IDENTIFICATION AND DOCKING ANALYSIS OF ALLATOSTATIN RECEPTOR IN *CARAUSIUS MOROSUS*

by

Burçin Duan

B.S., Molecular Biology and Genetics, Boğaziçi University, 2009

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

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2012

IDENTIFICATION AND DOCKING ANALYSIS OF ALLATOSTATIN RECEPTOR IN *CARAUSIUS MOROSUS*

APPROVED BY:

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

Cemalettin Bekpen, Ph.D. .....

Assoc. Prof. Osman Uğur Sezerman .....

DATE OF APPROVAL: 16.07.2012

To my family and my husband...

ACKNOWLEDGEMENTS

This thesis project is a collaborative product of Necla Birgül's Cancer Signaling Lab (CSL) in Boğaziçi University with Uğur Sezerman from Sabancı University and Hamdi Torun's BioAFM Laboratory in Boğaziçi University. The cloning and cell culture experiments were performed in CSL, atomic force microscopy was performed in BioAFM lab and bioinformatics tools were conducted with Uğur Sezerman.

For that she has accepted me in this worthful project I would like to thank to my supervisor Necla Birgül İyison, at first. She made a great leading effort, showed patience, supported me in every step and lend assistance even in my personal life. I am very grateful to perform this project as a part of her CSL team and being trained in wet-lab personally by her.

I would also like to thank to Uğur Sezerman for his efforts on training me about bioinformatics, for spending his time on analyses of this project and for opening my mind via asking significant questions. His M.Sc. student Onur Yükselen spent so much effort on my training in bioinformatics tools and most of the times he solved my biggest problems even in his thesis-writing period. Thus, I would like to say that I cannot pay back their assistance and patience during the project.

I am also very happy about being able to perform atomic force microscopy experiments in Hamdi Torun's BioAFM lab. He spent his dear time to perform these experiments and showed a great patience about my ignorance in biophysics.

I would like to add my appreciations to all CSL members; an undergraduate student Mustafa Uğur Baş for helping me in RACE reactions, Emine Dindar for assisting me in cloning parts and personal friendship, Tuncay Şeker for his helps in cell-culture procedures and his friendship with unlimited patience and criticism, Zeynep Özcan from İbrahim Yaman's lab for her general assistance in the lab and in my personal life, İzzet Akiva, Vahap Kapıkıran, Güneş Tunçgenç, İpek Even and Esra Şekerci for

making my lab life enjoyable. I am also grateful to have Damla Kaptan, Gizem Gül and Ayşe Güven for their support, collective sense and friendships even outside my academic life. I would also give my thanks to Arzu Çelik, Cemalettin Bekpen and Tuncay Altuğ whom we have lost on the 5th of July, for their efforts to support me, to make me available in scientific studies and for their personal helps outside my lab-life.

At last, I would like to say that I am very lucky to have such a family and such a patient husband. In each day of my education, my family supported me in a full manner that I cannot pay them back. My little sister Gizem Duan prepared my figures in this thesis in Adobe Photoshop software and could stand on me. And my husband was a great supporter when I felt like I was lost in my undergraduate years and even in my thesis-writing period, with his patience and love.

This wok was supported by Boğaziçi University Research Fund (BAP project no: 5043).

Thanks for all.

ABSTRACT

IDENTIFICATION AND DOCKING ANALYSIS OF ALLATOSTATIN RECEPTOR IN *CARAUSIUS MOROSUS*

Allatostatin receptors (AlstRs) with their ligands (ASTs) play role in insect development, by inhibiting the production of Juvenile hormone. This thesis project uses the advances in structure-based studies and single-molecule based approaches in order to obtain a detailed information about AlstR-AST interaction in a model laboratory organism *Carausius morosus*. The mRNA sequence of AlstR was first obtained via 5' and 3' RACE reactions with the help of specific primers designed on conserved transmembrane region previously found by Birgül *et al.* The longest ORF was predicted and identified as an Allatostatin Receptor type C of insects. As a result it was named as AlstR-C of *C. morosus* (CamAlstR-C) in this project. Then, evolutionary analysis was combined with bioinformatics modeling and docking tools, to obtain some features of CamAsltR-C interaction pattern. This interaction was verified *in vivo* via atomic force microscopy. This was the first time of AFM study on living cells, in Boğaziçi University. Binding force was detected in CamAlstR-C over-expressed cells. In addition, the binding pocket was predicted bioinformatically, to be composed of second, third and N-terminal extracellular loops of CamAlstR-C. And in more detail, a specific conserved binding motif (IXTXP) was proposed on the third extracellular loop, which may be identical in other AlstRs.

ÖZET

CARAUSIUS MOROSUS'TA ALLATOSTATİN RESEPTÖRÜN TANIMLANMASI VE LİGANDIN YERLEŞTİRİLMESİ

Allatostatin reseptörleri (AlstR) ve bunların peptid ligandları (AST) böceklerde Gençlik Hormonu üretimini durdurarak metamorfoz aşamasında görev yapmaktadır. Bu tezde AlstR-AST bağlanmasına dair ayrıntılı bir bilgi edinmek amacıyla yapıya-dayalı yöntemler tek-molekül yaklaşımları ile bir arada, bir laboratuvar model organizması olan *Carausius morosus* üzerinde kullanılmıştır. Daha önceden bulunmuş olan TM bölgesindeki bilgiye göre hazırlanmış özel primer çiftleri yardımıyla 3' ve 5' RACE reaksiyonları gerçekleştirilerek mRNA dizisi bulunmuştur. En uzun açık okuma penceresi tahmin edilmiş ve böceklerdeki Allatostatin reseptörü tip C olduğu belirlenmiştir. Bunun üzerine, ilgili reseptöre bu projede *C. morosus* Allatostatin C reseptörü (CamAlstR-C) adı verilmiştir. Daha sonra, CamAlstR-C bağlanma koşullarını anlamak için evrimsel analiz biyoenformatik modelleme ve yerleştirme araçları ile birleştirilmiştir. Bu bağlanma atomik kuvvet mikroskobu ile *in vivo* olarak da gösterilmiştir. Bu çalışma, üniversitemizde AFM'nin canlı hücreler üzerinde denendiği ilk çalışmadır. CamAlstR-C'nin aşırı anlatımının sağlandığı hücrelerde bağlanma kuvvetleri ölçülebilmiştir. Buna ek olarak, bağlanma cebinin CamAlstR-C'nin N-ucu, ikinci ve üçüncü hücre-dışı çevrimlerinden oluştuğu biyoenformatik yollarla tahmin edilmiştir. Hatta daha özel olarak, diğer böcek AlstR'lerinde de aynı özelliği gösterebilecek, üçüncü hücre-dışı çevriminde özel olarak korunmuş bir bağlanma motifi (IXTXP) önerilmiştir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv
ABSTRACT vi
ÖZET vii
LIST OF FIGURES xi
LIST OF TABLES xiv
LIST OF SYMBOLS xvi
LIST OF ACRONYMS/ABBREVIATIONS xviii
1. INTRODUCTION 1
1.1. Allatostatin 2
1.1.1. Types and Affinities 2
1.1.2. Allatostatin Receptor 4
1.1.3. AlstR-AST System in Neuroscience 6
1.2. Somatostatin Receptor 6
1.2.1. Physiology 6
1.2.2. Structural Features 7
1.2.3. Intracellular Signaling and Relation with Cancer Studies 8
1.2.4. Common Structural Features of GPCRs 10
1.3. Bioinformatics in Structural Studies 12
1.3.1. Protein Threading 12
1.3.2. Molecular Docking 14
2. PURPOSE 16
3. MATERIALS 17
3.1. General Kits, Enzymes and Reagents 17
3.2. Plasmids 18
3.3. Bacteria and Cell Lines 18
3.4. Peptides 18
3.5. Solutions and Media 19
3.5.1. DNA Gel Electrophoresis 19
3.5.2. Nuclease-Free Water 20

3.5.3. RNA Gel Electrophoresis Buffers 20
3.5.4. Bacterial Media 21
3.5.5. Complete Growth Media 21
3.6. Primers 21
3.7. Web Servers and Softwares 22
3.8. Equipment 23
4. METHODS 26
4.1. Identification of The Receptor 26
4.1.1. RNA Extraction from Insect Tissue 26
4.1.2. cDNA Synthesis for RACE Reactions 27
4.1.3. 3' and 5' RACE 28
4.1.4. PCR Product Purification 33
4.1.5. Preparation for Sequencing 34
4.1.6. PCR Amplification of the Open Reading Frame 35
4.1.7. Cloning of the Open Reading Frame 36
4.2. *in vivo* Studies 38
4.2.1. General Cell Culture Techniques 38
4.2.2. Transfection 39
4.2.3. Immunofluorescence 39
4.2.4. Atomic Force Microscopy 39
4.3. Bioinformatics 41
4.3.1. Identification of Protein Family 41
4.3.2. Alignment of N-Termini 42
4.3.3. Secondary Structure Prediction 45
4.3.4. 3D Modeling 45
4.3.5. Molecular Docking 46
5. RESULTS 47
5.1. Identification of the Receptor 47
5.2. Phylogenetic Analysis 55
5.3. Secondary Structure Prediction 56
5.4. Modeling 57
5.5. Docking Analysis 58

6. DISCUSSION 70
APPENDIX A: PLASMID MAPS AND MULTIPLE CLONING SITES 76
APPENDIX B: DNA LADDERS 79
APPENDIX C: TMHMM PROBABILITIES 80
APPENDIX D: ATOMIC FORCE MICROSCOPY FORCE CURVES 87
REFERENCES 89

LIST OF FIGURES

Figure 1.1. Location of JH Synthesis and AST Peptide Release. 3

Figure 1.2. Types of Allatostatin Peptides. 4

Figure 1.3. Major Secondary Messengers of GPCRs. 9

Figure 1.4. Examples for Downstream Pathways of SSTRs. 10

Figure 1.5. Flowchart in Protein Threading. 14

Figure 5.1. RNA Isolation. 47

Figure 5.2. 3' RACE Results. 48

Figure 5.3. 5' RACE Results. 48

Figure 5.4. Colony PCR for 3' RACE Fragment. 49

Figure 5.5. Open Reading Frame Amplification. 49

Figure 5.6. Open Reading Frame Sequence. 51

Figure 5.7. Open Reading Frame Translation. 51

Figure 5.8. Immunofluorescence. 52

Figure 5.9. Atomic Force Microscopy Raw Data. 53

Figure 5.10. Atomic Force Microscopy Force Curves. 54

Figure 5.11. Blastn Results of mRNA. 55

Figure 5.12. Blastp Results of ORF Translation. 56

Figure 5.13. Phylogenetic Analysis. 59

Figure 5.14. N-Terminal Loop Alignments. 60

Figure 5.15. N-terminal Alignment within Apis Genus. 61

Figure 5.16. N-terminal Alignment within Anopheles Genus. 61

Figure 5.17. N-terminal Alignment within Drosophila Genus. 62

Figure 5.18. Secondary Structure Prediction via PSIPRED. 63

Figure 5.19. Motifs. 64

Figure 5.20. Structural Alignments on Templates. 65

Figure 5.21. Ligand Modeling. 66

Figure 5.22. Docked AST-C Ligand on CamAlstR-C. 67

Figure 5.23. Surface Interactions of CamAlstR-C and AST-C. 67

Figure 5.24. Interacting Residues. 68

Figure 5.25. Binding Pocket of CamAlstR-C. 69

Figure A.1. pGEMT-Easy Vector. 76

Figure A.2. pcDNA3 Vector. 77

Figure A.3. pcDNA3/HA Vector. 78

Figure A.4. pEGFP-N2 Vector. 78

Figure B.1. DNA Ladders. 79

Figure D.1. AFM Force Curves 1. 87

Figure D.2. AFM Force Curves 2. 88

LIST OF TABLES

Table 3.1. Kits, Enzymes and Reagents. 17

Table 3.2. Gel Electrophoresis Buffers. 19

Table 3.3. Nuclease Free Water. 20

Table 3.4. RNA Buffers. 20

Table 3.5. Bacterial Media. 21

Table 3.6. Contents of Growth Media. 21

Table 3.7. Primers. 22

Table 3.8. Web Servers and Softwares. 22

Table 3.9. Equipments. 23

Table 4.1. 3' RACE First PCR Reaction Reagents. 29

Table 4.2. 3' RACE First Reaction Condition. 29

Table 4.3. 3' RACE Second PCR Reaction Reagents. 30

Table 4.4. 3' RACE Second Reaction Condition. 30

Table 4.5. 5' RACE First PCR Reaction Reagents. 31

Table 4.6. 5' RACE First Reaction Condition. 31

Table 4.7. 5' RACE Second PCR Reaction Reagents. 32

Table 4.8. 5' RACE Second Reaction Condition. 33

Table 4.9. PCR Amplification Reagents. 36

Table 4.10. PCR Amplification Condition. 36

Table 4.11. The N-Terminal Sequences Used in Alignments. 44

Table 5.1. RNA concentration. 47

Table 5.2. I-TASSER Models. 58

Table 5.3. Top Ten Templates in CamAlstR Modeling. 58

Table 5.4. Conservation on Binding Pocket. 59

Table C.1. TMHMM Probabilities, Part 1. 80

Table C.2. TMHMM Probabilities, Part 2. 81

Table C.3. TMHMM Probabilities, Part 3. 82

Table C.4. TMHMM Probabilities, Part 4. 83

Table C.5. TMHMM Probabilities, Part 5. 84

Table C.6. TMHMM Probabilities, Part 6. 85

Table C.7. TMHMM Probabilities, Part 7. 86

LIST OF SYMBOLS

A	Alanine
bp	Basepair
C	Cysteine
D	Aspartic Acid
E	Glutamic Acid
F	Phenylalanine
G	Glycine
h	Hours
H	Histidine
I	Isoleucine
K	Lysine
L	Leucine
M	Methionine
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
N	Asparagine
ng	Nanogram
nm	Nanometer
P	Proline
pN	PicoNewton
Q	Glutamine
R	Arginine
RMSD	Root mean square deviation
rpm	Revolutions per minute
S	Serine
sec	Second
T	Threonine

U	Selenocysteine
V	Valine
V	Potential difference
W	Tryptophan
X	Any Amino Acid
Y	Tyrosine
μg	Microgram
μl	Microliter
μm/s	Micrometer per second
μM	Micromolar

LIST OF ACRONYMS/ABBREVIATIONS

3D	3 Dimensional
AC	Adenylate cyclase
Acetyl Co-A	Acetyl coenzyme-A
AFM	Atomic Force Microscopy
AKT	Protein kinase B
AlstR	Allatostatin Receptor
AST	Allatostatin Peptide
bFGF	Basic fibroblast growth factor
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BPB	Bromophenol blue
CA	Corpora allata
$CaCl_2$	Calcium chloride
CamAlstR-C	*C. morosus* Allatostatin receptor type-C
cAMP	cyclic adenosine monophosphate
CC	Corpus cardiaca
cDNA	Complementary deoxyribonucleic acid
CNS	Central nervous system
CSL	Cancer Signaling Laboratory
CO_2	Carbondioxide
CXCR4	CXC Chemokine receptor type 4
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol
EC	Extra-cellular
ECL	Extra-cellular loop
EDTA	Ethylenediaminetetraacetate

ERLB	EtBr RNA loading buffer
EtBr	Ethidium bromide
EtOH	Ethanol
FBS	Fetal bovine serum
GDP	Guanosine diphosphate
GIRK	G protein-coupled inwardly-rectifying potassium channel
GPCR	G protein-coupled receptor
GTP	Guanosine triphosphate
GH	Growth Hormone
hSSTR	Human somatostatin receptor
IC	Intra-cellular
ICL	Intra-cellular loop
I-TASSER	Iterative Threading Assembly Refinement
JH	Juvenile Hormone
KCl	Potassium chloride
LB	Luria-Bertani
MAPK	Mitogen-activated protein kinase
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MOPS	Morpholino propane sulfonic acid
mRNA	Messenger ribonucleic acid
MTS	Mitochondria Targeting Sequence
NaCl	Sodium chloride
NaOAc	Sodium Acetate
NaOH	Sodium hydroxide
NCC	Nervi corpus cardiaci
OD	Optical density
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	ProteinDataBank

PI3K	Phosphoinositide-3 kinase
PKC	Protein kinase C
PLC	Phospholipase C
PMSF	Phenylmethylsulphonyl fluoride
PTP	Protein tyrosine phosphatase
PTX	Pertussis toxin
RACE	Rapid Amplified cDNA Ends
Rb	Retinoblastoma tumour suppressor protein
RhoA	Ras homolog gene family, A
RNA	Ribonucleic acid
RT	Room temperature
SP	Specific primer
SST	Somatostatin peptide
SSTR	Somatostatin Receptor
TAE	Tris-acetic acid EDTA
TBE	Tris Base EDTA
TE	Tris-EDTA
TM	Trans-membrane
UV	Ultraviolet
VEGF	Vascular endothelial growth factor

1. INTRODUCTION

Insect models are being used in a wide-range of medical or biological studies, due to their evolutionary relationships with humans. In most of these, molecular approaches are required. For instance, chemical insecticides may harm the environment. However the current scientific progress can target molecular developmental pathways of these pests without adding any chemicals to the farmlands. In addition, functional homology studies on metabolites and hormones of insects enables emergence of new treatment methods for human diseases. Furthermore, the homology in cellular mechanisms of insects can lead to finding model mechanisms and making generalizations in cellular mechanisms of humans. Largely known mechanisms of molecular pathways in insects may serve as model systems and as means of experimental designs.

Preliminary studies on *Drosophila* species revealed that some ancient cellular mechanisms of insects would facilitate cell progression studies on Mammalia. Allatostatin receptor (AlstR), in its' activated form, was found to be responsible for inhibition of growth in insects at a specific time point of their development. The beginning of this thesis project relies on the previous findings about AlstR of *Drosophila melanogaster* and its' homology to human Somatostatin receptor (hSSTR) [1]. In 2001, other AlstR expressions were detected in *Periplaneta americana* and *Carausius morosus*; but their products were not identified.

Carausius morosus was taken as the model organism and AlstR system was studied on this insect species. The common name of *C. morosus* is Indian stick insect. Because it is studied very widely, the zoologists also call it as "laboratory stick insect". It feeds on leaves of variable plants, gives damage to the farmlands in South Africa and may be a treat for gardeners in urban areas.

1.1. Allatostatin

Since 1990s developmental studies on Arthropods revealed various cellular mechanisms responsible for the growth of these animals. Short allatostatic neuropeptides were found in their nervous system, inhibiting specific growth hormones which in turn play role in metamorphosis. These inhibitory neuropeptides were called as Allatostatins (AST).

A developmental regulatory hormone which is called Juvenile Hormone (JH) is produced in *Corpora Allata* (CA) of arthropods and a reduction of its' synthesis is required at the larval instar stages. This hormone is produced through a pathway called Mevalonate Pathway in which the preliminary steps are conserved in both Arthropoda and Mammalia. It starts with Acetyl Co-A and leads to the production of cholesterol and some isoprenoids in Mammalia, which play role in cell signalling and carcinogenesis. However arthropods cannot express the same enzymes as in Mammalian pathway, and the reactions go towards JH as an end product [2]. AST peptides are known to inhibit this pathway at the first steps [3, 4]. As a result insects cannot produce cholesterol as a product of mevalonate pathway and this makes them "sterol-dependent". Peptide is released to CA from the nerve endings protruding mostly from *pars lateralis* cells in the brain (Figure 1.1).

In fact, ASTs are functionally pleiotropic. They are released to both CA and hemolymph, targeting different tissues. They were found to inhibit not only JH production but also muscular contraction of midgut, vitallogenesis and oogenesis in this phylum [5, 6].

1.1.1. Types and Affinities

Allatostatins of insects can be classified in three different groups. The first group is called the "A-type ASTs" (AST-A) and they represent a common motif of Y/FXFGL/I-amide on their C-terminus. They were first identified in cockroach *Diploptera punctata* [7]. In a total of 13 *D. punctata* ASTs were produced by en-

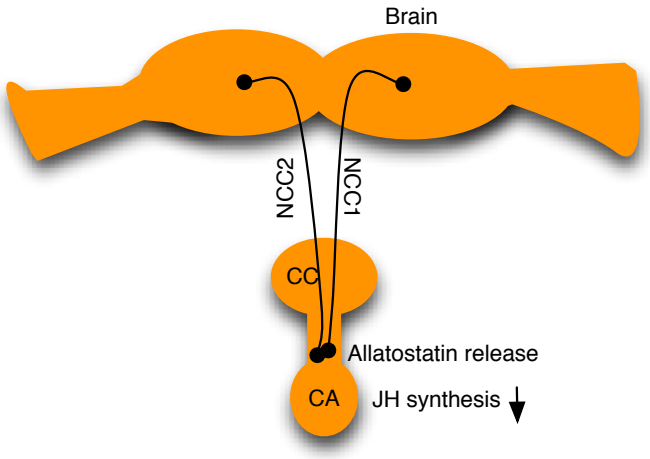

Figure 1.1. JH is synthesized in CA and the release of AST peptide from NCC nerves inhibits its' synthesis. CA: *corpora allata*; CC: *corpus cardiaca*; NNC1/2: *nervi corpus cardiaci* 1/2

doproteolytic cleavage of precursor proteins and they were all found to contain the same consensus (Y/FXFGL/Iamide) at their C-termini [8]. B-type ASTs (AST-B) are the members of a second group mostly active in crickets and which contain a C-terminal motif of W$(X)_6$W-amide [9]. Lastly, the third group (AST-C) is common in lepidopterans and they constitute a consensus like PISCF, without amidation at C-terminus [10, 11]. According to literature results, deletion studies on C-terminus of AST peptides resulted in weaker activation of the relevant receptors [12, 13]. This may confirm that C-terminal motives of these peptides would be the "active cores". In addition to the conservation of these motives on ASTs, full lengths can be identical within species. For instance, five *P. americana* AST peptides were found to be identical with *D. punctata* AST peptides [8].

These three groups of AST peptides activate different G protein coupled receptors (GPCR) called Allatostatin receptors (AlstR) in common. A specific type of ligand can activate more than one type of receptor and *vice versa*, also within different species in the phylum [14, 15]. Other identities or similarities and cross-reactivities of AST peptides were demonstrated so far in comparisons within insect classes or orders, with which we can expect sequence similarities between their relevant GPCRs. Allatostatic activity of these peptides is variable when they are cross-reacted intra-species [10].

Also the same peptide can show different affinities with different tissue samples [8]. Different types of ASTs can also have different allatostatic activity on JH inhibition. For instance, C-type of ASTs do not affect JH synthesis in *D. melanogaster* larva, but other types of ASTs may be responsible for this inhibition [9].

ASTs have effect on both JH synthesis and gut muscle contraction. They play different roles in hemolymph and on CA. According to the studies performed on 4 different target tissues (hindgut, midgut, brain and CA), their variable activity was supposed to be defined by the N-termini. In addition, the activity of AST peptides on JH inhibition is proposed to depend on amount of local release of the peptide, presence of degrading enzymes or sensitivity of CA to the peptide [8, 16].

Type	A-Type ASTs	B-Type ASTs	C-Type ASTs
Common motif	Y/FXFGL/Iamide	$W(X)_6$Wamide	PISCF
Origin	*Diploptera punctata*	*Gryllus bimaculatus*	*Manduca sexta*
Localization	PL cells of the brain	both PL and PI cells of the brain	innervating nerves on CA
Number in family	13	6	1

Figure 1.2. Types, motifs, first organism identified and number of Allatostatin types in these organisms are represented.

1.1.2. Allatostatin Receptor

ASTs of insects are produced in pars lateralis cells in brain, transported via NCC nerves ending at CA and finally interact with their allatostatin receptors (AlstR) on the CA triggering an intracellular signal transduction and resulting in inhibition of JH synthesis. These receptors are classified as GPCRs. They are composed of a highly conserved 7 trans-membrane (7TM) region, a glycosylated extracellular N-

terminus and generally phosphorylated intracellular C-terminus. If they are named according to their ligands, AlstRs can also be classified in three groups; AlstR-A group activated by A-type ASTs, AlstR-B group by B-type ASTs and AlstR-C group by C-type ASTs. Trans-membrane domains of AlstR-A, AlstR-B and AlstR-C groups were found to be closely related to vertebrate galanin receptors, bombesin receptors and somatostatin/opioid receptors respectively [5].

AlstRs have some common features within insect species. For instance, DRY motif on the intracellular loop between third and fourth TM regions (TM3 and TM4) is thought to be involved in the activation of the receptor [5]. It was shown that AlstRs also represent second messenger dependent kinase binding signals, Protein kinase C (PKC) and Protein kinase A phosphorylation sites [17]. In contrast to these conserved binding signals, each AlstR can show a tissue-specific, species-specific and stage-dependent pattern of expression and activation [18].

Birgül *et al.* identified the first *Drosophila* AlstR as an AlstR type A (AlstR-A) by reverse physiology studies. It was expressed in larva, pupa and the whole body of adult, but not in the embryos. As a result of electrophysiological studies with different ligands which are Mammalian and *Diploptera* homologs of ASTs, neither of Mammalian homologs could activate this receptor while *Diploptera* homologs could [1, 19]. Mammalian ligands do not share a sequence similarity with their functional analogs in insects. To sum up, it can be predicted that the activity of the receptors is not sequence-related but is ligand-specific. Auerswald and Birgül continued AlstR studies in *P. americana* and *C. morosus* and found other AlstR mRNA expressions in these species. By cloning of only the conserved 7TM regions of these receptors *P. americana* AlstR was found to be more related to AlstR-A of *Drosophila*. However two different mRNAs were found in *C. morosus*, one was more evolutionarily closer to AlstR-A but the other was closer to Somatostatin receptors (SSTR) of Mammalia [20]. Identification of this second receptor of *C. morosus* is taken as the goal of this thesis.

1.1.3. AlstR-AST System in Neuroscience

The field of neural circuitry and behaviour is still one of the most challenging studies in neuroscience. Evolution of complexity in the nervous system of humans makes these studies hard to succeed. This difficulty leads to seeking for similar, simpler and more familiar molecular systems and models in neuronal pathways, depending on the functional homology between humans and other earlier organisms. Tan *et al.* established a molecular system for studying neuronal circuitry in rodents. They used *Drosophila* A-type AST with its' cognate receptor in order to inactivate Mammalian neurons in a selective and rapid way *in vivo*. They infected rat and ferret cortical neurons with AlstR and then treated with the ligand. The results proved that this system could hyperpolarize neurons transduced and inhibit spiking even to optimized visual stimuli [21]. This neuronal silencing method is used and developed successfully by other scientists, to date [22, 23].

1.2. Somatostatin Receptor

The Mammalian homolog of AlstR was predicted as somatostatin receptor (SSTR) in convenience with some specific sequence motifs conserved in TM regions [1]. Both AlstR and SSTR represent the typical sequence motives of Family A Rhodopsin-like GPCRs.

1.2.1. Physiology

Somatostatin (SST) peptide was first described in ovine hypothalamus by Brazeau *et al.*, as an inhibitor of rat and human growth hormone (GH) secretion. They identified a 14 residue-long polypeptide (SST-14) which is conserved within vertebrates [24]. Prior to 2010, another active form of SST peptide (SST-28) was discovered, which is 28 residue-long and processed from the same precursor protein by post-translational cleavage [25].

Beside hypothalamic expression SSTs were detected in a wide variety of tis-

sues, such as gastrointestinal tract [26], nervous system [27], thyroid, pancreatic islets [28] etc. Compatible with peptide expression, distribution of receptors (SSTRs) was widespread in compatible tissues, such as CNS, adrenals, spleen, kidney, liver, lungs and islets. Four and five SSTR subtypes were described in fish and human respectively [29,30]. However the knowledge about ligand selectivity of the receptors remains unclear [31]. Klein and Sheridan proposed that and tissue-specific expression patterns of ligands and SSTRs would define their physiological effects and particular intracellular responses, in variable conditions [32]. For instance at gestation, SST release in hypothalamus reduces secretion of GH from the pituitary [33]. We now know that somatostatins do not only inhibit GH secretion but they also behave like neurotransmitters in brain by inhibiting release of dopamine, norepinephrine, corticotrophin-releasing hormone and thyroid-releasing hormone; in thyroid by inhibiting release of T3 and T4; on adrenals by inhibiting aldosteron secretion, on kidney by inhibiting renin secretion and water absorption; on gastrointestinal tract suppressing the motor activity and also in islet cells suppressing both synthesis and secretion of insulin and glucagon [26,29,32,34,35].

