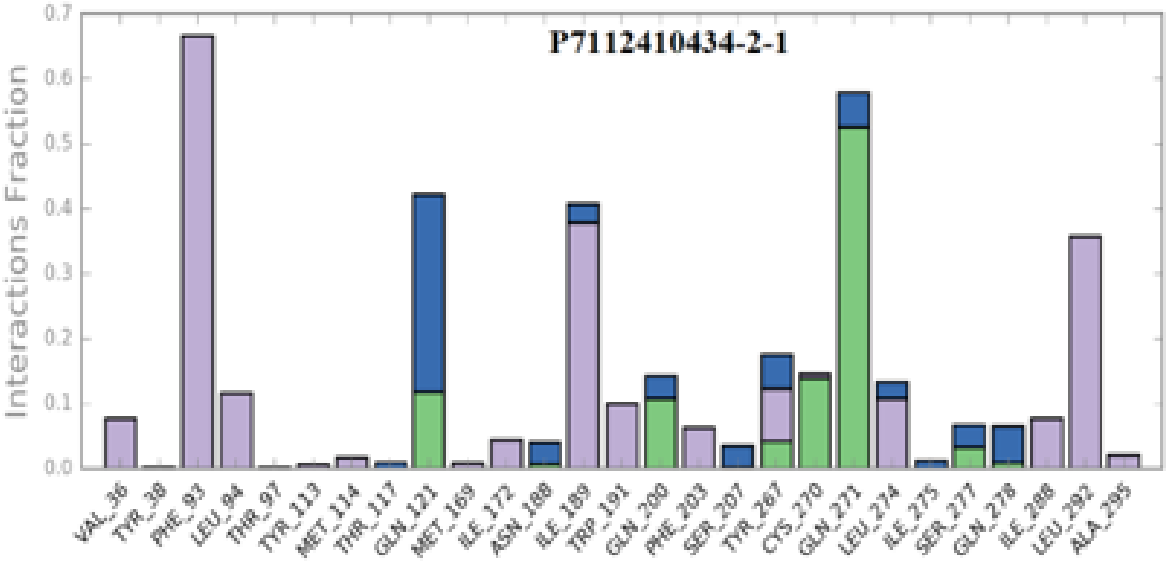


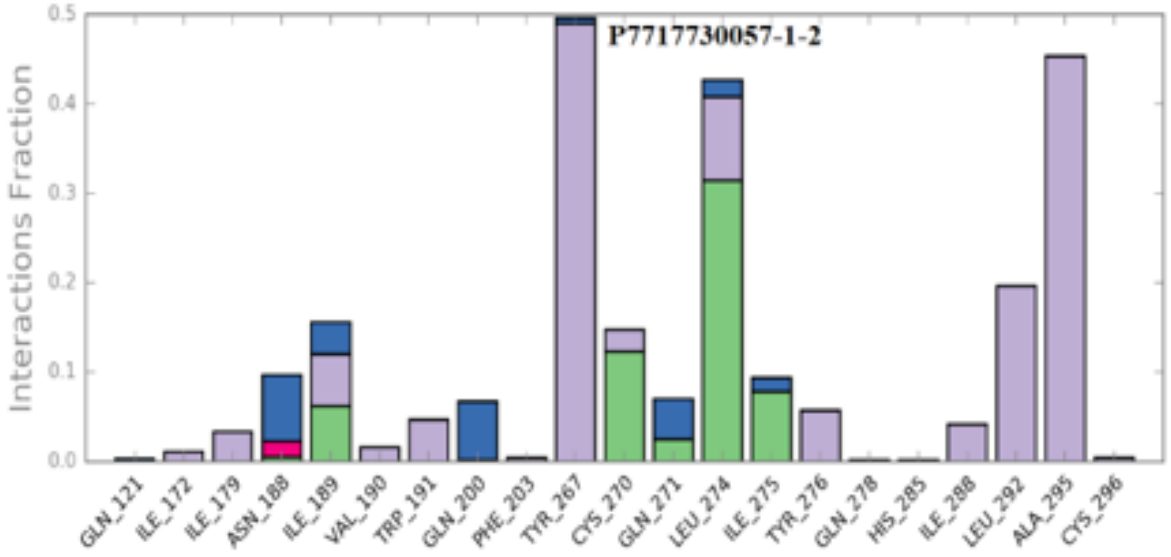


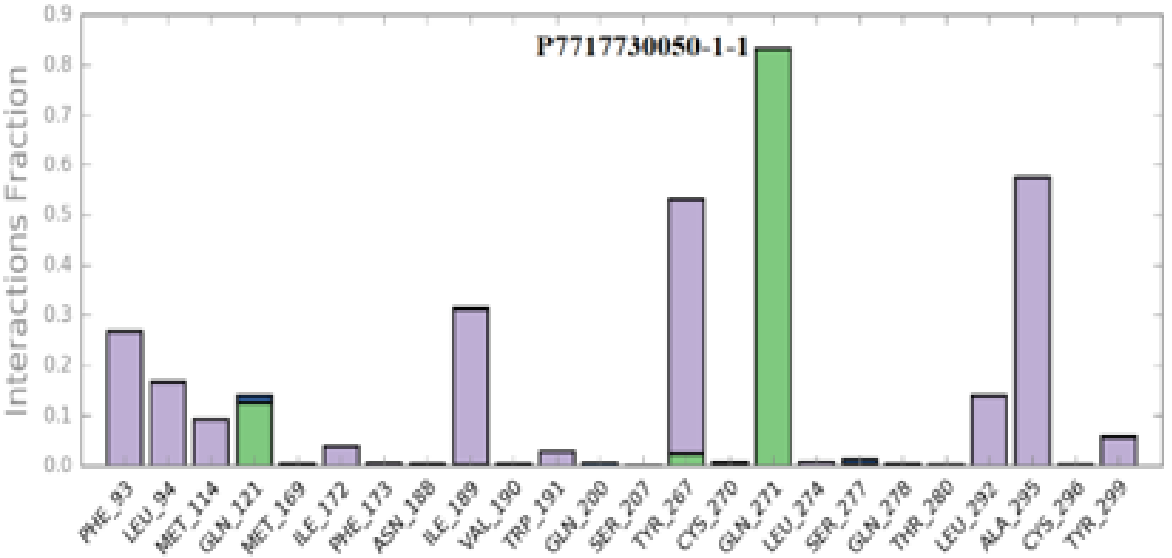

P7119812493-2
Interactions Fraction
2.0
1.5
1.0
0.5
0.0
THR_97
MET_98
MET_114
THR_117
GLY_118
GLN_121
PHE_122
VAL_168
ILE_172
LEU_178
ILE_179
ASN_188
ILE_189
VAL_190
TRP_191
GLN_200
PHE_203
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
GLN_278
ILE_288
LEU_292
ALA_295

