In Silico Characterization and Homology Modeling of Histamine Receptors

By

(Md. Nayem Zobayer)

A project submitted in partial fulfillment of the requirement for the degree of

Master of Science in Biomedical Engineering

Khulna University of Engineering & Technology (KUET)

Khulna 9203, Bangladesh

February, 2019

Declaration

This is to certify that the thesis work entitled **"*In Silico* Characterization and Homology Modeling of Histamine Receptors"** has been carried out by **Md. Nayem Zobayer** in the Department of Biomedical Engineering, Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

Signature of Supervisor

Signature of Candidate

Approval

This is to certify that the thesis work submitted by Md. Nayem Zobayer entitled "*In Silico* Characterization and Homology Modeling of Histamine Receptors" has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of M. Sc. Engineering in the Department of Biomedical Engineering, Khulna University of Engineering & Technology (KUET), Khulna, Bangladesh in February 2019.

BOARD OF EXAMINERS

1. 28.02.19 — Chairman (Supervisor)
Prof. Dr. A. B. M. Aowlad Hossain
Dept. of Electronics and Communication Engineering (ECE)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh

2. 28.02.19 — Member
Head of the Department
Dept. of Biomedical Engineering (BME)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh

3. Nasrin 28.02.19 — Member
Dr. Nasrin Akhter
Assistant Professor
Dept. of Biomedical Engineering (BME)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh

4. 28.02.19 — Member
Dr. Muhammad Muinul Islam
Assistant Professor
Dept. of Biomedical Engineering (BME)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh

5. 28/02/2019 — Member (External)
Prof. Dr. Khondoker Moazzem Hossain
Biotechnology and Genetic Engineering Discipline,
Khulna University, Khulna-9208, Bangladesh

Dedicated to

MY PARENTS
&
ALL 'SPECIAL CHILD' OF THE WORLD

Acknowledgement

In the ending of my thesis work for M.Sc. Eng. degree, first of all, I am mostly thankful to the Almighty Allah for giving me the strength and confidence to complete it successfully.

In this stage, I would like to thank my graduation committee. It is pleasure to express my deepest sense of gratitude and respect to my honorable thesis advisor, ***Prof. Dr. A. B. M. Aowlad Hossain***, Department of Electronics and Communication Engineering (ECE), KUET, who was my inspiration during my thesis work. His invaluable suggestions, proper guidance, and encouragement helped to finish my work. His activeness, friendly behavior, and motivation were really memorable. He is a true mentor of my educational carrier and personal life motivator as well. I have learnt a lot of things from him and it's my utmost desire to work with him throughout my research carrier. If I have to name a true inspiration of my life beside my parents, it will be my dearest thesis supervisor, my deepest gratitude to him.

In addition, I wish to express my sincere acknowledgement to **Dr. Muhammad Muinul Islam,** Head, Department of Biomedical Engineering for giving me many good pieces of advice. I am also much grateful to **Dr. Nasrin Akhter,** for her kind advice and suggestions during this research. It is my pleasure to acknowledge the roles of **Engr. Md. Asadur Rahman (**PhD), Department of BME who was with me from the beginning of my thesis work, helped me, and supported me. Not only the academic support but also, he provided continuous mental support to overcome different obstacles in the field of research. Truly, I must say, he is the brother from another mother for me.

Besides, I am very much grateful to my parents, who tolerated all my reckless decisions throughout my life and sacrificed a lot to provide me with the ample space to complete this thesis. I am grateful to all the BME faculty members, staffs, and postgraduate students who helped me directly or indirectly.

Author

February, 2019

Abstract

Histamine plays vital role in molecular mechanism of allergic reactions. Therefore, characterization and homology modeling of Histamine receptor is of great importance to design effective vaccines. In this thesis, different methods are applied to analyse biomolecular features of histamine receptors and design best models of these receptors. In addition to this, the study tried to identify potential B cell and T cell epitope based vaccine of an allergen and consequently, emphasized on to develop a B cell prediction tool. Identified four histamine receptors, such as Histamine H1, Histamine H2, Histamine H3 and Histamine H4 have been analysed through ProtParam to extract physiochemical properties and ClastalW algorithm has been applied to identify conserved regions. Motif and Transmembrane regions have been identified through MEME suit and TMHMM servers, respectively. For homology modeling, I-tasser has been used and generated models have been validated through RAMPAGE, ERRAT and PROCHECK. Targeted api m3 allergen then rendered through self-optimized prediction method with alignment for physiochemical feature extraction. NetCTL 1.2 has been applied to identify preliminary T cell epitope candidates and then scrutinized by Stabilized Matrix Base Method, relative to IC50 values. Predicted T cell epitopes have been further analysed for conservancy and population coverage via IEDB tools. B cell epitopes of api m3 allergen have been predicted through, BCPREDS, ABCpred, Bepipred and Bcepred. In addition, classifier based single interface B cell epitope prediction and/or validation tool has been developed through establishing efficient MATLAB algorithms to classify beta turn regions, hydrophilic regions, surface accessible regions and antigenic regions. Lastly, with superimposing graphical representation of these four criteria in a single interface graph plotted to identify B cell epitopes via this tool. Extracted results denotes that, Histamine receptors possess molecular weight around 55.7 KDa, theoretical pI 9.33-9.62, instability index 34.93-47.00, aliphatic index (AI) was above 90 and the receptors were hydrophobic except histamine H1 receptor. Moderately conserved region was found in 75-94 amino acid position. A profound motif has been identified from 84-149 amino acid position for four histamine receptors with significantly lower E-value. It has been identified that, these receptors are seven pass transmembrane protein and a gap between transmembrane helix number five and six was found in each histamine receptor except Histamine H2 receptor, which can be potential drug target candidate. Generated 3D models have been passed through every spheres of validation. Api m3 allergen has been found relative thermostable nature and only 10.46% of the overall secondary structure consisted of beta turn region. Five MHC class I T-cell epitopes were identified and scrutinized and YTEESVSAL found out as the best epitope. For MHC class II T-cell epitopes YPKDPYLYYDFYPLE and GGPLLRIFTKHMLDV have been found as most prominent T-cell epitopes of api m3 allergen. This study also revealed that, GDRIPDEKN and PHVPEYSSS, as the most effective B-cell epitopes of api m3. The proposed tool efficiently identified B cell epitopes and provided result in a single interface. The tool can aid in B cell research and vaccine development. Finally, the suggested potential drug targets can be applied in designing more sustainable antihistamines and relevant drugs in treating allergic diseases. Predicted T-cell and B-cell epitopes of api m3 allergen could help the researchers to test these vaccines further for immunoreactivity applying *in vivo* analysis. As still there is no report of T-cell and B-cell epitopes of *Apis mellifera*, this study can be the pioneer in finding effective vaccine against allergens of honeybee. This research also predicted potential B cell epitope regions from an antigenic protein. The most exciting feature of this part of the study is, it presents results of potential B cell epitopes on a single interface, so that, researchers don't need to search for every feature (*e.g.*, hydrophilicity, antigenicity, beta turn, surface accessibility etc.) separately. Finally, the study can certainly aid in B cell epitope-based vaccine design research.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviated Form	Elaboration
WHO	World Health Organization
WBC	White Blood Cell
GPCR	G Protein Coupled Receptor
NF-κB	Nuclear Factor – Kappa Beta
CNS	Central Nervous System
IgE	Immunoglobulin E
BBB	Blood Brain Barrier
REM	Rapid Eye Movement
GABA	Gamma Amino Butyric Acid
cAMP	Cyclic Adenylyl Mono Phosphate
DNA	Deoxy Ribonucleic Acid
RNA	Ribonucleic Acid
TP	Transmembrane Protein
TCR	T Cell Receptor
APCs	Antigen Presenting Cells
MHC	Major Histocompatibility Complex
ELISA	Enzyme Linked Immuno Sorbent Assay
Da	Dalton
T_H cells	Helper T Cells
BCRs	B Cell Receptors
SLOs	Secondary Lymphoid Organs
MZ	Marginal Zone
HRH1	Histamine H1 Receptor
HRH2	Histamine H2 Receptor
HRH3	Histamine H3 Receptor
HRH4	Histamine H4 Receptor
AI	Aliphatic Index
GRAVY	Grand Average of Hydropathicity
7TM	Seven Pass Transmembrane
CTL	Cytotoxic T Lymphocyte

MAPK	Mitogen-Activated Protein Kinases
SMM	Stabilized Matrix Base Method
HLA	Human Leukocyte Antigen
TAP	Transport Associated Proteins
HMM	Hidden Markov Model
RNN	Recurrent Neural Network
SVM	Support Vector Machine
NMR	Nuclear Magnetic Resonance
EM	Electronmicroscopy
MATLAB	Matrix Laboratory
HPLC	High Performance Liquid Chromatography

Chapter 1

Introduction

1.1 Introduction

Allergy is a medical state in which exposure to an allergen (such as, pollen, dust, specific type of food, insect bite, venom and others) triggers an unusual immune response. This type of undesired immune response, referred an allergic reaction, occurs because the immune system interacts with the foreign particle, which is normally doesn't abrupt with the immune system of common non-allergic population. The foreign particle that trigger the immune reaction are called allergens. The symptoms of an allergic reaction can be ranged from mild (itchy mouth, a few hives) to severe (throat tightening, difficulty breathing). Anaphylaxis is one of the most serious allergic reaction that is very sudden in onset and can lead to death [1], [2].

Upon invasion of an allergen immune system of human body responds, but when the response is aggressive then physiologically human body shows different malfunctions. The major molecule that triggers excessive immune response in presence of allergen is called histamine. There are four known type of histamine receptors, designated as Histamine H1 receptor, Histamine H2 receptor, Histamine H3 receptor and Histamine H4 receptor. These receptors play key roles in disease progression of allergy. The possible therapeutics of any allergen protein can be developed through B cell and T cell epitope mapping and simultaneous vaccine design.

1.2 Background and Motivation

Fact sheets from different public health organizations are enough to attain ample motivation to conduct intensive research on the genre of allergy and its probable sustainable medical treatments. Report from WHO suggest that, more than 170 food items cause allergic reaction. Most of the severe food allergens are milk, egg, peanut, tree nuts, wheat, soy, fish and crustacean shellfish, those are compulsory dietary elements of food all over the world [3]. Dust, pollen and chemical residues can provoke harmful allergic reaction and lead to anaphylactic reactions. On the other hand, of insect bites and venoms are can cause lethal allergic reaction, that can lead to paralysis, organ damage and even to death. Among different causal agents of allergy food borne allergens are most frequently found. The Centers for Disease Control & Prevention reports that the prevalence of food allergy in children

increased by 50 percent between 1997 and 2011 and in between 1997 and 2008, the prevalence of peanut or other tree nut allergy appears to have more than tripled in children [4]. Most shocking fact is that, in every three minutes, a food allergy reaction sends someone to the intensive care unit. Every year only in the U.S., 200,000 people require emergency medical care for allergic reactions to food. Childhood hospitalizations for food allergy tripled between the late 1990s and the mid-2000s and expected to increase more rapidly by 2020. About 40 percent of children with food allergies have experienced a severe allergic reaction such as anaphylaxis and chronic allergic reactions [3].

In molecular level, upon invasion of a foreign particle, redundancy of histamine can be found in patient's blood sera and one of the most critical factors in allergic reaction. As a consequence of an immunogenic response to foreign pathogens histamine is produced by basophils and mast cells, those are readily found in nearby connective tissues. Histamine enhances the permeability of the capillaries to White Blood Cells (WBC) in blood and some proteins, to permit them to engage pathogens in the infected tissues. These histamines are recognized by histamine receptors, proteins those situated in various parts of the body and bind with histamine to produce a specific to nonspecific allergic reaction cascade on the organism [5]. There are four known type of histamine receptors, designated H1 - H4. As these receptors play pivotal role in disease progression of allergy, we need to delve deeper for the biological features of these four receptors. This will encourage histamine receptor-based drug targeting researches.

One of the major ways to develop suitable drugs or vaccines against these allergen proteins are to find out B cell and T cell epitopes. Honey bee (*Apis mellifera*) is an economically important insect, which has diverse and prolong history of close interactions with human. As apiculture is a budding industry in our country, we need to assess its risk factors associated with the beekeepers. Honey bee venom can induce severe to acute allergic reactions in different parts of human body, such as the heart, kidney, nervous systems and eye [6]. Patients with venom allergy usually suffer from large local reactions. The proportion belonging to this group usually ranged from 2.4-26.4% of mass population and this range can be as high as 38% in beekeepers. Around 0.3-7.5% of studied population suffered from systemic anaphylaxis and severe systemic reaction was 14-43% among beekeepers. According to the study, mortality rate was 0.03-0.45 per million but this number might not be significant as many deaths caused by anaphylactic reactions to insect stings probably remained unreported [7]. So, it is of utmost importance to identify potential vaccine candidate

against honey bee venom allergen. With that note, best B cell and T cell epitope-based vaccine candidate design can bring new insights in venom allergen studies.

Whereas, *in silico* B cell epitope-based vaccine design concedes few limitations. Among the existing methods of epitope mapping (structural and functional approach) most of the wet lab protocols are expensive, laborious, time consuming, and often unable to identify properly all epitopes. Existing structural epitope mapping procedure determine the protein structure comprising residues in direct contact with an antibody but occasionally fail to reveal contribution of amino acids in accordance with the binding strength. But, effective *in silico* B cell epitope mapping tool can lead to overcome these issues. Current, *in silico* tools often lower in specificity, time consuming, and often unable to identify properly significant antigenic portions. Tool that can provide significant B cell epitope mapping data can be very useful in the field of *in silico* vaccinology research.

1.3 Problem Statements

As we have discussed in the motivation section, there are four type of histamine receptors, such as Histamine H1 receptor, Histamine H2 receptor, Histamine H3 receptor and Histamine H4 receptor. With the aid of these receptors Histamine provokes allergenic reactions, although functionally these receptors are different from each other.

Histamine H1 receptor is G Protein Coupled Receptor (GPCR), which up regulates NF-κB activity thereby eliciting inflammation. Crystal structure of Histamine H1 receptor has already been reported [8]. But, as histamine H1 receptor influences sleep wake state, antihistamines available for blocking this receptor stimulates dizziness & convulsion in patients. So, we need to investigate deeper in histamine H1 molecular pattern to find out effective H1 antihistamines that will generate lower side effects than traditional H1antihistamines. Histamine H2 receptor inhibits antibody production, neutrophil activation and chemotaxis [9]. Nonetheless, till to date no crystal structure has been reported. Structure wise comparison with Histamine H1 receptor can lead to overcome the issues with H1antihistamines, as we already know, Histamine H2 receptor doesn't influence sleep wake cycle. Histamine H3 receptor associated with Central Nervous System (CNS) and controls the synthesis and release of endogenous histamines through feedback mechanism [10] but, till to date no crystal structure or model of this receptor has been reported. Immunomodulatory functions such as T cell differentiation and dendritic cell activation are characteristics role of Histamine H4 receptor. Drug design of allergen is solely depending on these receptors

agonists and antagonists, thus knowledge on the physicochemical properties of these receptors are of utmost importance. As it is discussed earlier, crystal structure of Histamine H1 receptor is available in PDB database but for other histamine receptors no such data is available. In order to better understand the functions and molecular biology of these receptors we need to analyze their 3D structures. To overcome major pitfalls in antihistamine medications, this study may widen the bottleneck of current concerns of histamine and antihistamine research.