1.2.2. Structural Features

SSTRs are a member of the biggest GPCR family which is called as Family A Rhodopsin-like receptors. They share some common features such as 7 trans-membrane (7TM) helices, extracellular variable N-terminus and an intracellular C-terminus. In all vertebrate species, SSTR subtypes contribute to a sequence pattern in their 7TM region, YANSCANPI/VLY. The other signal on these Family A receptors is the DRY consensus sequence at their cytoplasmic end of 3rd trans-membrane helix (TM3). The second and third intracellular loops (ICL2, ICL3) are known to interact with G proteins. On their C-terminal loop they have a Cysteine amino acid to comprise an anchor with the membrane through this flanking region. Additionally, other consensus patterns were identified for phosphorylation of protein kinase A (PKA), protein kinase C (PKC) and calmodulin kinase II [34, 36].

1.2.3. Intracellular Signaling and Relation with Cancer Studies

Since SSTRs belong to Family A Rhodopsin-like receptors they are supposed to represent the same activation states as of other GPCRs. When the ligand interacts with the SSTR, a conformational change occurs and the G proteins, which are bound to the receptor, are activated. In case of rhodopsin, this conformational change begins with the rotation and tilting of 6th trans-membrane helix (TM6) relative to TM3. At this point the interaction between DRY motif of TM3 and an acidic residue of TM6 is disrupted. An exchange from GDP to GTP occurs and the heterotrimeric G proteins dissociate from the receptor. Gα and G$\beta\gamma$ subunits activate different secondary messenger systems.

Gα proteins can be classified in four types. Gα_s stimulates adenylate cyclase (AC) and leads to cAMP accumulation. G$\alpha_{i/o}$ inhibits AC activity leading to G protein-coupled inwardly-rectifying potassium channel (GIRK) activation. Gα_q generates phosholipase C (PLC) pathway and Gα12/13 proteins activate RhoA pathway (Figure 1.3).

SSTRs were shown to couple with pertussis toxin (PTX) sensitive G proteins which are classified in Gα_i group. They can activate GIRK channels and also modulate voltage-dependent calcium channels via Gα_o. They were found to recruit phosphotases via Tyrosine residues on their C-terminal. Florio *et al.* revealed the direct relation of these receptors with phospholipase C/A2/D (PLC/D/A2) activation [37]. Also other effectors of SSTR activation such as MAPKs and phosphoinositide-3 kinase (PI3K) were found. Similarly, insect homolog AlstRs were found to couple differentially to both PTX-sensitive and PTX-insensitive G proteins [15].

As a consequence of evidences in various pathways, the intracellular pathway to be activated can be figured by various factors. For instance, SSTR subtype can determine the type of G protein coupling and the active downstream pathway (Figure 1.4). The secondary messengers expressed in a tissue can represent the possible intracellular responses. In addition, regulation of the receptor by internalization and desensitization,

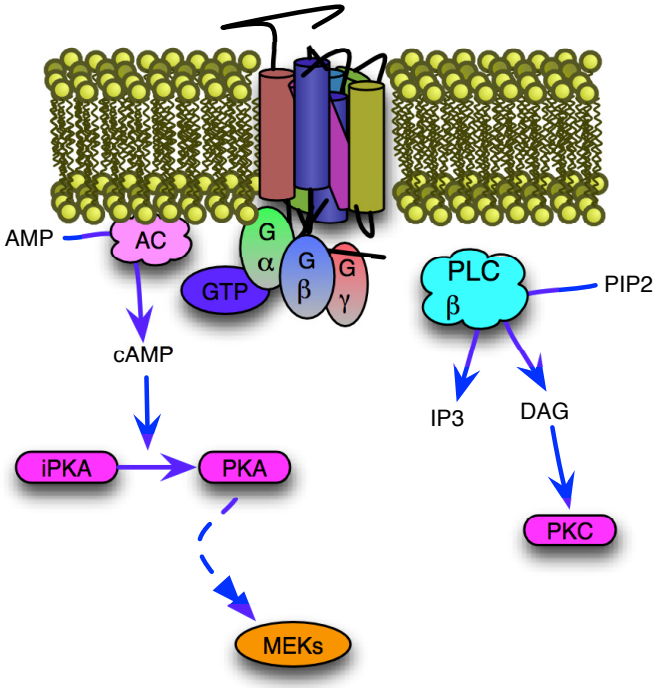

Figure 1.3. Differential G protein coupling defines the type of secondary messenger and activates different intracellular pathways. In the figure two major pathways activated by GPCRs are illustrated.

and the crosstalk with different signaling pathways can lead to variable consequences.

Being as popular cancer research systems since 1980s, SSTRs and their peptide ligands or agonists are also known for their anti-proliferative effects. Protein tyrosine phosphatases (PTP) were described as the major effectors in anti-proliferative effects of SSTRs. For instance, SSTR2 was shown to activate SHP-1, resulting in cell growth arrest [38]. With the activation of SSTR1 SHP-2 and c-src were recruited, and Ras and MAPK activity were inhibited [39]. In another evidence, SSTR2 inhibited activity of phosphatidyl inositol 3 kinase (PI3K) thus reduced activation of AKT which originally facilitates proliferation and survival [40]. SSTR5-stimulated phospholipase C pathway was able to modulate retinoblastoma tumor suppressor protein (Rb) and p21 [41]. In addition to their anti-proliferative effects, SSTR3 and SSTR2 activations were shown to modulate p53-dependent and independent apoptotic pathways respectively, in human pancreatic adenocarcinoma [42]. Finally their analogs (ex: ocreotide) were indirectly linked to regulation of vascular endothelial growth factor (VEGF) and basic fibroblast

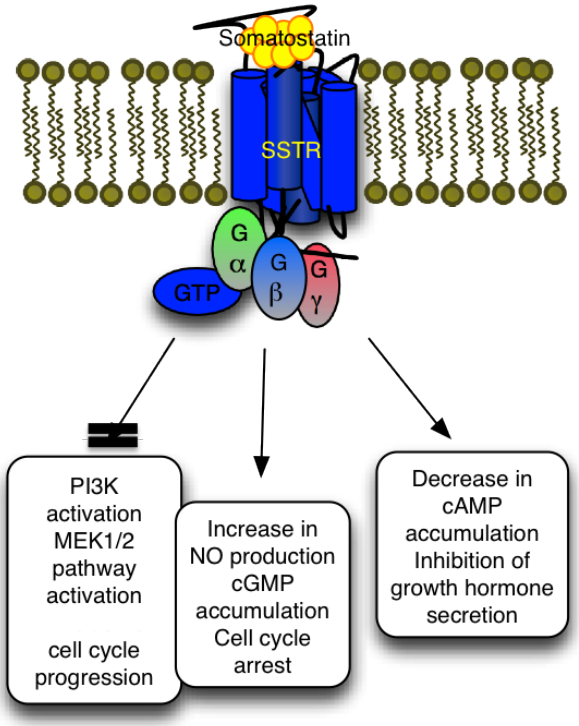

Figure 1.4. SSTR activation stimulates/inhibits activation of various secondary messenger pathways, resulting in various cellular responses.

growth factor (bFGF) secretion, so to angiogenesis [43]. All of these evidences and years of study about SSTRs resolved them as plausible means of cancer research and drug design studies [44–46].

1.2.4. Common Structural Features of GPCRs

GPCRs represent more than 800 proteins in humans and they are still denoted as "the largest *terra incognita*" for structural biologists [47].

The first GPCR which was solved via X-ray crystallography in 2000, was bovine rhodopsin [48]. Ligand binding pocket for 11-*cis* retinal ligand of rhodopsin and the conformational changes between active and inactive states of the receptor were discovered. Then opsin structure was obtained in 2008 [49]. Activation states were observed in more detail and the "ionic lock" between R135 and E247 was defined as the broken interaction during activation. The most conserved residues in TM1 are N55 and in

TM2 D83 and the NPxxY motif in TM7. The homology of rhodopsin to the other members of family A GPCRs is less than 25%, making it difficult for homology modeling at first. In addition, it is hard to understand the activation processes of the other GPCRs by rhodopsin, because it is activated by photons of light meaning no activating ligand-binding to the pocket [36, 50].

Still there are some common structural features of GPCRs defined in rhodopsin-like family. For instance, extracellular (EC) part is known to be responsible for ligand binding, while the intracellular (IC) part is responsible for G protein or arrestin interactions. Some conserved sequence motifs are DRY in TM3 forming the ionic lock, WxP motif in TM6 and NPxxY motif in TM7. Extracellular loops (ECLs) are less conserved than intracellular loops (ICLs). One common pattern on EC region is the disulfide bridge between ECL2 and TM3. This ECL has diverse features within GPCR families. For instance the β-hairpin structure in ECL2 of rhodopsin forms a closed pocket for the retinal ligand while the binding pockets in ECLs of other GPCRs are larger and more accessible for ligands, even if they contain a β-hairpin as in C-X-C chemokine receptor type 4 (CXCR4). In Palczewski's structure there is a β-hairpin on ECL2 of rhodopsin. This structure forms a pocket-like intrusion for retinal ligand binding and the N-terminus loop of the receptor behaves like a shelter on the ligand. This feature is supposed to be specific for rhodopsin GPCR only. In the same structure, there is a disulfide bridge between TM3 and ECL2, which is very common within other GPCR structures. This loop is proved to be responsible for opening of rhodopsin during activation [48]. The binding pockets of GPCRs which are activated by the same ligand show at least 50-60% identity. And these tend to show rigidity. On the other hand, ICL1 show similar backbone structure and also the same leucine residue at the same position. ICL2 and ICL3 are variable in length and sequence. ICL3 is proposed to interact with variable G proteins and arrestins or other downstream messengers. Some GPCRs have 8th helix which has a common F(RK)xx(FL)xxx(LF) motif. And the C-terminus of many GPCRs contain a palmitoylation site which plays role in anchoring of this terminus to the membrane again [47].

1.3. Bioinformatics in Structural Studies

With the advance in X-ray crystallography and nuclear magnetic resonance (NMR) techniques, structure based studies became popular in many areas such as drug design, target assessment for proteins, mutagenesis and also function determination for novel proteins. The largest structure database is ProteinDataBank (PDB). By July 2011, it contained 74140 structural entries, 68651 of which were proteins, peptides and viruses. The major part of these depositions are generated by X-ray crystallography and NMR (updates can be found at `http://www.rcsb.org/pdb/static.do?p=general_information/news\_publications/index.html`). Since July 2010, 20 protein or peptide structures were identified and deposited in PDB, per day. Despite the rate of structure depositions, sequence databases are growing much faster. This is mostly because there are still limitations in these structure determination techniques, such as some proteins are difficult to crystallize for X-ray or some are too big for NMR studies. Especially GPCRs are instable outside the membrane and are very difficult to crystallize. Consequently, *in silico* structure prediction and docking tools are getting more important in the field of protein structural biology, because they are quite cheaper and faster.

1.3.1. Protein Threading

Computational structure prediction tools can be grouped in three: homology (comparative) modeling, protein threading and *ab initio* methods [51, 52]. All of these methods are based on the idea that evolutionary related sequences introduce similar 3D structures. Thus, functionally similar patterns on the proteins are more conserved than their sequences. Both homology modeling and protein threading are template-based methods which have common major steps: i) the protein of unknown structure (target) is searched for homologous proteins which have experimental structural data (template) ii) target-template alignment is performed iii) the residues of the target protein was substituted into the coordinates of the aligned template residues and a 3D model is built for the target iv) the model accuracy is evaluated and model assessment is performed [53].

The sequence identity between target and the template proteins is the most important factor in template-based modeling. The identity of $\geq$30% can be predicted like that the target and the template are evolutionarily closely related and can share identical 3D conformations. Thus, the basic homology modeling via multiple sequence alignments and pairwise sequence alignments will be adequate for the best model. The results are sufficient enough to facilitate structure-based drug design studies. If the identity is below 15% the modeling process becomes more complicated, *ab initio* methods are required and the model produced cannot be reliable for drug design studies. When the identity is between 15% and 30% the low identity regions of the target can be recognized by threading methods. The models obtained are sufficient for site-directed mutagenesis experiments and assignment of protein functions [54].

Iterative Threading Assembly Refinement (I-TASSER) approach generated in Zhang Lab is a composite of these methods [55, 56]. It combines threading methods with *ab initio* methods, making the approach reliable with templates of low identity. It is based on 3 major steps: threading, fragment assembly and iteration (Figure 1.5). In threading step; the target sequence is predicted for secondary structures and aligned in profile-profile alignment, unlike the sequence-sequence alignment in homology modeling. At the beginning, evolutionary related structures are searched via PSI-BLAST and a profile is built predicting the secondary structure by PSIPRED. This profile and secondary structure predictions are threaded in LOMETS which uses combinations of different threading programs. The templates are evaluated and top hits are selected for assembly. However in homology modeling programs, generally only one template is selected for one target sequence. In fragment assembly step, aligned regions are taken from the templates and the unaligned regions are predicted via *ab initio* simulations. The continuous fragments are assembled by relative reorientation depending on energy calculations. These fragment decoys are clustered by Monte Carlo simulation and SPICKER, in order to identify low free-energy states. In the iteration step, cluster centroids are taken and fragment assembly is performed again. Structural restraints coming from PDB structures, from cluster centroid structures or from profile-to-profile alignments are evaluated and lowest energy structures are reselected. The resulting Cα backbone is given to REMO for addition of all atoms to the model [55, 56].

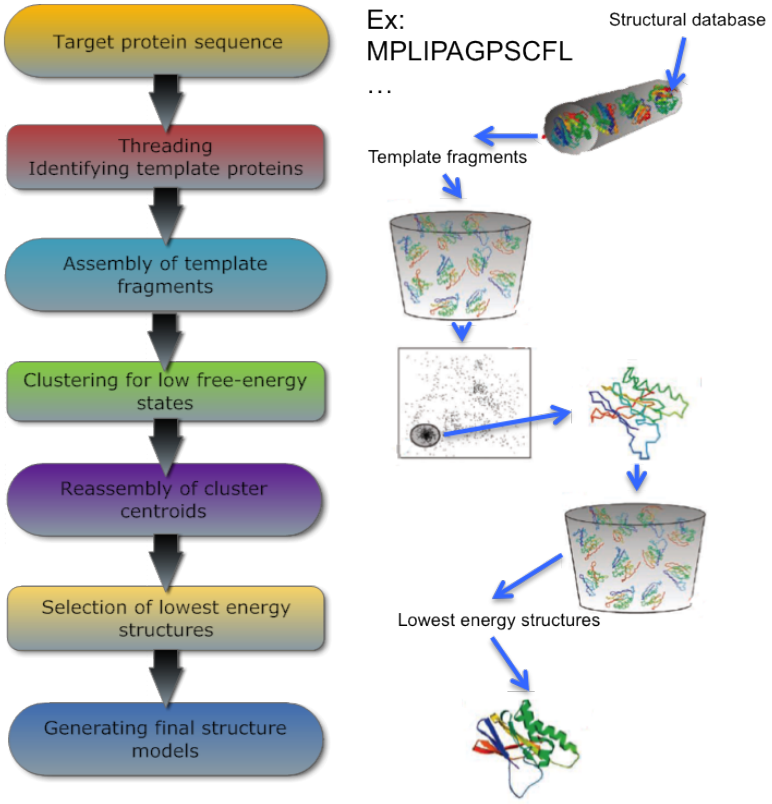

Figure 1.5. The main steps of I-TASSER is illustrated. These are: threading, fragment assembly and iteration which cover reassembly and model refinement.

As an advance of development of threading methods, full-length modeling of protein sequences with more reasonable structures and more reliable resolution values (root mean square deviation calculations) helps for further acceleration in drug design, functional and mutagenesis studies.

1.3.2. Molecular Docking

In drug discovery studies, usually molecular docking of the target protein comes after structural modeling. Proteins, ligands, nucleic acids, chemicals and even carbohydrates and lipids can be docked in *in silico* softwares. These programs predict relative orientations of molecules, interacting residues, strength of association and binding affinities, and try to find the best-fit conditions. They are generally used between a small molecule and a bigger one, such as in ligand-receptor docking. The most cited docking tools include AutoDock from The Scripps Research Institute, DOCK from UCSF [57], FlexX (BioSolveIT) [58], GOLD (CCDC) [59] and ICM (Molsoft) [60]. All

softwares or servers for docking are difficult to compare, so that the researcher should try and find the best docking program for the specific proteins of interest. The success of a docking program is generally interpreted in terms of root mean square deviations (RMSD). If the RMSD values are lower than $2\mathring{A}$ the docking would be successful [61]. In all of the protocols the ligand and the target proteins are prepared individually according the requirements of the program, such as identification of rotatable and rigid bonds, identification of binding region on the target, addition of Hydrogens or removal of solvent around the proteins [62, 63]. In docking step, the programs search the pre-defined sample space or make random changes in the ligand to find the limitations defined by the user. At last the program uses some scoring functions, mostly force field calculations, to define the best-fit docking prediction.

2. PURPOSE

The purpose of this thesis project is to identify a somatostatin-like receptor in *Carausius morosus* and to find its' ligand binding sites bioinformatically.

3. MATERIALS

3.1. General Kits, Enzymes and Reagents

General kits and other reagents used in the project and their companies are tabulated below:

Table 3.1. List of kits, enzymes and reagents.

Name	**Company**
High Pure RNA Isolation Kit	Roche Applied Science, USA
RNeasy MinElute Cleanup Kit	QIAGEN, Hilden, GERMANY
DNase I (Recombinant)	Roche Applied Science, USA
RNase A	Fermentas, CANADA
Trizol Reagent	Invitrogen, USA
High Pure PCR Product Purification Kit	Roche Applied Science, USA
5'/3' RACE Kit 2nd Generation	Roche Applied Science, USA
Taq DNA Polymerase	Fermentas, CANADA
QIAquick Gel Extraction Kit	QIAGEN, Hilden, GERMANY
T4 DNA Ligase	New England Biolabs Inc., MA, USA.
QIAprep Miniprep Kit	QIAGEN, Hilden, GERMANY
Paq5000 Hotstart DNA Polymerase	Agilent Technologies, CA, USA
XbaI and EcoRI Restriction Enzymes	Fermentas, CANADA
FuGENE HD Transfection Reagent	Promega, WI USA
Trypsin	HyClone, CANADA
1X PBS	HyClone, CANADA
HA monoclonal antibody	Pierce, CANADA
Goat anti-mouse IgG, FITC conjugate	Pierce, CANADA

All laboratory chemicals were analytical grade from Sigma (St. Louis, MO, USA)

and Merck (Schucdarf, Germany), unless stated otherwise in the text. DNA molecular weight markers and deoxyribonucleotides were purchased from Fermentas (Burlington, Canada).

3.2. Plasmids

pGEMT Easy vector was commercially obtained from Promega, Wisconsin, USA. pcDNA3 was commercially obtained from Invitrogen, CA, USA. pEGFP-N2 ws obtained commercially from Clontech, CA, USA. pcDNA3/HA vector was provided as previously modified by Nesrin Özören, Boğaziçi University, TURKEY.

3.3. Bacteria and Cell Lines

Bacterial strain used in this study was *E. coli* TOP10 (genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG).

Huh7 cell lines were provided by Mehmet Öztürk, Bilkent University TURKEY.

3.4. Peptides

Peptides and hormones such as Allatostatin IV and Insulin were obtained commercially from Sigma-Aldrich MO, USA. Sequence of Allatostatin IV peptide is DRLYSFGL-amide.

3.5. Solutions and Media

3.5.1. DNA Gel Electrophoresis

Table 3.2. Preparation of DNA gel electrophoresis buffer

Solution	**Content**
50X Tris-acetic acid EDTA (TAE)	2M Tris-acetate 50mM ethylenediaminetetraacetic acid pH 8.5
TE Buffer	10mM Tris-HCl 1mM EDTA, pH 8.0
Ethidium bromide (EtBr)	10 mg/ml
10X Tris Base EDTA (TBE)	108 g Tris base 55 g Boric acid 9.3 g EDTA Distilled water up to 1 L
Loading buffer	For 10ml: 2.4 ml dH_2O 0.1 ml 1M Tris-HCl, pH 7.6 0.3 ml 1% Bromophenol Blue (BPB) 6 ml 100% glycerol 1.2 ml 0.5M EDTA

3.5.2. Nuclease-Free Water

Table 3.3. Preparation of nuclease free water

Content	**Procedure**
1 ml 0.1% (v/v) Diethylpyrocarbonate	
1000 ml distilled water	let for 1h at RT autoclave cool down to RT

3.5.3. RNA Gel Electrophoresis Buffers

Table 3.4. Preparation of RNA buffers

Solution	**Content**
Diethylpyrocarbonate treated	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 1L. pH 7.00
EtBr RNA loading buffer	0.72 ml formamide 0.16 ml 10X MOPS 0.26 ml formaldehyde 0.18 ml DEPC treated water (DPH) 0.1 ml 80% glycerol 0.08 ml Bromophenol blue 50 μg EtBr

3.5.4. Bacterial Media

Table 3.5. Contents of bacterial medium

Solution	**Content**
Luria-Bertani medium (LB)	10 g tryptophan 5 g yeast extract 10 g NaCl Distilled water up to 1 L, autoclaved
Ampicillin stock	10 g tryptophan 5 g yeast extract 5 g NaCl 15 g Agar Distilled water up to 1 L, autoclaved

3.5.5. Complete Growth Media

Table 3.6. Contents of complete DMEM

Solution	**Content**
Complete DMEM	DMEM, Gibco, USA 10% FBS, Gibco, USA 1X Penicillin-Streptomycin, Gibco, USA

3.6. Primers

Primers used in RACE and PCR reactions, sequencing and cloning were purchased from Harvard University MGH DNA Sequencing Core (Boston, USA).

Table 3.7. Primers

Primer Name	**Sequence**	**Restriction site**
SP1	GAAGCGGAGCACCACGTAGAT	
SP2	GAGGCCGTTGCCGAGCAG	
SP3	CACGGATCCGACGATGCACACGGCCGCGTA	
SP4	CTGCACATGGAGAACAGCGTG	
SP5	AGGGAATTCTTCCCGCGGCGCCACCACCGG	
SSTR-ATG	AAACTAGTCAAAATGTCTGTGGAACAAGTGACG	SpeI
SSTR11 (without Stop codon)	TTTGAATTCTTGGATCCTCACCTGGGTCGGCTG	EcoRI, BamHI
SSTR12 (with Stop codon)	TTTGAATTCTTGGATCCTCTACACCTGGGTCGGCTG	EcoRI, BamHI
universal SP6	ATTTAGGTGACACTATAG	
universal T7	TAATACGACTCACTATAGGG	
actin57b-for	CCAGATCATGTTTGAGACCT	
actin57b-rev	TCTTCATCAGGTAGTCGGTC	

3.7. Web Servers and Softwares

Table 3.8. List of web servers and softwares used.

Server/Software	**Web Site**	**Company**
AutoDock Vina v. 1.1.2		Molecular Graphics Lab, The Scripps Research Institute, USA
EasyPRED	`http://www.cbs.dtu.dk/biotools/EasyPred/`	CBS, DENMARK
I-TASSER	`http://zhanglab.ccmb.med.umich.edu/I-TASSER/`	Zhang Lab, UMICH, USA
MEGA5		Biodesign Institute, USA
MGL Tools		Molecular Graphics Lab, The Scripps Research Institute, USA
ModWeb	`https://modbase.compbio.ucsf.edu/scgi/modweb.cgi`	Sali Lab, UCSF, USA

MolSoft Browser Pro		Molsoft LLC, USA
MotifScan	http://myhits.isb-sib.ch/cgi-bin/motif_scan	MyHits, SIB, SWISS
PDBsum	http://www.ebi.ac.uk/pdbsum/	EMBL EBI, ENGLAND
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	UCL, ENGLANG
Quantity One		BIO-RAD, ITALY
ScanProsite	http://prosite.expasy.org/scanprosite/	Expasy, SIB, SWISS
TMHMM, v. 2.0	http://www.cbs.dtu.dk/services/TMHMM/	CBS, DENMARK
VMD		University of Illinois, USA
WhatIF	http://swift.cmbi.ru.nl/whatif/	RUMC, NETHERLANDS

3.8. Equipment

Table 3.9. List of equipments used.

Equipment	**Company**
Atomic Force Microscope	Dimension Edge Atomic Force Microscope System, Bruker
Autoclave	Midas 55, Prior Clave, UK
Balances	DTBH 210, Sartorius, GERMANY Electronic Balance VA 124, Gec Avery, UK
Cantilevers	Bruker AFM Probes, Bruker
Carbon dioxide tank	2091, Habaş, TURKEY
Cell culture incubators	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 Microfuge, Cole Parmer, Illinois, USA
Deepfreezers	−20°C, Arçelik, TURKEY −70°C Freezer, Harris, UK −86°C ULT Freezer, ThermoForma, USA
Documentation System	Gel Doc XR System, Bio-Doc, ITALY
Electrophoresis System	BIORAD Mini-Sub Cell GT, USA
Heat blocks	DRI-Block DB-2A, Techne, UK
Homogenizer	MagNa Lyser, Roche, USA MagNa Lyser Green Beads, Roche, USA
Laminal flow cabinet	Labcaire BH18, UK
Magnetic Stirrers	M221 Elektro-mag, TURKEY Clifton Hotplate Magnetic Stirrer, HS31, UK
Micropipettes	Finnpipette, Thermo, USA
Microscopes	Inverted Microscope, CKX41, Olympus, JAPAN Fluorescence Microscope, Observer.Z1, Zeiss, GERMANY Confocal microscope, True Confocal TCS, Leica, GERMANY
Microwave oven	M1733N, Samsung, MALAYSIA
pH meter	WTW, GERMANY
Pipettor	Pipetus-akku, Hirschmann Laborgeräte, 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	Quantity One, Bio-Rad, ITALY
Spectrophotometer	Agilent 8453, USA NanoDrop 1000, USA
Thermocyclers	AB Applied Biosystems 2720 Thermal Cycler, USA
Vacuum pump	BIORAD DNA Engine Peltier Thermal Cycler, USA KNF Neuberger, USA
Vortex	Vortexmixer VM20, Chiltern Scientific, UK
Water baths	TE-10A, Techne, UK
Water purification	WA-TECH, GERMANY

4. METHODS

4.1. Identification of The Receptor

4.1.1. RNA Extraction from Insect Tissue

The adult insects were killed in liquid nitrogen and total bodies were stored at -80°C. 66 mg Stick insect tissue was measured and taken into a clean eppendorf tube (1,5 mL). 1 mL Trizole reagent (Invitrogen) was added. The mixture was taken into tubes containing MagNa Lyser Green Beads (Roche) and homogenization was performed in MagNa Lyser (Roche) for 60 sec at 6500 rpm. The homogenate was taken into a clean eppendorf tube. 200 μl chloroform was added and mixed by inverting. The tube was incubated at room temperature (RT) for 15 min. The tube was centrifuged at 12000 g for 15 min at +4°C. The aqueous upper phase was taken into a clean eppendorf tube. 500 μl isopropanol was added and mixed by inverting. The sample was incubated at RT for 10 min. The tube was centrifuged at 12000 g for 10 min at +4°C. Supernatant was discarded. 1 ml 70% EtOH was added and mixed by vortexing until the pellet dissociated from the tube wall. The tube was centrifuged at 7500 g for 5 min at +4°C. Supernatant was discarded. 1 ml 70% EtOH was added again and mixed by vortexing until the pellet dissociated from the tube wall. The tube was centrifuged at 7500 g for 5 min at +4°C. Supernatant was discarded. The pellet was left for 5 min to dry. Dried pellet was resuspended in 50 μl DEPC water.

Total RNA concentration was measured in nanodrop. DNase treatment was performed as follows; maximum 50 μg RNA was mixed with 5 μl DNase Incubation Buffer of High Pure RNA Isolation Kit (Roche), 1 μl DNase I (recombinant) (Roche) and filled with DEPC water until the total volume was 50 μl. This mixture was incubated at 37°C for 30 min. 350 μl RLT buffer of RNeasy MinElute Cleanup Kit (QIAGEN) and 1% β-mercaptoethanol (βME) were added and mixed. 350 μl 70% EtOH was added and mixed by inverting. The solution was taken onto RNeasy MinElute Spin Column (QIAGEN) in a collection tube. The column was centrifuged at 10000 g for 1 min.

The flow-through was discarded. 700 μl RW1 buffer (QIAGEN) was added and the tube was centrifuged at maximum speed for 1 min. The flow-through was discarded. 500 μl RPE buffer (QIAGEN) was added and the tube was centrifuged at maximum speed for 1 min. The flow-through was discarded. This step was repeated again and the flow-through was discarded. The column was taken into a clean eppendorf tube. 50 μl nuclease free water was added to the middle of the column and the tube was centrifuged at maximum speed for 1 min. Total RNA concentration was measured in nanodrop.