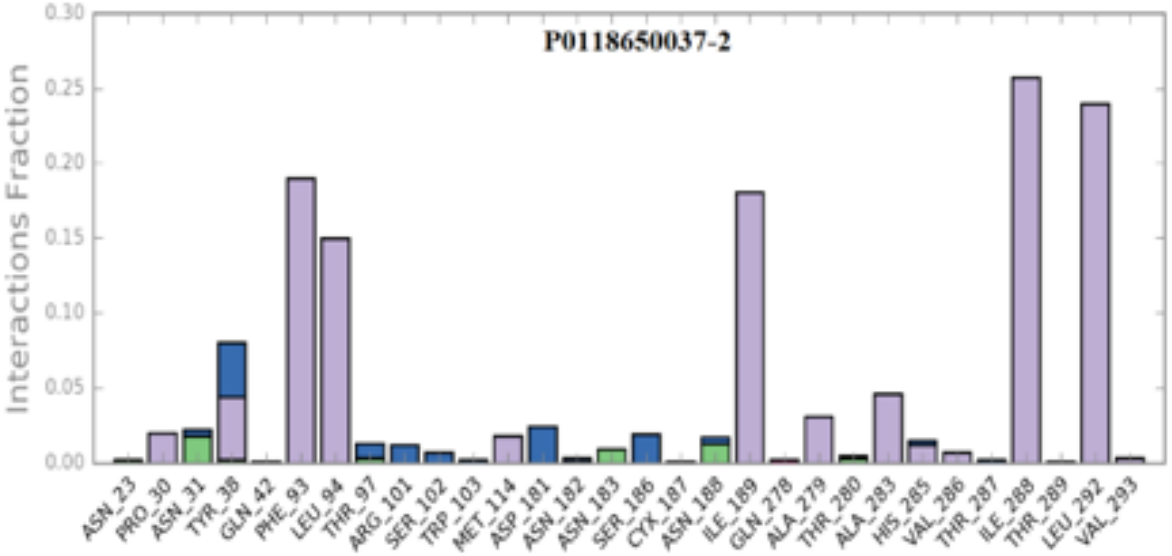

P7119812967-1
Interactions Fraction
1.2
1.0
0.8
0.6
0.4
0.2
0.0
TYR_38
PHE_93
LEU_94
THR_97
MET_98
MET_114
THR_117
GLN_121
MET_169
ILE_172
ASN_188
ILE_189
TRP_191
GLN_200
PHE_203
SER_207
TYR_267
TRP_268
CYS_270
GLN_271
LEU_274
GLN_278
ILE_288
LEU_292
ALA_295
CYS_296
TYR_299

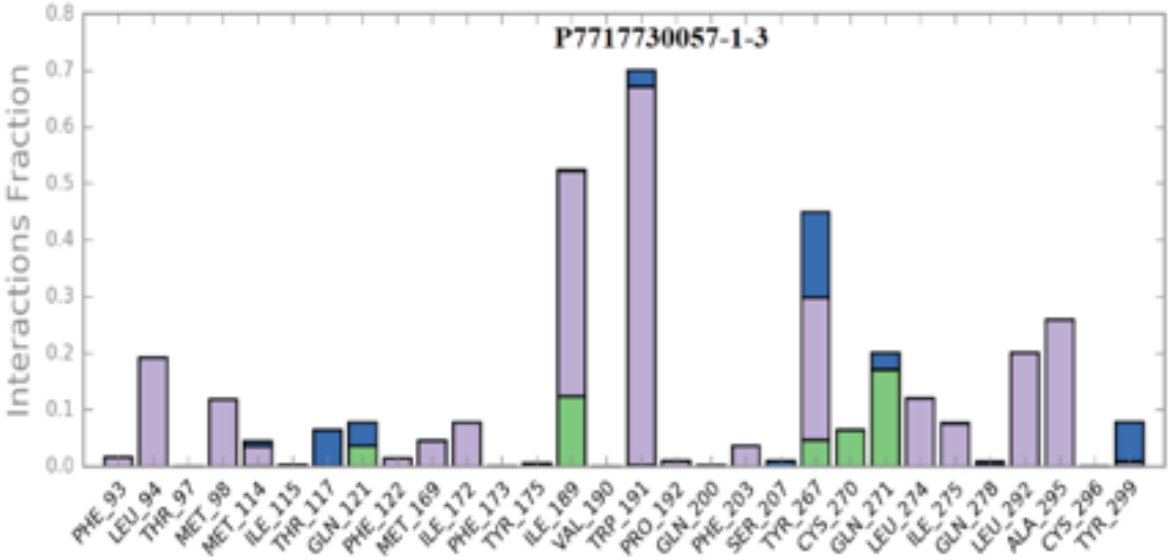

P7717730050-1-4
Interactions Fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
LEU_94
MET_98
MET_114
GLN_121
PHE_122
MET_169
ILE_172
ILE_179
ASN_188
ILE_189
VAL_190
TRP_191
GLN_200
PHE_203
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
SER_277
GLN_278
LEU_292
ALA_295
CYS_296
TYR_299