Earlier in the motivation section, it has been discussed that the possible therapeutics of this allergen protein can be developed through B cell and T cell epitope mapping and simultaneous vaccine design. Major honey bee allergens include hyaluronidase, phosphatase and phospholipase-A2, these enzymes activate cells of immune system and via the mast cell degranulation influence histamine release and produce Immunoglobulin E (IgE) can induce lgE-mediated hypersensitivity reactions can be induced by stings of hymenoptera in venom allergic patients and this reaction might range from local up to severe systemic reactions and even acute anaphylaxis. Traditional medicines are available for treating honey bee venom allergy. The most frequently used medicine is epinephrine and usually provided through injection. Different types of histamine blocker or anti-histamines are also used. Ranitidine, famotidine and cimetidine are histamine type 2 blockers used to enhance the effects of diphenhydramine. Corticosteroids like prednisone or methylprednisolone are often applied to reduce the hyper immune responses. Venom specific immunotherapy is one of the most potential treatments of honey bee venom allergy as it is said to be effective and enhance quality life of patients. Moreover, venom specific immunotherapy economic analysis showed that it is not cost effective for preventing future systemic reactions.

However, there has been no report of Api m3 B-cell and T-cell epitopes spectra till to date. Therefore, successful design of Api m3 B-cell and T-cell epitope may aid value through providing devising new therapeutic modalities for immunotherapy of honey bee venom allergy. The experimental procedures developed to identify the B cell epitopes can grossly be divided into structural and functional studies. The X-ray crystallography can precisely identify the position of epitope within the protein structure but is laborious, time consuming, costly, technically difficult, and not applicable for all antigens [11]. One of the frequently used protocol for functional epitope mapping are screening of antigen-derived proteolytic fragments or peptides for antibody binding and testing the Antigen antibody reactivity of

mutants (site-directed mutagenesis or randomly mutated) [11]. Other techniques like display technologies and mimotope analysis have also become more acceptable alternative for epitope mapping due to their relative cheapness, flexibility, and speed [12], [13]. Current methods that can conduct epitope mapping usually have some drawbacks. In order to design discontinuous B cell epitopes *in silico*, a protein's antigenicity, surface accessibility, flexibility, beta turn, and hydrophobicity need to be predicted. For studying B cell epitope, the first step is to classify potential regions and non-significant regions of each scale. Successful identification and classification of protein's antigenic, surface accessible, hydrophilic, flexible and beta turn regions are of utmost importance to select suitable B cell epitopes. These scales are based on different principles and thus, often it becomes much complex to find out potential B cell candidates in a protein sequence. In order to better understand and better identification of B cell epitopes better approaches can be considered to identify potential B cell candidate at a time, which will enhance the accuracy of the current methods and reduce identification time as well.

1.4 Objectives

The objectives of this study are to attain profound knowledge on biological activities and justification of histamine as potential drug targets against allergic diseases. Secondly, to design efficient vaccine against most pathogenic honey bee allergen, Api m3 and lastly, to eradicate bottlenecks of B cell epitope mapping and simplifying classification through different algorithms. The objectives are divided into following step by step approaches:

- Investigation on the physicochemical properties that elicit histamine as a hypersensitivity provoking biomolecule.
- 3D structure prediction and model validation of histamine receptors.
- Identification of potential T cell & B cell epitopes and building up 3D models of Api m3 allergen.
- Algorithm based discontinuous B cell epitope prediction by using four classifiers.

1.5 Contributions of this Thesis

- **Homology Modeling of Histamine Receptors:** This study aims to design 3D structures of Histamine H1, Histamine H2, Histamine H3 and Histamine H4 receptors. Among these, only Histamine H1 receptor has crystal structure and it will

be compared with the predicted model for Histamine H1 receptor, and accuracy of the prediction will be determined.

- **Vaccine Candidate Against api m3:** This research also brings new insights on B cell and T cell epitope-based vaccine design of api m3 allergen of honey bee. Study also performed homology modelling of api m3 allergen of *Apis mellifera.*
- **Classifier based B cell epitope prediction:** Using algorithm based classifier this study incorporates four different features of B cell epitope into a single interface.

1.6 Thesis Outline

The overall study can be outlined as follows

- **Introduction:** The introductory information about the project work is presented in Chapter 1. In addition, the previous works related to this project has widely examined and their important limitations are reported in this chapter. The objective of this project is also clarified in this chapter.
- **Overview on Histamine Receptors and Vaccine Design Against Api m3 Allergen:** The basics of histamine receptors biological roles, importance of homology modeling of these receptors are discussed. The principles of *in silico* vaccine design and its importance over wet lab vaccine development are also discussed. A brief introduction on T cell and B cell epitope and shortcomings of current B cell epitope mapping are added in Chapter 2, as well.
- **Methodology:** The Chapter 3 of this study is focused on different methods involved in finding out physiochemical properties of histamine receptors and design homology models. Plus, methods for designing a vaccine for an allergen have been thoroughly discussed. The protocols for classifier-based B cell epitope design are also discussed.
- **Result Analysis and Discussion:** In Chapter 4, results of physiochemical properties of histamine receptors and respective models have been discussed. Probable vaccine candidates of api m3 allergen and functionality of our classifier-based B cell epitope prediction tool also been discussed.
- **Conclusions and Future Perspectives:** The total thesis work has been concluded in Chapter 5. Here, the future perspectives have also been a concerning issue that will help the future researchers to work more efficiently with the task related to this project work.

Chapter 2

Overview on Histamine Receptors and Vaccine Design Against Api m3 Allergen

2.1 Introduction

In this chapter, basics biomolecular functions of different histamine receptors have been discussed. Protein motif, transmembrane analysis, multiple sequence alignment and homology modelling of a protein also been discussed in this chapter. There are four type of histamine receptor, histamine H1 receptor, histamine H2 receptor, histamine H3 receptor and histamine H4 receptor. The antagonists available to date against histamine H1 receptor possess sedative features, which is one of the major limitations of first-generation allergy drugs [5]. On the other hand, activation of histamine H2 receptor can lead to vasodilation of smooth muscle, gastric acid secretion, smooth muscle relaxation, inhibits antibody secretion, cytokine production and T cell proliferation, inhibition of neutrophil activation and chemotaxis. Although, histamine H2 receptor's traditional therapeutics are not reported to be associated with drowsiness or other side effects. Histamine H3 receptor draws interests as it is causal agent of H3R disorders and for its position in the Central Nervous System (CNS). Histamine H4 receptor is newly identified histamine receptor and scientific community still searching for more information on histamine H4 receptor [9], [10]. Among these receptors, only histamine H1 receptor has crystal structure and to know more about other three receptors, homology modelling is an essential part. In the case of where, no crystal of protein or ligand or receptors structure is present, then homology modelling is performed. It is also known as comparative modelling of protein or ligand or receptors, refers to constructing an atomic-resolution model of the "target" protein from its primary structure and an experimental three-dimensional structure of a related homologous protein. Another crucial feature of protein is motif. Motif provides primary concepts on protein structure and relative connectivity among the secondary structural elements. Protein motif plays important part in a protein structure because the relationship between primary structure and tertiary structure of a protein is not straightforward, two biopolymers may share the same motif yet lack appreciable primary structure homology. It is easier to find out sequence homology and also ancestry prediction can be carried out to analyse the protein sequences' common evolutionary origins by multiple sequence alignment. Multiple sequence alignment is often used to analyse sequence conservation among protein, finding out protein domain, tertiary and secondary structures, and even individual amino acids or nucleotides. Altogether, homology modelling

of these four receptor and other features analysis can lead to know more about molecular biology of histamine receptors.

2.2 Basics of Histamine Receptor

2.2.1 Histamine H1 Receptor

The histamine H1 receptor is a member of G-protein-coupled receptors (GPCR) family, which is traditionally referred as rhodopsin like protein. The histamine H1 receptor generally activated through the presence of indigenous amine histamine. This receptor normally found in lines of smooth muscles, on vascular endothelial cells, in the heart, and in the Central Nervous System (CNS) as well. This receptor is associated to different signalling cascades, such as an intracellular G-protein (G_q) that stimulates phospholipase C and the Inositol Triphosphate (IP3). Antihistamines, those are effective to nullify this receptor, are used as anti-allergic drugs. The crystal structure of the receptor has already been determined and applied to identify novel histamine H1 receptor ligands in structure-based *in silico* screening studies.

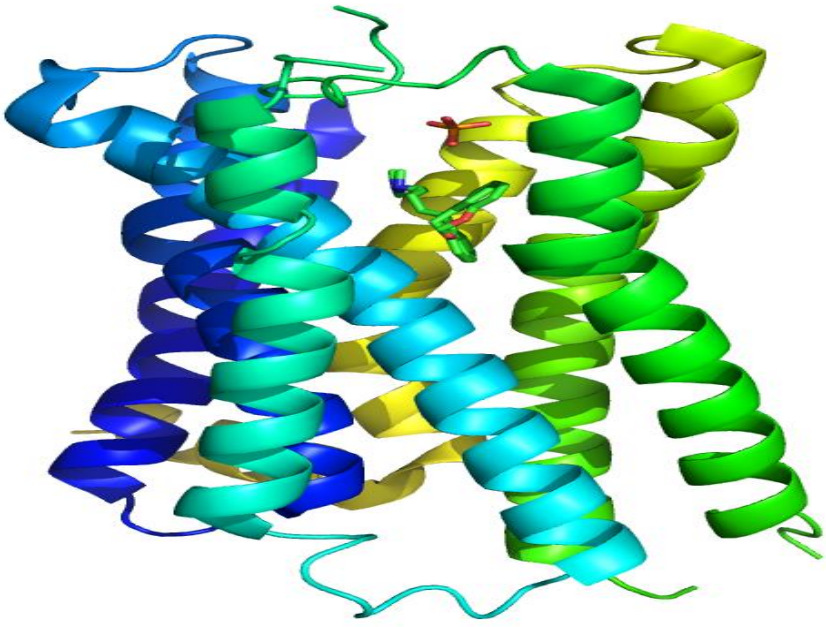

Figure 2.1: Histamine H1 Receptor with Doxepin as ligand [8].

It normally triggered by biogenic histamine, excreted by neurons that consist their cell bodies in the tuberomammillary nucleus of the hypothalamus. The histaminergic neurons of the tuberomammillary nucleus convert active during the 'wake' cycle of a sleep-wake cycle, firing at around 2 Hz; during slow wave sleep, this firing proportion drips to approximately 0.5 Hz. Lastly, during REM sleep, histaminergic neurons stop firing altogether. It has been

identified that histaminergic neurons possess the highest wake-selective firing pattern of all known neuronal types [10].

The tuberomammillary nucleus is a dominant histaminergic nucleus that highly controls the sleep-wake cycle. The H1 antihistamines that penetrates the Blood-Brain Barrier (BBB) inhibit H_1 receptor function on neurons that project from the tuberomammillary nucleus. This action is responsible for the drowsiness side effect associated with these drugs.

2.2.2 Histamine H2 Receptor

As it is already known that, histamine is a ubiquitous messenger molecule secreted from mast cells, enterochromaffin-like cells, and neurons [9]. The actions of histamine are mediated by histamine receptors H1, H2, H3 and H4. The histamine H2 receptor belongs to the rhodopsin-like family of G Protein-Coupled Receptors. It is an integral membrane protein and stimulates gastric acid secretion. It also controls gastrointestinal motility and intestinal secretion. Histamine H2 receptor is thought to be responsible for regulating cell growth and differentiation.

Histamine H2 receptors are expressed in different tissue system of human body, such as Neutrophils, Mast cells, Gastric parietal cells, Vascular Smooth Muscle, Heart and Uterus. The receptor also expressed in central nervous system tissues namely Caudate-Putamen, Cerebral Cortex, Hippocampal formation and Dentate nucleus of the cerebellum. Activation of histamine H2 receptor can lead to vasodilation of smooth muscle, gastric acid secretion, smooth muscle relaxation, inhibits antibody secretion, cytokine production and T cell proliferation, inhibition of neutrophil activation and chemotaxis.

2.2.3 Histamine H3 Receptor

Histamine H3 receptors are largely expressed in the Central Nervous System (CNS) and to a lower extent the peripheral nervous system. In these two systems, they act as auto receptors in presynaptic histaminergic neurons, and efficiently controls histamine production by a mechanism termed as feedback inhibition [14]. This receptor has also been revealed to carry the capabilities of pre-synaptically inhibit the release of a number of other neurotransmitters including, a wide range of neurotransmitters which includes but probably not limited to dopamine, GABA, acetylcholine, noradrenaline, histamine, and serotonin.

The gene sequence for Histamine H3 receptors expresses only about 22% and 20% sequence similarity with both H1 and H2 receptors, respectively. Currently, there is a lot of interest in

the histamine H3 receptor as a potential therapeutic target because of its involvement in the neuronal mechanism behind many cognitive H3R-disorders and especially its location in the Central Nervous System [15], [16].

Like the other histamine receptors, histamine H3 receptor is a member of G-Protein Coupled Receptor (GPCR). The H3 receptor is tied to the G_i G-protein, so this receptor influences to inhibit the formation of cAMP. Also, the β and γ subunits react with N-type voltage gated calcium channels, to minimize action potential mediated influx of calcium ions and hence downregulate neurotransmitter release. Histamine H3 receptor function as presynaptic auto receptors on pre-existed histamine bearing neurons. The diverse expression of histamine H3 receptors throughout the cortex and subcortex regions indicate its capability to modulate the release of a large number of neurotransmitters. Histamine H3 receptors are thought to play a part in the control of satiety [17].

2.2.4 Histamine H4 Receptor

Histamine H4 receptor is highly expressed in bone marrow, it controls the release and activity of neutrophil [18] and it also mediates change in shape of eosinophil as well as its migration and mast cell degranulation. This occurs via the βγ subunit acting at phospholipase C to cause actin polymerisation and eventually chemotaxis.

Immunomodulatory functions such as T cell differentiation and dendritic cell activation are characteristics role of Histamine H4 receptor. It is also expressed in the colon, liver, lung, small intestine, spleen, testes, thymus, tonsils, and trachea. It was also found that histamine H4 receptors shows a uniform expression pattern in the human oral epithelium [19].

The 3D structure of the H4 receptor has not been solved yet due to the difficulties of GPCR crystallization. Some attempts have been made to develop structural models of the H4 receptor for different purposes, but remain unsuccessful.

2.3 Structural Features of Protein

2.3.1 Protein Motif

Motif provides primary concepts on protein structure and relative connectivity among the secondary structural elements. Protein motif plays important part in a protein structure because the relationship between primary structure and tertiary structure of a protein is not straightforward, two biopolymers may share the same motif yet lack appreciable primary structure homology [20]. That’s the issue with the motif keeps researchers interested to

understand the biomolecular functions of motifs. A structural motif does not necessarily have to be associated with a sequence motif. Also, the presence of a sequence motif does not necessarily imply a distinctive structure. In most DNA motifs, for example, it is assumed that the DNA of that sequence does not deviate from the normal "double helical" structure. But in case of protein it not true at all.

The biomolecular functions of a protein motif are of utmost importance. Tim hunt stated protein motif as a shorter part of a protein sequence, mainly involved in recognition and targeting functions, often isolate from other functional characteristics of the molecule in which they belong. Usually motifs are linear, in the sense that 3D orientation is not required to fetch rear fragments of the molecule united together to build the recognizable part. There are two type of motifs usually found in sequences, some are highly conserved and some allows certain substitution across the motif.

Motif plays crucial roles in each biological signalling cascades as they have signal transduction capabilities, regulatory functions and protein-protein interactions. It functions as an interacting system that are recognized by biomolecules. Additionally, protein motifs can be classified into two broad categories, these are ligand binding sites and modification sites. Motif that helps in binding a ligand to it or to a specific molecule is commonly termed as ligand binding motifs and motif that aid in any type of modification of a site, such as deletion, alteration, substitution is refereed as modification motifs.