Both the RNA samples before DNase treatment and after DNase treatment were checked in electrophoresis as follows; 2 μg RNA sample was mixed with 2X loading dye (Fermentas) and nuclease free water until the total volume was 20 μl. This sample was incubated at 95°C for 5 min and loaded onto MOPS gel. Electrophoresis was performed at 90V for 30 min.

4.1.2. cDNA Synthesis for RACE Reactions

First strand cDNA synthesis for 3' RACE was performed by 5'/3' RACE Kit 2nd Generation (Roche). cDNA synthesis buffer (5X), dNTP mix (each 10mM), oligo dT-anchor primer, 1 μg RNA and Reverse Transcriptase enzyme were mixed in the required amounts. Double-distilled water was added until the total volume was 20 μl. The mixture was mixed and spinned down. The tube was incubated at 55°C for 1 h. Enzyme inactivation was performed at 85°C for 5 min.

First strand cDNA synthesis for 5' RACE was performed with the same kit. cDNA synthesis buffer (5X), dNTP mix (each 10mM), SP1 primer (12.5 μM), 1 μg RNA and Reverse Transcriptase enzyme were mixed in the required amounts. Double-distilled water was added until the total volume was 20 μl. The mixture was mixed and spinned down. The tube was incubated at 55°C for 1 h. Enzyme inactivation was performed at 85°C for 5 min. Secondly the cDNA obtained was purified with High Pure PCR Product Purification Kit (QIAGEN) as follows. 100 μl of Binding Buffer was added onto cDNA sample and mixed well. The sample was added to the High

Pure Filter Tube inside a collection tube. The tube was centrifuged at 8.000 g for 30 sec in a table-top centrifuge. The flow-through was discarded and 500 μl Wash Buffer was added onto the filter tube. The tube was centrifuged at 8.000 g for 30 sec. The flow-through was discarded and 200 μl Wash Buffer was added onto the filter tube. The tube was centrifuged at maximum speed for 2 min. The flow-through was discarded and the filter tube was taken into a sterile eppendorf tube (1.5 ml). 50 μl Elution Buffer was added to the middle of the filter tube and the tube was centrifuged at 8.000 g for 30 sec. The filter tube was discarded and the purified cDNA sample for 5' RACE was stored at -20°C. Thirdly, poly(A) tailing of this cDNA was performed according to the protocol in 5'/3' RACE Kit 2nd Generation (Roche). 19 μl purified cDNA was mixed with 2.5 μl of 10X Reaction Buffer and 2.5 μl of 2 mM dATP in a sterile eppendorf tube. The mixture was incubated for 3 min at 94°C and immediately chilled on ice. The tube was spinned down and 1 μl of Terminal Transferase enzyme (80 U/μl) was added. The mixture was incubated at 37°C for 20 min. The enzyme inactivation was performed at 70°C for 10 min. This dA-tailed cDNA was immediately taken on ice and stored at -20°C.

4.1.3. 3' and 5' RACE

PCR Amplification for 3': 3' amplification of cDNA was performed as an ordinary nested PCR reaction. The buffers of Taq DNA Polymerase (recombinant, Fermentas) were used. In the first round of PCR for 3' region; 1 μl of unpurified first strand cDNA synthesized for 3' RACE was used. In a thin-wall PCR tube 10X Taq buffer with KCl, 25 mM $MgCl_2$, PCR anchor primer from 5'/3' RACE Kit 2nd Generation (Roche), SP4 primer, Taq DNA polymerase, dNTP mix and distilled water were mixed according to the amounts given below. cDNA was added and mixed by the tip of pipette. The mixtures were spinned down and taken into ThermoCycler. The reaction conditions were adjusted as given below.

Table 4.1. 3' RACE first PCR reaction reagents and amounts.

Reagent	Final Concentration	Amount
cDNA		1 μl
10X Reaction buffer	1X	5 μl
25 mM MgCl_2	1.25 mM	2.5μl
12.5 μM oligo dT-anchor primer (Roche)	0.25 μM	1 μl
10 μM SP4 primer	0.25 μM	1.25 μl
dNTP mix (10 mM each)	0.2 mM	1 μl
5U/μl Taq DNA polymerase	2.5 U	0.5 μl
distilled water		37.75 μl
TOTAL		50 μl

Table 4.2. 3' RACE first reaction conditions.

	Temperature (°C)	Duration	Number of cycles
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	10
Annealing	55	30 sec	
Elongation	72	40 sec	
Denaturation	94	15 sec	25
Annealing	57	30 sec*	
Elongation	72	40 sec	
Final Elongation	72	7 min	1

* Each cycle was elongated by 20 sec.

In the second round of PCR for 3' region; 1 μl of PCR product of first round of 3' PCR was used. In a thin-wall PCR tube 10X Taq buffer with KCl, 25 mM MgCl_2, PCR anchor primer from 5'/3' RACE Kit 2nd Generation (Roche), SP5 primer, Taq DNA polymerase, dNTP mix and distilled water were mixed according to the optimum amounts given below. PCR product of first round of 3' was added and mixed by the tip

of pipette. The mixtures were spinned down and taken into ThermoCycler. Optimizations for obtaining specific annealing were tried both for the reagent concentrations and for thermo-cycling conditions. The optimum reaction conditions were adjusted as given below.

Table 4.3. 3' RACE second PCR reaction reagents and amounts.

Reagent	Final Concentration	Amount
1/10 diluted 1st PCR product		1 μl
10X Reaction buffer	1X	5 μl
25 mM $MgCl_2$	1.25 mM	2.5 μl
12.5 μM PCR anchor primer (Roche)	0.25 μM	1 μl
10 μM SP5 primer	0.25 μM	1.25 μl
dNTP mix (10 mM each)	0.2 mM	1 μl
5U/μl Taq DNA polymerase	2.5 U	0.5 μl
distilled water		37.75 μl
TOTAL		50 μl

Table 4.4. 3' RACE second reaction conditions.

	Temperature (°C)	Duration	Number of cycles
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	10
Annealing	55	30 sec	
Elongation	72	40 sec	
Denaturation	94	15 sec	25
Annealing	59	30 sec*	
Elongation	72	40 sec	
Final Elongation	72	7 min	1

* Each cycle was elongated by 20 sec.

5' RACE Reaction: For the PCR amplification of 5', dA-tailed cDNA was amplified with Taq DNA Polymerase (recombinant, Fermentas). In the first round of PCR

for 5' region; 10X Taq buffer with KCl, 25 mM $MgCl_2$, Oligo dT-anchor primer from 5'/3' RACE Kit 2nd Generation (Roche), SP2 primer, Taq DNA polymerase, dNTP mix and distilled water were mixed according to the amounts gven below, in a thin-wall PCR tube. dA-tailed cDNA was added and mixed by the tip of pipette. The mixtures were spinned down and taken into ThermoCycler. The reaction conditions were adjusted as given below.

Table 4.5. 5' RACE first PCR reaction reagents and amounts.

Reagent	Final Concentration	Amount
cDNA		1 μl
10X Reaction buffer	1X	5 μl
25 mM $MgCl_2$	1.25 mM	2.5μl
12.5 μM oligo dT-anchor primer (Roche)	0.25 μM	1 μl
12.5 μM SP2 primer	0.25 μM	1 μl
dNTP mix (10 mM each)	0.2 mM	1 μl
5U/μl Taq DNA polymerase	2.5 U	0.5 μl
distilled water		38 μl
TOTAL		50 μl

Table 4.6. 5' RACE first reaction conditions.

	Temperature (°C)	Duration	Number of cycles
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	10
Annealing	55	30 sec	
Elongation	72	40 sec	
Denaturation	94	15 sec	10
Annealing	60	30 sec*	
Elongation	72	40 sec	
Final Elongation	72	7 min	1

* Each cycle was elongated by 20 sec.

In the second round of PCR for 5' region; 1 μl product of first round of 5' PCR was used. In a thin-wall PCR tube 10X Taq buffer with KCl, 25 mM MgCl_2, PCR anchor primer from 5'/3' RACE Kit 2nd Generation (Roche), SP5 primer, Taq DNA polymerase, dNTP mix and distilled water were mixed according to the optimum amounts in Table. PCR product of first round of 5' was added and mixed by the tip of pipette. The mixtures were spinned down and taken into ThermoCycler. Optimizations for obtaining specific annealing were tried both for the reagent concentrations and for thermo-cycling conditions. The optimum reaction conditions were adjusted as in Table 4.8.

Table 4.7. 5' RACE second PCR reaction reagents and amounts.

Reagent	Final Concentration	Amount
1st PCR product		1 μl
10X Reaction buffer	1X	5 μl
25 mM MgCl_2	1.25 mM	2.5μl
12.5 μM PCR anchor primer (Roche)	0.25 μM	1 μl
12.5 μM SP3 primer	0.25 μM	1 μl
dNTP mix (10 mM each)	0.2 mM	1 μl
5U/μl Taq DNA polymerase	2.5 U	0.5 μl
distilled water		38 μl
TOTAL		50 μl

Table 4.8. 5' RACE second reaction conditions.

	Temperature (°C)	Duration	Number of cycles
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	
Annealing	60	30 sec	10
Elongation	72	40 sec	
Denaturation	94	15 sec	
Annealing	65	30 sec*	10
Elongation	72	40 sec	
Final Elongation	72	7 min	1

* Each cycle was elongated by 20 sec.

4.1.4. PCR Product Purification

The second PCR products from both 3' and 5' reactions were run in 1% Agarose gel electrophoresis for 30 min at 90 V. The defined fragments were cut and taken into clean eppendorf tubes. The amplified fragments were extracted from the gel with QIAquick Gel Extraction Kit (QIAGEN). The gel slices were weighed. 1 mg gel weight was supposed as 1 μl of gel volume. 3 volumes of Buffer QG was added onto each volume of gel slice and the tube was incubated at 50°C for 10 min. After dissolution of gel, 1 gel volume of isopropanol was added and mixed. The sample mix was applied into the QIAquick spin column and centrifuged at maximum speed for 1 min. The flow-through was discarded and 0.75 ml Buffer PE was added onto the column. The tube was centrifuged and the flow-through was discarded. The column inside the collection tube was centrifuged for additional 1 min. The column was taken into a sterile eppendorf tube. 20 μl prewarmed (50°C) Buffer EB was added on the middle of column. The tube was left for 1 min at RT and centrifuged at maximum speed for 1 min. These extracted PCR fragments were measured for their concentrations in nanodrop and stored at -20°C.

4.1.5. Preparation for Sequencing

Cloning onto pGEM-T Easy: The purified fragments were ligated onto pGEM-T Easy vector (Promega) as follows. All of the reagents were thawed and spinned down. 5 μl of 2X Rapid Ligation Buffer, 0.5 μl of pGEM-T Easy Vector (25 ng) and 3.5 μl of purified PCR fragment were mixed in a 0.5 eppendorf tube. 1 μl T4 DNA ligase (3 U/μl) was added and mixed by pipetting. The mixture was incubated at RT for 1 h. The ligation products were transformed into CaCl_2 competent TOP10 bacteria. 50 μl competent bacteria was mixed with 10 μl of ligation reaction and incubated in ice for 50 min. Heat shock was performed for the bacteria at 42°C for 45 sec. The transformants were again taken into ice for 2 min. 1 ml LB was added in sterile condition and the transformants were incubated at 37°C for 1 h in a shaker (200 rpm). 200 μl of the transformants were spread onto Ampicillin plate and left for O/N incubation at 37°C. The colonies were selected and used for colony PCR.

In a 1.5 ml eppendorf tube, PCR mix for colony PCR was prepared. 10X Taq buffer with KCl, 25 mM MgCl_2, universal T7 primer, universal SP6 primer, Taq DNA polymerase, dNTP mix and distilled water were mixed according to the amounts in Table. The mixture was divided into thin-wall PCR tubes, as each will contain 25 μl mixture. Colonies were picked with pipette tips and added into each PCR tube. The mixtures were spinned down and taken into ThermoCycler. The reaction conditions were adjusted as in the Table. The products were run in 1% Agarose gel electrophoresis. The colonies that contained ligated inserts were selected and grown in ampicillin containing 5 ml liquid LB at 37°C O/N.

Plasmid Isolation and Sequencing: Plasmids were isolated with QIAprep Miniprep Kit (QIAGEN). The bacteria tube was centrifuged at 13000 rpm for 5 min. The supernatant was discarded and the cells were resuspended in 250 μl Buffer P1. The suspension was transferred into an eppendorf tube and 250 μl P2 was added and mixed by inverting. 350 μl Buffer N3 was added and mixed by inverting immediately. The tube was centrifuged at 13000 rpm for 10 min in a table-top microcentrifuge. The supernatant was applied onto QIAprep spin column and centrifuged for 1 min. The

flow-through was discarded. 0.75 ml Buffer PE was added onto the column and centrifuged for 1 min. The flow-through was discarded and centrifuged for additional 1 min. The column was put into a clean eppendorf tube. 50 μl prewarmed (50°C) Buffer EB was applied to the center of column and left for 1 min at RT. The column was centrifuged for 1 min. The extracted plasmids were stored at -20°C and measured for concentrations in nanodrop. Samples were selected and sent to Macrogen Korea for sequencing with unversal primers SP6 and T7.

4.1.6. PCR Amplification of the Open Reading Frame

The fragment sequences from 3' and 5' RACE reactions and sequence of TM region were put in the order and the longest open reading frame was predicted in a six-frame translation web server. New primers were designed for the first ATG codon (SSTR-ATG) and the Stop codon (SSTR12). SpeI restriction site was added before ATG codon and EcoRI site was added after Stop codon (Figure 3.9). Another primer (SSTR11) was designed without Stop codon. For the PCR amplification of open reading frame (ORF) cDNA sample prepared for 3' RACE reactions was used. All the PCR reagents were from Paq5000 Hotstart DNA Polymerase (Agilent Technologies). In a thin-wall PCR tube 10X Paq buffer with $MgCl_2$, SSTR-ATG, SSTR11/SSTR12, Paq DNA polymerase, dNTP mix and distilled water were mixed according to the amounts given below. cDNA was added and mixed by the tip of pipette. The mixtures were spinned down and taken into ThermoCycler. In order to obtain specific single band of the predicted length amounts of $MgCl_2$, primers, cDNA sample and conditions such as the annealing temperature and cycle number were optimized. The optimum reaction conditions were adjusted as given below.

Table 4.9. PCR amplification reaction reagents and amounts for ORF.

Reagent	Final Concentration	Amount
cDNA		1 μl
10X Paq reaction buffer	1X	5 μl
10 μM SSTR-ATG	0.4 μM	2 μl
10 μM SSTR11 or SSTR12	0.4 μM	2 μl
dNTP mix (10 mM each)	0.2 mM	1 μl
Paq5000 DNA polymerase		0.5 μl
DMSO	5%	2 μl
distilled water		36.5 μl
TOTAL		50 μl

Table 4.10. PCR amplification conditions for ORF.

	Temperature (°C)	Duration	Number of cycles
Initial denaturation	94	2 min	1
Denaturation	94	15 sec	35
Annealing	57	30 sec	
Elongation	72	1 min	
Final Elongation	72	7 min	1

The PCR product was run in 1% Agarose gel electrophoresis for 30 min at 90 V. The fragment was cut from the gel and taken into a sterile eppendorf tube for gel extraction protocol. Gel extraction protocol was performed as mentioned above in PCR Product Purification section. The extracted PCR product was measured for their concentrations in nanodrop and stored at -20°C.

4.1.7. Cloning of the Open Reading Frame

The purified product and modified pcDNA3/HA vector were cut by XbaI (produces compatible flanking sequences with SpeI) and EcoRI enzymes (Fermentas). The

restriction reaction mixtures were prepared according to the Table. The mixture was incubated at 37°C for 2 h. Enzyme inactivation was performed at 65°C for 20 min. The cut samples were run in 1% Agarose gel electrophoresis for 40 min at 90 V. The restricted bands were cut from the gel and extractions were performed as mentioned above in "PCR Product Purification" section. The purified fragments were measured for their concentrations in nanodrop and stored at -20°C.

The restricted plasmid and insert was taken into ligation reaction. Protocol of T4 DNA Ligase (New England BioLabs) was performed for 1:3 fraction. 2 μl 10X T4 DNA Ligase Buffer, 100 ng plasmid, 60 ng ORF insert and distilled water until the total volume was 20 μl were mixed in an eppendorf tube. 1 μl T4 DNA ligase was added and mixed. The mixture was incubated at RT for 20 min. Enzyme inactivation was performed at 65°C for 10 min. And the product was diluted with 80 μl distilled water.

The ligation products were transformed into CaCl_2 competent TOP10 bacteria. 100 μl competent bacteria was mixed with 30 μl of ligation reaction and incubated in ice for 50 min. Heat shock was performed for the bacteria at 42°C for 45 sec. The transformants were again taken into ice for 2 min. 1 ml LB was added in sterile condition and the transformants were incubated at 37°C for 1 h in a shaker (200 rpm). Total of the transformants were spread onto Ampicillin plate and left for O/N incubation at 37°C. The colonies were selected and used for colony PCR. Colony PCR reactions were performed as mentioned above in "Preparation for Sequencing" section, except the primers were SSTR-ATG primer and SSTR11 or SSTR12 primers. The colony PCR products were run in 1% Agarose gel electrophoresis. The colonies that contained ligated inserts were selected and grown in ampicillin containing 5 ml liquid LB at 37°C O/N. *"Plasmid Isolation and Sequencing"* part of "Preparation for Sequencing" section was performed again for ORF ligated pcDNA3 plasmids. And the purified plasmids were sent to MacroGen Korea for sequencing of the full-length ORF of the receptor.

In order to facilitate overexpression studies in cell culture and because of the absence of commercial antibodies against this novel receptor, the ORF was tagged in

the same cloning procedure. New reverse primer, which was designed without Stop codon (SSTR11), was used. It contained the same EcoRI restriction site. The ORF amplified via ATG primer and SSTR11 was cloned onto modified pcDNA3/HA vector which contain an N-terminal HA epitope tag. The same cloning procedure was repeated for the HA-tagged version of the receptor.

4.2. *in vivo* Studies

4.2.1. General Cell Culture Techniques

In order to obtain efficient transfection and show the effects of CamAlstR on cell cycle progression in mammalian cells, Huh7 cell lines were chosen.

Huh7 cell lines are hepatocellular carcinoma cells which have enormous cell proliferation capacity and strong adhesive properties. These cell lines were grown in Dulbecco's Modified Eagle Medium (DMEM by Gibco) containing 10% Fetal Bovine Serum (FBS) and 1X Penicillin-Streptomycin (Gibco). The cells were incubated in 5% CO_2 conditioned 37°C incubator in 10 cm cell-culture Petri plates. All treatments were performed under a sterile hood and the equipments were wiped with 70% EtOH prior to use. The cells were passaged every 3-day before reaching 90% confluency, as follows. The medium was removed and the cells were washed with 1X PBS. PBS was removed and trypsin-EDTA solution was added onto cells. The plate was incubated at 37°C for 2 min. Complete DMEM was added onto the cells to inactivate trypsin and the cells were taken into a clean falcon tube. The tube was centrifuged at 600 g for 4 min. The supernatant was removed and the cells were resuspended in complete DMEM. This suspension was spread into a clean 10 cm plate and left for incubation in 37°C incubator.

Each cell line was stored in 10% DMSO containing complete growth medium at -80°C refrigerators.

4.2.2. Transfection

Overexpression studies were performed on Huh7 cell lines due to their high transfection capacity. The cells were thawed and passaged prior to transfection. In a sterile tube plasmid DNA and serum free medium was mixed. In a 6-well plate experiment up to 6 μg of plasmid DNA was used. Prewarmed (room temperature) FuGENE HD Transfection Reagent (Promega) was added in a 3:1 (reagent volume:DNA amount) ratio. At each step the mixture was vortexed and spinned down. Reagent-DNA mixture was incubated at RT for at most 15 min. This mixture was added onto the cells and the plate was swirled. The cells were left for incubation at 37°C for 24 h. HA-tagged CamAlstR was expressed in Huh7 cell lines as plasmid pcDNA3/HA-CamAlstR. Each transfection was verified for efficiency with an EGFP containing plasmid (pEGFP-N2) and for transfection toxicity with an empty plasmid (modified pcDNA3/HA).

4.2.3. Immunofluorescence

The cell growth and transfection were performed on sterile coverslips in the plate wells. Transfected cells were washed with PBS twice. Ice-cold 4% paraformaldehyde was added onto the cells and the plate was incubated for 20 min on a shaker in RT. Washing was performed with PBS three times for 5 min. Blocking was performed in 2% BSA solution for 30 min on a shaker in RT. The cells were incubated with HA antibody (Pierce) in a 1:300 ratio at +4 °C for overnight on a shaker. Washing was performed with PBS three times for 5 min. The cells were incubated in secondary antibody (FITC anti-mouse, 1/300 ratio) for 4 h at RT. Washing was performed in PBS three times for 10 min. The coverslips were taken onto clean slides and covered with PBS. These samples were observed under confocal microscope.

4.2.4. Atomic Force Microscopy

In order to define and measure an interaction between proteins, protein and a surface or proteins and ligands, atomic force microscopy (AFM) is used as the direct method in single molecule level since 1990s. It can detect an interaction as low as 1

pN force and 0.1 $\mathring{A}$ in vertical distance [64, 65]. The cantilever of AFM deflects laser when interaction occurs. In this thesis project, this method was used to define receptor ligand interaction on living cells.

Atomic force microscopy experiments were performed in BioAFM Lab of Hamdi Torun in Electrical and Electronical Engineering Department of Boğaziçi University. The cells were grown and transfected on sterile coverslips in 12-well plates. Experiments were performed with and without fixation of the cells. OTR4 silicon nitride cantilever (k = 80 pN/nm, Bruker) was incubated in 1 mg/ml Allatostatin IV peptide for 15 min at RT. A few drops of 1X PBS was added onto the cells and the sample coverslips were taken onto sample stage. Laser was positioned as it strikes the cantilever end, before the sample was placed. After sample placement, piezoelectric translator head was lowered in scanning tunneling microscope (STM) mode until a water meniscus was formed between peptide solution and PBS on the cells. Than the distance between cantilever tip and sample was adjusted to 1 mm via focusing of a single cell. The Sum signal was maximized here and the piezoelectric translator head was allowed to engage with a speed of 0.25 μm/step. Set point was adjusted for 0.3 V. On each sample 65 recordings were counted. The experiment was calibrated in an empty area with no cell. The control groups were set as a sample with no CamAlstR expression, a sample which cantilever penetrated the cell membrane and another sample washed with excess peptide to saturate CamAlstR receptors. The unbinding events were evaluated against non-adhesive events and these were taken for force calculations.

The AFM apparatus measured the laser deflection coming from cantilever tip as voltage units and we produced force-distance curves from potential-distance curves, according to the following formulas:

$$F = k\ \delta_z$$

Here, k is the spring constant of the cantilever and δ_z is the cantilever deflection change in z axis. This deflection was calculated by the following formula from calibration curve:

$$\delta_z = \mathrm{V}_{tot}\ 1/\mathrm{S}$$

Here, S is optical lever sensitivity and obtained from the slope of linear part of calibration curve on a stiff surface which is assumed to be "indefinitely hard". V_{tot} is the potential difference when the cantilever moves one step towards and against the sample [66]. Finally the force differences in unbinding events were assumed as the interaction forces between our receptor and its' ligand.

4.3. Bioinformatics

4.3.1. Identification of Protein Family

Protein sequences of Allatostatin receptors (AlstR) from different insect species were taken into evaluation. The ones that contained only partial protein sequences were excluded and only one receptor type was taken into account in one genus. The genera included in this analysis were *Apis* and *Nasonia* from Hymenoptera, *Aedes*, *Anopheles* and *Drosophila* from Diptera, *Tribolium* from Coleoptera, *Bombyx*, *Spodoptera* and *Manduca* from Lepidoptera, *Acyrthosiphon* from Hemiptera, *Daphnia* from Crustaceans, and *Periplaneta* from Dictyoptera. Sequences from 8 Allatostatin A receptors (AlstR-A) from 8 species, 7 Allatostatin B receptors (AlstR-B) from 7 species and 9 Allatostatin C receptors (AlstR-C) from 9 species were aligned with our receptor

at first and checked for the maximum coverage of similarity. The most conserved regions were included and the variable regions (mostly N-terminus and C-terminus) were deleted. Each type of AlstR was aligned between species in Biology Workbench 3.2 [67] and the resulting profiles were saved. These type-specific profiles were taken into profile-to-profile alignment via PAM matrix in ClustalWProf and saved. Our AlstR of *C. morosus* was aligned with the final profile of 24 proteins, in sequence-to-profile via ClustalWProf and the alignment was exported. The alignment file was edited in MEGA5 software [68]. The flanking regions of our sequence were deleted in order to obtain identical numbers of residues. The evolutionary relatedness was inferred by using Maximum Likelihood method based on the JTT matrix-based model with 500 bootstrap replica, in MEGA5. When the number of common sites was ¡ 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

4.3.2. Alignment of N-Termini

The first 54 residues of our receptor sequence were included in the analyses on N-terminus loop. At the beginning, 10 insect genera were taken into alignment with the total CamAlstR-C query, at most two genus in each order within Neoptera. However only 3 genera were included in ClustalW analysis. Seven genera (ex. *Solenopsis* species and *Bombus* species) were excluded from the multiple alignment because of the facts that more than one species did not include a related GPCR of known sequence, there were no functional homology with our sequence, coding mRNA was not known for the related GPCR or the alignment results of proteins were not significant with our sequence (Table 4.11). Two genera within the same order and another genus from another order were included in order to validate if variation/conservation is order-specific or genus-specific. As a result 11 *Drosophila* species, 3 *Anopheles* species and 2 *Apis* species were aligned with CamAlstR-C and their corresponding aligned regions were taken as N-termini. The longest N-terminus was 167 amino acids in *Drosophila* species. The shortest one was 39 amino acids in *Anopheles* species. N-termini regions of each genus were aligned in ClustalW with BLOSUM62 matrix. The alignment profiles

of 3 different genera were saved in Workbench3.2 and they were aligned in CulstalWProf (Gonnet matrix) in order to compare inter-genera conservations. Finally N-terminal 54 amino acids of our receptor were aligned to this existing profile via ClustalWProf.

Table 4.11. Protein sequences from different orders used in N-terminal alignments are given.

Common names for the order	**Order name**	**Genus name**	**Species name**	**Protein Accession Code**	**Max. identity**	**Cover.**	**Condition**	**Inc./Exc.**	**Other orthologs taken within the genus**
bees, ants, wasps	Hymenoptera	Apis	*Apis mellifera*	XP_396335	70%	89%	Alignment with functionally homologous receptors	Included	XP_003698610 (A. Florea)
bees, ants, wasps	Hymenoptera	Nasonia	*Nasonia vitripennis*	XP_001600654	60%	92%	No other species with a known related GPCR	Excluded	
butterflies	Lepidoptera	Manduca	*Manduca sexta*	ADX66345	68%	75%	No other species with a known related GPCR	Excluded	
butterflies	Lepidoptera	Bombyx	*Bombyx mori*	NP_001127736	71%	74%	No other species with a known related GPCR	Excluded	
mosquitos, flies	Diptera	Anopheles	*Anopheles gambiae*	XP_320274	70%	73%	Significant alignments with more than one species	Included	EFR29674 (A. darlingi), AAT88068 (A. stephensi)
mosquitos, flies	Diptera	Drosophila	*Drosophila pseudoobscura pseudoobscura*	XP_001353984	63%	79%	Significant alignments with more than one species	Included	XP_002008619 (D. mojavensis), XP_001985522 (D. grimshawi), NP_649039.4 (D. melanogaster), XP_001972840 (D. erecta), XP_002042582 (D. sechellia), XP_002094901 (D. yakuba), XP_002067937 (D. willistoni), XP_002085370 (D. simulans), XP_001958236 (D. ananassae)
mantids, cockroaches	Dictyoptera	Periplaneta	*Periplaneta americana*	AAK52473	28%	81%	No other species with a known related GPCR	Excluded	
aphids	Hemiptera	Acyrthosiphon	*Acyrthosiphon pisum*	XP_001950448	60%	92%	No other species with a known related GPCR	Excluded	
beetles	Coleoptera	Tribolium	*Tribolium castaneum*	EFA10792	71%	74%		Excluded	

4.3.3. Secondary Structure Prediction

Transmembrane helices of CamSSTR sequence was predicted in TMHMM server v. 2.0 [69], PSIPRED [70] and PDBSUM [71], and compared. Protein identification motifs were obtained in ScanProsite (ExPASy) [72]. The other motifs that can relate to functional determination were scanned in Motif Scan (MyHits, SIB, Switzerland) [73].