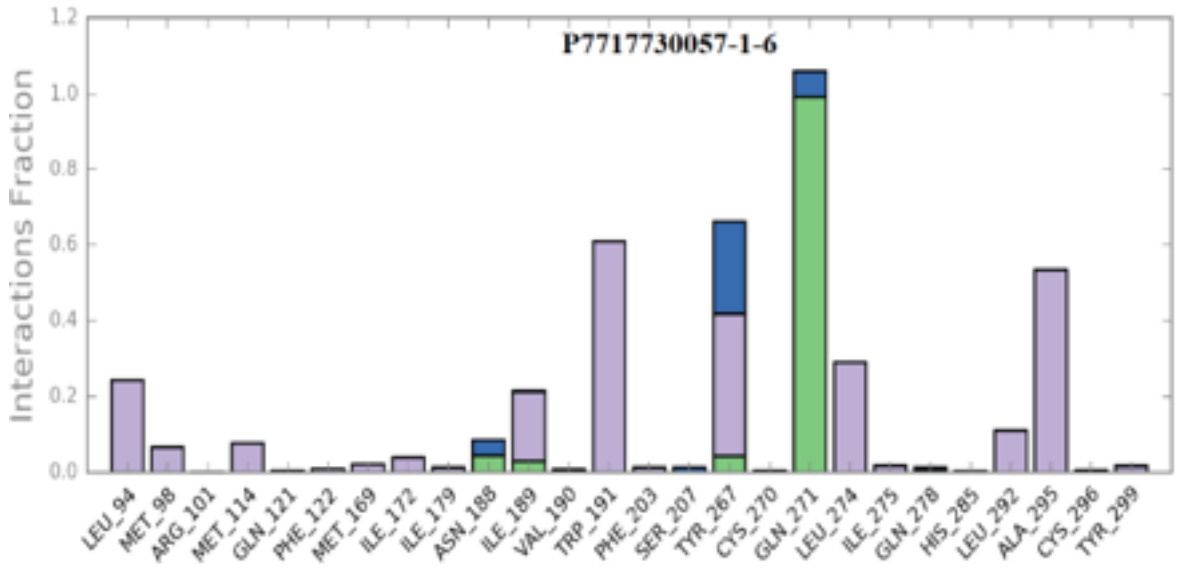

P7112410434-2-1
Interactions Fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
VAL_36
TYR_38
PHE_93
LEU_94
THR_97
TYR_113
MET_114
THR_117
GLN_121
MET_169
ILE_172
ASN_188
ILE_189
TRP_191
GLN_200
PHE_203
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
SER_277
GLN_278
ILE_288
LEU_292
ALA_295


P7717730057-1-2
Interactions Fraction
0.0
0.1
0.2
0.3
0.4
0.5
GLN_121
ILE_172
ILE_179
ASN_188
ILE_189
VAL_190
TRP_191
GLN_200
PHE_203
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
TYR_276
GLN_278
HIS_285
ILE_288
LEU_292
ALA_295
CYS_296


P7717730050-1-1
Interactions Fraction
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
PHE_93
LEU_94
MET_114
GLN_121
MET_169
ILE_172
PHE_173
ASN_188
ILE_189
VAL_190
TRP_191
GLN_200
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
SER_277
GLN_278
THR_280
LEU_292
ALA_295
CYS_296
TYR_299


P0118650037-2
Interactions Fraction
0.30
0.25
0.20
0.15
0.10
0.05
0.00
ASN_23
PRO_30
ASN_31
TYR_38
GLN_42
PHE_93
LEU_94
THR_97
ARG_101
SER_102
TRP_103
MET_114
ASP_181
ASN_182
ASN_183
SER_186
CYX_187
ASN_188
ILE_189
GLN_278
ALA_279
THR_280
ALA_283
HIS_285
VAL_286
THR_287
ILE_288
THR_289
LEU_292
VAL_293


P7717730057-1-3
Interactions Fraction
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PHE_93
LEU_94
THR_97
MET_98
MET_114
ILE_115
THR_117
GLN_121
PHE_122
MET_169
ILE_172
PHE_173
TYR_175
ILE_189
VAL_190
TRP_191
PRO_192
GLN_200
PHE_203
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
GLN_278
LEU_292
ALA_295
CYS_296
TYR_299


P7717730057-1-6
Interactions Fraction
1.2
1.0
0.8
0.6
0.4
0.2
0.0
LEU_94
MET_98
ARG_101
MET_114
GLN_121
PHE_122
MET_169
ILE_172
ILE_179
ASN_188
ILE_189
VAL_190
TRP_191
PHE_203
SER_207
TYR_267
CYS_270
GLN_271
LEU_274
ILE_275
GLN_278
HIS_285
LEU_292
ALA_295
CYS_296
TYR_299