Liner motif operated protein-protein interaction has led new dimensions in drug development in recent years [21]. Potential MDM2 protein motif analog Nutlin-3 and integrin aiming RGD-mimetic Cilengitide: Nutlin-3 antagonises the interaction of MDM2's SWIB domain with p53 thus stabilising p53 and inducing senescence in cancer cells [22]. Because of this, Cilengitide prevents integrin-dependent signalling, triggering the disintegration of cytoskeleton, cellular detachment, introduction of apoptosis in endothelial and glioma cells [23], [24]. On the other hand, peptides targeting the Grb2 and Crk SH2/ SH3 adaptor domains are currently also under intensive studies. [25], [26].

2.3.2 Multiple Sequence Alignment of Proteins

When sequence alignment for two or more protein, DNA or RNA sequence are performed in a quest for finding sequence homology and evolutionary origins, is termed as multiple sequence alignment. Most of the cases, input of multiple sequence alignment is a set of sequence data of different proteins expected to possess an evolutionary relationship by that mean these regions might share a connection and might have been evolved from a common ancestor.

Multiple sequence alignment also refers to the process of aligning such a sequence set. Because three or more sequences of biologically significant length can be hard and are almost always time-consuming to align by hand, computational algorithms are used to produce and analyse the alignments. Multiple sequence alignments require more sophisticated methodologies than pairwise alignment because they are more computationally complex. Most of the multiple sequence alignment programs use heuristic methods rather than global optimization because identifying the optimal alignment between more than a few sequences of moderate length is prohibitively computationally expensive.

It is easier to find out sequence homology and through multiple sequence alignment data it is easier to find out evolutionary origins. Multiple sequence alignment is often used to analyse sequence conservation among protein, finding out protein domain, tertiary and secondary structures, and even individual amino acids or nucleotides.

There are a lot of functional uses of multiple sequence alignment. It can be used in identifying active sites or binding sites in a protein or a site that regulates any certain feature of a protein function. Multiple sequence alignment data can be analysed via different spheres, such as it can provide ample idea on sequence homology and identity. Sequence identity stands for complying with same residues in respective positions. Likewise, sequence similarity means complying with only quantitively similar residues. In DNA sequences, pyrimidines are considered identical to each other. This is also true for purines. Most importantly, similarity is a special type of homology, the sequences these are similar in residual features, the more homologous they usually tend to be. This similarity can certainly aid in find out their evolutionary origins.

2.3.3 Transmembrane Protein

The integral membrane protein that passes through the entirety of the cell membrane to which it is permanently attached in called transmembrane protein. Major proportion of the transmembrane proteins act as entryways to permit the carriage of specific substances across the attached membrane. Frequently, these transmembrane proteins go through different conformational changes to transport a substance through the membrane.

One of the major features of transmembrane proteins is that its polytopic nature, that aggregate and precipitate in water. The region of the protein sequence that spans through the membrane is widely hydrophobic in nature, permitting for transmembrane prediction methods such as the hydropathy plot. Polarity of amino acids varies along the length of transmembrane sequence. This polarity largely depends location of the amino acids in the spanning peptides or the opening is situated in the lipid portion or in the aqueous portion [27].

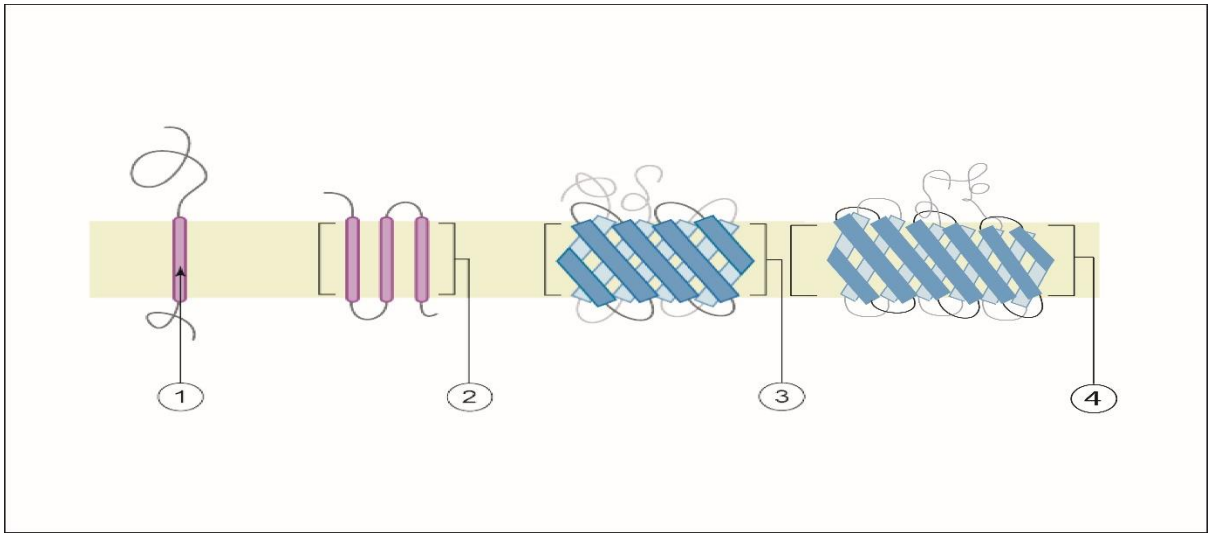

Figure 2.2: Cell membrane spanning transmembrane protein. 1. Is a single pass transmembrane protein, 2. & 3. is multi pass transmembrane protein and 4. Is a seven pass transmembrane protein.

Other type of integral membrane protein is the integral monotopic protein. This integral monotopic protein is also permanently attached to the cell membrane but does not pass through the cell membrane [28].

Generally, two basic types of transmembrane proteins are known [29], Such as, alpha-helical transmembrane proteins and beta-barrels transmembrane proteins of spanning sequences. Commonly, alpha-helical transmembrane peptide sequences are located in the inner membranes of bacterial cells or the plasma membrane of eukaryotes, and sometimes in the outer surface of membranes. This is the most predominant group of transmembrane

proteins, as an example, in humans, around 27% of all proteins have been estimated to be alpha-helical membrane proteins [30].

Other type of classification of these transmembrane proteins can be done through the spanning number of these peptides, like how many times these transmembrane sequence or peptide of the protein passed the bilayer, if it is one then it will be called as single pass transmembrane protein and if it is passed seven times then it will be called as seven pass transmembrane protein. This transmembrane-ship can be extended from single pass to seven passes transmembrane protein.

2.4 Homology Modeling

In the case of where, no crystal of protein or ligand or receptors structure is present, then homology modelling is performed. It is also known as comparative modelling of protein or ligand or receptors, refers to constructing an atomic-resolution model of the "target" protein from its primary structure and an experimental three-dimensional structure of a related homologous protein, commonly termed as template. If there is lack of probable template, then global template can be applied. Homology modelling depends on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps of amino acid residues in the query sequence to amino acid residues in the template sequence. It has been established concept that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very dissimilar structure from the template [31].

Evolutionarily related proteins have similar amino acid sequences and naturally occurring homologous proteins have similar protein structure. In one study, it has been shown that three-dimensional protein structure is evolutionarily more conserved than would be expected on the basis of sequence conservation alone [32].

The multiple sequence alignment and template structure are then used to produce a simple structural model of the target. Because protein structures are more conserved than DNA sequences or RNA sequences, detectable levels of sequence similarity usually imply significant structural similarity [33].

The quality of the homology model may vary on the quality of the sequence alignment and template structure or probability of finding a good template structure. The approach can be complicated by the presence of alignment gaps, which is very common in multiple sequence

alignment (commonly called indels), that indicate a structural region present in the target but not in the template, and by structure gaps in the template that arise from poor resolution in the experimental procedure (usually X-ray crystallography) used to solve the structure. Model quality drops with reducing sequence identity; a typical model has ~1–2 Å root mean square deviation between the matched C^{α} atoms at 70% sequence identity but only 2–4 Å agreement at 25% sequence identity, which is a huge variation in deviation when quality of model is in concern. However, the errors are significantly higher in the loop regions of a protein sequence, where the amino acid sequences of the target and template proteins may be completely different.

Loop modelling is used where a template structure is totally absent, are generally much less accurate than the models predicted based on a template structure. Errors in side chain folding and position of amino acid also increase with decreasing identity, and variations in these packing configurations have been suggested as a important reason behind the poor model quality at low identity [34]. Overall, these various atomic-position errors are very important and impede the use of homology models for purposes that require crystallization or atomic-resolution data, such as drug design and protein–protein interaction analysis; even the quaternary structure of a protein may be very hard to predict from homology models of its subunit(s). Nevertheless, homology models can be useful in reaching qualitative summary about the biochemistry and biomolecular behaviour of the query sequence, especially in formulating hypotheses about why certain residues are conserved, which may in turn lead to experiments to test those hypotheses. For example, the spatial arrangement of conserved residues may suggest whether a particular residue is conserved to stabilize the folding, to participate in binding some small molecule, or to foster association with another protein or nucleic acid.

Homology modelling can generate high-quality structural models when the target and template are closely related, which has inspired the formation of a structural genomics consortium dedicated to the production of representative experimental structures for all classes of protein folds [35]. The chief inaccuracies in homology modelling, which worsen with lower sequence identity, derive from errors in the initial sequence alignment and from improper template selection [36]. These inaccuracies can be normalized by different structure validation tools, which are complementary and fix a standard of being a high quality model. Like other methods of structure prediction, current practice in homology modelling is

assessed in a biennial large-scale experiment known as the Critical Assessment of Techniques for Protein Structure Prediction.

2.5 Honey Bee and Allergens

2.5.1 Honey Bee (Apis mellifera)

Honey bee (*Apis mellifera*) is a flying insect fall within the genus of *Apis* in the bee group. They are widely recognised for building of perennial shells from wax, for the larger extent of their colonies, and for their surplus production and storage of honey, distinguishing their hives as a prized foraging target of many animals, including honey badgers, bears and human hunter gatherers. In twenty first century, only seven species of honey bee *(Apis mellifera)* are recognized, with around 44 subspecies [37], although historically 7-11 species of honeybee's are recognized. There is some other type of honeybee species, those are recognized for their honey producing and storage capabilities, but interestingly, only the members of genus *Apis*, are actual honey bees.

Honey bee uses venom to protect thyself from the enemy. Thus, the honey bee sting and related venom sac of are modified so as to pull free of the body once lodged, and the sting apparatus has its own musculature and ganglion, which permits honey bee to keep sending venom once separated. The honey bee gland which secretes the alarm pheromone is also simultaneously associated with the sting apparatus. The embedded stinger emits additional alarm pheromone after it has torn loose; other defensive workers are thereby attracted to the sting site. The worker immediately dies after the sting becomes lodged and is subsequently torn loose from the bee's abdomen. The honey bee's venom commonly referred as apitoxin, carries several active components, the most abundant of which is melittin [38], and the most biologically active are enzymes, particularly phospholipase A2 [39].

Honey Bee venom is under extensive studies for its potential features and uses in reducing risks for adverse events from diverse venom therapy [40], also for rheumatoid arthritis [41], and efficacy as potential immunotherapy for attaining protection against insect allergen [42]. Bee venom products are currently being marketed in many countries, but, as of 2019, there are no approved clinical uses for these products which carry various warnings for potential allergic reactions [43].

2.5.2 Allergen of Apis mellifera

There are twelve known allergens of *Apis mellifera*, these allergens are chronically numbered from api 1 to api 12. These allergens have different biochemical features and biomolecular actions as well. Briefly, these allergens of honey bee are described as follows:

- ***Api m 1*:** The biochemical name of Api m 1 allergen is phospholipase A2. Molecular weight is about 16 Da and it has mild allergenicity in accordance with its severity among the patients. The function of phospholipase A2 is the secretion of this protein into venom follows a seasonal pattern. This variation is synchronized with melittin variation, i.e. their production increase in the same months. There are few metal binding sites for binding calcium ion and active sites in this allergen protein. Api m 1 also responsible for arachidonic acid secretion, lipid catabolic process and phospholipid metabolic process. Api m 1 is considered as mild allergen among the twelve allergens of honey bee.
- ***Api m 2*:** The biochemical name of Api m 2 allergen is Hyaluronidase. Molecular weight is about 39 Da and it has mild to moderate allergenicity in accordance with its blood level assay. The function of Hyaluronidase is to Hydrolyze high molecular weight hyaluronic acid to produce small oligosaccharides. There is one active site in this allergen protein which acts as proton donor. Api m 2 also responsible for defence response and carbohydrate metabolic process. Api m 2 is considered as weak allergen among the twelve allergens of honey bee.
- ***Api m 3*:** The biochemical name of Api m 3 allergen is acid phosphatase. Molecular weight is about 43 Da and it has acute allergenicity in accordance with immunoassays. Recombinant Api m 3 is specifically recognized by IgE antibodies of pooled serum in Western blots and by 37% of the individual sera of honeybee venom-sensitized patients in ELISA analysis. Chemical name of Api m 3 is known as Venom acid phosphatase Acph-1, which has catalytic activity. There are two active sites of this allergen positioned in 26 & 273 of its protein sequence. First one function as nucleophile and second one serves as proton donor. Functionally, Api m 3 has enzymatic activity similarly as hydrolase. This allergen is expressed by venom gland in honey bee. Api m 3 is considered most acute allergen among the twelve allergens of honey bee and can cause death if treatment can not be ensured at preliminary level.
- ***Api m 4*:** The biochemical name of Api m 4 allergen is Melittin. Molecular weight is about 3 Da and it has moderate allergenicity in accordance elevated serum IgE

antibodies. Melittin were found in 7 of 24 honeybee venom-sensitive persons and in 5 of 20 non-sensitive beekeepers, measured by RAST. - None of the 13 bee-allergic patients tested showed upregulation of the activation marker CD203c on basophils in response to Api m 4. One of the major toxins of bee venom with strong hemolytic activity. Forms a pore in the cell membrane by inserting into lipid bilayers in an alpha-helical conformation and has multiple effects, probably, as a result of its interaction with negatively charged phospholipids. It inhibits well known transport pumps such as the Na^+-K^+-ATPase and the H^+-K^+-ATPase. It increases the permeability of cell membranes to ions, particularly Na^+ and indirectly Ca^{2+}, because of the Na^+-Ca^{2+}-exchange. It acts synergistically with phospholipase A2.

- ***Api m 5*:** The obsolete name of Api m 5 allergen is allergen C and the biochemical name is Dipeptidylpeptidase IV. Molecular weight is about 100 KDa and it is the most weighted molecule of honey bee venom proteins. Venom dipeptidyl-peptidase which removes N-terminal dipeptides sequentially from polypeptides having unsubstituted N-termini provided that the penultimate residue is proline. May process promelittin into its active form and/or modulate the chemotactic activity of immune cells after the insect sting. Api m 5 has aminopeptidase activity and serine type peptidase activity. Api m 5 is considered as one of the weak allergens among the twelve allergens of honey bee.

- ***Api m 6*:** The Molecular weight of Api m 6 is about 8 Da and it has mild allergenicity in accordance with its blood level assay of 43 sera from Apis mellifera-allergic patients, 18 (42%) showed IgE binding to Api m 6 on immunoblot of bee venom extract. Purified Api m 6 was also recognized on immunoblot.
- ***Api m 7*:** The biochemical name of Api m 7 allergen is CUB serine protease. Molecular weight is about 39 Da and it has mild to moderate allergenicity in accordance with its blood level assay. The function of CUB serine protease is similar to hydrolase, protease and serine protease. This allergen is expressed by venom gut of honey bee.
- ***Api m 8*:** The biochemical name of Api m 8 allergen is Carboxylesterase. Molecular weight is about 70 KDa and it has mild to moderate allergenicity in accordance with its blood level assay. 46% of the sera from 28 bee venom allergic individuals had IgE that reacted with rApi m 8.0101 in ELISA. Api m 8 has similar to activity as

hydrolase, protease and serine protease. Api m 8 is considered as weak allergen among the twelve allergens of honey bee.