4.3.4. 3D Modeling

In order to predict the 3D configuration of a protein which has no X-ray or NMR data known, homology modeling is widely used. One of the most common used comparative modeling software is MODELLER program. It is possible to create a 3D model of a protein which has high identity ($\geq$ 25%) with a known structure. However in most cases this identity cannot be reached. If the protein of interest has lower identities with multiple templates, iterative modeling increases the accuracy of modeling from multiple templates. I-TASSER is one of the most widely used iterative modeling programs. Therefore, servers such as ModWeb, EasyPRED (which was not published yet) and I-TASSER were tried and compared according to template proteins they have included in modeling. I-TASSER server was chosen for 3D modeling as a final decision.

Longest ORF translation of our receptor was loaded into I-TASSER server in the default conditions. Each model output was evaluated according to the placement of N-terminal loop, TM-score and C-score values. The models were also evaluated in build/check/repair model tool of WhatIF web interface [74].

The ligand sequence *Manduca sexta* Allatostatin C sequence (EVRFRQCYFNPISCF) was also modeled in I-TASSER, but with a restraint. It is known that the C7 and C14 on this sequence form a disulfide bridge which originally is 2.05 $\AA$ in distance. So, the ligand modeling was restrained for a distance of 2.05 $\AA$ between Sulphur atoms of Cysteine residues. The resulting models were evaluated according to distance between Sulphur atoms of C7 and C14, C-scores and TM-scores.

4.3.5. Molecular Docking

Docking analysis was performed in AutoDock Vina version 1.1.2 [57]. The receptor was taken from I-TASSER model and Hydrogens were added. Third ligand model from I-TASSER was loaded onto the receptor and flexible bonds were detected. A grid to bind the ligand was determined on receptor molecule. Exhaustiveness of docking analysis was set for 8. Grid values were adjusted as the following: center_x = -19.861 center_y = -2.222 center_z = -12.194 size_x = 58 size_y = 60 size_z = 66

The program was run and the outputs were evaluated according to their approaching to the conserved sites of the receptor. The refined output was taken into LIGPLOT [75]. LIGPLOT was run in PDBSUM server to obtain the possible interactions between the two proteins. The results were again refined according to conserved sites on both the receptor and ligand molecules.

5. RESULTS

5.1. Identification of the Receptor

We performed 3' and 5' RACE reactions in order to find 3' and 5' sequences of full-length open reading frame (ORF) of our receptor. Trans-membrane (TM) region was previously determined by Birgül *et al.* [20], so that we could design specific primers from TM region. At the beginning RNA isolation was performed from *Carausius morosus* tissues (Figure 5.1 and Table 5.1). Than cDNA was produced and verified by actin control primers in PCR reactions (Figure 5.5).

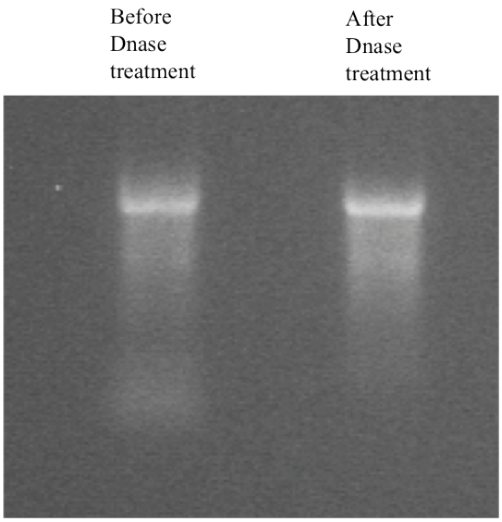

Figure 5.1. Total RNA isolated from Stick insect tissue. 2 μg RNA samples were run in MOPS gel.

Table 5.1. Total RNA extraction results.

	Before DNase treatment	After DNase treatment
Concentration (ng/μl)	1731.10	902.80
Absorbance	20.06	9.41
260/280	2.09	2.17
260/230	2.16	2.40

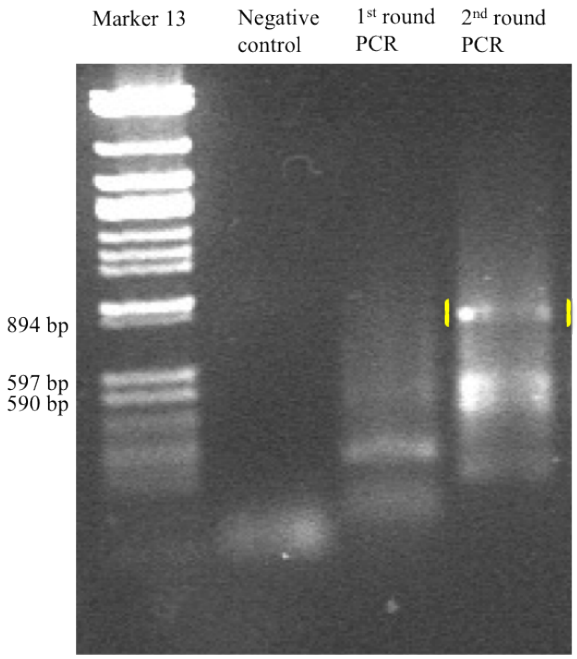

Figure 5.2. Products of 3' RACE reactions from 1st round and 2nd round were run in 1% Agarose gel. The Negative Control sample contained no cDNA. The amplified fragment in yellow brackets was the 3' end of the receptor mRNA.

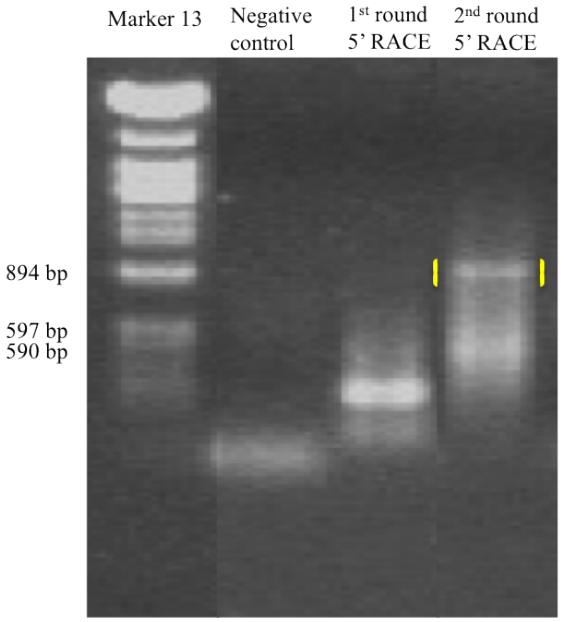

Figure 5.3. Products of 5' RACE reactions from 1st round and 2nd round were run in 1% Agarose gel. The Negative Control sample contained no cDNA. The amplified fragment in yellow brackets was the 5' end of the receptor mRNA.

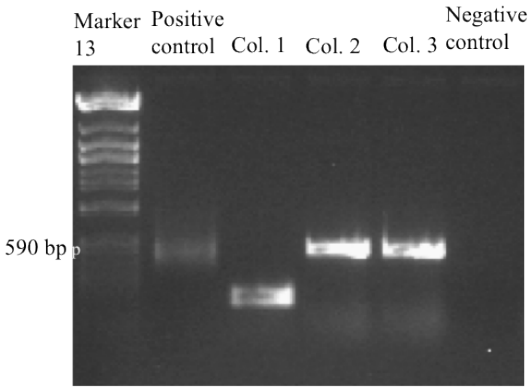

Figure 5.4. Colony PCR confirmation of colonies containing 3' RACE fragment. Positive control represents direct PCR amplification of 3' RACE fragment. Negative control represents colony PCR in absence of colonial DNA. 2nd and 3rd colonies contained the fragment of interest on the vector.

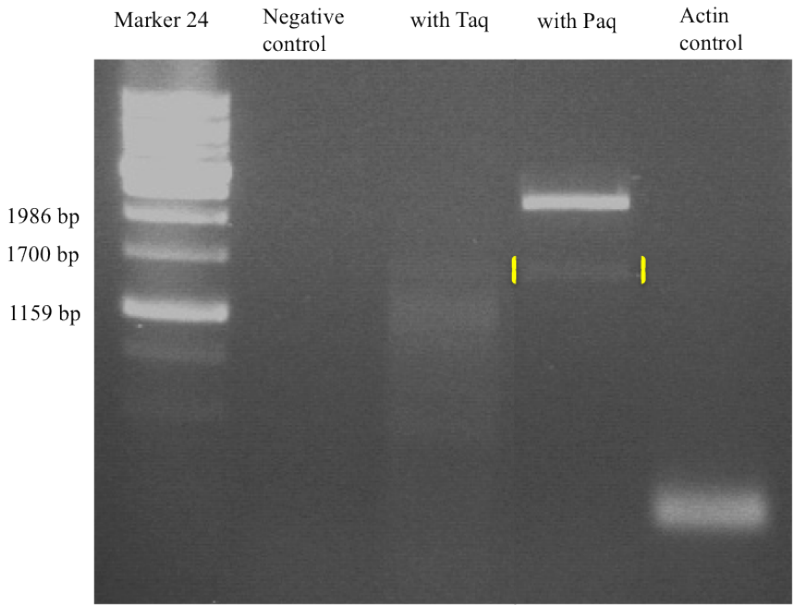

Figure 5.5. Products of ORF amplification with Taq and Paq DNA polymerases. The Negative Control sample contained no cDNA. Drosophila actin primers were used as positive control. The amplified fragment in yellow brackets was the specific banding pattern of our receptor.

3' RACE reaction was performed as an ordinary nested PCR in order to optimize the reaction. As can be seen in Figure 5.4 and 5.3, there were so many unspecific amplification patterns. Various conditions were tried for optimization, but the unspecific amplification could not be overcome. All strong bands were taken for sequence analysis, only one fragment was chosen as the original terminal sequence of the receptor, according to presence of homologous sequences with other receptors and previously determined TM region sequence (fragments in yellow brackets in Figure 5.4 and 5.3).

Both 3' and 5' RACE sequences were combined with TM sequence. The longest ORF sequence is illustrated in Figure 5.6 and the predicted translation of this mRNA is given in Figure 5.7. The translated protein is predicted as 441 amino acids in length, having 7 TM helices. With the amplification of longest ORF via new primers, new sequencing samples were sent to Macrogen, Korea. Twelve cloning products from different PCR samples were sent for sequencing and one of them, which was predicted to be unmutated, was selected for further analysis.

The cloned ORF was verified for translation in Huh7 cell lines. Immunofluorescence experiments showed a uniform distribution of the receptor when over-expressed in cells (Figure 5.8). Expression was not detected in the nucleus. In over-expression studies the membrane might be saturated with the receptor, so that we could not see only surface-expression. Further verification for surface expression of the receptor was performed in Atomic Force Microscopy (AFM) experiments.

AFM studies showed us that our receptor was interacting with Allatostatin IV peptide of *Manduca sexta*, when over-expressed in Huh7 cell lines (Figure 5.9b). In order to see if the interaction was specific for the receptor of interest, non-transfected Huh7 cells were used, which were known not to express this receptor. Figure 5.9a shows that there was no retracting force due to an interaction. The other verification was performed with excess peptide in the medium. This high concentrated soluble ligand was assumed to block the receptors and the ligand on the tip of cantilever would no more have an interaction with the receptor, as confirmed in Figure 5.9c. All of these samples and control experiments were repeated 65 times per cell and on more than one

ATGTCTGTGGAACAAGTGACGGTGATGGTGACGTCATGGTCGACCCTGGAGGACTTGAAC
ACGTCCTCGACGACGGACCCGCTGCTGGAGGCGGGGAACGGCTCGGCGCACGGCAACATG
TCGCTGGACCACCACGAGTGCGACACAGACCAGCCGACGCTGCTGCTGATCACACAGGTG
CTGTGCGCGGCCGTGTGCATCGTCGGCCTGCTCGGCAACAGCCTCGTCATCTACGTGGTG
CTCCGCTTCTCCAAGATGCAGACAGTCACCAACATGTACATCGTGAACCTCGCCGTCGCT
GACGAGTGCTTCCTCATCGGCGTGCCCTTCCTGCTCGCCACCATGAACCTGCGCATCTGG
CCCTTCGGCAGCATCATGTGCAAGGCGTACATGACGACGACGAGCATCAACCAGTTCACG
AGCAGCATCTTCCTGACGATCATGAGCGCGGACCGCTACATAGCCGTGTGCCACCCCATC
TCCGCGCCCAAGGTCCGCACGCCCCTCATCTCCAAGGTGGTGTCCGCGTCGGCGTGGACC
GCCAGCGCCCTGCTCATGATCCCCATCTTCATATACGCCAACATCATGACGAACGAGAAC
GTCGTGAGCTGTAACATCCTGTGGCCGGAGAGCGAGAACGTGAGCAGCCAGGGCGCCTTC
ACGCTCTACACCTTCATACTCGGCTTCTCCATCCCGCTCATGCTCATCTCCGTGTTCTAC
ATCCTCGTGATCAGGAAGCTGAAGACCGTCGGGCCCAAGAACAAGTCCAAGGAGAAGAAG
AAGTCGCACCGCAAGGTGACGAAGCTGGTGCTGACGGTGATCACGGTGTACGTGCTGTGC
TGGCTGCCGTACTGGCTCACGCAGATGGCGCTCATCTTCACGCCGCCCAAGCAGTGCCAG
TCGAGGGTGTCCATCACAGTGTCCCTGCTGGCGGGGTGCCTCAGCTACTCCAACTCCGCC
ATGAACCCCATCCTGTACGCCTTCCTCAGCGACAACTTCAAGAAGAGCTTCCTGAAGGCG
TGCACGTGCGCCACGGGCCAGGACGTCAACAACGTGCTGCACATGGAGAACAGCGTGTTC
CCGCGGCGCCACCACCGGCAAGGAGGCTCCGAGAGGCAGCGCCCCAACAAGATGGCGACC
TTGGAGACGGGGGAGGGCGGCGAGAAGGGCCCATTGGTCAGCAAGGCCGACAGCTCCGTC
ACCACCATGACCATGACCTCGCGCTCCAACATCACCGTCAGCACCGACGCCAAGGACCAG
GCGAGCAAGGACGCTGCTGCCATGGAGACGGCCCTCAAGAACGGCCAGCAGCCGACCCAG
GTGTAG

Figure 5.6. ORF mRNA. The sequence is 1326 bp, begining with ATG and ending with TAG stop codon.

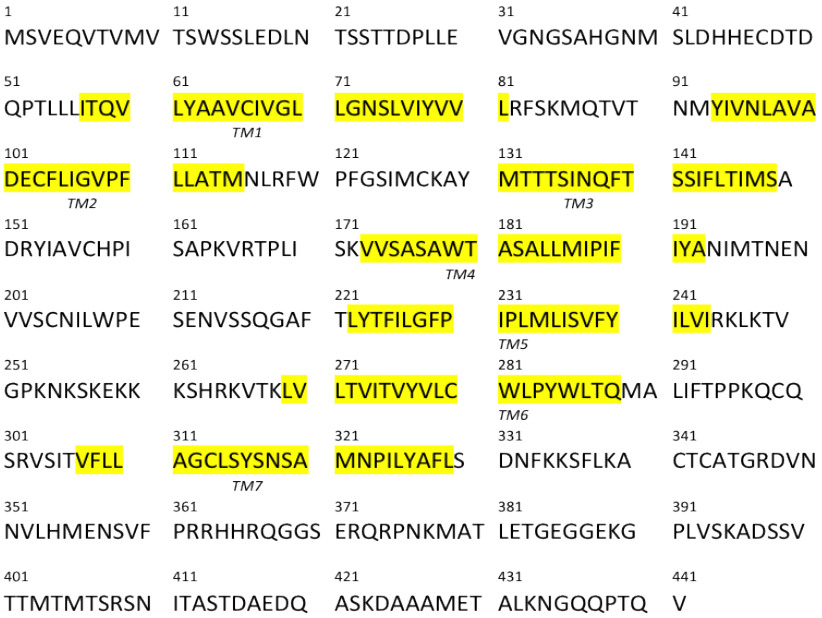

Figure 5.7. Longest ORF was translated (441 aa). The yellow bars indicate the TM regions predicted in TMHMM. The numbers indicate the residue number.

cell. The interaction cases were counted manually. At least 25% of approach/retract trials on one transfected cell showed peptide interaction pattern. However in non-transfected controls and excess ligand group, this ratio was 0%.

You can see a typical adhesion force-distance curve in Figure 5.10a with a cantilever speed of 10 μm/s. 0 point on the Piezo displacement axis was the starting point of cantilever. The blue plot was the force exerted on cantilever when it was approaching to the surface of the cell. After approximately 3.7 μm of approach the tip reaches the surface and deflection from it increases, meaning that a force was exerted on the tip. The red plot showed the retraction force. As it can be seen, the force exerted on cantilever decreases with retraction on the surface. When it reached the surface border at about 3.7 μm again, it showed an adhesion force on the ligand, decreasing the force curve towards minus area. The second minus peak of this retraction showed typical peptide unbinding force of approximately 250 pN. In Figure 5.10b and Appendix D, magnified versions of force-displacement curves were given. The unbinding force incresed with increasing the cantilever speed. For instance the upper figure of 5.10b showed about 100 pN unbinding force with 1 μm/s speed, whereas the lower figure showed approximately 1000 pN with 20 μm/s speed.

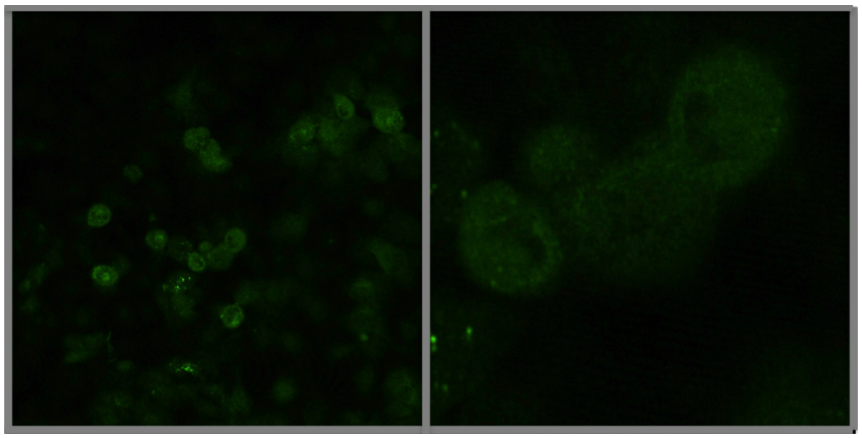

Figure 5.8. Location of expression was confirmed via immunofluorescence. The right figure is magnification of two cells showing uniform distribution of the receptor in cell.

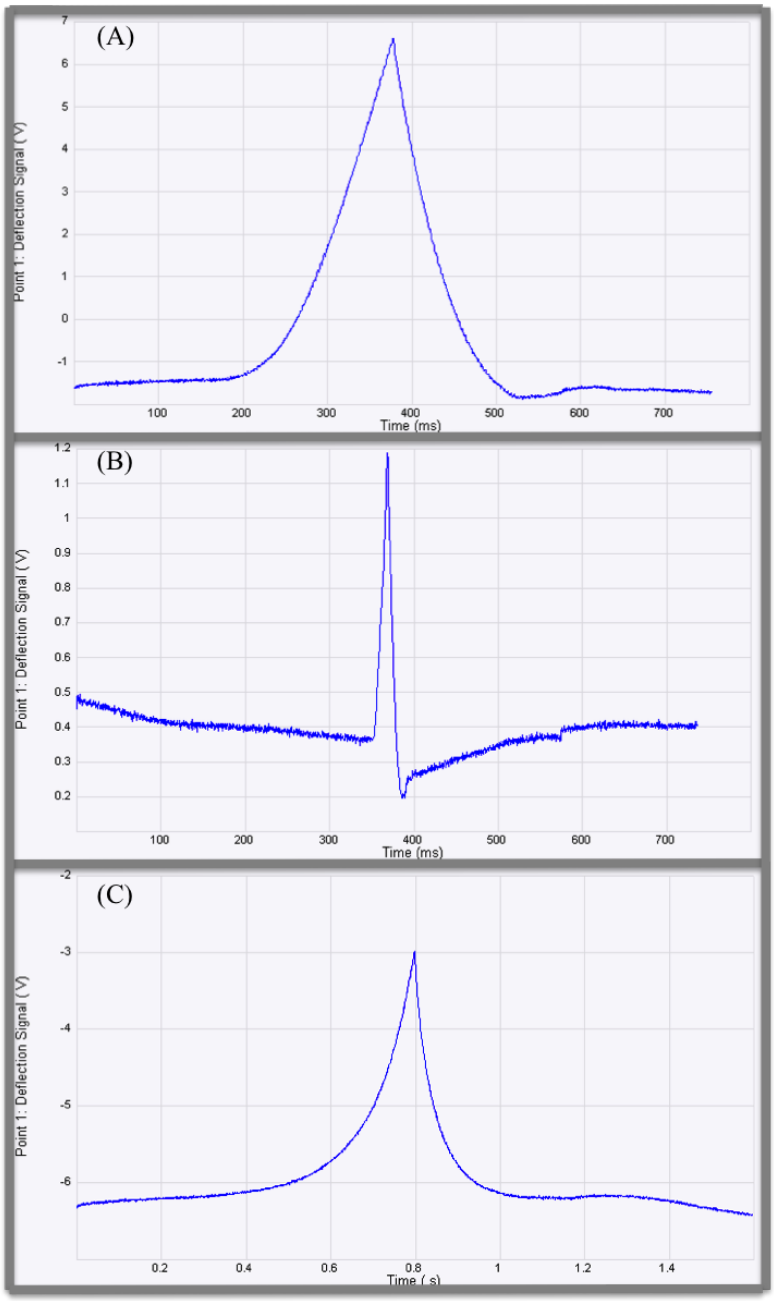

Figure 5.9. AFM deflection signals. (A) Huh7 cells which were not transfected, (B) pcDNA3/HA-CamAlstR-C transfected cells and (C) interaction on transfected cells after washing with excess ligand.

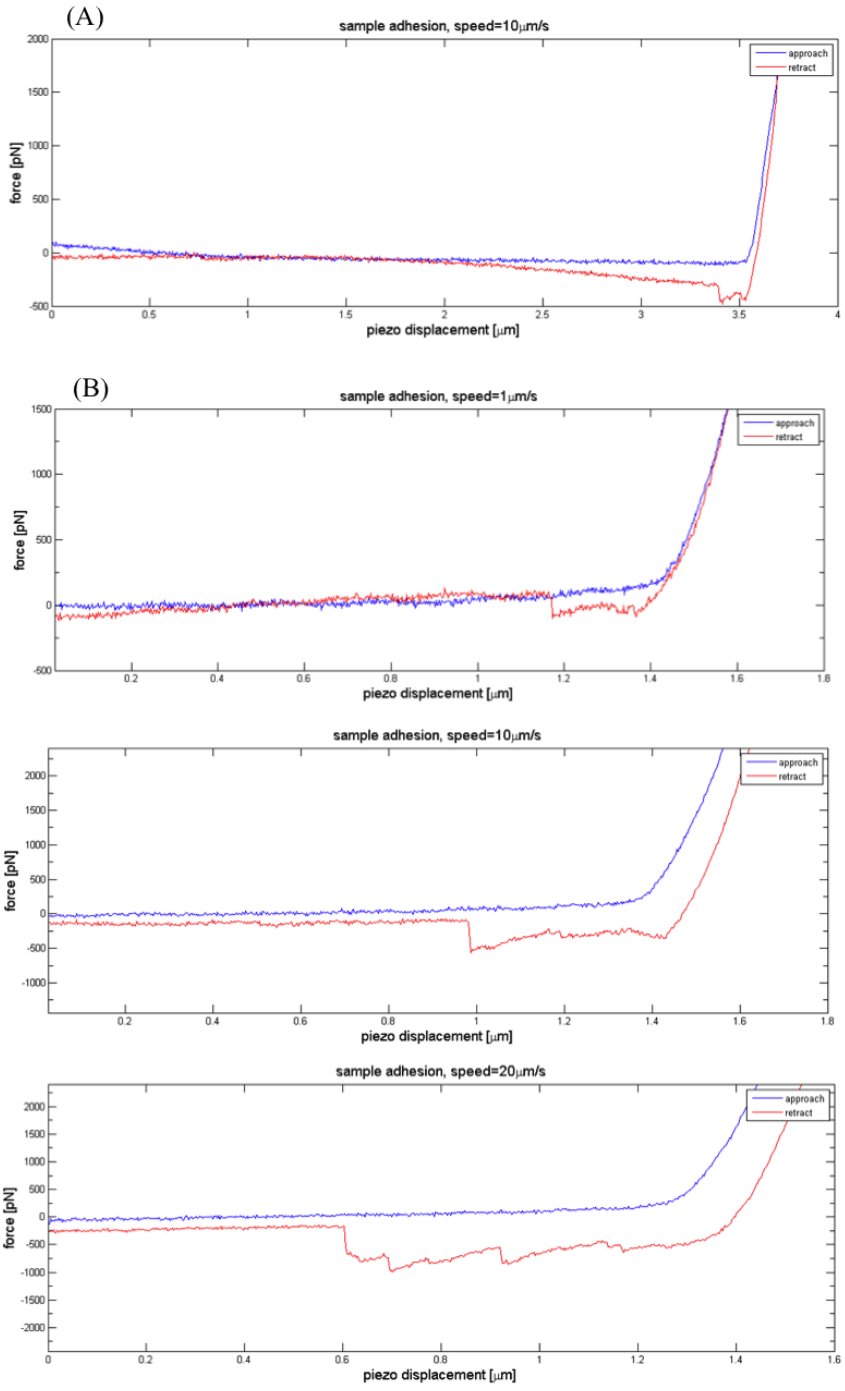

Figure 5.10. AFM force curves. (A) Typical adhesion force curve obtained on a transfected Huh7 cell, (B) magnified examples for adhesion force curves in different speeds on the same cell.

5.2. Phylogenetic Analysis

Bioinformatics studies on the receptor began with the determination of the protein family. In basic Blastn search within nucleotide collection databases, ORF sequence was found to be mostly related to *Acyrthosiphon pisum* (pea aphid) somatostatin receptor type 2-like, transcript variant 2 mRNA (Figure 5.11). Blastp alignment on translation of this ORF gave the closest identity with *Bombus terrestris* (bumblebee) somatostatin receptor type 4-like protein (Figure 5.12). As a result of these, our receptor of interest was defined as a somatostatin-like receptor, which corresponds to Allatostatin receptor (AlstR) family in Arthropoda.

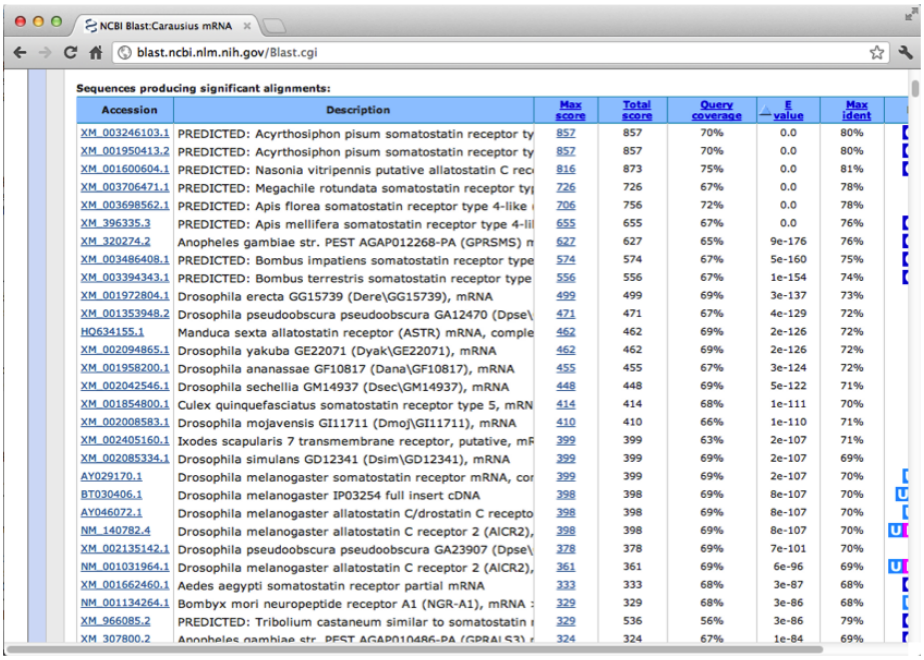

Figure 5.11. NCBI Blastn server alignment in default condition. The alignment was performed within nucleotide collection.

Then, all types (A, B and C types) of Allatostatin receptors were taken into Maximum Likelihood analysis in order to identify the type of our receptor of interest. The tree with the highest log likelihood (-8446.5855) was drawn (Figure 5.13). The branch nodes, which lead to receptor type differentiations, showed significantly high probabilities (bootstrap values of 99 and 100). This meant that the different paralogs were distributed in a reliable way. And our receptor CamSSTR was placed within the AlstR type C family branch. According to the evolutionary relatedness of TM regions

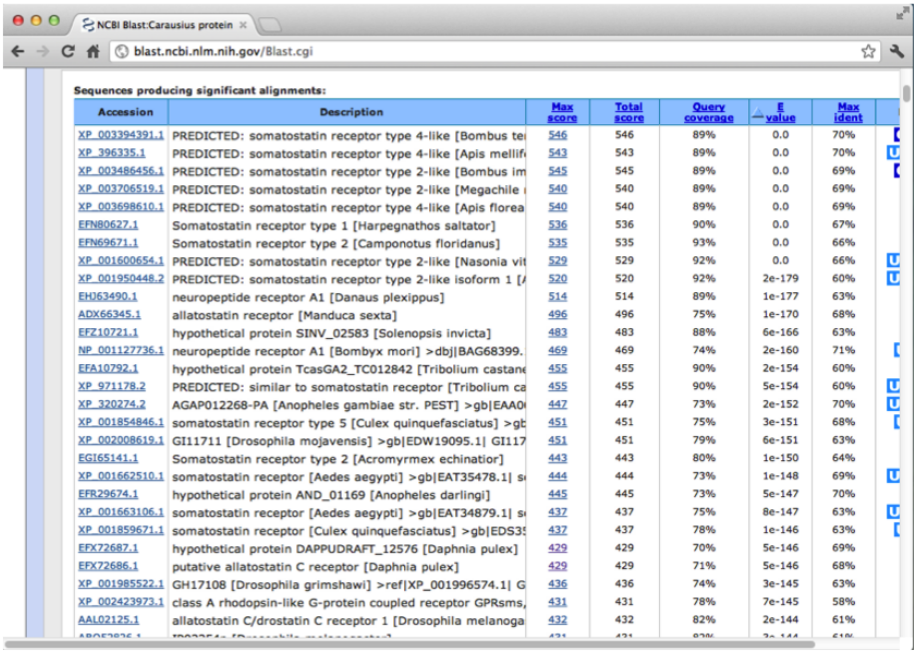

Figure 5.12. NCBI Blastp server alignment in default condition. The alignment was performed within non-redundant protein sequences.

with Allatostatin receptor type C of other insects, our protein of interest was named as *Carausius morosus* Allatostatin receptor type C, and abbreviated as CamAlstR-C.

In the alignments, N-terminus of the receptor was very variable and showed no conservation (Figure 5.14). However, conservation was detected within different species of a genus (Figures 5.15, 5.16 and 5.17).

5.3. Secondary Structure Prediction

In order to understand the conserved domains and regions on CamAlstR-C we predicted its' secondary structure and motifs in various web servers. PSIPRED results of TM helix predictions (Figure 5.18) were consistent with the probability results of TM regions in TMHMM server (Appendix C). Figure 5.7 shows the predicted TM regions. As both secondary structure prediction and alignment tools showed the same result; our receptor of interest belonged to GPCRs which contain 7 trans-membrane helices. The first trans-membrane helix (TM1) begins from 57th residue, ends at 82nd residue. TM2 begins with 93rd residue, ends with 115th residue. TM3 spans from 131st residue

until 149th residue. TM4 begins with 173rd residue ends at 193rd residue. TM5 spans from 222nd until 244th residue. TM6 spans from 268th until 288th residue. And the last TM (TM7) spans from 307th until 329th residue. According to PSIPRED results CamAlstR-C has another 8th helix which is located on the intracellular C-terminus, spanning between 420th and 433rd residues (Figure 5.18).

In other parts CamAlstR-C showed common signs of family A (rhodopsin-like) GPCRs. For instance, DRY motif at the end of TM3 was typical. NPILY and WLP motifs on CamAlstR-C correspond to conserved NPXXY and WXP motifs of this family respectively (Figure 5.19). β-hairpin structure in ECL2 and disulfide bridge between ECL2 and TM3 were detected in CamAlstR-C similar to rhodopsin structure. Some other motifs obtained from MotifScan were summarized in Figure 5.19 and these gave us clues about its' functional homology to mammalian SSTRs.

5.4. Modeling

The template similarities obtained in 3D modeling trials of ModWeb and EasyPRED were 39% and 18.4%, respectively. However, only the region between 52nd and 337th residues was covered in ModWeb results. As an alternative, I-TASSER server was a threading method in which the coverage and similarity rate would increase. The models for both CamAlstR-C and AST-C peptide were compared according to their confidence scores (C-score). Higher the C-score means the more confident the model depending on the significance of threading template alignments. However, some other conditions such as given in Table 5.2 were evaluated with C-scores. Receptor model1 showed the highest C-score and reliable TM-score with any inconvenient 3D conformation for GPCRs. The structural alignments of model with the threaded template proteins are illustrated in Figure 5.20. Some templates could not be covered in N-terminus alignments. The names of structurally threaded template proteins were given in Table 5.3. Model1 of CamAlstR was also checked in WhatIF. Root mean square Z-score for bond angles was 1.526, which was close enough to a high resolution X-ray structure value. The backbone conformation was also appropriate to well-refined structures (Ramachandran Z-score: -0.051). Bond length Z-score (0.997) was also very

close to 1, meaning they deviated normally from standard bond lengths. However in ligand models, model1 and 2 were excluded because of our constraint in disulfide bridge length and model3 was included in docking analysis. In Figure 5.21, bond distances are illustrated on each ligand model.

Table 5.2. Top five models obtained from I-TASSER.

Receptor:	**C-score**	**TM Score**	**RMSD**	**Condition**
Model 1	-2.11	0.46±0.15	12.1±4.4Å	*Included in docking*
Model 2	-2.30	nd*	nd	N-terminal loop extends through membrane
Model 3	-2.43	nd	nd	N-terminal loop extends through membrane
Model 4	-2.92	nd	nd	
Model 5	-2.30	nd	nd	C-terminal loop behaves like TM
Ligand:	**C-score**	**TM Score**	**RMSD**	**Condition**
Model 1	-0.65	0.63±0.14	1.7±1.4Å	Extended disulfide bond lenght
Model 2	-2.58	nd	nd	Extended structure
Model 3	-3.11	nd	nd	*Included in docking*
Model 4	-2.42	nd	nd	Extended disulfide bond lenght
Model 5	-1.85	nd	nd	Extended disulfide bond lenght

* nd: not defined by the program

Table 5.3. Top ten templates used in I-TASSER for CamAlstR modeling.

PDB code	**Identity 1**	**Identity 2**	**Coverage**	**Name of protein**
3emlA	0.22	0.19	0.65	Human A2A adenosine receptor
2ks9A	0.23	0.24	0.81	NK1R receptor
4ea3A	0.37	0.25	0.63	N/OFQ opioid receptor
4djhA	0.36	0.24	0.64	Human kappa opioid receptor
2ks9A	0.23	0.24	0.81	
3emlA	0.22	0.19	0.65	
4ea3A	0.37	0.25	0.63	
4djhA	0.37	0.24	0.64	
4djhA	0.37	0.24	0.64	
3v2yA	0.17	0.17	0.68	Lipid G protein-coupled receptor

5.5. Docking Analysis

In Figure 5.22 and 5.23, docked molecules were represented. The cartoon representation revealed that AST-C was closer to ECL2 and ECL3 of CamAlstR-C. After detailed analysis in PDBsum (Figure 5.24), interacting residues were labeled in Figure 5.25. These residues were also verified for their conservation in other AlstRs of insects (Table 5.4).

Table 5.4. Conservation analysis of residues in proposed binding pocket, within insect AlstRs.

Region	Sequence	Conservation
13-26	*WSTLEDLNTSSTTD*	very low
210-222	*ESENVSSQGAFTL*	next to a very conserved region
292-298	*IFTPPKQ*	conserved IXTXP motif

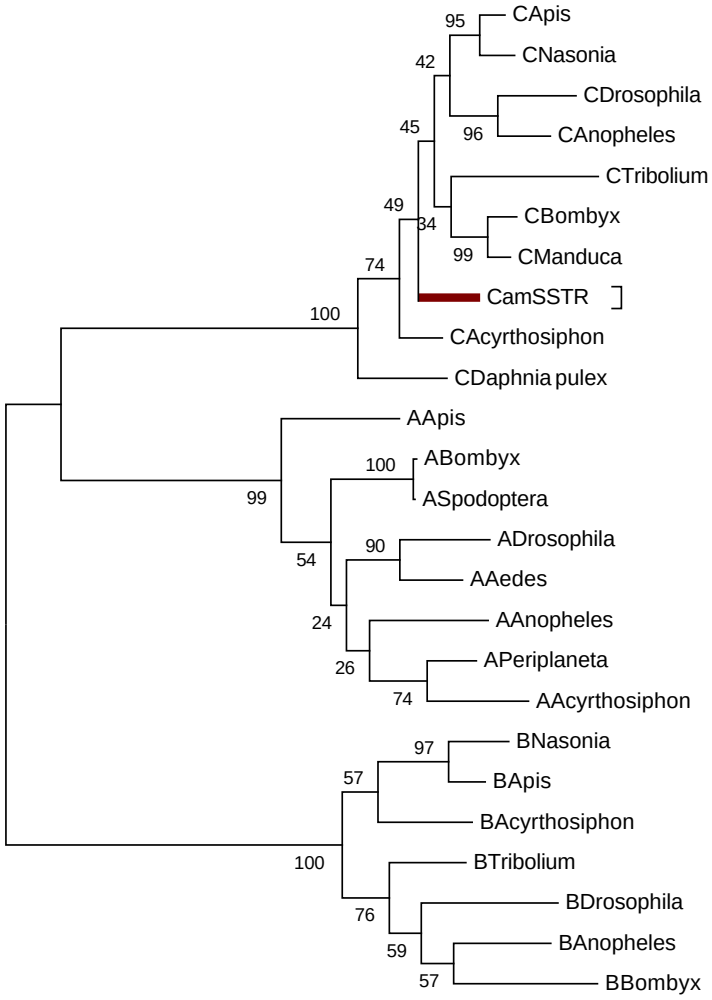

Figure 5.13. The evolutionary analysis of CamSSTR was inferred by using the Maximum Likelihood method in MEGA5. The letters (A, B and C) before the genus names correspond to the types of receptors. Ex; CDrosophila: AlstR type C of *Drosophila*

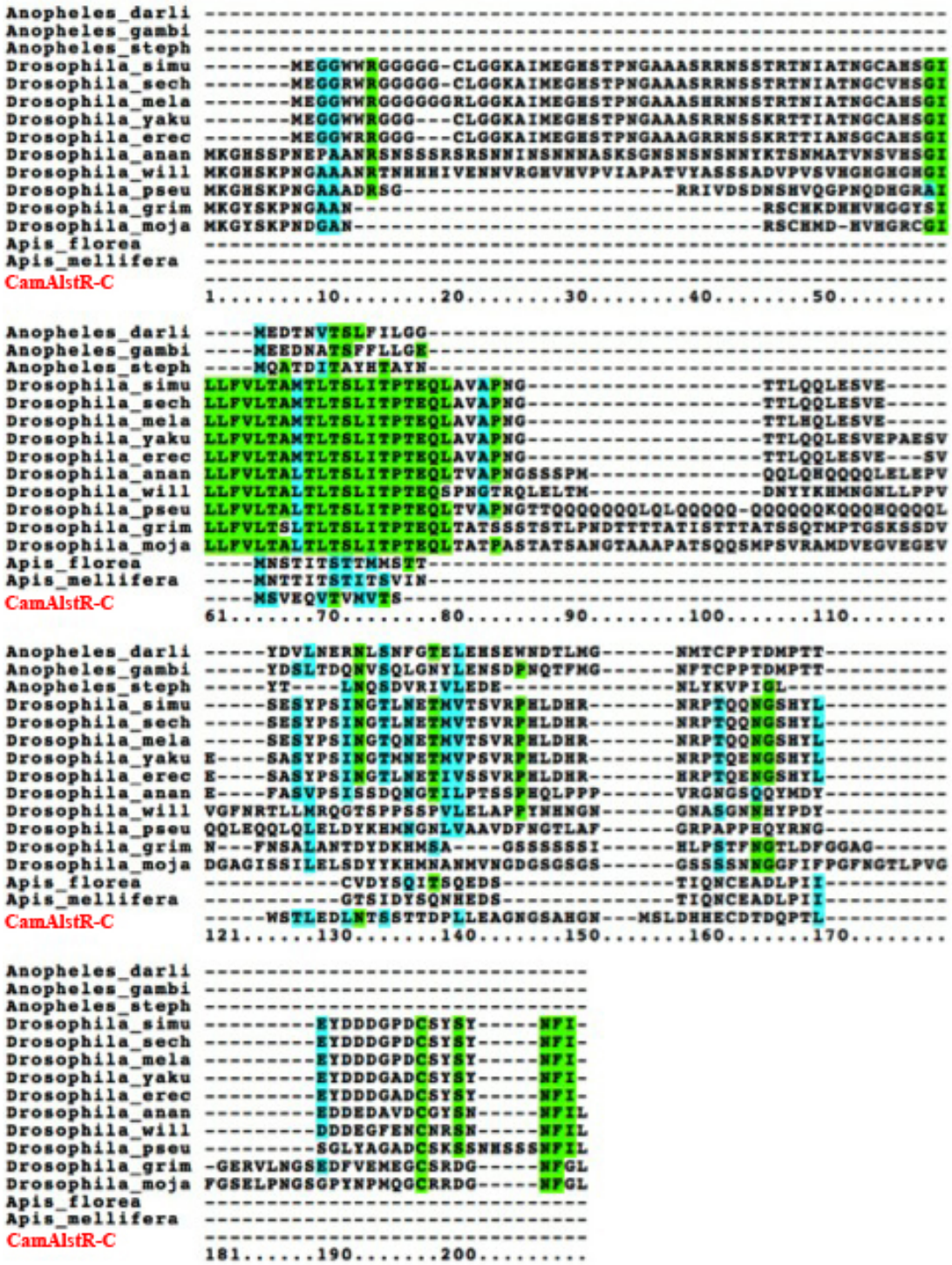

Figure 5.14. N-terminal loop alignments of AlstR-C from 3 different genera with CamAlstR-C.

```
CLUSTAL 2.1 multiple sequence alignment


A.mellifera      MNTTITSTITSVINGTSIDYSQN--HEDSTIQNCEADLPII
A.florea         MNSTITS--TTMMSTTCVDYSQITSQEDSTIQNCEADLPII
                 **:****  *:::. *.:****   :***************
```

Figure 5.15. N-terminal alignment of AlstR-C within *Apis* genus. "*" corresponds to 100% conservation, ":" and "." correspond to lower conservation.

```
CLUSTAL 2.1 multiple sequence alignment


A.gambiae        MEEDNATSFFLLGEYDSLTDQNVSQLGNYLENSDPNQTFMGNFTCPPTDMPTT
A.darlingi       MEDTNVTSLFILGGYDVLNERNLSNFGTELEHSEWNDTLMGNMTCPPTDMPTT
A.stephensi      MQATDITAYHTAYNYT-LNQSDVRIV---LEDENLYKVPIG-L----------
                 *:  : *: .    *  *.: ::  .   **..:  .. :* :
```

Figure 5.16. N-terminal alignment of AlstR-C within *Anopheles* genus. "*" corresponds to 100% conservation, ":" and "." correspond to lower conservation.

```
CLUSTAL 2.1 multiple sequence alignment

D.sechellia         MEG-------GRWRGGGGG-CLGGKAIMEGHSTPNGAAASRRNSSTRTNIATNGCVH-SG
D.simulans          MEG-------GWWRGGGGG-CLGGKAIMEGHSTPNGAAASRRNSSTRTNIATNGCAH-SG
D.melanogaster      ------------------------------------------------------------
D.melanogasterB     MEG-------GWWRGGGGGGRLGGKAIMEGHSTPNGAAASHRNNSTRTNIATNGCAH-SG
D.erecta            MEG-------GWRR--GGG-CLGGKAIMEGHSTPNGAAAGRRNSSKRTTIANSGCAH-SG
D.yakuba            MEG-------GWWR--GGG-CLGGKAIMEGHSTPNGAAASRRNSSKRTTIATNGCAH-SG
D.ananassae         MKGHSSPNEPAANRSNSSSRSRSNNINSNNNASKSGNSNSNSNNYKTSNMATVNSVH-SG
D.willistoni        MKGHSKPNGAAANRTNHHH-IVENNVRGHVHVPVIAPATVYASSSADVPVSVHGHGHGHG
D.mojavensis        MKG-----------------------YSKPNDGAN-----RSCHMD--HVHGRCG-----
D.grimshawi         MKG-----------------------YSKPNGAAN-----RSCHKDH-HVHGGYS-----
D.pseudoobscura     MKG----------------------HSKPNGAAADRSGRRIVDSDNSHVQGPNQDHGRA


D.sechellia         ILLFVLTAMTLTSLITPTEQLAVAPNGTT--LQQLESVE---------SESYPSINGTLN
D.simulans          ILLFVLTAMTLTSLITPTEQLAVAPNGTT--LQQLESVE---------SESYPSINGTLN
D.melanogaster      --------MTLTSLITPTEQLAVAPNGTT--LHQLESVE---------SESYPSINGTQN
D.melanogasterB     ILLFVLTAMTLTSLITPTEQLAVAPNGTT--LHQLESVE---------SESYPSINGTQN
D.erecta            ILLFVLTAMTLTSLITPTEQLAVAPNGTT--LQQLESVE------SVESASYPSINGTLN
D.yakuba            ILLFVLTAMTLTSLITPTEQLAVAPNGTT--LQQLESVEP---AESVESASYPSINGTMN
D.ananassae         ILLFVLTALTLTSLITPTEQLTVAPNGSSSPMQQLQHQQQQLELEPVEFASVPSISSDQN
D.willistoni        ILLFVLTALTLTSLITPTEQ---SPNGTR----QLELTMDN-YYKHMNGNLLPPVVG-FN
D.mojavensis        ILLFVLTALTLTSLITPTEQLTATPASTATSANGTAAAPAT---SQQSMPSVRAMDVEGV
D.grimshawi         ILLFVLTSLTLTSLITPTEQLTATSSSTSTLPNDTTTTATI---STTTATSSQTMPTGSK
D.pseudoobscura     ILLFVLTALTLTSLITPTEQLTVAPNGTTQQQQQQQLQLQQ---QQQQQQQQQKQQQHQQ
                           :**********   :. .:

D.sechellia         ETMVT---SVRP--HLDHRNRPTQQ---------------------------NGS-----
D.simulans          ETMVT---SVRP--HLDHRNRPTQQ---------------------------NGS-----
D.melanogaster      ETMVT---SVRP--HLDHRNRPTQQ---------------------------NGS-----
D.melanogasterB     ETMVT---SVRP--HLDHRNRPTQQ---------------------------NGS-----
D.erecta            ETIVS---SVRP--HLDHRHRPTQE---------------------------NGS-----
D.yakuba            ETMVP---SVRP--HLDHRNRPTQE---------------------------NGS-----
D.ananassae         GTILP---TSSP--HQLP--PPVRG---------------------------NGSQ----
D.willistoni        RTLLMRQGTSPPSSPVLELAPPYNH---------------------------NGNGNASG
D.mojavensis        EGEVDG-AGISSILELSDYYKHMNANMVNGDGSGSGSGSSSSNNGGFIFPGFNGTLPVG-
D.grimshawi         SSDVN----FNSALANTDYDKHMSA----------GSSSSSSIH---LPSTFNGTLDFGG
D.pseudoobscura     QQLQQ----LEQQLQLELDYKHMNG-----------NLVAAVD-------FNGTLAFGR
                                                                      **.

D.sechellia         --HYLEYDDD-----GPDCSYSYNFI------
D.simulans          --HYLEYDDD-----GPDCSYSYNFI------
D.melanogaster      --HYLEYDDD-----GPDCSYSYNFI------
D.melanogasterB     --HYLEYDDD-----GPDCSYSYNFI------
D.erecta            --HYLEYDDD-----GADCSYSYNFI------
D.yakuba            --HYLEYDDD-----GADCSYSYNFI------
D.ananassae         --QYMDYEDDE---DAVDCGYSNNFIL-----
D.willistoni        NNHYPDYDDDE---GFENCNRSNNFIL-----
D.mojavensis        -FGSELPNGSGPYNPMQGCRRDGNFGL-----
D.grimshawi         AGGERVLNGSEDFVEMEGCSRDGNFGL-----
D.pseudoobscura     PAPPHQYRNGSGLYAGADCSKSSNHSSSNFIL
                           ..       .*  . *.
```

Figure 5.17. N-terminal alignment of AlstR-C within *Drosophila* genus. "*" corresponds to 100% conservation, ":" and "." correspond to lower conservation.



Figure 5.18. Secondary structure prediction in PSIPRED server. The explanation of abbreviations were given at the upper left corner.

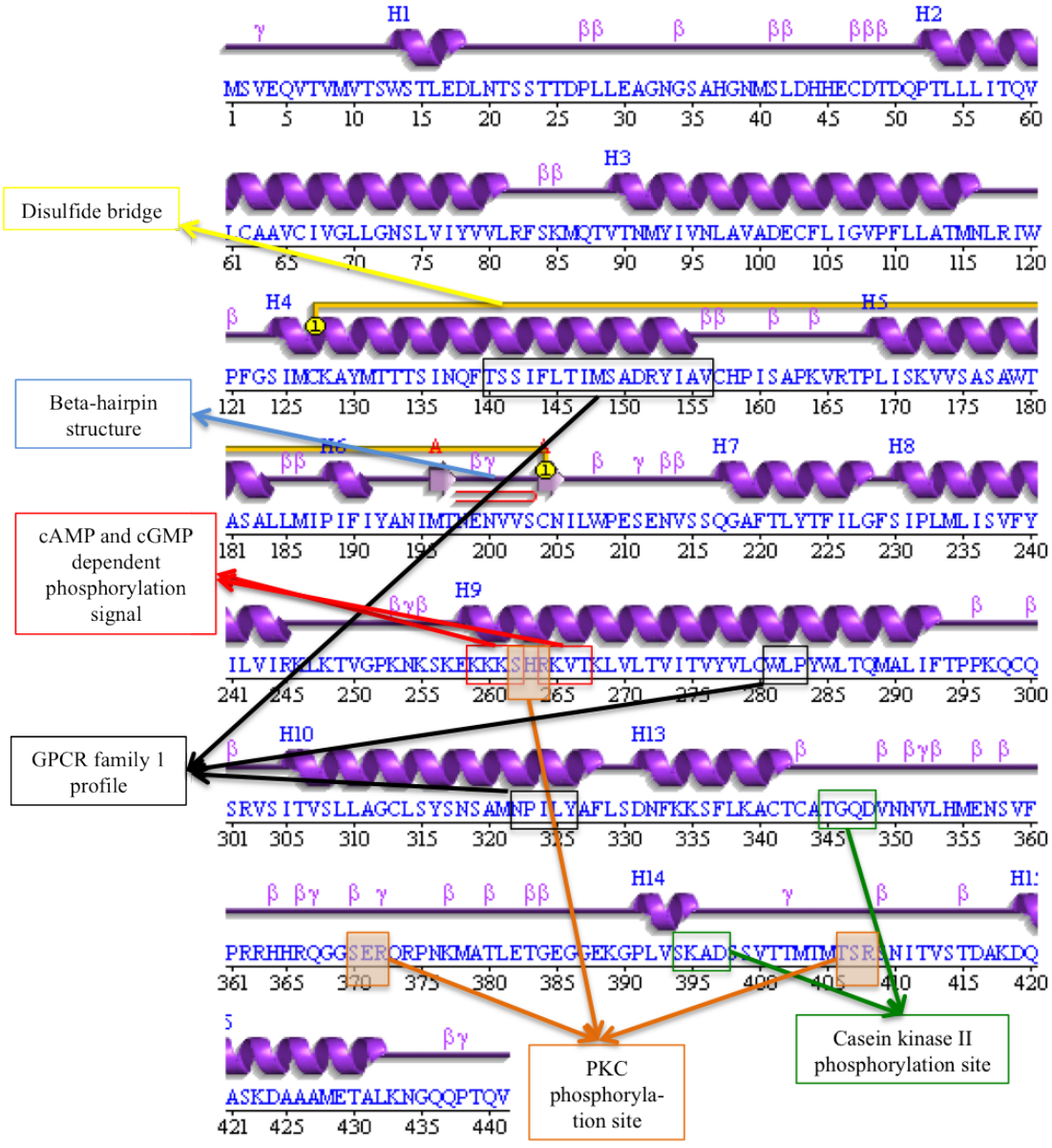

Figure 5.19. The sequence motifs that were detected in combination of PDBsum and MotifScan. The figure was modified from output picture of PDBsum. H: helix, β: beta turn, γ: gamma turn, arrow: strand

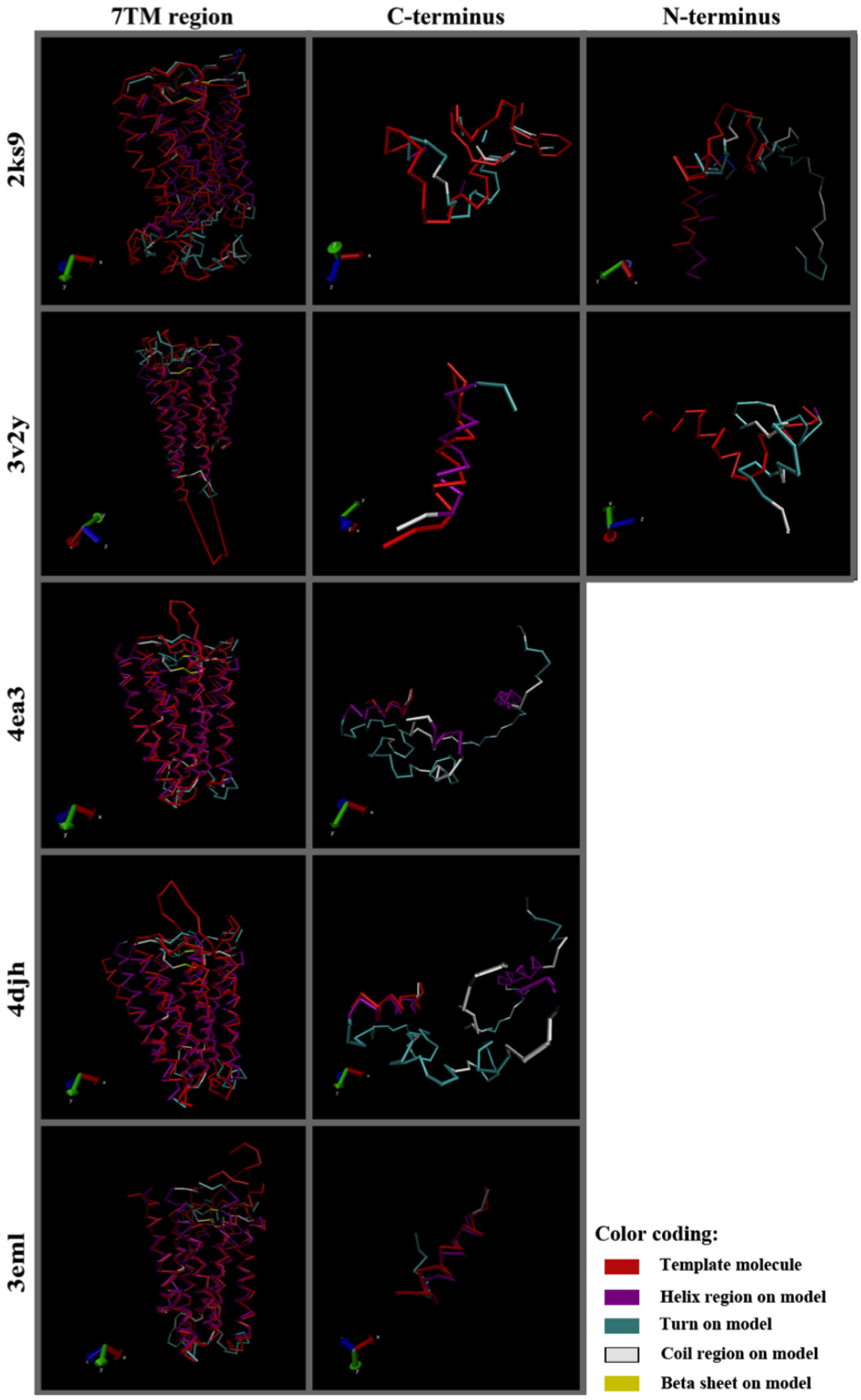

Figure 5.20. The threaded templates were structurally aligned with CamAlstR-C model by I-TASSER and visualized in VMD. The codes on the left side are PDB codes for template proteins.