- ***Api m 9*:** The biochemical name of Api m 9 allergen is Serine carboxypeptidase. Molecular weight is about 60 KDa and it has mild to moderate allergenicity in accordance with its blood level assay. There are three active sites in this allergen protein. The function of Api m 9 is similar to hydrolase, protease and serine protease. Api m 9 is considered as weakest allergen among the twelve allergens of honey bee.
- ***Api m 10*:** The biochemical name of Api m 1 allergen is icarapin variant 2, carbohydrate-rich protein. Molecular weight is about 50-55 KDa and it has mild allergenicity in accordance with its severity among the patients. Api m 10 forms two isoforms by alternative splicing. There is a signal peptide and a chain of icarapin. Api m 10 is considered as mild allergen among the twelve allergens of honey bee. Still much information needs to be attained to know more about this protein.
- ***Api m 11*:** The biochemical name of Api m 11 allergen is major royal jelly protein. Molecular weight is about 50 KDa and it has mild allergenicity in accordance with its severity among the patients. There is a sinal peptide of Api m 11 which is 16 amino acid long and a chain which is 399 amino acid long.
- ***Api m 12:*** The biochemical name of Api m 12 allergen is Vitellogenin. Molecular weight is about 200 KDa and it has mild to moderate allergenicity in accordance with its severity among the patients. Functionally, Api m 12 is a Precursor of the egg-yolk proteins that are sources of nutrients during embryonic development (By similarity). Involved in the differentiation of honeybee larvae into queens. The function of Vitellogenin is transporting lipid and reserving nutrients. This is a power house protein of all honey bee protein which serve as nutrient reservoir. One of the most unique features of this protein is it's a serine rich venom.

2.6 Fundamentals of T Cell Lymphocytes

2.6.1 T Cell

T cell, sometime referred as T lymphocyte, is a kind of lymphocyte that plays key role in cell-mediated immunity. T cells can be differentiated from other lymphocytes, by the presence of specific type of receptor, which is called T cell receptor. These cells are termed as *T cells* because they mature in the thymus gland [44] (although some T cells also mature in the tonsils [45]). Usually, there are several types of T cells and each type have unique function in the immune system. The major proportion of the human T cell lymphocytes are

alpha beta T cells (αβ T cells). This type of T cell reorders theirs alpha (α) and beta (β) chains on the cell specific receptor and are the major part of adaptive immune system. Specialized gamma delta T cells are part of innate immune system (a small minority of T cells in the human body, more frequent in ruminants), have invariant T-cell receptors with limited diversity, that can effectively present antigens to other T cells [46].

2.6.2 Types of T cell

- **Effector:** Effector cells are most important type of T cells that actively respond immediately to a stimulus, such as co-stimulation. Effector T cell includes helper T cell, killer, regulatory, and potentially other T cell types. Memory cells are their opposite counterpart that are longer lived to target future infections as required.
- **Helper:** Helper T Cells (T_H cells) contribute other White Blood Cells (WBC) in immunologic processes, including maturation of B cells into plasma cells and memory B cells, and activation of cytotoxic T cells and macrophages. These cells are also known as $CD4^+$ T cells because they express the CD4 glycoprotein on their surfaces. Helper T cells become activated when they are presented with peptide antigens by MHC class II molecules, which are expressed on the surface of Antigen Presenting Cells (APCs).
- **Cytotoxic T Cell:** Cytotoxic T cells (T_C cells, CTLs, T-killer cells, killer T cells) destroy virus- infected cells and tumor cells, and are also implicated in transplant rejection. These cells are also known as $CD8^+$ T cells since they express the CD8 glycoprotein at their surfaces. These cells recognize their targets by binding to antigen associated with MHC class I molecules, which are present on the surface of all nucleated cells. Through IL-10, adenosine, and other molecules secreted by regulatory T cells, the $CD8^+$ cells can be inactivated to an anergic state, which prevents autoimmune diseases.

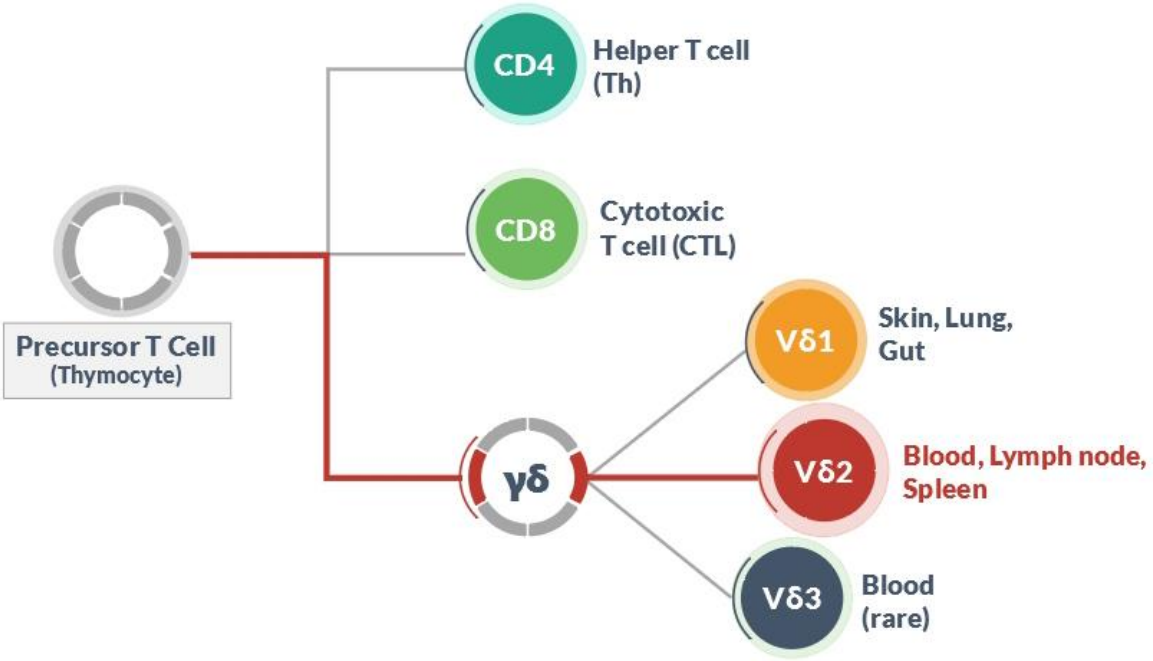

Figure 2.3: Different types of T Cell and its differentiation [46].

- **Memory**: Antigen-presenting T cells increase and differentiate into memory and effector T cells after they meet their cognate antigen within the context of an MHC molecule on the surface of an Antigen Presenting Cell (e.g. a dendritic cell). The prerequisite of the process is the appropriate co-stimulation with the antigen encounter.
- **Regulatory T cells:** Regulatory T cells (suppressor T cells) are very crucial for the maintenance of immunological tolerance. Their principle role is to shut down T cell-mediated immunity toward the end of an immune reaction and to suppress autoreactive T cells that escaped the process of negative selection in the thymus. Suppressor T cells along with Helper T cells can collectively be called Regulatory T cells due to their regulatory functions.
- **Natural Killer T cells:** Natural Killer T cells join the adaptive immune system with the innate immune system. The conventional type T cells that commonly recognize peptide antigens presented by Major Histocompatibility Complex (MHC) molecules, natural killer cells identify glycolipid antigen presented by a molecule named CD1d. Once this natural killer T cells are activated, these T cells can accomplish functions ascribed to both T_h and T_c cells. Natural killer (NK) T cells are also capable to identify and eliminate some tumor cells and the cells infected with herpes virus [47].

- **Mucosal Associated Invariant**: Mucosal associated invariant cells display innate, effector like abilities [48], [49]. In humans, mucosal associated invariant cells are commonly found in different part of body, such as blood, lungs, liver and mucosa, which defends against infection and microbial activity [48].
- **Gamma Delta T Cells:** Gamma delta T cells (γδ T cells) represent a small subset of T cells that possess a distinct T cell receptor (TCR) on their surfaces. A majority of T cells have a T cell receptor composed of two glycoprotein chains called α- and β- T cell receptor chains. However, in Gamma delta T cells, the T cell receptor is composed of one γ-chain and one δ-chain. This group of T cells is much less common in humans and mice (about 2% of total T cells); and are found mostly in the gut mucosa, within a population of lymphocytes known as intraepithelial lymphocytes. The antigenic molecules that activate Gamma delta T cells are still widely unknown.

2.6.3 T Cell Epitope

Generally, T cell epitopes are expressed on the cellular surface of an Antigen Presenting Cell (APC), where it binds with major histocompatibility complex molecules. Antigen presenting cells in humans are specialized to present MHC class II peptides, alternatively most nucleated somatic cells present major histocompatibility complex (MHC) class I peptides. The usual length of peptides that is presented by T cell epitopes by MHC class I molecules are 8 and 11 amino acid length, whereas MHC class II molecules present longer peptides, 13-17 amino acids in length [50], and non-classical MHC molecules also present non-peptidic epitopes such as glycolipids.

In wet lab analysis, epitopes can be mapped using protein microarrays, and with the ELISPOT or Enzyme Linked Immuno Sorbent Assay (ELISA) techniques. Another wet lab technique involves high-throughput mutagenesis, an epitope mapping method developed to improve rapid mapping of conformational epitopes on structurally complex proteins [51]. There is several *in silico* methods that helps in reducing cost and labour of analysis wet lab methods. MHC class I and II epitopes can be reliably predicted by computational means alone [52], although not all *in silico* T cell epitope prediction algorithms are equivalent in their accuracy [53].

Epitopes are often used in proteomics and the study of other gene products. Using recombinant DNA techniques genetic sequences coding for epitopes that are

recognized by common antibodies can be fused to the gene. Following synthesis, the resulting epitope tag allows the antibody to find the protein or other gene product enabling lab techniques for localisation, purification, and further molecular characterization. Common epitopes used for this purpose are Myc-tag, HA-tag, FLAG-tag, GST-tag, 6xHis [54], V5-tag and OLLAS [55]. Peptides can also be bound by proteins that form covalent bonds to the peptide, allowing irreversible immobilisation [56]. These strategies have also been successfully applied to the development of "epitope-focused" vaccine design [57], [58].

According to the molecular biology, T-cells present on their surface a specific receptor commonly known as T cell receptor (TCR) that activates the identification of antigens when they are displayed on the surface of antigen-presenting cells (APCs) bound to major histocompatibility complex (MHC) molecules. T-cell epitopes are presented by class I (MHC I) and II (MHC II) MHC molecules that are recognized by two distinct subsets of T-cells, CD8 and CD4 T-cells, respectively. Subsequently, there are CD8 and CD4 T-cell epitopes. CD8 T-cells become cytotoxic T lymphocytes (CTL) following T CD8 epitope recognition. Meanwhile, primed CD4 T-cells become helper (Th) or regulatory (Treg) T-cells [59]. Th cells amplify the immune response, and there are three major subclasses: Th1 (cell-mediated immunity against intracellular pathogens), Th2 (antibody-mediated immunity), and Th17 (inflammatory response and defense against extracellular bacteria)

Recognizing epitopes in antigens is of utmost interest for various practical reasons, including gaining more knowledge on disease etiology, immune system monitoring, developing diagnosis assays, and designing epitope-based vaccines. B-cell epitopes can be identified by different methods including solving the 3D structure of antigen-antibody complexes, peptide library screening of antibody binding or performing functional assays in which the antigen is mutated and the interaction antibody-antigen is evaluated [60], [61]. On the other contrary, experimental identification of T-cell epitopes is carried out using MHC multimers and lymphoproliferation or ELISPOT assays, among others [62], [63]. Traditional epitope identification has depended entirely upon experimental techniques, being costly and time-consuming. Thereby, scientists have developed and implemented epitope prediction methods that facilitate epitope identification and decrease the experimental load associated with it.

2.7 Fundamentals of B Cell Lymphocytes

2.7.1 B Cell

B cells, commonly referred as B lymphocytes, are a specific type of White Blood Cell (WBC) of the lymphocyte subtype [64]. The major function of B cells is taking part in humoral and innate immunity and these are the pivotal component of the adaptive immune system by secreting different antibodies [64]. Consequently, B cells present antigen and act as antigen presenting cells and secrete cytokines [64]. In mammals, B cells mature in the bone marrow, which is positioned at the core of almost all bones [65]. B cells, unlike the other two classes of lymphocytes, T cells and natural killer cells, express B Cell Receptors (BCRs) on their cell membrane. B cell receptors allow the B cell to bind to a specific antigen, against which it will initiate an antibody response [64].

2.7.2 Development of B cell

B cells commonly go through two major types of selection processes while its development in the bone marrow to ensure proper development. There are two types of B cell selection process, one is positive selection and another is negative selection. In case of Positive selection, antigen-independent signalling involving both the pre-BCR and the BCR [66], [67]. If these receptors do not bind to their corresponding ligand, B cells do not attain the proper signals and eventually ease to develop [66], [67]. On the other hand, negative selection occurs through the binding of self-antigen with the BCR; If the BCR can bind strongly to self-antigen, then the B cell undergoes different fates. There are four commonly known fates of B cell development, such as clonal deletion, receptor editing, anergy, or ignorance [67]. This negative selection process leads to a state of central tolerance, in which the mature B cells don't bind with self-antigens those are presented in the bone marrow [68].

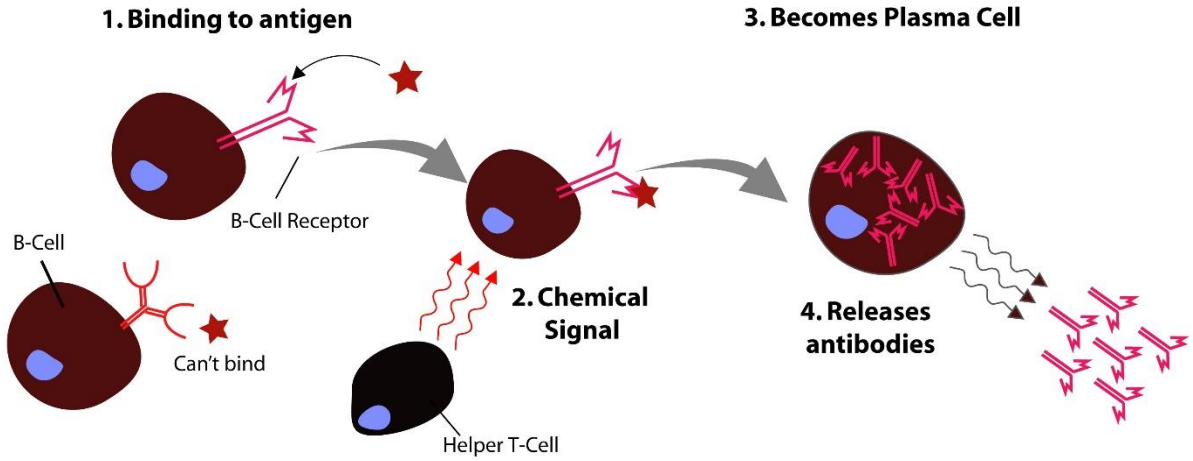

Figure 2.4: Biomolecular actions of B cell

To complete the development process, immature B cells migrate from the bone marrow into the spleen as transitional B cells, passing through two transitional stages, these transitional stages are called as T1 and T2 [69]. Throughout B cells migration to the spleen and after spleen entry, they are considered T1 B cells [70]. Within the spleen, T1 B cells transition to T2 B cells [70]. T2 B cells differentiate into either follicular (FO) B cells or Marginal Zone (MZ) B cells depending on signals received through the BCR and other receptors [71]. Once differentiated, they are considered as mature B cells, or naive B cells [70].