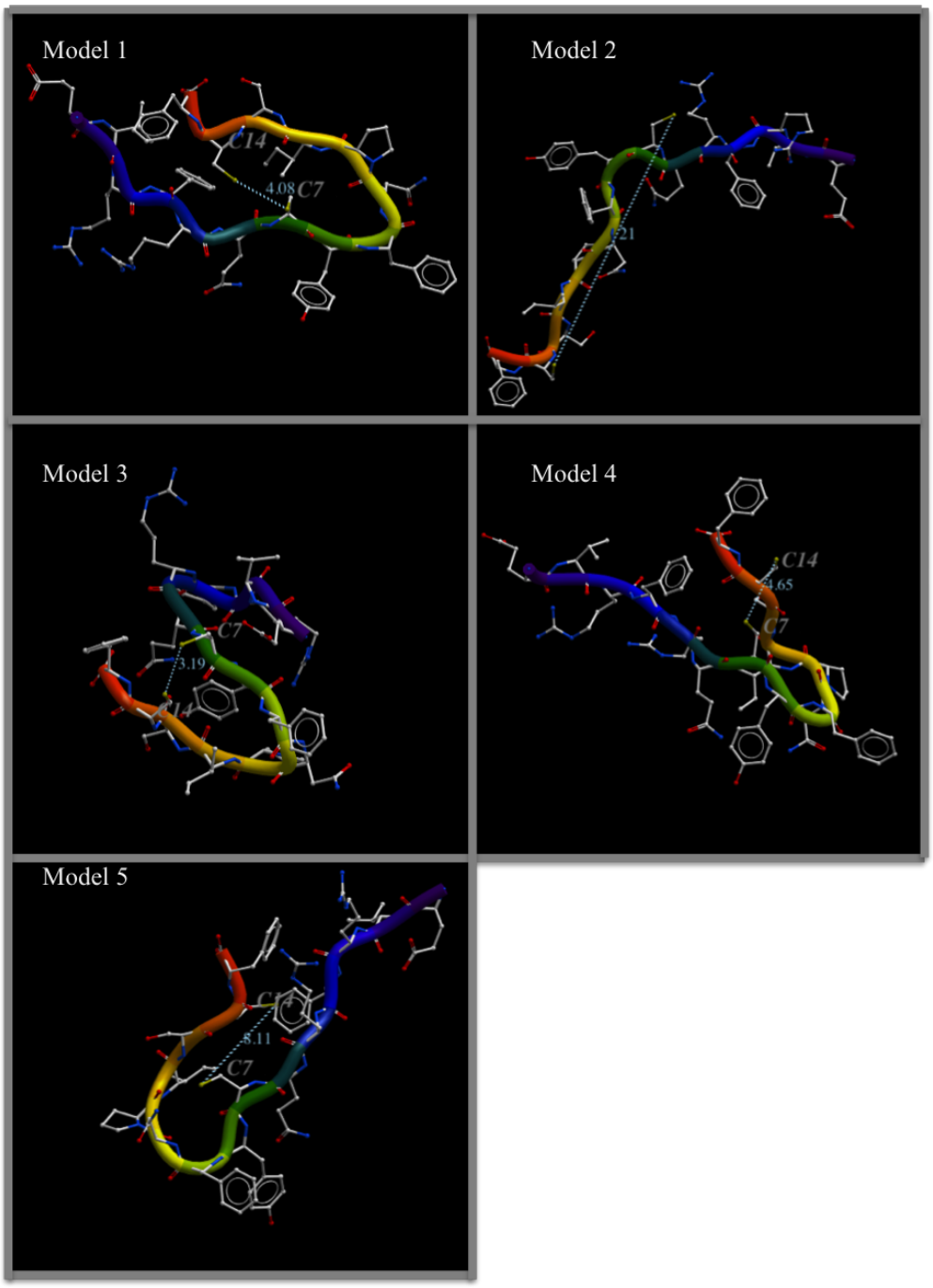

Figure 5.21. Five refined models for ligand were evaluated for selection of best model. The disulfide bridge distances were measured in Molsoft Browser Pro.

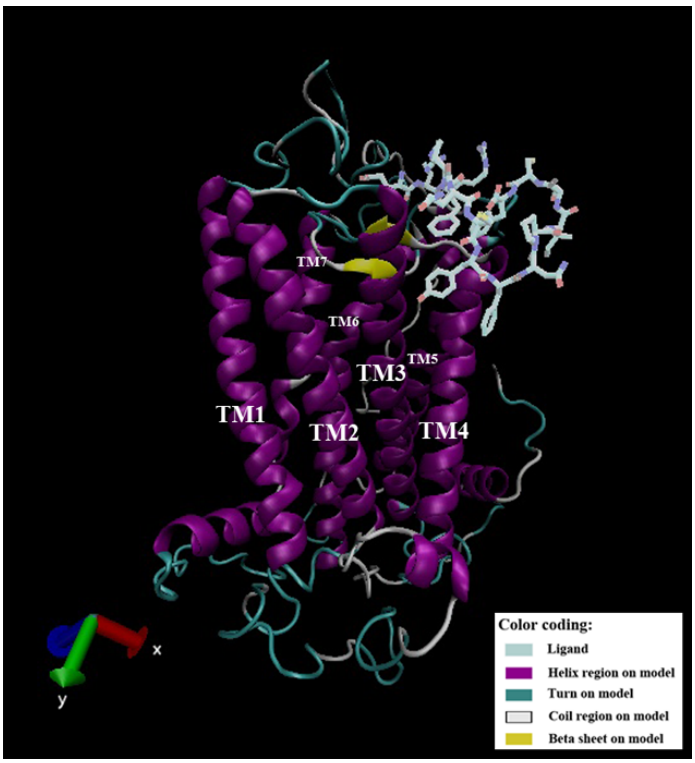

Figure 5.22. Docked AST-C ligand on CamAlstR-C. CamAlstR-C is represented in the cartoon form, while AST-C ligand is represented in bond form, in VMD.

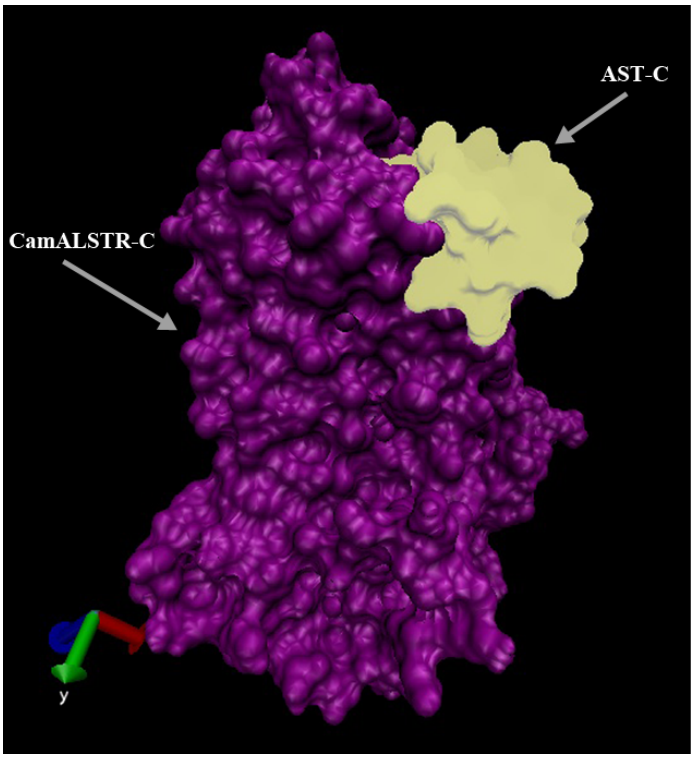

Figure 5.23. Docked AST-C on CamAlstR-C, represented in surface form in VMD.

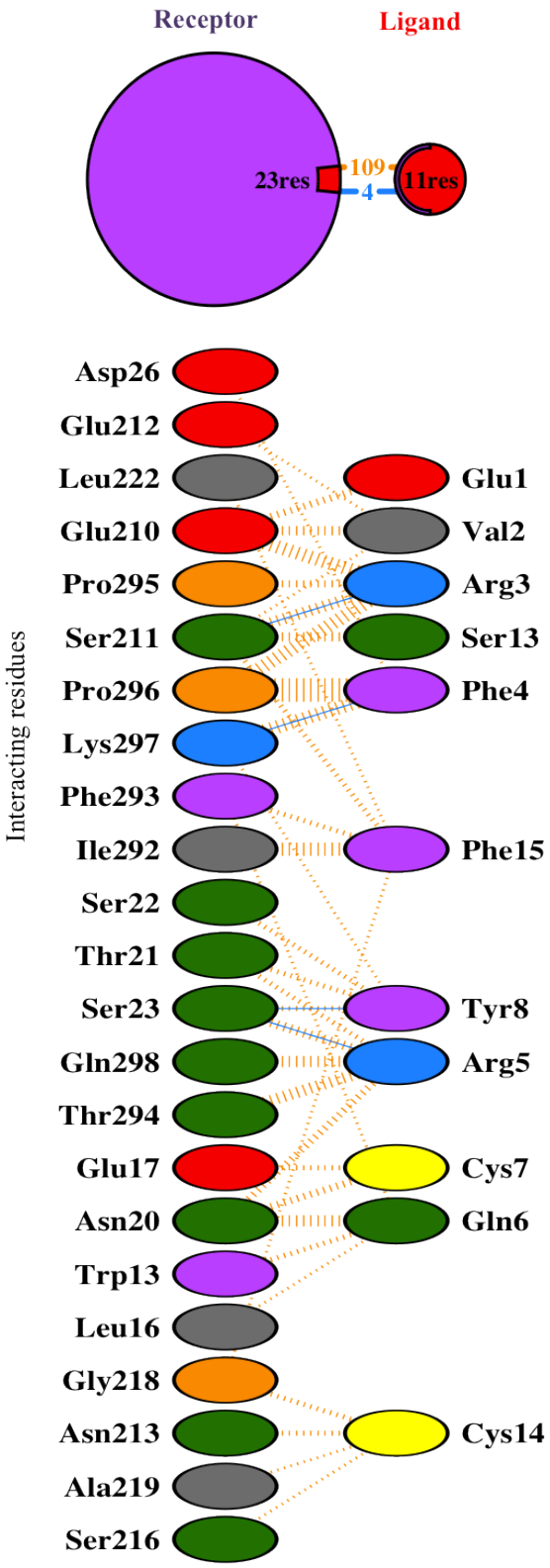

Figure 5.24. The interacting residues between receptor and ligand were extracted in PDBsum. Blue lines represent H-bonds, yellow dashes represent non-bonded contacts.

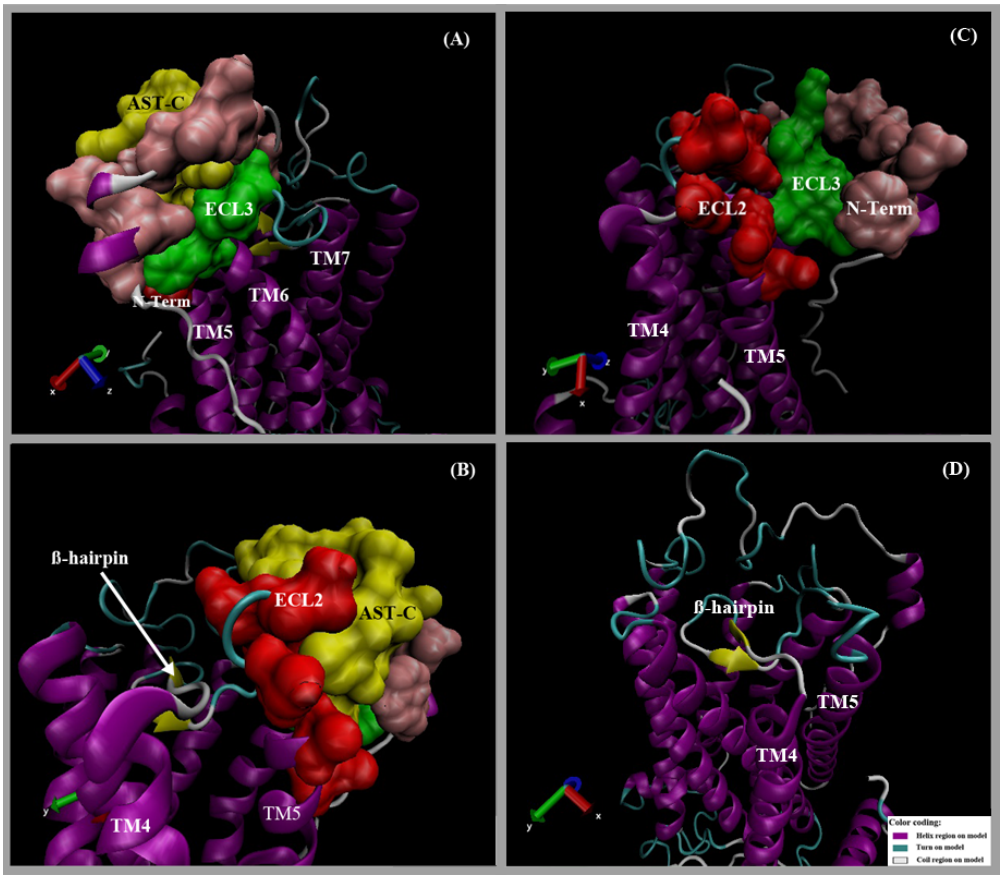

Figure 5.25. The binding pocket is represented according to the interaction data obtained in PDBsum. (A), (B) and (C) are different views at the same pocket, only (C) does not include AST-C ligand. β-hairpin structure on ECL2 is represented in (D).

6. DISCUSSION

In most of disease cases that result from differential expression of variable proteins; solution and treatment strategies are very complicated. This is because any factor (ex. protein, metabolite, structural changes etc.) in the cell can behave differentially in a time-dependent and amount-dependent manner, or can cross-react with other factors. After some effort, this phenomenon becomes like an unpredictable "soup" for medical scientists. However, it is easier to add an extrinsic factor into this "soup" and predict its effects, mostly in disease treatment studies. In recent days of scientific progress in drug design, structure based studies became very popular for the hot spot needs of medical approaches. And with the advance in successive bioinformatics tools and sensitive single-molecule approaches, interpretation of functional genomics is getting more tangible. In this thesis, both structure-based predictions and single-molecule dynamics are combined to understand a membrane molecule that would also facilitate and form a basis for physiological and drug design studies for human health. Combining the needs in drug design and advances in these research approaches, Allatostatins are valuable extrinsic factors which may not cross-activate unwanted pathways in human cells because they are insect-specific. For instance, their understanding can answer why insects do not produce cholesterol [2]. Their extrinsic expression can facilitate silencing studies in neurons [21]. Also synthetic modifications on these receptors and ligands may replace somatostatin analogs used in cancer treatments. As a result, studying nonhuman-specific factors with the help of structure-based and single-molecule based approaches represents the worth of this thesis project.

In the progress of this thesis project a somatostatin-like receptor was detected in *C. morosus*. Previously amplified conserved TM region was taken as the template in RACE reactions. As depicted in Results section, 5' and 3' end amplifications of this cDNA exhibited unspecific fragments. The strongest three bands in 3' RACE were cloned and sequenced Only the longest amplification was found to be a part of 7TM GPCR sequence, according to Blastn search. The other two fragments gave no GPCR hits in their alignments. The length of this fragment was close to 800 bp. In

5' RACE reactions thes strongest two bands were cloned and sequenced. According to Blastn search only the longest fragment (nearly 800 bp) was found to be a part of 7TM GPCRs. After combining these fragments, the longest ORF was predicted and untranslated regions were removed. At this point, it was also difficult to obtain the sequence with no mutation. The cloning procedure for ORF was repeated 12 times to compare the sequences coming from different sample tubes, via various PCR reactions. The consensus of these products was taken and given in the Results section. One of the PCR amplifications which was known to include no mutation with regard to the 1326 bp-long consensus ORF sequence, was cloned for the following *in vivo* studies.

After obtaining the longest wild-type ORF, the translated product could be compared with the other insect receptors in sequence databases. Phylogenetic tree of AlstR types (Figure 5.13), which was obtained only from their conserved TM regions, showed differentiation of A, B and C types within Insecta class. In this analysis *D. pulex* AlstR-C was included as an out-group, from Crustaceans. A and C-type of AlstRs were more closer and B-type AlstRs has diverged before A and C-type differentiation. The tree was compatible with the closer relationships within suborders. As an example, *Drosophila* and *Anopheles* homologs have diverged in the same branch, as they were both Dipterans. The other branches were also compatible with taxonomical classifications. According to this tree, CamAlstR-C has diverged before the evolution of other Neopteran suborders and just after evolution of Hemipteran *Acyrthosiphon* AlstR-C. Being a Crustacean, *D. pulex* AlstR-C was totally different from Neopteran AlstR-Cs. However, divergence patterns of A-type and B-type AlstRs were different, but also compatible with the taxonomical suborders. For example, B-type receptors have differentiated into two main branches of Hemiptera-Hymenoptera and Coleoptera-Diptera-Lepidoptera at the beginning. A-type AlstRs differentiated into Hymenopteran receptors at first and than the others diverged from the other branch. Thus we could not say that different AlstR paralogs co-evolved and diverged in the same order. *C. morosus* belongs to Phasmatodea order, being closer to Dictyopteran *Periplaneta* and Hemipteran *Acyrthosiphon*. Within AlstR-C homologs, CamAlstR-C seemed to evolve after divergence of *Acyrthosiphon* AlstR-C. The reason of absence of *Periplaneta* AlstR-C in this analysis was its' absence in sequence databases. In addition, we could assume

that if an A-type AlstR was found in *C. morosus*, its' evolution would be represented in the branch of APeriplaneta and AAcyrthosiphon.

Multiple alignment studies on N-terminal region of AlstR-A resulted in another notable conservation. Figure 5.14 represented a genus-specific conservation pattern between 3 genera, and CamAlstR-C showed a significant diversity in this profile. Previous studies on AST ligands showed that these peptides could cross-activate their relevant receptors in different species. So we could conclude that N-terminal regions of these receptors might not define ligand-specific activity, but they must assign for 3D conformation and rigidity of binding pocket so that become responsible in the efficiency of ligand interaction.

In order to understand structural characteristics of CamAlstR-C, its' secondary structural features were illustrated in Figure 5.19. It exhibited a DRY motif in TM3, NPILY in TM7 and WLP motif in TM6, which were conserved within Family 1 rhodopsin-like GPCRs. Also the disulfide bridge between C127 of TM3 and C204 of ECL2 was typical for this family. Rhodopsin structure contains a β-hairpin structure on ECL2 and this forms a closed pocket for retinal ligand. The same structure was conserved in CamAlstR-C, forming a pocket-like vacancy on the ECL region, such as in CXCR4. The other secondary structure features showed some functional motifs in accordance with SSTRs. For example; cAMP and cGMP dependent phosphorylation signals, PKC phosphorylation signals and Casein kinase II phosphorylation signals were detected in the intracellular regions of CamAlstR-C. These motifs would be clues for future studies on downstream pathways. SSTRs were known to couple with pertussis toxin-sensitive G proteins which inhibited adenylate cyclase activity and lead to inhibition of cAMP accumulation. As Kumar reviewed for SSTR motifs, they could also activate PKC-dependent and Casein kinase II-dependent pathways [34].

In 3D model construction of CamAlstR-C, the most difficult part was to find a significant homology for N-terminal region. In Figure 5.20, threaded templates of I-TASSER showed insignificant alignments for N-terminus. This meant that this program used *ab initio* energy calculations to model this region. Within the model outputs

of CamAlstR-C the ones in which N-terminal loops did not disrupt membrane region and formed more globular shape, were chosen for docking. In comparison with the sheltering N-terminal region of Palczewski's rhodopsin structure, CamAlstR-C N-terminus behaved similar to this shelter on top of the receptor but much more freely flanking towards the membrane. The beginning portion of N-terminal loop came closer to ECL2 and ECL3, and played role in interaction with the ligand (Figure 5.25c). As a result of its' low conservation within insect species and in combination with docking results, this loop can be assumed to have a role in strengthening the binding pocket.

AST-C ligands were previously proved to include a disulfide bridge [76]. Our ligand model was prepared according to this finding, and the distance between two Cysteine residues was restricted in the server. The output model with the closest distance was taken for docking analysis. In this docking analysis ligand was supposed to behave like a rigid molecule while binding to the pocket.

In the results of docking analysis (Figure 5.24), the binding pocket was found to include some residues from ECL2, ECL3 and N-terminus (Figure 5.25a, b and c). At least four H-bonds were detected; two of these were formed with N-terminal polar Serine residue and with Lysine and Serine of ECL3. N-terminal part of ligand constituted these stronger interactions (Figure 5.24). However Phenylalanine of C-terminal part of ligand seemed as important for strength of this interaction. If we compare AST-A (FGLamides) ligands with our AST-C (PISCF) ligand we could also predict that the C-terminal Phenylalanine played an efficient role in both ligands. In accordance with this prediction, we were able to bind Allatostatin IV ligand, which was AST-A type, with our receptor in single molecule level (Figure 5.10). Allatostatin IV of *M. sexta* showed a interaction force of approximately 100 pN at 1 μm/s, which was in the normal range of peptide interactions. Thus, Allatostatin interactions might depend on some functional groups other than on C-terminal motifs.

After obtaining the interaction results on CamAlstR-C, these regions were analyzed for conservation. Table 5.4 summarizes conservation of regions. Only the residues between 292 and 298 on ECL3 contributed to a common pattern. For example, I292,

T294 and P296 on this region were approximately 100% conserved within AlstR-C types of other insects. The residues between 210 and 222 on ECL2 were less conserved. Totally conserved residues on ECL2 were N205, I206, W208 and P209 (NIXWP), but they were not interacting in docking analysis. Thus this conserved NIXWP motif was supposed to have an important role in conformational exposition of the following residues (ESENVSSQGAFTL) for interaction. N-terminal interaction region from 13th to 26th residues showed very low conservation as could be predicted from N-terminal alignments within insect AlstRs. In conclusion, IXTXP motif on ECL3 region of CamAlstR-C could be predicted as a conserved binding motif for other insect AlstRs.

Besides an understanding on CamAlstR-C interaction with its' ligand, this thesis project was also successful about using a novel application system of atomic force microscopy (AFM). Direct measurement of receptor-ligand interactions is still greatly limited because of the difficulty in working membrane proteins. Mostly indirect biophysical approaches were in use. Finally in 2009, Shi *et al.* tried to fractionate the membrane and used it in AFM to show the oligomerization state of Ste2p protein [77]. However, in our study we were able to measure a binding case directly on the membrane of a living cell. The raw data gave us the potential coming from the diffraction of laser from the bending cantilever. We converted this potential change into the form of binding force between CamAlstR-C receptor and AST-A ligand. As illustrated in Figure 5.10, the binding forces were compatible with the previous results of typical receptor-ligand interactions measured *in vitro* [78]. Because there were other receptor types on the cell membrane, we tried to be sure if these interaction curves were specific for our receptor. We checked the interaction results for non-transfected cells and saw no adhesion curves. We also checked the interaction after saturating the receptors with excessive ligands and saw no adhesion curves (Figure 5.9). Still there are other questions about working on living cells, such as whether the flexibility of membrane might change the interaction curves or whether the ligand could extend like a spring while applying the retraction force etc. Despite disruption in the adhesion curves may occur due to these unanswered questions, we were able to get the significant evidence of interaction between these two molecules on the surface of a living cell. As a re-

sult, CamAlstR-C and AST-A interaction became a candidate system to study AFM directly in living systems.

For further studies on CamAlstR-C, the significance of proposed binding pocket can be analyzed in AFM experiments. Performing mutations/deletions on these predicted sites can help to ensure its' binding properties. And detailed understanding on binding characteristics of AlstRs can facilitate synthetic drug productions. If this project is combined with other molecular studies on effects of this receptor-ligand system on human cells, a plausible treatment for cancer can be also proposed.

APPENDIX A: PLASMID MAPS AND MULTIPLE CLONING SITES

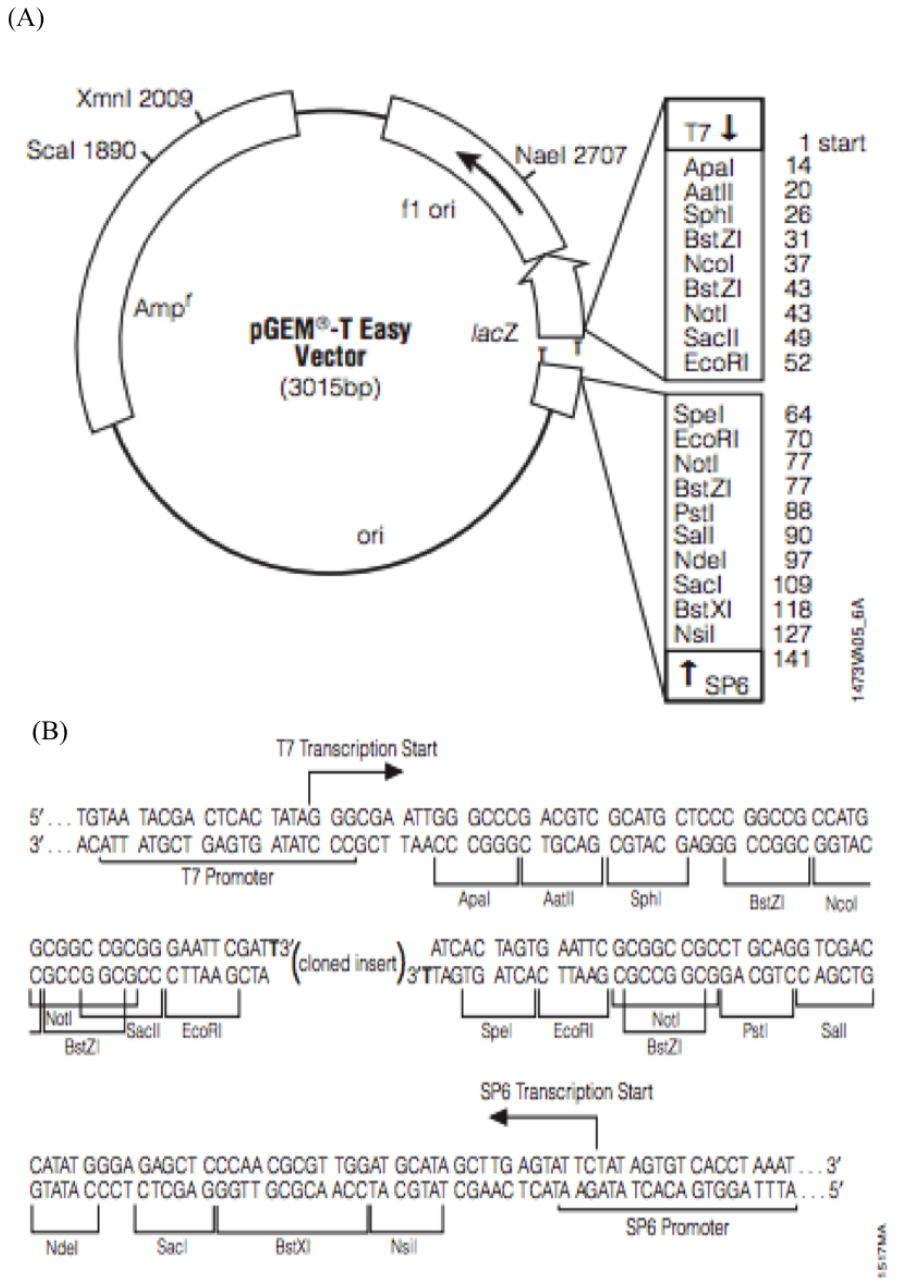

Figure A.1. pGEMT-Easy vector map (A) and multiple cloning site (B).

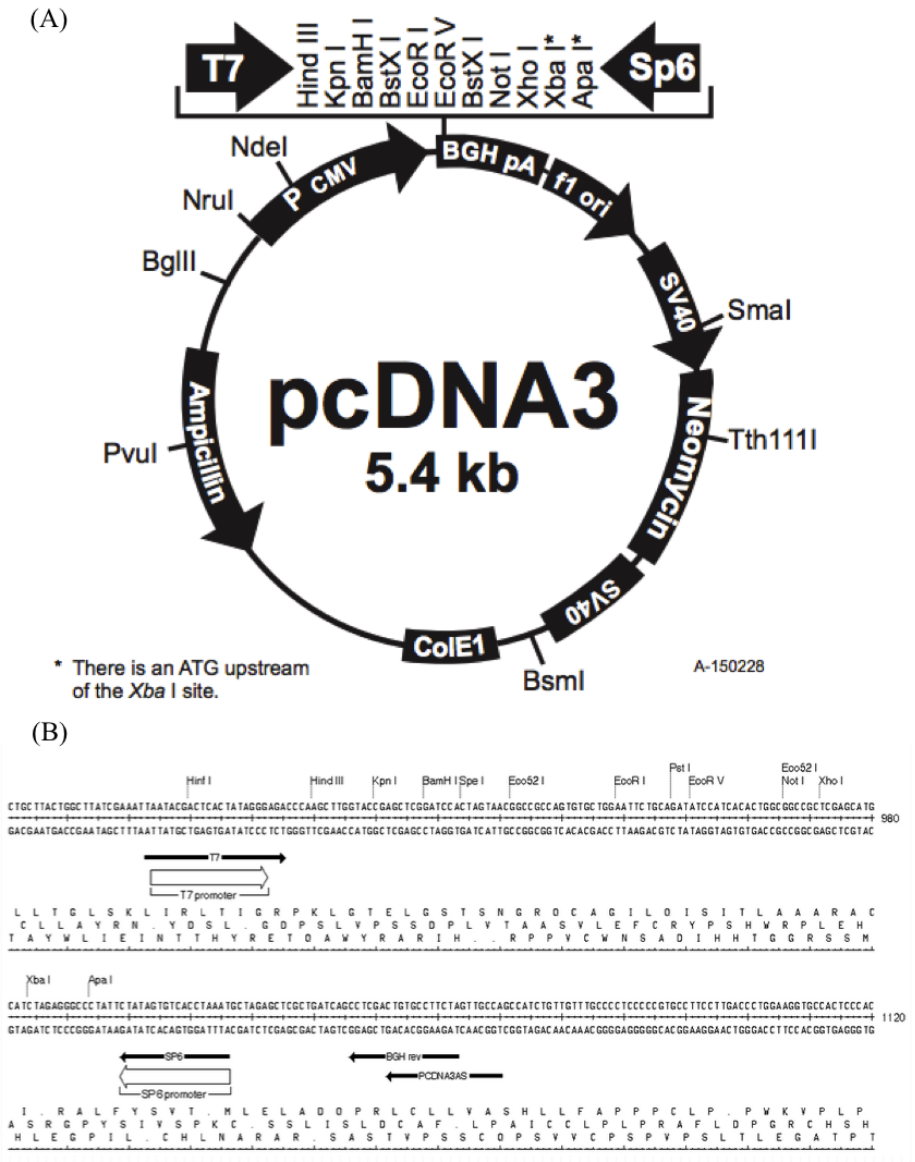

Figure A.2. pcDNA3 vector map (A) and multiple cloning site (B).

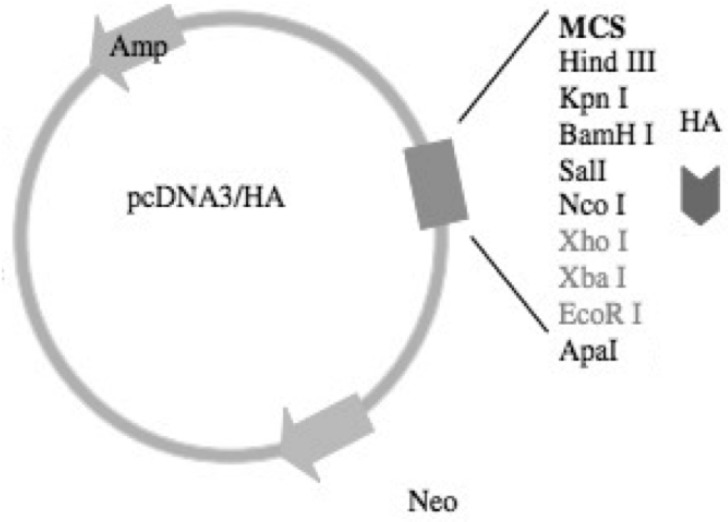

Figure A.3. Modified pcDNA3/HA vector map

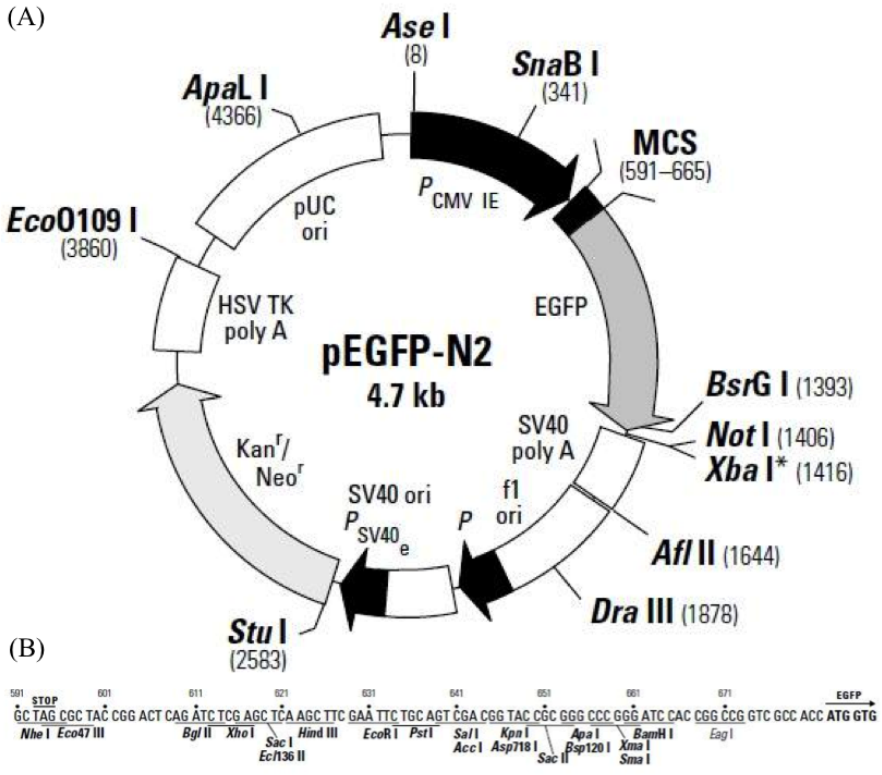

Figure A.4. pEGFP-N2 vector map (A) and multiple cloning site (B).

APPENDIX B: DNA LADDERS

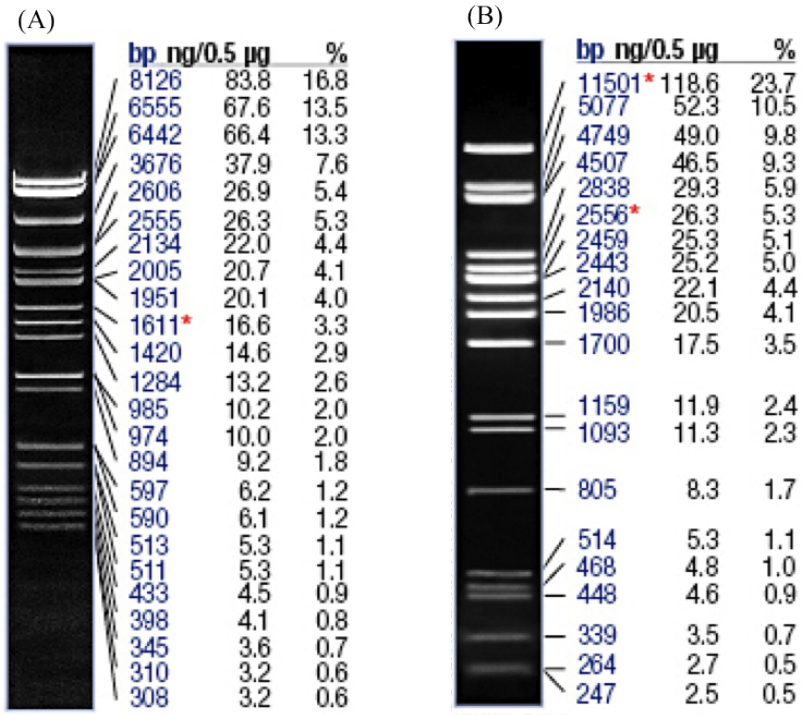

Figure B.1. (A) Marker 13 from Fermentas, (B) Marker 24 from Fermentas.

APPENDIX C: TMHMM PROBABILITIES

Table C.1. TMHMM probabilities.

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
1	M	0.065370	0.000000	0.934630	33	N	0.065340	0.000000	0.934650
2	S	0.065360	0.000010	0.934630	34	G	0.065340	0.000000	0.934650
3	V	0.065360	0.000010	0.934630	35	S	0.065340	0.000000	0.934650
4	E	0.065360	0.000010	0.934630	36	A	0.065340	0.000000	0.934650
5	Q	0.065360	0.000010	0.934630	37	H	0.065340	0.000000	0.934650
6	V	0.065360	0.000010	0.934630	38	G	0.065340	0.000000	0.934650
7	T	0.065350	0.000020	0.934630	39	N	0.065340	0.000000	0.934650
8	V	0.065350	0.000020	0.934630	40	M	0.065340	0.000000	0.934650
9	M	0.065350	0.000020	0.934630	41	S	0.065340	0.000000	0.934650
10	V	0.065350	0.000020	0.934630	42	L	0.065340	0.000000	0.934650
11	T	0.065350	0.000020	0.934630	43	D	0.065340	0.000000	0.934650
12	S	0.065350	0.000020	0.934630	44	H	0.065340	0.000000	0.934650
13	W	0.065350	0.000020	0.934630	45	H	0.065340	0.000000	0.934650
14	S	0.065350	0.000020	0.934630	46	E	0.065340	0.000000	0.934650
15	T	0.065350	0.000020	0.934630	47	C	0.065340	0.000010	0.934650
16	L	0.065350	0.000020	0.934630	48	D	0.065340	0.000040	0.934610
17	E	0.065350	0.000020	0.934640	49	T	0.065340	0.001030	0.933640
18	D	0.065350	0.000020	0.934640	50	D	0.065340	0.002830	0.931840
19	L	0.065350	0.000020	0.934640	51	Q	0.065320	0.008720	0.925960
20	N	0.065340	0.000020	0.934640	52	P	0.064960	0.023100	0.911940
21	T	0.065340	0.000020	0.934640	53	T	0.063860	0.045470	0.890670
22	S	0.065340	0.000010	0.934640	54	L	0.060530	0.084740	0.854730
23	S	0.065340	0.000010	0.934640	55	L	0.055470	0.212700	0.731830
24	T	0.065340	0.000010	0.934640	56	L	0.052910	0.283400	0.663690
25	T	0.065340	0.000010	0.934640	57	I	0.051250	0.530310	0.418440
26	D	0.065340	0.000010	0.934650	58	T	0.050340	0.617060	0.332590
27	P	0.065340	0.000010	0.934650	59	Q	0.046550	0.811520	0.141920
28	L	0.065340	0.000010	0.934650	60	V	0.026510	0.917900	0.055590
29	L	0.065340	0.000010	0.934650	61	L	0.004280	0.980990	0.014720
30	E	0.065340	0.000010	0.934650	62	C	0.002710	0.988690	0.008590
31	A	0.065340	0.000000	0.934650	63	A	0.001830	0.995070	0.003100
32	G	0.065340	0.000000	0.934650	64	A	0.001010	0.998550	0.000440

Table C.2. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
65	V	0.000650	0.999220	0.000120	98	A	0.052200	0.927570	0.020230
66	C	0.000540	0.999380	0.000080	99	V	0.050820	0.929330	0.019850
67	I	0.000480	0.999490	0.000030	100	A	0.049910	0.930740	0.019360
68	V	0.000480	0.999490	0.000020	101	D	0.049550	0.931180	0.019280
69	G	0.000480	0.999490	0.000020	102	E	0.040700	0.941480	0.017820
70	L	0.000490	0.999490	0.000020	103	C	0.023610	0.967560	0.008840
71	L	0.000540	0.999430	0.000030	104	F	0.017490	0.978010	0.004500
72	G	0.002840	0.997020	0.000140	105	L	0.017120	0.981620	0.001260
73	N	0.012160	0.987440	0.000400	106	I	0.017070	0.982200	0.000730
74	S	0.016090	0.983430	0.000480	107	G	0.017070	0.982230	0.000690
75	L	0.018000	0.981450	0.000560	108	V	0.017070	0.982250	0.000680
76	V	0.022480	0.976610	0.000910	109	P	0.017240	0.981230	0.001540
77	I	0.042830	0.954590	0.002580	110	F	0.017320	0.980260	0.002420
78	Y	0.110180	0.882190	0.007630	111	L	0.017660	0.975440	0.006900
79	V	0.127700	0.861410	0.010890	112	L	0.019210	0.956190	0.024600
80	V	0.355890	0.630180	0.013920	113	A	0.023050	0.913430	0.063520
81	L	0.546300	0.436690	0.017010	114	T	0.027830	0.855820	0.116350
82	R	0.866600	0.109900	0.023500	115	M	0.036080	0.507630	0.456300
83	F	0.904540	0.055540	0.039920	116	N	0.037840	0.165100	0.797060
84	S	0.927480	0.009510	0.063010	117	L	0.038830	0.131040	0.830130
85	K	0.933840	0.001400	0.064770	118	R	0.039890	0.124350	0.835760
86	M	0.934610	0.000650	0.064740	119	I	0.040080	0.144420	0.815500
87	Q	0.934920	0.000360	0.064720	120	W	0.044960	0.116260	0.838780
88	T	0.933540	0.001960	0.064510	121	P	0.061900	0.086840	0.851270
89	V	0.921100	0.015270	0.063640	122	F	0.063440	0.092640	0.843920
90	T	0.910660	0.027060	0.062280	123	G	0.071420	0.084090	0.844490
91	N	0.889950	0.050880	0.059170	124	S	0.072410	0.085390	0.842200
92	M	0.586850	0.364950	0.048200	125	I	0.073930	0.117960	0.808110
93	Y	0.187990	0.767410	0.044600	126	M	0.077420	0.116330	0.806250
94	I	0.103750	0.853390	0.042860	127	C	0.079720	0.115220	0.805060
95	V	0.088630	0.869080	0.042290	128	K	0.081290	0.127830	0.790880
96	N	0.086030	0.871960	0.042020	129	A	0.081350	0.318610	0.600040
97	L	0.061210	0.902930	0.035860	130	Y	0.081370	0.498080	0.420550

Table C.3. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
131	M	0.081430	0.516910	0.401660	164	K	0.994320	0.000400	0.005270
132	T	0.081520	0.523790	0.394690	165	V	0.994270	0.000460	0.005270
133	T	0.081610	0.530320	0.388070	166	R	0.994450	0.000280	0.005270
134	T	0.081660	0.696170	0.222180	167	T	0.993210	0.001510	0.005270
135	S	0.081870	0.761310	0.156810	168	P	0.960650	0.034080	0.005270
136	I	0.081970	0.782440	0.135590	169	L	0.692800	0.301950	0.005250
137	N	0.084730	0.788770	0.126500	170	I	0.632270	0.362500	0.005230
138	Q	0.086710	0.817840	0.095450	171	S	0.613680	0.381100	0.005220
139	F	0.088330	0.874650	0.037020	172	K	0.607540	0.387250	0.005210
140	T	0.097940	0.882900	0.019160	173	V	0.281110	0.714300	0.004590
141	S	0.101100	0.883880	0.015020	174	V	0.131240	0.864390	0.004380
142	S	0.111840	0.876840	0.011320	175	S	0.073480	0.922230	0.004290
143	I	0.113620	0.880390	0.005990	176	A	0.030100	0.965830	0.004070
144	F	0.115550	0.879130	0.005320	177	S	0.012950	0.983060	0.003990
145	L	0.127980	0.866880	0.005140	178	A	0.006190	0.989920	0.003880
146	T	0.146790	0.848090	0.005120	179	W	0.002980	0.993170	0.003840
147	I	0.167700	0.827190	0.005110	180	T	0.001580	0.994590	0.003830
148	M	0.285510	0.709390	0.005100	181	A	0.000710	0.995460	0.003830
149	S	0.443240	0.551650	0.005100	182	S	0.000480	0.995690	0.003830
150	A	0.525580	0.469310	0.005110	183	A	0.000320	0.995850	0.003830
151	D	0.576330	0.418550	0.005120	184	L	0.000250	0.995930	0.003830
152	R	0.588480	0.406400	0.005110	185	L	0.000160	0.996000	0.003840
153	Y	0.600850	0.394040	0.005110	186	M	0.000150	0.995970	0.003880
154	I	0.610910	0.383980	0.005110	187	I	0.000080	0.995700	0.004220
155	A	0.626290	0.368580	0.005130	188	P	0.000070	0.993680	0.006250
156	V	0.639210	0.355630	0.005160	189	I	0.000050	0.992300	0.007650
157	C	0.883570	0.111200	0.005230	190	F	0.000020	0.988990	0.010990
158	H	0.919000	0.075760	0.005240	191	I	0.000020	0.954490	0.045490
159	P	0.921580	0.073180	0.005240	192	Y	0.000060	0.692540	0.307410
160	I	0.921850	0.072910	0.005240	193	A	0.000140	0.569920	0.429950
161	S	0.946290	0.048460	0.005240	194	N	0.000230	0.495440	0.504330
162	A	0.974340	0.020410	0.005250	195	I	0.000290	0.483690	0.516030
163	P	0.993280	0.001460	0.005260	196	M	0.001020	0.215140	0.783850

Table C.4. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
197	T	0.001310	0.075660	0.923030	230	S	0.000000	0.999980	0.000010
198	N	0.001370	0.021880	0.976750	231	I	0.000000	1,000000	0.000000
199	E	0.001440	0.004140	0.994420	232	P	0.000000	1,000000	0.000000
200	N	0.001440	0.002440	0.996120	233	L	0.000000	1,000000	0.000000
201	V	0.001440	0.002250	0.996310	234	M	0.000000	1,000000	0.000000
202	V	0.001450	0.001930	0.996610	235	L	0.000000	1,000000	0.000000
203	S	0.001460	0.000660	0.997880	236	I	0.000010	0.999990	0.000000
204	C	0.001460	0.000400	0.998130	237	S	0.000090	0.999910	0.000000
205	N	0.001460	0.000350	0.998190	238	V	0.000180	0.999820	0.000000
206	I	0.001460	0.000340	0.998190	239	F	0.000850	0.999150	0.000000
207	L	0.001460	0.000300	0.998230	240	Y	0.006200	0.993790	0.000000
208	W	0.001460	0.000170	0.998360	241	I	0.017550	0.982440	0.000010
209	P	0.001470	0.000130	0.998400	242	L	0.056830	0.943110	0.000060
210	E	0.001470	0.000020	0.998500	243	V	0.100580	0.899280	0.000130
211	S	0.001470	0.000020	0.998500	244	I	0.278420	0.721310	0.000270
212	E	0.001470	0.000020	0.998510	245	R	0.907480	0.091130	0.001380
213	N	0.001470	0.000150	0.998370	246	K	0.985900	0.012650	0.001460
214	V	0.001470	0.000440	0.998090	247	L	0.988150	0.010390	0.001460
215	S	0.001470	0.001150	0.997380	248	K	0.997520	0.001010	0.001470
216	S	0.001470	0.002290	0.996240	249	T	0.998250	0.000280	0.001470
217	Q	0.001470	0.007060	0.991480	250	V	0.998490	0.000040	0.001470
218	G	0.001450	0.020820	0.977730	251	G	0.998530	0.000010	0.001470
219	A	0.001390	0.074230	0.924380	252	P	0.998530	0.000000	0.001470
220	F	0.001320	0.127980	0.870690	253	K	0.998530	0.000000	0.001470
221	T	0.001220	0.284770	0.714010	254	N	0.998530	0.000000	0.001470
222	L	0.000260	0.825010	0.174720	255	K	0.998530	0.000000	0.001470
223	Y	0.000160	0.918850	0.080980	256	S	0.998530	0.000000	0.001470
224	T	0.000140	0.938990	0.060870	257	K	0.998530	0.000000	0.001470
225	F	0.000030	0.988750	0.011220	258	E	0.998530	0.000000	0.001470
226	I	0.000010	0.998030	0.001970	259	K	0.998530	0.000000	0.001470
227	L	0.000000	0.999490	0.000510	260	K	0.998530	0.000000	0.001470
228	G	0.000000	0.999860	0.000140	261	K	0.998530	0.000000	0.001470
229	F	0.000000	0.999940	0.000050	262	S	0.998390	0.000150	0.001470

Table C.5. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
263	H	0.995910	0.002630	0.001470	296	P	0.001560	0.000270	0.998170
264	R	0.985590	0.012970	0.001430	297	K	0.001570	0.000030	0.998390
265	K	0.787790	0.210790	0.001430	298	Q	0.001570	0.000580	0.997840
266	V	0.664250	0.334390	0.001350	299	C	0.001570	0.000940	0.997500
267	T	0.642070	0.356610	0.001310	300	Q	0.001570	0.002240	0.996200
268	K	0.622810	0.375900	0.001290	301	S	0.001560	0.007110	0.991330
269	L	0.283980	0.715320	0.000700	302	R	0.001560	0.014900	0.983550
270	V	0.015740	0.983810	0.000460	303	V	0.001510	0.049930	0.948560
271	L	0.006940	0.992820	0.000240	304	S	0.001460	0.070240	0.928310
272	T	0.006190	0.993690	0.000120	305	I	0.001190	0.099670	0.899130
273	V	0.001140	0.998830	0.000020	306	T	0.000980	0.167740	0.831270
274	I	0.000270	0.999720	0.000000	307	V	0.000250	0.597900	0.401850
275	T	0.000160	0.999840	0.000000	308	S	0.000190	0.634720	0.365090
276	V	0.000120	0.999870	0.000000	309	L	0.000180	0.636660	0.363160
277	Y	0.000110	0.999890	0.000000	310	L	0.000170	0.638530	0.361300
278	V	0.000100	0.999890	0.000000	311	A	0.000170	0.640460	0.359360
279	L	0.000100	0.999890	0.000000	312	G	0.000170	0.641290	0.358530
280	C	0.000100	0.999890	0.000010	313	C	0.000170	0.641590	0.358230
281	W	0.000100	0.999880	0.000020	314	L	0.000180	0.641750	0.358070
282	L	0.000100	0.999790	0.000110	315	S	0.000190	0.641770	0.358040
283	P	0.000110	0.999350	0.000540	316	Y	0.000220	0.641780	0.357990
284	Y	0.000110	0.998290	0.001600	317	S	0.000540	0.641470	0.357990
285	W	0.000110	0.994960	0.004930	318	N	0.001410	0.640600	0.357980
286	L	0.000160	0.922920	0.076910	319	S	0.003320	0.638710	0.357970
287	T	0.000650	0.852610	0.146750	320	A	0.005800	0.636230	0.357970
288	Q	0.000800	0.557180	0.442030	321	M	0.011060	0.630970	0.357970
289	M	0.000830	0.453930	0.545250	322	N	0.012470	0.629550	0.357980
290	A	0.000830	0.436660	0.562510	323	P	0.013490	0.628530	0.357980
291	L	0.000830	0.427690	0.571470	324	I	0.013740	0.628280	0.357980
292	I	0.000890	0.312520	0.686590	325	L	0.016180	0.625830	0.357990
293	F	0.001080	0.021530	0.977390	326	Y	0.045990	0.595970	0.358030
294	T	0.001330	0.005690	0.992980	327	A	0.071030	0.570880	0.358090
295	P	0.001480	0.001860	0.996660	328	F	0.089970	0.551680	0.358350

Table C.6. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
329	L	0.149930	0.491500	0.358570	362	R	0.640630	0.000000	0.359370
330	S	0.592950	0.047770	0.359280	363	R	0.640630	0.000000	0.359370
331	D	0.639700	0.000940	0.359360	364	H	0.640630	0.000000	0.359370
332	N	0.640040	0.000610	0.359360	365	H	0.640630	0.000000	0.359370
333	F	0.640170	0.000470	0.359360	366	R	0.640630	0.000000	0.359370
334	K	0.640600	0.000050	0.359360	367	Q	0.640630	0.000000	0.359370
335	K	0.640610	0.000030	0.359360	368	G	0.640630	0.000000	0.359370
336	S	0.640600	0.000050	0.359350	369	G	0.640630	0.000000	0.359370
337	F	0.640600	0.000050	0.359350	370	S	0.640630	0.000000	0.359370
338	L	0.640600	0.000050	0.359350	371	E	0.640630	0.000000	0.359370
339	K	0.640620	0.000030	0.359350	372	R	0.640630	0.000000	0.359370
340	A	0.640620	0.000030	0.359350	373	Q	0.640630	0.000000	0.359370
341	C	0.640620	0.000030	0.359350	374	R	0.640630	0.000000	0.359370
342	T	0.640620	0.000030	0.359350	375	P	0.640630	0.000000	0.359370
343	C	0.640620	0.000030	0.359350	376	N	0.640630	0.000000	0.359370
344	A	0.640620	0.000030	0.359350	377	K	0.640630	0.000000	0.359370
345	T	0.640620	0.000030	0.359350	378	M	0.640630	0.000000	0.359370
346	G	0.640620	0.000030	0.359350	379	A	0.640630	0.000000	0.359370
347	Q	0.640620	0.000020	0.359350	380	T	0.640630	0.000000	0.359370
348	D	0.640620	0.000020	0.359350	381	L	0.640630	0.000000	0.359370
349	V	0.640620	0.000020	0.359350	382	E	0.640630	0.000000	0.359370
350	N	0.640620	0.000020	0.359350	383	T	0.640630	0.000000	0.359370
351	N	0.640620	0.000020	0.359350	384	G	0.640630	0.000000	0.359370
352	V	0.640620	0.000020	0.359350	385	E	0.640630	0.000000	0.359370
353	L	0.640620	0.000020	0.359350	386	G	0.640630	0.000000	0.359370
354	H	0.640630	0.000010	0.359350	387	G	0.640630	0.000000	0.359370
355	M	0.640630	0.000010	0.359350	388	E	0.640630	0.000000	0.359370
356	E	0.640630	0.000000	0.359370	389	K	0.640630	0.000000	0.359370
357	N	0.640630	0.000000	0.359370	390	G	0.640630	0.000000	0.359370
358	S	0.640630	0.000000	0.359370	391	P	0.640630	0.000000	0.359370
359	V	0.640630	0.000000	0.359370	392	L	0.640630	0.000000	0.359370
360	F	0.640630	0.000000	0.359370	393	V	0.640630	0.000000	0.359370
361	P	0.640630	0.000000	0.359370	394	S	0.640630	0.000000	0.359370

Table C.7. TMHMM probabilities, continued

Res. number	Res. name	Inside	Helix	Outside	Res. number	Res. name	Inside	Helix	Outside
395	K	0.640630	0.000000	0.359370	419	D	0.640630	0.000000	0.359370
396	A	0.640630	0.000000	0.359370	420	Q	0.640630	0.000000	0.359370
397	D	0.640630	0.000000	0.359370	421	A	0.640630	0.000000	0.359370
398	S	0.640630	0.000000	0.359370	422	S	0.640630	0.000000	0.359370
399	S	0.640630	0.000000	0.359370	423	K	0.640630	0.000000	0.359370
400	V	0.640630	0.000000	0.359370	424	D	0.640630	0.000000	0.359370
401	T	0.640630	0.000000	0.359370	425	A	0.640630	0.000000	0.359370
402	T	0.640630	0.000000	0.359370	426	A	0.640630	0.000000	0.359370
403	M	0.640630	0.000000	0.359370	427	A	0.640630	0.000000	0.359370
404	T	0.640630	0.000000	0.359370	428	M	0.640630	0.000000	0.359370
405	M	0.640630	0.000000	0.359370	429	E	0.640630	0.000000	0.359370
406	T	0.640630	0.000000	0.359370	430	T	0.640630	0.000000	0.359370
407	S	0.640630	0.000000	0.359370	431	A	0.640630	0.000000	0.359370
408	R	0.640630	0.000000	0.359370	432	L	0.640630	0.000000	0.359370
409	S	0.640630	0.000000	0.359370	433	K	0.640630	0.000000	0.359370
410	N	0.640630	0.000000	0.359370	434	N	0.640630	0.000000	0.359370
411	I	0.640630	0.000000	0.359370	435	G	0.640630	0.000000	0.359370
412	T	0.640630	0.000000	0.359370	436	Q	0.640630	0.000000	0.359370
413	V	0.640630	0.000000	0.359370	437	Q	0.640630	0.000000	0.359370
414	S	0.640630	0.000000	0.359370	438	P	0.640630	0.000000	0.359370
415	T	0.640630	0.000000	0.359370	439	T	0.640630	0.000000	0.359370
416	D	0.640630	0.000000	0.359370	440	Q	0.640630	0.000000	0.359370
417	A	0.640630	0.000000	0.359370	441	V	0.640630	0.000000	0.359370
418	K	0.640630	0.000000	0.359370					

APPENDIX D: ATOMIC FORCE MICROSCOPY FORCE CURVES

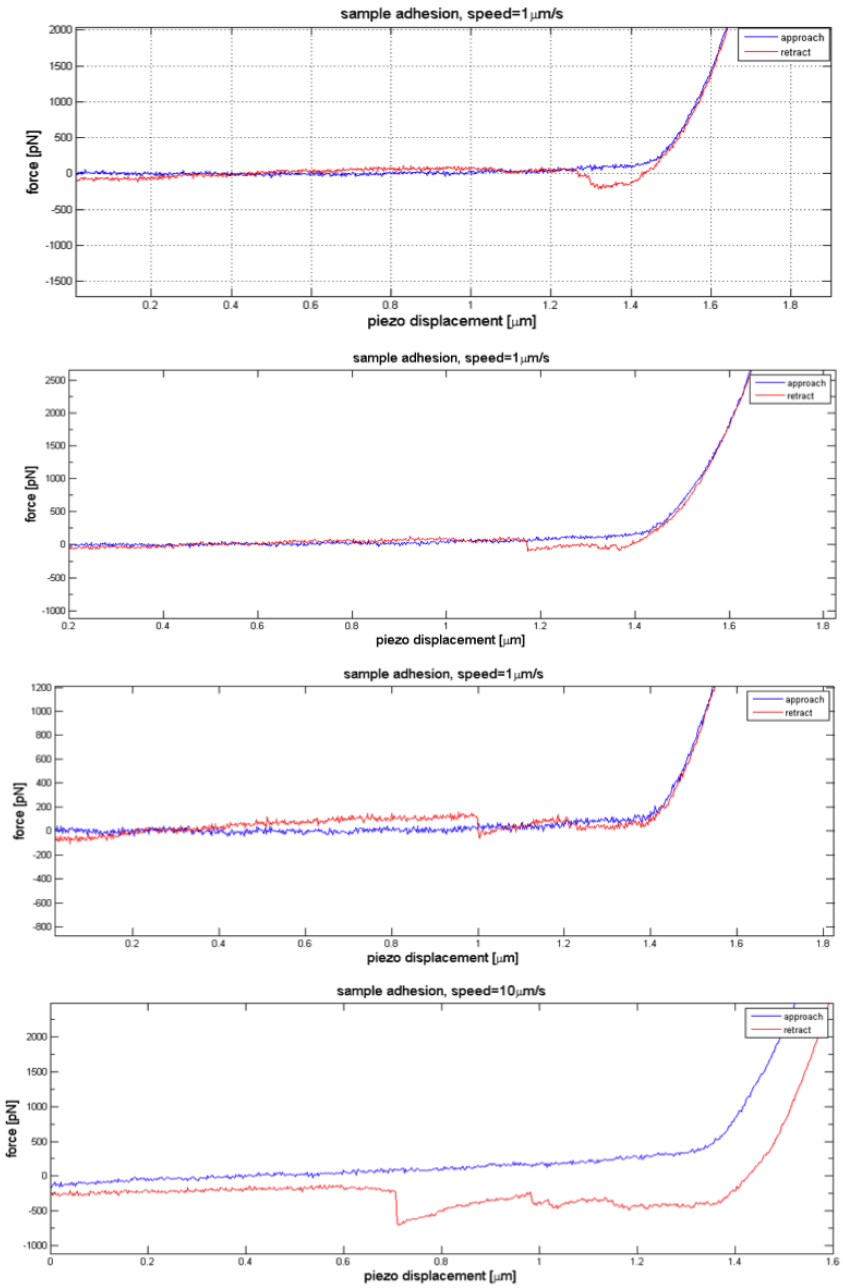

Figure D.1. Adhesion force curves obtained by different speeds of cantilever approach on the same cell.