2.7.3 Activation of B cell:

Commonly, B cell activation occurs in the Secondary Lymphoid Organs (SLOs), such as the spleen and lymph nodes [72]. After B cells mature in the bone marrow, they migrate through the blood to Secondary Lymphoid Organs (SLOs), which get a constant supply of antigen through circulating lymph. At the secondary lymphoid organs (SLOs), B cell activation begins when the B cell binds to an antigen via its B cell receptor. Although the events taking place immediately after activation have yet to be completely determined, it is believed that B cells are activated in accordance with the kinetic segregation model, initially determined in T lymphocytes.

- **T Cell Dependent Activation:** Antigens that stimulate B cells with the help of T-cell are referred as T cell-dependent antigens and include foreign proteins [72]. They are commonly termed as such cause of their inability to induce a humoral response in organisms that lack T cells [72]. B cell response to these antigens takes multiple days, though antibodies generated have a higher affinity and are more functionally versatile than those generated from T cell-independent activation [72].

- **T Cell Independent Activation:** Antigens that induce B cells without T cell help are known as T cell-independent antigens [72] and include foreign polysaccharides and unmethylated CpG DNA [73]. They are named as such cause of their ability to induce a humoral response in organisms that lack T cells [72]. B cell response to these antigens is very rapid, though antibodies generated tend to have much lower affinity and are less functionally versatile than those generated from T cell-dependent activation [72].

  As with T cell dependent antigens, B cells activated by T cell independent antigens need additional signals to complete activation, but instead of receiving them from T cells, they are provided either by recognition and binding of a common microbial

constituent to Toll-Like Receptors (TLRs) or by higher crosslinking of B cell receptors to repeated epitopes on a bacterial cell [72]. B cells activated by T cell independent antigens go on to proliferate outside lymphoid follicles but still in Secondary Lymphoid Organs (SLOs), possibly undergo immunoglobulin class switching, and differentiate into short-lived plasma blasts that produce early, weak antibodies mostly of class IgM, but also some populations of long-lived plasma cells [74].

- **Memory B Cell Activation:** Memory B cell activation starts with the identification and binding of their target antigen, which is shared by their parent B cell [75]. Some of the memory B cells can be activated without the help of T cell, such as certain virus-specific memory B cells, but others need T cell help [76]. Upon antigen binding, memory B cell takes up the antigen through receptor-mediated endocytosis, degrades it, and presents it to T cells as peptide pieces in complex with MHC-II molecules on the cell membrane [75]. Memory T helper (T_H) cells, typically memory follicular T helper (T_{FH}) cells, that were derived from T cells activated with the same antigen recognize and bind these MHC-II-peptide complexes through their T cell receptor [75]. Following the T cell receptor of MHC-II peptide binding and the relay of other signals from the memory T_{FH} cell, the memory B cell is activated and differentiates either into plasmablasts and plasma cells via an extrafollicular response or enter a germinal center reaction where they generate plasma cells and more memory B cells [75], [76]. It is unclear whether the memory B cells undergo further affinity maturation within these secondary GCs [75].

2.7.4 Types of B Cell

- **Plasmablast:** A very short life spanned, proliferating antibody-secreting cell those are normally generated from B cell differentiation [72]. Plasmablasts are secreted early stage of an infection and their antibodies tend to have a lower affinity towards their target antigen compared to plasma cell [73].
- **Plasma Cells:** Plasma cells usually have much longer life than Plasmablast, non-proliferating antibody-secreting cell, those are generally generated from B cell differentiation [72]. A study concluded that, B cells first differentiate into a plasmablast-like cell, after that it re-differentiate into a plasma cell [73].

- **Lymphoplasmacytoid Cell:** A cell with a mixture of B lymphocyte and plasma cell morphological features that is thought to be closely related to or a subtype of plasma cells. This cell type is found in pre-malignant and malignant plasma cell dyscrasias that are associated with the secretion of Immunoglobulin M (IgM) monoclonal proteins.
- **Memory B Cell:** Memory B cell is a type of dormant B cell arising from B cell differentiation [72]. The function of memory B cell is to circulate through the body and initiate a stronger, more rapid antibody response if they find any foreign particle or antigen that had activated their parent B cell [77].
- **B-2 cell - FO B cells and MZ B cells:** Follicular (FO) B Cell (also known as a B-2 cell) - Most common type of B cell and, when not circulating through the blood, is found mainly in the lymphoid follicles of secondary lymphoid organs (SLOs) [73]. They are responsible for generating the majority of high-affinity antibodies during an infection [72]. Marginal Zone (MZ) B cell is commonly found in the marginal zone of the spleen and serves as a first line of defense against blood-borne pathogens, as the marginal zone gets large amounts of blood from the general circulation [78]. B-1 cell: B-1 cell normally arises from a developmental pathway different from Follicular B cells and Marginal zone B cells [79]. Regulatory B (Breg) cell - An immunosuppressive B cell type that inhibits the expansion of pathogenic, pro-inflammatory lymphocytes through the secretion of IL-10, IL-35, and TGF-β [80]. Also, it promotes the generation of regulatory T (Treg) cells by directly interacting with T cells to skew their differentiation towards Tregs [80]. No common Breg cell identity has been described and many Breg cell subsets sharing regulatory functions have been found in both mice and humans [80]. It is currently unknown if Breg cell subsets are developmentally linked and how exactly differentiation into a Breg cell occurs [80]. There is evidence showing that nearly all B cell types can differentiate into a Breg cell through mechanisms involving inflammatory signals and BCR recognition [80].

2.8 Classification Based B Cell Epitope Mapping

Identifying B cell epitopes is a prerequisite step for designing epitope-based vaccines, therapeutic antibodies, and tools for diagnostics. Currently, epitopes-based vaccines development is state of art technique discovering promising group of biopharmaceuticals. In the last few decades, several studies have been conducted to categorize experimentally

obtained epitopes with the aid of different tools. Different *in silico* tools are developed and employed in quest of predict B cell epitopes based on protein sequence or structural features. The focal points of identifying B cell epitope is to replace an antigen in the process of immunization, serodiagnosis, and antibody production. The precise determination of B cell epitopes still presents crucial challenges for vaccinologists. There are remarkable advances in the field of B cell epitope mapping and computational prediction and these advances lead molecular insights into the process of biorecognition and formation of antigen-antibody complex and this will successfully aid in localizing B cell epitopes.

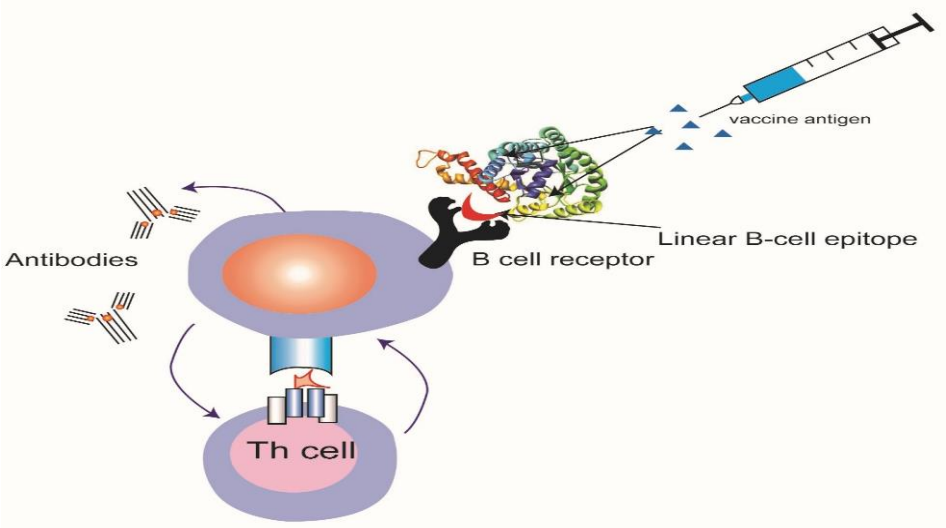

Figure 2.5: Molecular function of B cell epitope-based vaccine

There are some key features for identifying a potential B cell epitope. By evaluating some characteristics of a protein, it might be possible to determine suitable vaccine candidate. From the experimental data it is commonly known that, B cell epitope situated in a protein's most accessible region and this region are hydrophilic in nature. Also, B cell epitopes are part of beta turn region of a protein and that portion is usually antigenic.

2.8.1 Beta turn region of protein

Beta turn is a special characteristic of secondary structure of proteins. These beta turns are not regular feature of proteins and responsible for change in direction of secondary structure of the polypeptides. These motifs are very common in proteins and polypeptides. Each consists of four amino acid residues (labelled i, i+1, i+2 and i+3). They can be defined in two ways, firstly, by the possession of an intra-main-chain hydrogen bond between the CO of residue i and the NH of residue i+3; Alternatively, having a distance of less than 7Å between the Cα atoms of residues i and i+3. The hydrogen bond criterion is the one most appropriate

for everyday use, partly because it gives rise to four distinct categories; the distance criterion gives rise to the same four categories but yields additional turn types.

2.8.2 Surface Accessibility of Protein

For mapping B cell epitope, it is prerequisite to identify most surface accessible region in a protein sequence. Surface accessibility and beta turn region of a protein play pivotal role as it directly involves with protein folding. Understanding the folding of proteins remains one of the major scientific challenges. One way to explore this complex problem is to get information from the protein structures themselves, such as secondary structures, side chain interactions, H-bonds, and other parameters describing the solvent accessibility.

The folding process of soluble proteins decreases the surface in contact with the solvent. This is related to the secondary structures of proteins. Accurate knowledge of residue accessibility would thus aid the prediction of secondary structures. Different methods of prediction are based on the use of protein structure databases and on multiple sequence alignments. They have various efficiencies, notably depending on the number of relative accessibility states (i.e., exposed, buried, and in-between) [81]- [84].

Further, because active sites of proteins are often located at the surface of the protein, greater insight into residue accessibility would be important in understanding and predicting structure/function relationships with B cell epitopes.

2.8.3 Antigenicity of Protein

Antigenic property of a protein is crucial part of a protein, specifically in case of mapping an epitope, because epitope should be antigenic in nature. The term antigenicity refers to the ability of a protein surface region to be potentially antigenic, while immunogenicity' refers to the ability of any antigenic site to elicit such response under particular circumstances (immunization protocol, genetic constellation of the organism, etc.). The fact that bona fide antigenic sites may not be immunogenic in one (or several) experimental situations, and that the experimenter may not know the reason why the site is silent, makes antigenicity research a challenging discipline. Even in situations when one does obtain an antibody response to a protein antigen, conceptual difficulties abound. Identification of epitopes can only be based on indirect experimental procedures (such as comparative strength of binding of the same specific antibody to homologous proteins with a small number of amino acid replacements) and thus the antigenicity data is never complete in principle. Other factors introducing

uncertainty include varying immunogenic potentials of selected amino acid replacements in a set of homologous proteins, or an apparently random selection of monoclonal antibodies of a given specificity.

2.8.4 Hydrophilicity of Protein

Naturally, epitopes are hydrophilic in nature and B cell epitopes are mapped in hydrophilic zones of a protein sequence. So, epitopes can be mapped through hydrophilicity plot. Generally, hydropathicity graph is a quantitative analysis of the degree of hydrophobicity or hydrophilicity of amino acids of a protein. It is used to characterize or identify possible eater affinity related structure or domains of a protein.

The plot has amino acid sequence of a protein on its x-axis, and degree of hydrophobicity and hydrophilicity on its y-axis. There is a number of methods to measure the degree of interaction of polar solvents such as water with specific amino acids. For instance, the Kyte-Doolittle scale indicates hydrophobic amino acids, whereas the Hopp-Woods scale measures hydrophilic residues [85]. This can be also applicable for histamine receptors [86], [87].

Analyzing the shape of the plot gives information about partial structure of the protein. For instance, if a stretch of about 20 amino acids shows positive for hydrophobicity, these amino acids may be part of alpha-helix spanning across a lipid bilayer, which is composed of hydrophobic fatty acids. On the converse, amino acids with high hydrophilicity indicate that these residues are in contact with solvent, or water, and that they are therefore likely to reside on the outer surface of the protein.

Based on these features derived from experimental B cell epitopes, there is an opportunity to develop a single interface tool that can identify potential therapeutic vaccines via determining B cell epitopes of other proteins, those are causal agent of different diseases.

2.9 Chapter Summary

Honey bee is a valuable economic insect with potential threat to bee keepers as well as mass population for its venom allergen. Api m 3 is considered as the highest prominent causal agent among other venom allergen of honey bee. If it is possible to develop B cell and T cell epitope-based vaccine against Api m 3 allergen of honey bee, then we can safeguard one of the most vulnerable community of our country. Along with this, for mapping B cell epitope in a simplified and accurate way a classifier-based algorithm may be handy to eradicate few pitfalls in the area of B cell epitope-based vaccine prediction.

Chapter 3

Methodology

3.1 Introduction

In this *in silico* study, we plan to characterize and model of Histamine receptors. We focus on the physiochemical properties, motif, domain, transmembrane region and three-dimensional structure of *Homo sapiens.* Histamine receptors to better understand histamines biological events. Physicochemical properties of histamine receptors can provide significant knowledge on biological events of these G protein coupled receptors. *In silico* computational study can be useful to attain desired information. Furthermore, sequence alignment provides insight on homology, dissimilarities and conserved region of a group of protein [88]. In this regard, multiple sequence alignment data is more effective in order to acquire clarified information of sequence homology. Protein motif is crucial part of its feature and can be analysed through computational. Finally, we planned to develop 3D model which can be helpful to study complex structure of protein and target specific region of a targeted protein to design effective drugs or vaccines. The main objective of epitope prediction is to design and bring up with hypoallergenic molecules that can replace venom related treatments and other traditional healing techniques. Therefore, the outcome of our current study may aid their value through providing devising new therapeutic modalities for immunotherapy of honeybee venom allergy.

3.2 Materials & Methods

The research work of this study has three folds to fulfil the objectives (Figure 3.1). Different resources were applied to analyze biological features of histamine receptors and various techniques were employed to design models for histamine H1 receptor, Histamine H2 receptor, Histamine H3 receptor and Histamine H4 receptor. Here, three major model validation techniques are used. These methods are discussed in the first fold of methodology section and different techniques for designing vaccine against an allergen are described in section two of the methodology section. In the last section of methodology, we discussed techniques for developing a classifier-based B cell epitope prediction.

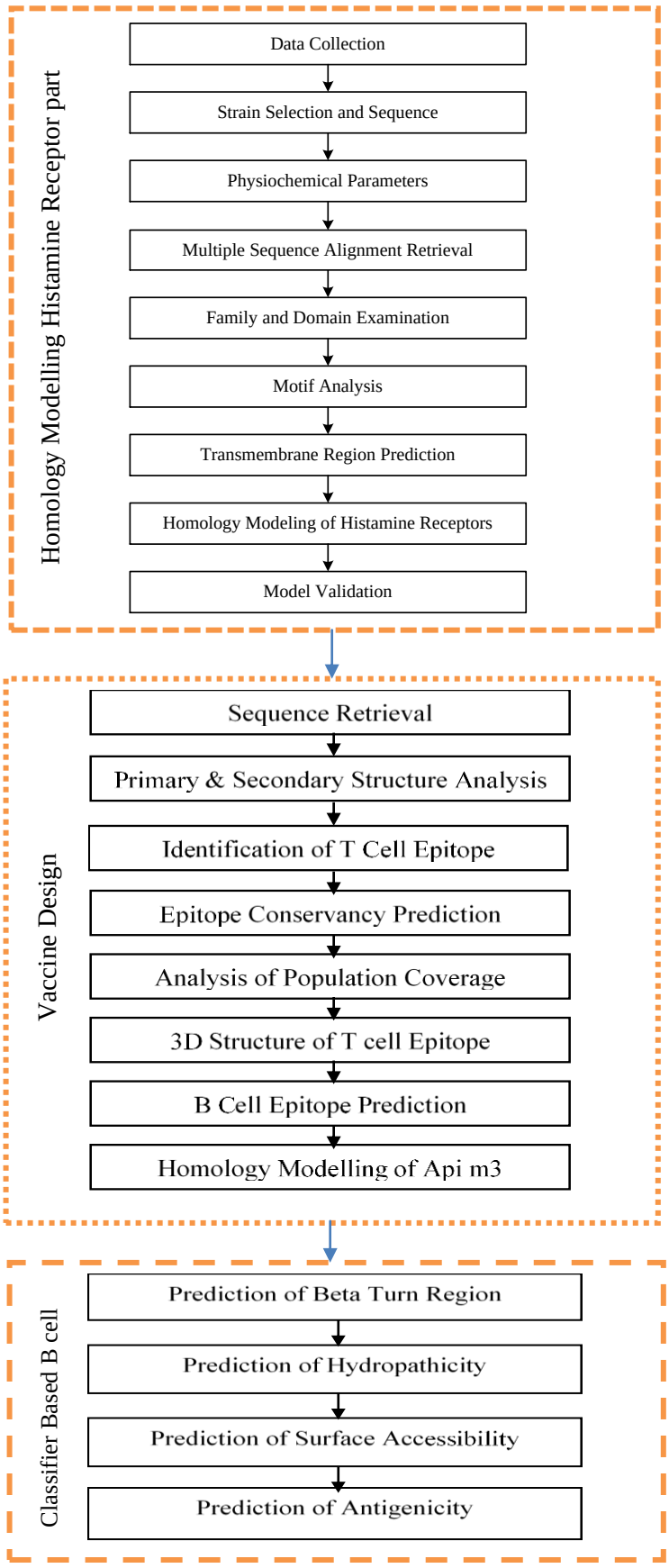

Figure 3.1: Schematic Outline of Methodology

3.3 Methods Involved in Homology Modeling of Histamine Receptors

3.3.1 Data Collection and Sequence Retrieval

Four histamine receptors of Human were retrieved from UniProtKB database [89]. The accessions of different histamine receptors are as follows: P35367 (HRH1_HUMAN Histamine H1 receptor), P25021 (HRH2_HUMAN Histamine H2 receptor), Q9Y5N1 (HRH3_HUMAN Histamine H3 receptor) and Q9H3N8 (HRH4_HUMAN Histamine H4 receptor). Full FASTA sequences of these receptors were collected from the UniProtKB database.

3.3.2 Analysis of Physiochemical Parameters

Physiochemical parameters such as: molecular weight (M.Wt), theoretical pI (Isoelectric Point), most redundant amino acids, instability index (II), Aliphatic Index (AI) and grand average of hydropathicity(GRAVY) were calculated by using ExPASy ProtParam tool [90]. Isoelectric point [pI] was computed in order to determine the acidic or basic nature of protein [91]. From the amino acid distribution chart most redundant amino acid was obtained. From the chart, top two highly occurring amino acids were counted to analyze the redundant amino acids pattern. Th instability index denotes the stability of the enzyme in *in vitro* condition. Instability index below 40 is generally regarded as the enzyme is stable whereas greater than 40 are considered as unstable [92]. Aliphatic index defines thermal stability based on position occupied and redundancy of amino acids alanine, valine and leucine of globular proteins [93].

3.3.3 Multiple Sequence Alignment

In order to analyze sequence similarities of histamine receptors, multiple sequence alignment was performed. ClustalW [94] was used for sequence alignment of histamine receptors. Data attained from ClustalW further analyzed with Jalview 2 [95] in a quest of finding conserved or core region of histamine receptors.

3.3.4 Prediction of Histamine Family and Domain

Evolutionary pattern can be exposed for a group of protein by identifying their family and domain. The family and domain of a protein possess signature of its function. Protein family denotes on clan of this protein and domain illustrates its distinctive functions. InterPro was used for predicting family and domain of histamine receptors [96].

3.3.5 Motif Analysis

Protein motifs are vital signature of the belonging domain. In order to predict motifs of histamine receptors, MEME suit [97] was used.

3.3.6 Prediction of Transmembrane Region

Prediction of transmembrane helices is of utmost importance in functional analysis of protein. TMHMM Server v. 2.0 [98] was applied for predicting transmembrane helices in histamine receptors.

3.3.7 Homology Modeling of Histamine Receptors

Homology modeling of histamine receptors was conducted by using I-tasser using software [99]. As previously mentioned, crystal structure of histamine H1 receptor is already available, but we performed homology modeling of all four receptors in order to cross check the validation of our predicted models with crystal structure of histamine H1 receptor. This approach can justify our prediction of other histamine receptors (HRH2, HRH3 & HRH4) 3D model.

3.3.8 Model Validation

Predicted models usually contain some errors in their primitive structure. Trouble shooting step is prerequisite to overcome with these issues. Therefore, validation of predicted models was performed by using different software those are frequently used for model verification, such as RAMPAGE, ERRAT, PROCHECK [100, 101, 102].

3.4 Methods for Vaccine development of Api M3

3.4.1 Sequence Retrieval

The protein sequence of Api m3 was retrieved from UniProtKB database [100]. The accession number of this protein assigned as AAY57281.1. Lastly, the sequence was saved in FASTA format for further analysis.

3.4.2 Primary and Secondary Structure Analysis

Physiochemical properties of Api M3 were analyzed by using ExPASy ProtParam tool and another ExPASy tool, self- optimized prediction method with alignment (SOPMA) [103] was used to analyse secondary structure. ProtParam tool was applied to attain data on different physiological and chemical properties of Api M3. Amino acid composition, molecular weight, theoretical isoelectric point (pI), grand average hydropathicity (GRAVY), estimated half-life, extinction coefficient [104], instability index [105], aliphatic index [106] of the protein were estimated using the default parameters by protparam tool. SOPMA was applied to analyze properties such as, helical pattern, globular regions,

transmembrane helices, solvent accessibility, bend region, random coil and coiled-coil region.

3.4.3 Identification of T cell Epitope

Identification of Consistent CTL epitopes is of utmost importance for contiguous vaccine design. Computational analysis of CTL epitopes reduce strenuous efforts of wet lab experiments also mitigates unnecessary cost. A web-based server, NetCTL-1.2 [107] was used in our study for predicting human CTL epitopes. This epitope prediction tool provides combined scores based on three parameters, such as proteasomal cleavage, TAP transport efficiency, and MHC class I affinity. For identifying potential T cell epitopes, we set the threshold value at for our current study which have sensitivity and specificity of 0.89 and 0.94, respectively. Five epitopes carrying highest combined score were selected for further experimentation. We used MHC-1 binding prediction [108], the Stabilized Matrix Base Method (SMM) [109] to calculate IC50 values of peptide binding to different class 1 MHC molecules. Prior to this analysis, peptide length was set to 9 amino acids for both frequent and non- frequent allele. The resulted alleles having binding affinity IC50 less than 200 nm were considered for further analysis.

MHC class II epitopes were predicted by using Immune Epitope Database 3.0 from IEDB T cell prediction tool [110]. In our current study, we have predicted MHC II HLA-DQA1*01:01/ DQB1*05:01, HLA- DRB1*11:01, HLA-DRB1*03:01 and HLA-DRB1*15:01 restricted T-cell epitopes by using this tool. This tool asses antigenic portion of a query amino acid sequence by providing a score. Higher score enhances greater probability of that region forming T cell epitope.

3.4.4 Epitope Conservancy Prediction

Epitope conservancy plays pivotal role in selecting effective epitopes. Epitope conservancy can be defined as the part of the protein sequences that restrain the epitope measured at or exceeding a specific level of identity [111].

3.4.5 Analysis of Population Coverage

Epitopes those were selected on the basis of IC50 value, further analyzed with IEDB population coverage tool for predicting population coverage for each epitope. IEDB population coverage tool is a web-based tool that works on the basis of MHC binding and/or T cell restriction data. This tool is highly efficient in determining the portion of individuals' response to a set of epitopes with known MHC restrictions. In our current study, we applied this tool for attaining the following information on population coverage: (1) prediction of

population coverage, (2) recognition of HLA combinations by the population, and (3) HLA combinations recognized by 90% of the population (PC90). One by one, five epitopes and their MHC-I molecules were pasted in the query box and population coverage area selected before submission to the server for analysis [111].

3.4.6 Design of the Three-Dimensional (3D) T Cell Epitope Structure

In this study, we used PEP-FOLD server [112] for designing 3D structures of T cell epitopes. PEP-FOLD applies a *de novo* approach to design 3D structures from a given amino acid sequence. This server is capable of designing 3D structure of 9-36 amino acid long peptide residues. PEP-FOLD worked on the principle of Structural Alphabet (SA) letters to explain the structural annotations of four consecutive amino acid residues, couples the predicted series of Structural Alphabet (SA) letters to a greedy algorithm as well as a coarse- grained force field [113], [114]. PEP-FOLD server provides five models as proposed 3D structure and best model was selected for further analysis.

3.4.7 B Cell Epitope Prediction

B-cell epitope prediction tools of IEDB were applied to find out linear B-cell epitopes. For predicting quality B-cell epitopes some significant features are always counted. Most crucial features of B-cell properties are flexibility, antigenicity, surface accessibility, hydrophilicity [115]. In our current study, four immunoinformatics tools, such as BCPREDS [116], ABCpred [117], BepiPred [118] and Bcepred [119] were applied for predicting linear the B-cell epitopes through full protein sequence of Api m3 by using their default threshold values. Outputs from four B-cell prediction tools were aligned and most common result was predicted as potential B-cell epitope. Lastly, flexibility, antigenicity, surface accessibility, hydrophilicity of predicted epitope was checked by using the Karplus and Schulz flexibility prediction [120], Kolaskar and Tongaonkar antigenicity scale [121], Emini surface accessibility prediction [122], Parker hydrophilicity prediction [123] and beta turn regions [124], respectively of IEDB B-cell analysis resource.

3.4.8 Homology Modeling & Validation of Api m3

Homology modeling of Api m3 was conducted by using SWISS-MODEL [125]. Predicted model usually contains some errors in their primitive structure. Trouble shooting step is prerequisite to overcome with these issues. Therefore, validation of predicted model was performed by using different software those are frequently used for model verification, such as RAMPAGE, ERRAT, PROCHECK [100] – [102]. Lastly, the model is illustrated with its T-cell and B-cell epitope with the Pymol 2.0 software.

3.5 Methods for Classifier based B Cell Epitope Prediction

In this research, we have used MATLAB programming to classify different parameters of protein to predict potential B cell epitope from a protein or a group of proteins. Here we used protein sequence bearing the accession AAY57281.1 from UniProtKB database for testing our algorithm.

3.5.1 Prediction of Beta Turn Region

Chou and Fasman secondary structure prediction scale for protein was used to develop an algorithm. Generally, a protein's secondary structure is composed of alpha helix, beta turn regions, and coil-coil regions. Among these regions, only beta turn region was our only point of concern. So, the developed algorithm should efficiently classify beta turn regions from alpha helix and coil-coil regions. Firstly, the plain protein sequence will be inserted onto the query box of the MATLAB programme and it will wait for the command of user about the length of the window. Then, the designed MATLAB programme calculates the total number of amino acids from the query sequence. finds the total number of windows can be created with one shift according to the user command on the window length. Initially, *i*=0 and programme will calculate mean weight of amino acids based on Chou and Fasman theorem [124] and store the mean weight until $i > N$. After terminating the process, it will automatically calculate the total mean weight and will be compared with mean weight. Finally, the programme will generate a resulting plot based on the outcomes of the program.

3.5.2 Prediction of Hydropathicity

From the HPLC analysis of twenty amino acid residues a hydropathicity scale was proposed by Parker and colleagues [123]. This scale of amino acids was used in our MATLAB programme to identify potential hydrophilic regions from hydrophobic regions of query protein according to the command of user on window length. For initial state, i=0 and programme will calculate mean weight of amino acids based on Parker's amino acid hydropathicity index and store the mean weight until $i > N$. After terminating the process, it will automatically calculate the total mean weight and will be compared with mean weight. Finally, the programme will represent the output as graphical plots.

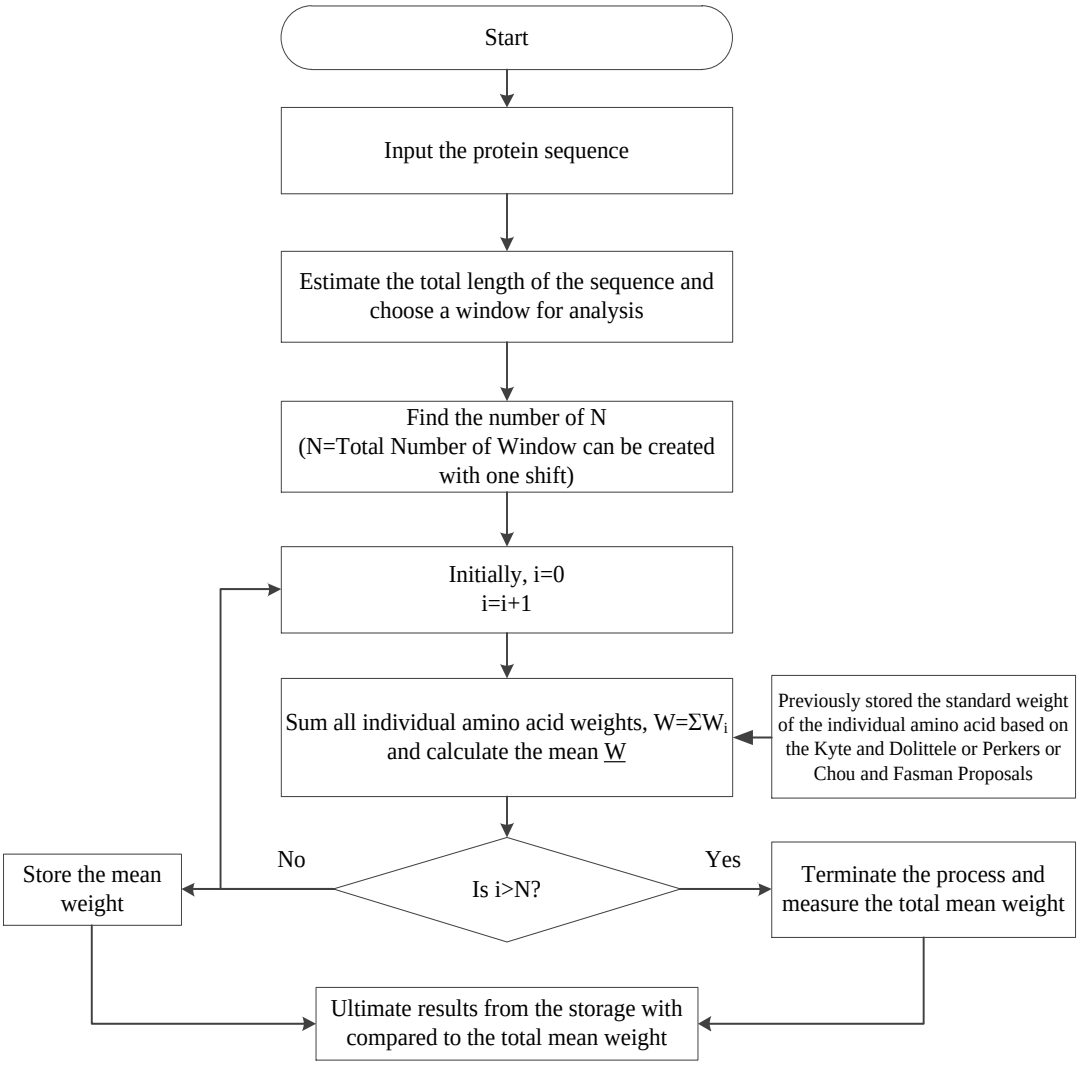

Figure 3.2: Block diagram of algorithm development for predicting B cell epitope of a protein antigen.