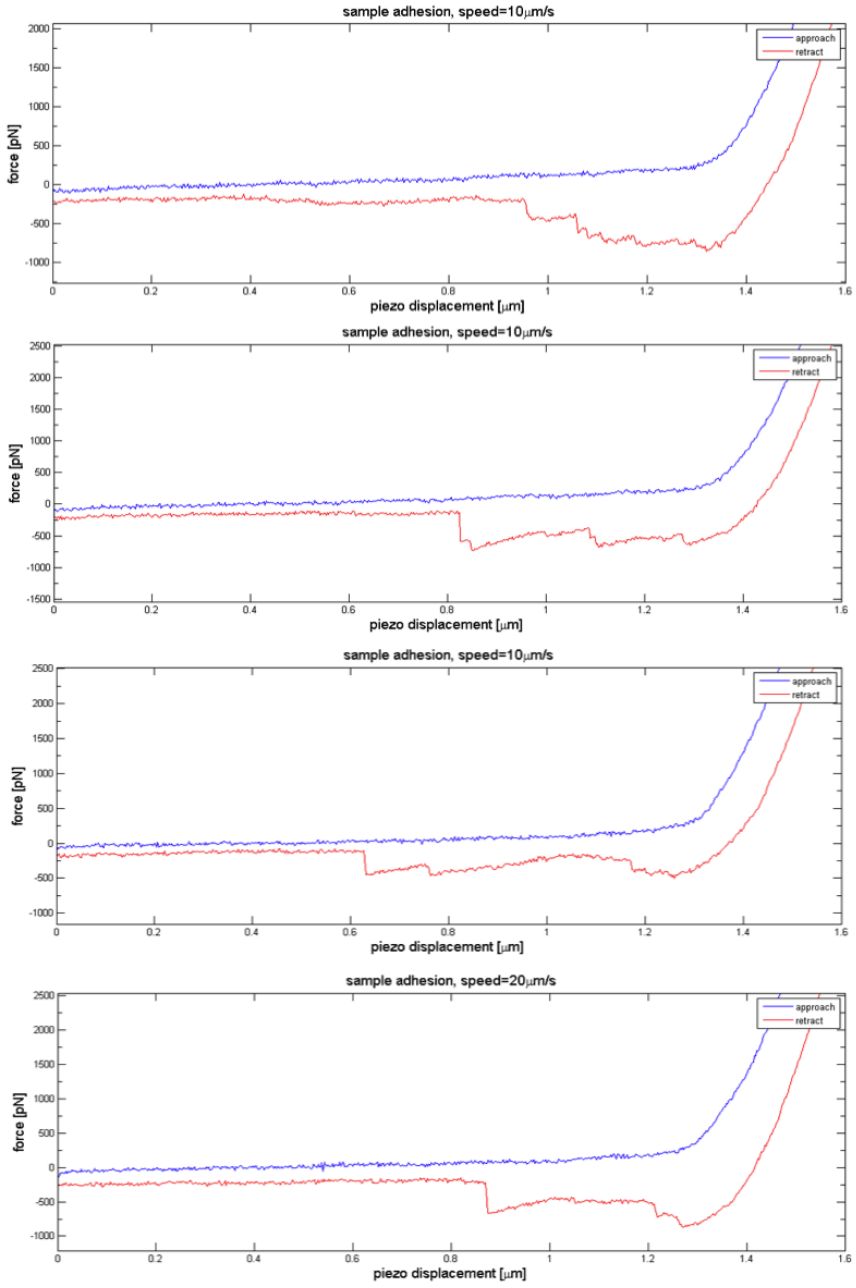

Figure D.2. Adhesion force curves obtained by different speeds of cantilever approach on the same cell.

REFERENCES

1. 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. 5892–900, 1999.

2. Bellés, X., D. Martín, and M.-D. Piulachs, "The Mevalonate Pathway and the Synthesis of Juvenile Hormone in Insects.", *Annual Review of Entomology*, Vol. 50, pp. 181–99, 2005.

3. Martín, D., M. D. Piulachs, and X. Bellés, "Inhibition of Vitellogenin Production by Allatostatin in the German Cockroach.", *Molecular and Cellular Endocrinology*, Vol. 121, No. 2, pp. 191–196, 1996.

4. Li, Y., S. Hernandez-martinez, F. Fernandez, and J. G. Mayoral, "Biochemical, Molecular, and Functional Characterization of PISCF-Allatostatin, a Regulator of Juvenile Hormone Biosynthesis in the Mosquito Aedes aegypti", *Journal of Biological Chemistry*, Vol. 281, No. 45, pp. 34048–34055, 2009.

5. Mayoral, J. G., M. Nouzova, A. Brockhoff, M. Goodwin, S. Hernandez-Martinez, D. Richter, W. Meyerhof, and F. G. Noriega, "Allatostatin-C Receptors in Mosquitoes.", *Peptides*, Vol. 31, No. 3, pp. 442–50, 2010.

6. Audsley, N., H. J. Matthews, N. R. Price, and R. J. Weaver, "Allatoregulatory Peptides in Lepidoptera, Structures, Distribution and Functions.", *Journal of insect physiology*, Vol. 54, No. 6, pp. 969–80, 2008.

7. Donly, B. C., Q. Ding, S. S. Tobe, and W. G. Bendena, "Molecular Cloning of the Gene for the Allatostatin Family of Neuropeptides from the Cockroach Diploptera punctata.", *Proceedings of the National Academy of Sciences of the United States*

of America, Vol. 90, No. 19, pp. 8807–11, 1993.

8. Bendena, W. G., C. S. Garside, C. G. Yu, and S. S. Tobe, "Allatostatins: Diversity in Structure and Function of an Insect Neuropeptide Family.", *Annals Of The New York Academy Of Sciences*, Vol. 814, pp. 53–66, 1997.

9. Stay, B. and S. S. Tobe, "The Role of Allatostatins in Juvenile Hormone Synthesis in Insects and Crustaceans.", *Annual review of entomology*, Vol. 52, pp. 277–99, 2007.

10. 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 of the United States of America*, Vol. 88, No. 21, pp. 9458–62, 1991.

11. Audsley, N., J. Matthews, and R. J. Weaver, "Neuropeptides Associated with the Frontal Ganglion of Larval Lepidoptera.", *Peptides*, Vol. 26, No. 1, pp. 11–21, 2005.

12. Matthews, H. J., N. Audsley, and R. J. Weaver, "Alanine Substitution and Deletion Analogues of Manduca sexta Allatostatin: Structure-Activity Relationship on the Spontaneous Contractions of the Foregut of Larval Lacanobia oleracea.", *Journal of insect physiology*, Vol. 52, No. 2, pp. 128–35, 2006.

13. Audsley, N., R. J. Weaver, and J. P. Edwards, "Metabolism of Manduca sexta Allatostatin by Hemolymph of Larvae of the Tomato Moth, Lacanobia oleracea.", *Peptides*, Vol. 23, No. 4, pp. 717–23, 2002.

14. Veenstra, J. a., "Allatostatin C and Its Paralog Allatostatin Double C: The Arthropod Somatostatins.", *Insect biochemistry and molecular biology*, Vol. 39, No. 3, pp. 161–70, 2009.

15. Larsen, M. J., K. J. Burton, M. R. Zantello, V. G. Smith, D. L. Lowery, and T. M. Kubiak, "Type A Allatostatins from Drosophila melanogaster and Diplotera

puncata Activate Two Drosophila allatostatin Receptors, DAR-1 and DAR-2, Expressed in CHO Cells.", *Biochemical and biophysical research communications*, Vol. 286, No. 5, pp. 895–901, 2001.

16. Lloyd, G. T., a. P. Woodhead, and B. Stay, "Release of Neurosecretory Granules Within the Corpus Allatum in Relation to the Regulation of Juvenile Hormone Synthesis in Diploptera punctata.", *Insect biochemistry and molecular biology*, Vol. 30, No. 8-9, pp. 739–46, 2000.

17. Lungchukiet, P., B. C. Donly, J. Zhang, S. S. Tobe, and W. G. Bendena, "Molecular Cloning and Characterization of an Allatostatin-Like Receptor in the Cockroach Diploptera punctata.", *Peptides*, Vol. 29, No. 2, pp. 276–85, 2008.

18. Lungchukiet, P., J. Zhang, S. S. Tobe, and W. G. Bendena, "Quantification of Allatostatin Receptor mRNA Levels in the Cockroach, Diploptera punctata, Using Real-Time PCR.", *Journal of insect physiology*, Vol. 54, No. 6, pp. 981–7, 2008.

19. Kreienkamp, H.-J., H. J. Larusson, I. Witte, T. Roeder, N. Birgul, H.-H. Honck, 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.", *The Journal of biological chemistry*, Vol. 277, No. 42, pp. 39937–43, 2002.

20. Auerswald, L., N. Birgül, G. Gäde, H. J. Kreienkamp, and D. Richter, "Structural, Functional, and Evolutionary Characterization of Novel Members of the Allatostatin Receptor Family from Insects.", *Biochemical and biophysical research communications*, Vol. 282, No. 4, pp. 904–9, 2001.

21. Tan, E. M., Y. Yamaguchi, G. D. Horwitz, S. Gosgnach, E. S. Lein, M. Goulding, T. D. Albright, and E. M. Callaway, "Selective and Quickly Reversible Inactivation of Mammalian Neurons in vivo using the Drosophila Allatostatin Receptor.", *Neuron*, Vol. 51, No. 2, pp. 157–70, 2006.

22. Wehr, M., U. Hostick, M. Kyweriga, A. Tan, A. P. Weible, H. Wu, W. Wu, E. M. Callaway, and C. Kentros, "Transgenic Silencing of Neurons in the Mammalian Brain by Expression of the Allatostatin Receptor (AlstR).", *Journal of Neurophysiology*, Vol. 102, No. 4, pp. 2554–2562, 2009.

23. Ikrar, T., Y. Shi, T. Velasquez, M. Goulding, and X. Xu, "Cell-Type Specific Regulation of Cortical Excitability Through the Allatostatin Receptor System.", *Frontiers in neural circuits*, Vol. 6, No. January, p. 2, 2012.

24. Brazeau, P., W. Vale, R. Burgus, N. Ling, M. Butcher, J. Rivier, and R. Guillemin, "Hypothalamic Polypeptide That Inhibits the Secretion of Immunoreactive Pituitary Growth Hormone.", *Science*, Vol. 179, No. 68, pp. 77–79, 1973.

25. Pradayrol, L., H. Jörnvall, V. Mutt, and A. Ribet, "N-Terminally Extended Somatostatin: The Primary Structure of Somatostatin-28.", *FEBS Letters*, Vol. 109, No. 1, pp. 55–58, 1980.

26. Arimura, A., H. Sato, A. Dupont, N. Nishi, and A. V. Schally, "Somatostatin: Abundance of Immunoreactive Hormone in Rat Stomach and Pancreas.", *Science*, Vol. 189, No. 4207, pp. 1007–1009, 1975.

27. Trent, D. F. and G. C. Weir, "Heterogeneity of Somatostatin-Like Peptides in Rat Brain, Pancreas, and Gastrointestinal Tract.", *Endocrinology*, Vol. 108, No. 6, pp. 2033–2038, 1981.

28. Yamada, Y., S. R. Post, K. Wang, H. S. Tager, G. I. Bell, and S. Seino, "Cloning and Functional Characterization of a Family of Human and Mouse Somatostatin Receptors Expressed in Brain, Gastrointestinal Tract, and Kidney.", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 89, No. 1, pp. 251–255, 1992.

29. Nelson, L. E. and M. A. Sheridan, "Gastroenteropancreatic Hormones and Metabolism in Fish.", *General and Comparative Endocrinology*, Vol. 148, No. 2,

pp. 116–124, 2006.

30. Patel, Y., M. Greenwood, R. Panetta, N. Hukovic, S. Grigorakis, L. Robertson, and C. Srikant, "Molecular Biology of Somatostatin Receptor Subtypes.", *Metabolism Clinical And Experimental*, Vol. 45, No. 8, pp. 31–38, 1996.

31. Hannon, J. P., C. Nunn, B. Stolz, C. Bruns, G. Weckbecker, I. Lewis, T. Troxler, K. Hurth, and D. Hoyer, "Drug Design at Peptide Receptors: Somatostatin Receptor Ligands.", *Journal of molecular neuroscience MN*, Vol. 18, No. 1-2, pp. 15–27, 2002.

32. Klein, S. E. and M. a. Sheridan, "Somatostatin Signaling and the Regulation of Growth and Metabolism in Fish.", *Molecular and cellular endocrinology*, Vol. 286, No. 1-2, pp. 148–54, 2008.

33. Goodyer, C. G., C. L. Branchaud, and Y. Lefebvre, "In vitro Modulation of Growth Hormone (GH) Secretion from Early to Midgestation Human Fetal Pituitaries by GH-Releasing Factor and Somatostatin: Role of Gs-Adenylate Cyclase-Gi Complex and Ca2+ Channels.", *The Journal of Clinical Endocrinology & Metabolism*, Vol. 76, No. 5, pp. 1259–1264, 1993.

34. Kumar, U. and M. Grant, "Somatostatin and Somatostatin Receptors", *Results and Problems in Cell Differentiation*, Vol. 50, No. 1968, pp. 137–184, 2010.

35. Ahrén, B., M. Ericsson, P. Hedner, S. Ingemansson, and U. Westgren, "Somatostatin Inhibits Thyroid Hormone Secretion Induced by Exogenous TSH in Man.", *The Journal of Clinical Endocrinology & Metabolism*, Vol. 47, No. 5, pp. 1156–1159, 1978.

36. Kobilka, B. K., "G Protein Coupled Receptor Structure and Activation", *Biochim Biophys Acta.*, Vol. 1768, No. 4, pp. 794–807, 2007.

37. Florio, T., S. Thellung, S. Arena, a. Corsaro, a. Bajetto, G. Schettini, and P. J.

Stork, "Somatostatin Receptor 1 (SSTR1)-Mediated Inhibition of Cell Proliferation Correlates with the Activation of the MAP Kinase Cascade: Role of the Phosphotyrosine Phosphatase SHP-2.", *Journal of physiology, Paris*, Vol. 94, No. 3-4, pp. 239–50, 2000.

38. Pagès, P., N. Benali, N. Saint-Laurent, J. P. Estève, A. V. Schally, J. Tkaczuk, N. Vaysse, C. Susini, and L. Buscail, "sst2 Somatostatin Receptor Mediates Cell Cycle Arrest and Induction of p27(Kip1). Evidence for the Role of SHP-1.", *The Journal of Biological Chemistry*, Vol. 274, No. 21, pp. 15186–15193, 1999.

39. Buscail, L., N. Delesque, J. P. Estève, N. Saint-Laurent, H. Prats, P. Clerc, P. Robberecht, G. I. Bell, C. Liebow, and a. V. Schally, "Stimulation of Tyrosine Phosphatase and Inhibition of Cell Proliferation by Somatostatin Analogues: Mediation by Human Somatostatin Receptor Subtypes SSTR1 and SSTR2.", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 91, No. 6, pp. 2315–9, 1994.

40. Cervia, D. and P. Bagnoli, "An Update on Somatostatin Receptor Signaling in Native Systems and New Insights on Their Pathophysiology.", *Pharmacology therapeutics*, Vol. 116, No. 2, pp. 322–341, 2007.

41. Sharma, K., Y. C. Patel, and C. B. Srikant, "C-Terminal Region of Human Somatostatin Receptor 5 is Required for Induction of Rb and G1 Cell Cycle Arrest.", *Molecular endocrinology Baltimore Md*, Vol. 13, No. 1, pp. 82–90, 1999.

42. Sharma, K., Y. C. Patel, and C. B. Srikant, "Subtype-Selective Induction of Wild-Type p53 and Apoptosis, But Not Cell Cycle Arrest, by Human Somatostatin Receptor 3.", *Molecular endocrinology Baltimore Md*, Vol. 10, No. 12, pp. 1688–1696, 1996.

43. Villaume, K., M. Blanc, G. Gouysse, T. Walter, C. Couderc, M. Nejjari, C. Vercherat, M. Cordier-Bussat, C. Roche, and J.-Y. Scoazec, "VEGF Secretion by Neuroendocrine Tumor Cells is Inhibited by Octreotide and by Inhibitors of the

PI3K/AKT/mTOR Pathway.", *Neuroendocrinology*, Vol. 91, No. 3, pp. 268–278, 2010.

44. Ferjoux, G., C. Bousquet, P. Cordelier, N. Benali, F. Lopez, P. Rochaix, L. Buscail, and C. Susini, "Signal Transduction of Somatostatin Receptors Negatively Controlling Cell Proliferation.", *Journal of physiology, Paris*, Vol. 94, No. 3-4, pp. 205–10, 2000.

45. Pyronnet, S., C. Bousquet, S. Najib, R. Azar, H. Laklai, and C. Susini, "Antitumor Effects of Somatostatin.", *Molecular and cellular endocrinology*, Vol. 286, No. 1-2, pp. 230–7, 2008.

46. Ruscica, M., M. Arvigo, F. Gatto, E. Dozio, D. Feltrin, M. D. Culler, F. Minuto, M. Motta, D. Ferone, and P. Magni, "Regulation of Prostate Cancer Cell Proliferation by Somatostatin Receptor Activation.", *Molecular and cellular endocrinology*, Vol. 315, No. 1-2, pp. 254–62, 2010.

47. Katritch, V., V. Cherezov, and R. C. Stevens, "Diversity and Modularity of G Protein-Coupled Receptor Structures.", *Trends in pharmacological sciences*, Vol. 33, No. 1, pp. 17–27, 2012.

48. Palczewski, K., "Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor", *Science*, Vol. 289, No. 5480, pp. 739–745, 2000.

49. Scheerer, P., J. H. Park, P. W. Hildebrand, Y. J. Kim, N. Krauss, H.-W. Choe, K. P. Hofmann, and O. P. Ernst, "Crystal Structure of Opsin in Its G-Protein-Interacting Conformation.", *Nature*, Vol. 455, No. 7212, pp. 497–502, 2008.

50. Costanzi, S., J. Siegel, I. G. Tikhonova, and K. A. Jacobson, "Rhodopsin and the Others: a Historical Perspective on Structural Studies of G Protein-Coupled Receptors", *Current Pharmaceutical Design*, Vol. 15, No. 35, pp. 3994–4002, 2009.

51. Pierri, C. L., G. Parisi, and V. Porcelli, "Computational Approaches for Protein

Function Prediction: A Combined Strategy from Multiple Sequence Alignment to Molecular Docking-Based Virtual Screening.", *Biochimica et biophysica acta*, Vol. 1804, No. 9, pp. 1695–712, 2010.

52. Lee, D., O. Redfern, and C. Orengo, "Predicting Protein Function from Sequence and Structure", *Nature Reviews Molecular Cell Biology*, Vol. 8, No. 12, pp. 995–1005, 2007.

53. Cavasotto, C. N. and S. S. Phatak, "Homology Modeling in Drug Discovery: Current Trends and Applications.", *Drug discovery today*, Vol. 14, No. 13-14, pp. 676–83, 2009.

54. Hillisch, A., L. F. Pineda, and R. Hilgenfeld, "Utility of Homology Models in the Drug Discovery Process.", *Drug discovery today*, Vol. 9, No. 15, pp. 659–69, 2004.

55. Zhang, Y., "Prediction Report Template-Based Modeling and Free Modeling by I-TASSER in CASP7", *Proteins*, Vol. 69 Suppl 8, No. September, pp. 108–117, 2007.

56. Roy, A., A. Kucukural, and Y. Zhang, "I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction.", *Nature protocols*, Vol. 5, No. 4, pp. 725–38, 2010.

57. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility.", *Journal of Computational Chemistry*, Vol. 30, No. 16, pp. 2785–2791, 2009.

58. Rarey, M., B. Kramer, T. Lengauer, and G. Klebe, "A Fast Flexible Docking Method using an Incremental Construction Algorithm.", *Journal of Molecular Biology*, Vol. 261, No. 3, pp. 470–489, 1996.

59. Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor,

"Improved Protein-Ligand Docking using GOLD.", *Proteins*, Vol. 52, No. 4, pp. 609–623, 2003.

60. Abagyan, R., M. Totrov, and D. Kuznetsov, "Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation", *Journal of Computational Chemistry*, Vol. 15, No. 5, pp. 488–506, 1994.

61. Morris, G. M. and M. Lim-Wilby, "Molecular Docking.", *Methods In Molecular Biology Clifton Nj*, Vol. 443, pp. 365–382, 2008.

62. Teodoro, M. L. and G. N. Phillips, "Molecular Docking : A Problem With Thousands of Degrees of Freedom", *IEEE International Conference on Robotics and Automation*, pp. 960–966, 2001.

63. Teodoro, M. L., G. N. Phillips, and L. E. Kavraki, "Understanding Protein Flexibility Through Dimensionality Reduction.", *Journal of computational biology a journal of computational molecular cell biology*, Vol. 10, No. 3-4, pp. 617–634, 2003.

64. Cappella, B. and G. Dietler, "Force-Distance Curves by Atomic Force Microscopy", *Surface Science Reports*, Vol. 34, No. 1-3, pp. 1–104, 1999.

65. Hinterdorfer, P. and Y. F. Dufrêne, "Detection and Localization of Single Molecular Recognition Events Using Atomic Force Microscopy", *Nature Methods*, Vol. 3, No. 5, pp. 347–355, 2006.

66. Lee, C.-K., Y.-M. Wang, L.-S. Huang, and S. Lin, "Atomic Force Microscopy: Determination of Unbinding Force, Off Rate and Energy Barrier for Protein-Ligand Interaction.", *Micron Oxford England 1993*, Vol. 38, No. 5, pp. 446–461, 2007.

67. Subramaniam, S., "The Biology Workbench–A Seamless Database and Analysis Environment for the Biologist.", *Proteins*, Vol. 32, No. 1, pp. 1–2, 1998.

68. Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.", *Molecular Biology and Evolution*, Vol. 28, No. 10, pp. 2731–2739, 2011.

69. Krogh, A., B. Larsson, G. Von Heijne, and E. L. Sonnhammer, "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes.", *Journal of Molecular Biology*, Vol. 305, No. 3, pp. 567–80, 2001.

70. Buchan, D. W. A., S. M. Ward, A. E. Lobley, T. C. O. Nugent, K. Bryson, and D. T. Jones, "Protein Annotation and Modelling Servers at University College London", *Nucleic Acids Research*, Vol. 38, No. Web Server issue, pp. W563–W568, 2010, `http://discovery.ucl.ac.uk/168110/`.

71. Laskowski, R. A., E. G. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones, and J. M. Thornton, "PDBsum: A Web-Based Database of Summaries and Analyses of All PDB Structures", *Trends in Biochemical Sciences*, Vol. 22, pp. 488–490, 1997.

72. Sigrist, C. J. A., L. Cerutti, E. De Castro, P. S. Langendijk-Genevaux, V. Bulliard, A. Bairoch, and N. Hulo, "PROSITE, A Protein Domain Database for Functional Characterization and Annotation", *Nucleic Acids Research*, Vol. 38, No. Database issue, pp. D161–D166, 2010.

73. Pagni, M., V. Ioannidis, L. Cerutti, M. Zahn-Zabal, C. V. Jongeneel, J. Hau, O. Martin, D. Kuznetsov, and L. Falquet, "MyHits: Improvements to an Interactive Resource for Analyzing Protein Sequences", *Nucleic Acids Research*, Vol. 35, No. Web Server issue, pp. W433–W437, 2007.

74. Vriend, G., "WHAT IF: A Molecular Modeling and Drug Design Program.", *Journal of Molecular Graphics*, Vol. 8, No. 1, pp. 52–56, 29, 1990.

75. Wallace, A. C., R. A. Laskowski, and J. M. Thornton, "LIGPLOT: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions", *Protein Engineer-*

ing, Vol. 8, No. 2, pp. 127–134, 1995.

76. Ma, M., T. M. Szabo, C. Jia, E. Marder, and L. Li, "Mass Spectrometric Characterization and Physiological Actions of Novel Crustacean C-Type Allatostatins.", *Peptides*, Vol. 30, No. 9, pp. 1660–8, 2009.

77. Shi, C., M. F. Paige, J. Maley, and M. C. Loewen, "In vitro Characterization of Ligand-Induced Oligomerization of the S. cerevisiae G-Protein Coupled Receptor, Ste2p.", *Biochimica et biophysica acta*, Vol. 1790, No. 1, pp. 1–7, 2009.

78. Hinterdorfer, P. and A. van Qijen (editors), *Handbook of Single-Molecule Biophysics*, Springer Science+Business Media, London, 2009.