3.5.3 Prediction of Surface Accessibility

Our developed algorithm will calculate surface accessibility based on a given amino acid sequence, such as a point or frame for sequence number n will be a normalized product of the surface probabilities of amino acids in positions n - 2 to n + 3. By using the empirical amino acid accessible surface probabilities according to Janin and colleagues which are fractional probabilities (0.26 to 0.97) determined for an amino acid found on the surface of a protein. A surface residue is defined as one with >20 A' (2.0 nm') of water-accessible surface. With these fractional surface probabilities for amino acids, a surface probability (S) at sequence position n can be defined as follows:

$$S_n = [\prod_{i=1}^{6} \delta_{n+4-1}] * (0.37)^{-6} \quad (3.1)$$

In (3.1), δ_x is the fractional surface probability for the amino acid at position *x*. The S_n for a random hexapeptide sequence = 1.0, with probabilities greater than 1.0 indicating an increased probability for being found on the surface. After calculating Normalized surface accessible values for amino acids, the steps illustrated in Figure 3.2 will be carried out to determine most surface accessible area in a protein sequence.

3.5.4 Prediction of Antigenicity

Kolaskar and Tongaokar developed a semi-empirical method which utilizes physicochemical properties of amino acid residues and their probabilities or frequencies of occurrence in experimentally known segmental epitopes was developed to predict antigenic determinants on proteins [121]. Application of this method to a large number of proteins has shown by the **Kolaskar and Tongaonkar that the method can predict antigenic determinants with about 75% accuracy which is better than most of the known methods.** Firstly, the plain protein sequence will be inserted onto the query box of the MATLAB programme and it will wait for the command of user about the length of the window. Then, Programme will automatically calculate the number of amino acids in the query sequence and find the total number of windows can be created with one shift according to the user command on the window length. Initially, i=0 and programme will calculate mean weight of amino acids based on **Kolaskar and Tongaonkar** proposal and store the mean weight until $i > N$. After terminating the process, it will automatically calculate the total mean weight and will be compared with mean weight. Finally, the programme will generate a plot based on the output.

3.6 Chapter Summary

This Chapter has discussed about methods involved in designing and validation of four histamine receptors and also to get information on biological information about histamine receptors. It also discussed on the methods for development of vaccine against api m3 allergen of *Apis mellifera*. Finally, it discussed on classifier-based B cell epitope Identification.

Chapter 4

Result Analysis and Discussions

4.1 Introduction

This chapter discuss on different biological and structural features of histamine receptors and examine predicted models' authenticity. Secondly, acceptability of T cell and B cell epitopes predicted are analysed and further discussed to better understand its functions. Finally, Correlation between B cell epitope localization and physicochemical properties (e.g., hydrophilicity, solvent accessibility, flexibility, turns, polarity, antigenicity, and surface exposure) features are extracted to develop algorithms to predict and cross check B cell epitopes.

4.2 Analysis of Physicochemical properties and Homology models of Histamine Receptors

4.2.1 Physicochemical Properties of Histamine Receptors

Analysis through ExPASy ProtParam tools, it has been found that molecular weight of histamine receptors ranged from 40KDa-55KDa. Among the histamine receptors, HRH1 weighed highest (55.784 KDa) and HRH2 weighted lowest (40.098 KDa). Molecular weight varied with their relative amino acid number. Higher number of amino acid residues in the protein sequence resulted higher molecular weight as shown in Table 4.1. The isoelcetric point of these receptors has been found in the range of 9.33-9.62. *In vitro* stability of histamine receptors were also studied and found that, most of these receptors scored more than 40, except histamine H2 receptor (HRH2). This indicates relative instability of these receptors in *in vitro* conditions. Aliphatic index of these receptors were well above 90. GRAVY value was calculated as -0.086 for histamine H1 receptor (HRH1). In contrast, other histamine receptors showed positive GRAVY value. Finally, it was observed that, these four histamine receptors were rich in leucine as shown in Table 4.1.

Table 4.1: Analysis of Physicochemical Properties of Human Histamine Receptors

Receptors	Accession	No. of A.A	Molecular Weight	Theoretical pI	Instability Index	Aliphatic Index	GRAVY	Two Most Redundant AA %
HRH1	P35367	487	55784.12	9.33	47.00	90.06	-0.086	Leu (L) 11.3%, Ser (S)8.8%
HRH2	P25021	359	40098.12	9.36	34.93	113.51	0.396	Leu (L) 13.1%, Val (V) 8.9%
HRH3	Q9Y5N1	445	48671.37	9.43	42.23	89.3	0.166	Leu (L) 12.1%, Ala (A)11.2%
HRH4	Q9H3N8	390	44495.89	9.62	45.36	100.21	0.314	Leu (L) 11.8%, Ser (S)13.6%

Histamine receptors are crucial G protein coupled receptor for endogenous and exogenous histamine receive and transfer [103]. From our study, it is found that molecular weight of histamine H1 receptor (HRH1) is around 55.7KDa which is relatively higher than other receptors and histamine H2 receptor's (HRH2) molecular weight was lowest (40Kda) among four histamine receptors of human. This indicates that there might be heavy amino acid side chains in its tertiary structure. Also, it can be seen in Figure 4.4 that, 3D structure of Histamine receptor 1 (HRH1) is heavily loaded with alpha helix. All four receptors are basic in nature as the isoelectric point hits the value near 9.5. From the theory, instability index value greater than 40 regarded as the protein is unstable [104]. From the instability index, we found that histamine H2 receptor (HRH2) is stable and other three receptors value are marginally over 40, those have a higher tendency of becoming stable from slightly unstable conditions. Aliphatic index states relative volume poised by aliphatic side chains such as alanine, valine, isoleucine, and leucine. It also denotes thermal stability of a protein [105]. Aliphatic index of histamine receptors are ranging from 89.3 to 113.5, which indicates that the tendency of having a wide range of temperature sensitivity as aliphatic index (AI). This aliphatic index is well supported by data of most two redundant amino acids in the histamine receptors sequences. GRAVY value of protein

denotes hydropathicity of a protein and indicates weather the protein side chains are hydrophilic or hydrophobic in nature [85]. Leaving the histamine H1 receptor (HRH1), all other receptors are hydrophobic in nature and histamine H1 receptor (HRH1) can be interpreted as very slightly hydrophilic to hydrophobic.

4.2.2 Conservancy Analysis of Histamine receptor

No significant similarities have been observed among four histamine receptors. Partially conserved regions have been found from amino acid residue no. 75-94 (Figure 4.1A), which is about 18 amino acids long and this is the most conserved region found in multiple sequence alignment. Another partially conserved region was found 14 amino acids long, from residue number 477-490(Figure 4.1D). Two other sites (Figure 4.1B, 4.1C) lacking conservancy quality and consensus value comparing to sites are shown in Figure 4.1A & Figure 4.1D.

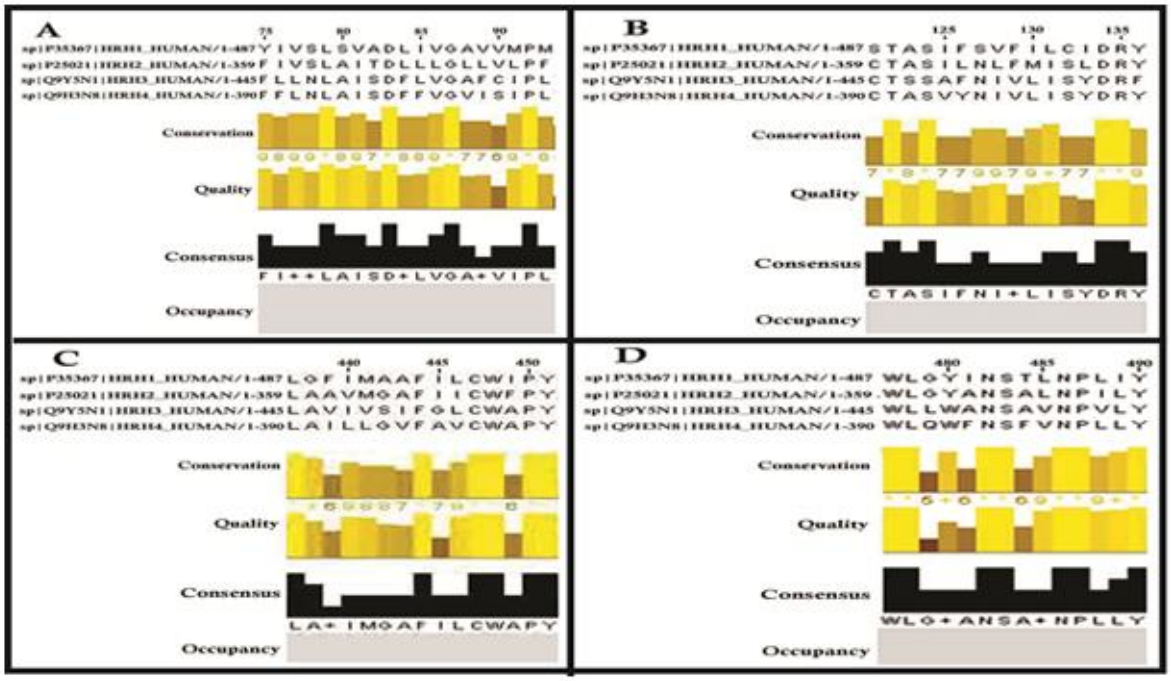

Figure 4.1: Multiple sequence alignment and conservancy analysis: (A-D), In four different sites,partially conserved regions have been found. From the figure, A & D sites carry most conservancy among the four notable sites found from multiple sequence alignment data.

4.2.3 Family and Domain of Histamine receptors

All of these four histamine receptors belong to major protein family, G protein-coupled receptor. Functional domain of four receptors was same as well, which is GPCR, rhodopsin-like, 7TM (Seven Pass Transmembrane). From the family analysis of these receptors, it can be said that, these are members of 7TM family and these receptors are

integral component of cellular membrane. Histamine H1 receptor regulates vascular permeability, up regulates vasoconstriction and responsible for eosinophil chemotaxis [8]. On the other hand, Histamine H2 receptor positively influences vasoconstriction and it stimulates gastric acid secretion [9]. As it has been already discussed, Histamine H3 receptor is associated with Central Nervous System (CNS), it regulates release of some neurotransmitter and also controls the level in cellular matrix upon activating feedback mechanisms. Lastly, Histamine H4 receptor plays pivotal role in inflammation response, regulates MAPK cascade and positively regulates cytosolic calcium ion concentration.

4.2.4 Motif Prediction

Motifs of histamine receptors were predicted by using MEME suite. By default set up, it can predict up to three motifs and distribution of motifs can be selected in three different parameters. For this experiment, we selected number of motifs as five and site distribution as any number of repetitions. All other parameters were left as provided in default value. MEME suite automatically predicts the width and occurrence number of motifs, in order to minimize the E-value of predicted motif.

We found four motifs for four histamine receptors with lower E value. We found lowest E value of 3E-23 for motif 1 with four sites and frame width of 50. Highest width was observed in motif 1 and lowest was found in motif 3. All motifs consisted of four sites and E values were ranged from 3.00E-23 to 1.80E-09. From our observation, motif 1 and motif 4 showed compactness as potential conserved regions of histamine receptors. Details of motif analysis are presented in Table 4.2.

Table 4.2: Motif analysis of four histamine receptors

Motif No.	Width	E- Value	Sites	Position in HRH1	Position in HRH2	Position in HRH3	Position in HRH4
1	50	3.00E-23	4	93-142	84-133	100-149	80-129
2	41	2.20E-20	4	446-486	266-306	352-374	297-319
3	23	2.10E-10	4	409-431	228-250	332-374	297-319
4	49	1.80E-09	4	41-89	32-80	48-96	29-77

In addition, Motif of a protein carries significant importance in proteomics as it denotes function of a protein domain and conserved regions. For designing primer, it is prerequisite to find out the conserved region of the protein. MEME suite especially designed for predicting Motifs in different types of sequences [97]. E- Value of the predicted motif

designates its statistical significance. Motifs with a lower E-value possess the probability of finding out an equally well conserved region in the test sequence [126]. From our analysis, we found total four motifs with lower E-value. Among these, motif no. 1 has lower E value, moderate width and four sites of occurrence.

4.2.5 Transmembrane Region

It is found that all four Histamine receptors are seven pass transmembrane protein. Transmembrane locations were different in each protein structure. From our observations, Histamine H2 receptor contained evenly distributed transmembrane helix than other three Histamine receptors. Histamine H1 receptor, Histamine H3 receptor and Histamine H4 receptor posses 100-200 amino acid residue gap between helix number 5^{th} and 6^{th} whereas, this gap is only 28 amino acid long in case of Histamine H2 receptor as shown in Figure 4.2. With this signature, Histamine H2 receptor functionally different from other receptor and this might aid in binding endogenous Histamine molecule with Histamine H2 receptor.

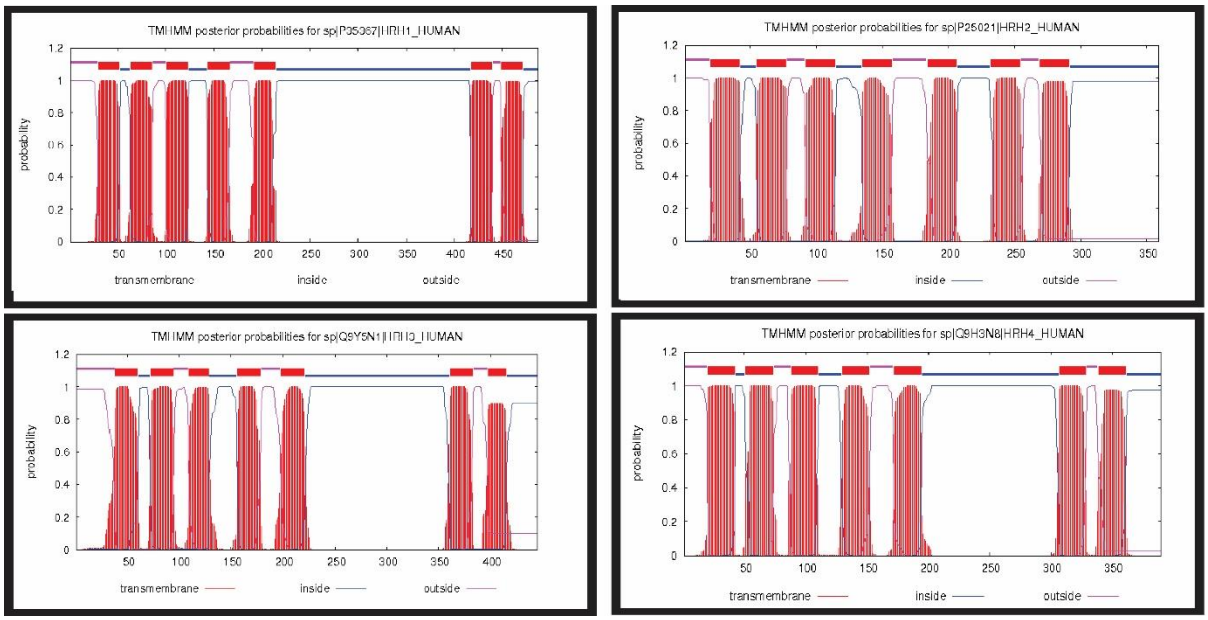

Figure 4.2: Transmembrane regions of histamine H1, H2, H3 and H4 receptor. Transmembrane regions are showed in thick red lines.

These receptors are also member of G protein coupled receptor and contain seven pass transmembrane regions. These transmembrane regions position vary from each other but the notable thing is approximately 200 amino acid residue gap between 5^{th} and 6^{th} transmembrane helix. This phenomenon is very common in histamine H1, H3 and H4 receptors but not observed in histamine H2 receptor.

4.2.6 Homology Modeling and Validation

The homology modeling has been performed by using I-tasser. For Histamine H1 receptor and Histamine H2 receptor, five models have been generated separately (Figure 4.3 & Figure 4.4) and for histamine H3 receptor and histamine H4 receptor, four models have been produced separately (Figure 4.5 & Figure 4.6).

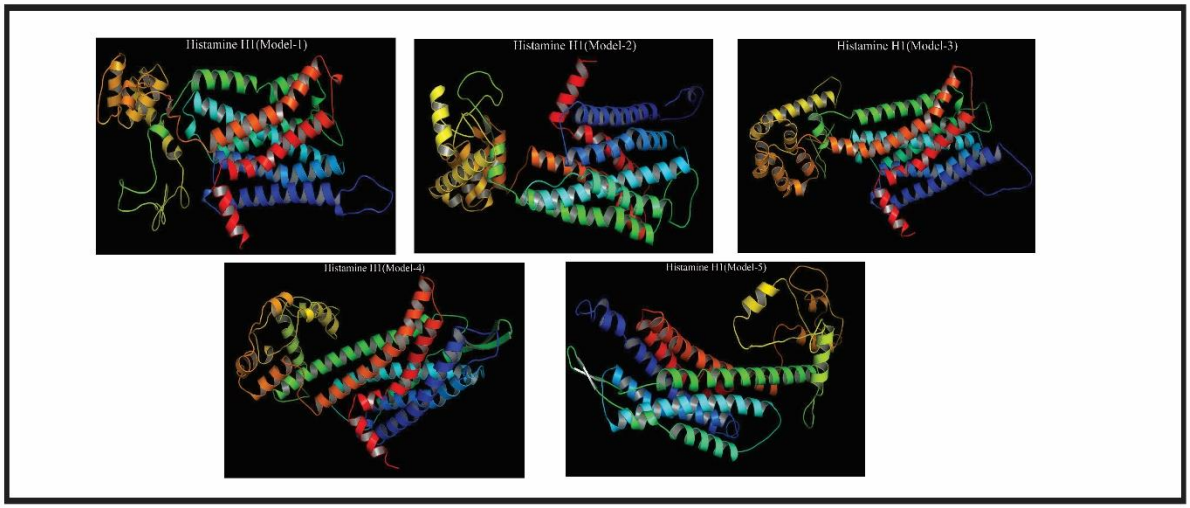

Figure 4.3: Homology modeling of histamine H1 receptor by I-tasser. Five models were generated by I-tasser. All models are heavily loaded with alpha helix and coil-coil regions. No sign of beta sheets in primitive structures. From these models of histamine H1 receptor one model will be selected as the best model after analyzing with model validation tools, i.e., ERRAT, RAMPAGE and PROCHECK.

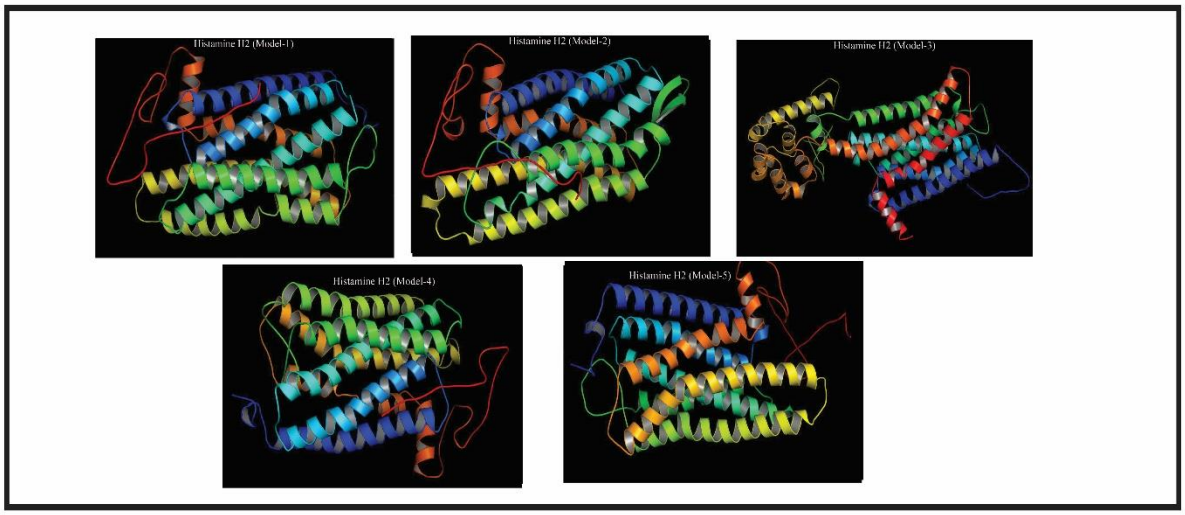

Figure 4.4: Homology modeling of histamine H2 receptor by **I**-tasser. Five models were generated by **I**-tasser. All models are heavily loaded with alpha helix and coil-coil regions. From these models of histamine H2 receptor one model will be selected as the best model after analyzing with model validation tools, i.e.,ERRAT, RAMPAGE and PROCHECK.

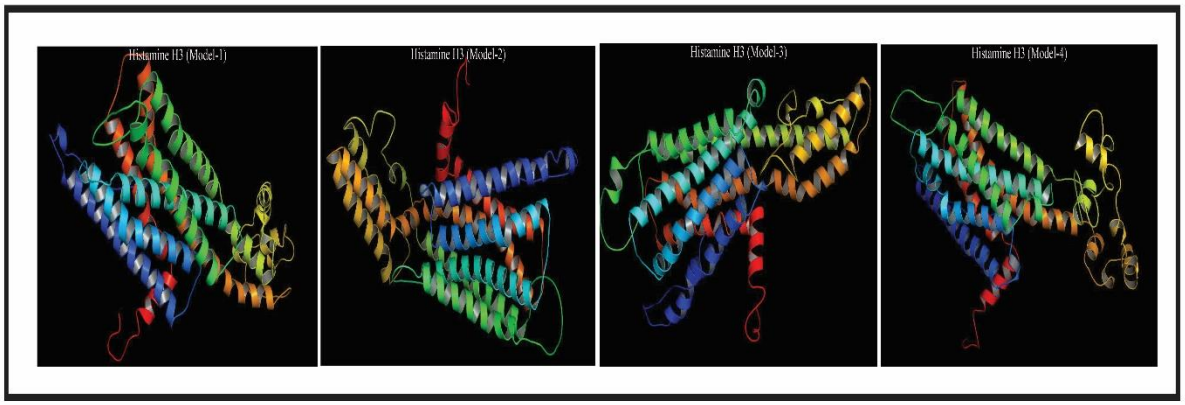

Figure 4.5: Homology modeling of histamine H3 receptor by **I**-tasser. Four models were preliminary generated by **I**-tasser. Redundancy of alpha helix also observed in 3D structures of these models. From these models of histamine H3 receptor one model will be selected as the best model after analyzing with model validation tools, i.e., ERRAT, RAMPAGE and PROCHECK.

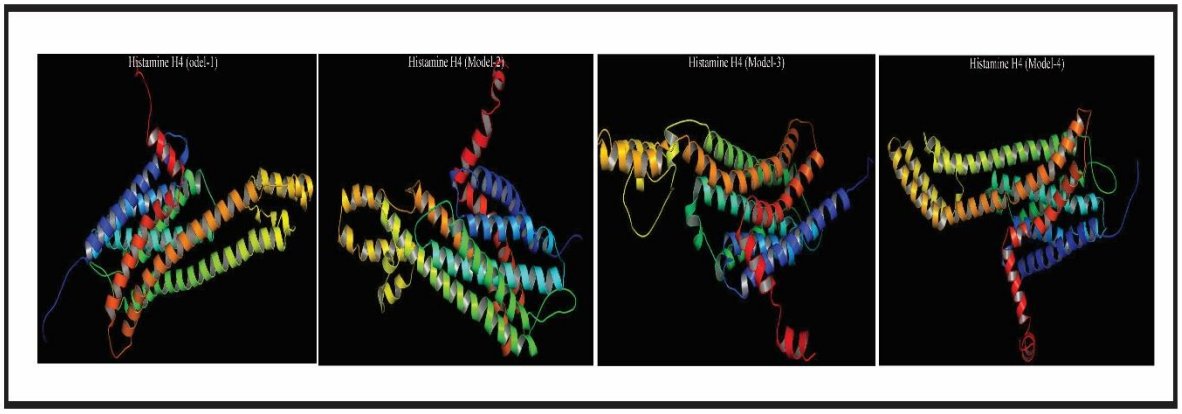

Figure 4.6: Homology modeling of histamine H4 receptor by **I**-tasser. Four models were generated by **I**-tasser. Redundancy of alpha helix also observed in 3D structures of these models. Some coiled coil regions also observed in histamine H4 receptor primitive structure. From these models of histamine H4 receptor one model will be selected as the best model after analyzing with model validation tools, i.e., ERRAT, RAMPAGE and PROCHECK.

These models were then validated by using ERRAT, Rampage and Ramachandran plot. Validation scores attained from ERRAT, Rampage and Ramachandran plot are showed in Table 4.3. Data from Table 4.3 suggests that, model 2 of histamine H1 receptor, model 4 of histamine H2 receptor, model 1 of histamine H3 receptor and model 2 of histamine H4 receptor were the best models according to model validation scores.

Table 4.3: Model Validation Scores by Different Tools e.g., ERRAT, RAMPAGE & PROCHEK

Receptor	Model No.	ERRAT* %	RAMPAGE** %	RAMACHA-NDRAN PLOT*** %
Histamine H1 Receptor (HRH1)	Model 1	86.221	75.7	69.9
	Model 2	93.946	83.5	77.2
	Model 3	88.703	80.4	74.7
	Model 4	87.891	81.6	75.2
	Model 5	90.337	75.1	69.9
Histamine H2 Receptor (HRH2)	Model 1	86.895	81.5	76.7
	Model 2	92.308	79.8	76.1
	Model 3	92.000	83.8	77.6
	Model 4	92.877	83.2	77.9
	Model 5	94.017	81.2	76.4
Histamine H3 Receptor (HRH3)	Model 1	90.618	82.6	82.5
	Model 2	85.981	83.5	80.1
	Model 3	82.710	76.7	75.3
	Model 4	90.337	86.5	78.0
Histamine H4 receptor (HRH4)	Model 1	86.649	80.7	79.5
	Model 2	92.147	87.4	83.7
	Model 3	90.814	81.7	77.3
	Model 4	93.651	85.8	81.7

*****Good high-resolution structures generally produce values around 95% or higher. Low resolution structures produce values around 91%

****** Amino acid residues in most favoured region. Expected or standard of good quality model around 98% or higher.

******* Residues in most favoured regions. a good quality model would be expected to have over 90% in the most favoured regions.

ERAAT validates models by statistical relation of non-bonded interactions among different atom types based on characteristic atomic interaction [100]. It assesses overall quality of a model at 0.01 and 0.05 level of significance and presents result as overall quality factor. Standard high-resolution structures generally produce values around 95% or higher. Low resolution structures produce values around 91%. Table 4.3 illustrates ERRAT score of our predicted models, which ranges from 82.710% to 94.017%. This range suggests significance of our predicted models according to the algorithm of ERRAT software. Best model from each histamine receptor group also scored above 90% (Figure 4.7).

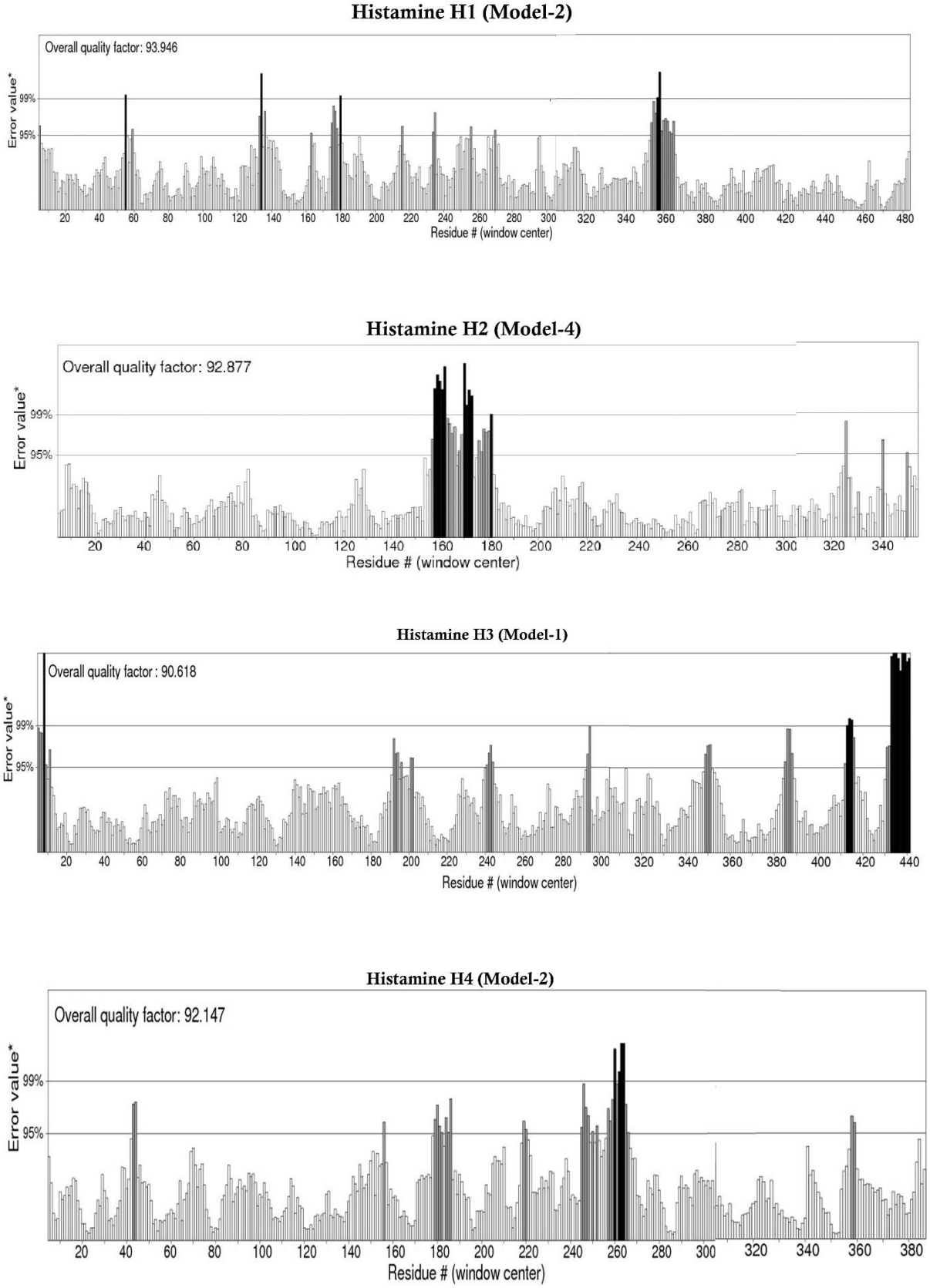

Figure 4.7: ERRAT score of best models for Histamine H1, H2, H3 and H4 receptor.

RAMPAGE is another 3D model validation tool, which presents result based on amino acids geometry and deviation. It provides result in three main categories, such as number of residues in favored region (expected value ~98.0%), number of residues in allowed region (~2.0% expected) and number of residues in outlier region. Table 4.3 illustrates RAMPAGE score in percentage, which ranges from 75.1% to 86.5%. The best models form each histamine group that was selected on the basis of RAMPAGE scored more than 82% (Figure 4.8). This range suggests quality models were predicted by using I-Tasser.

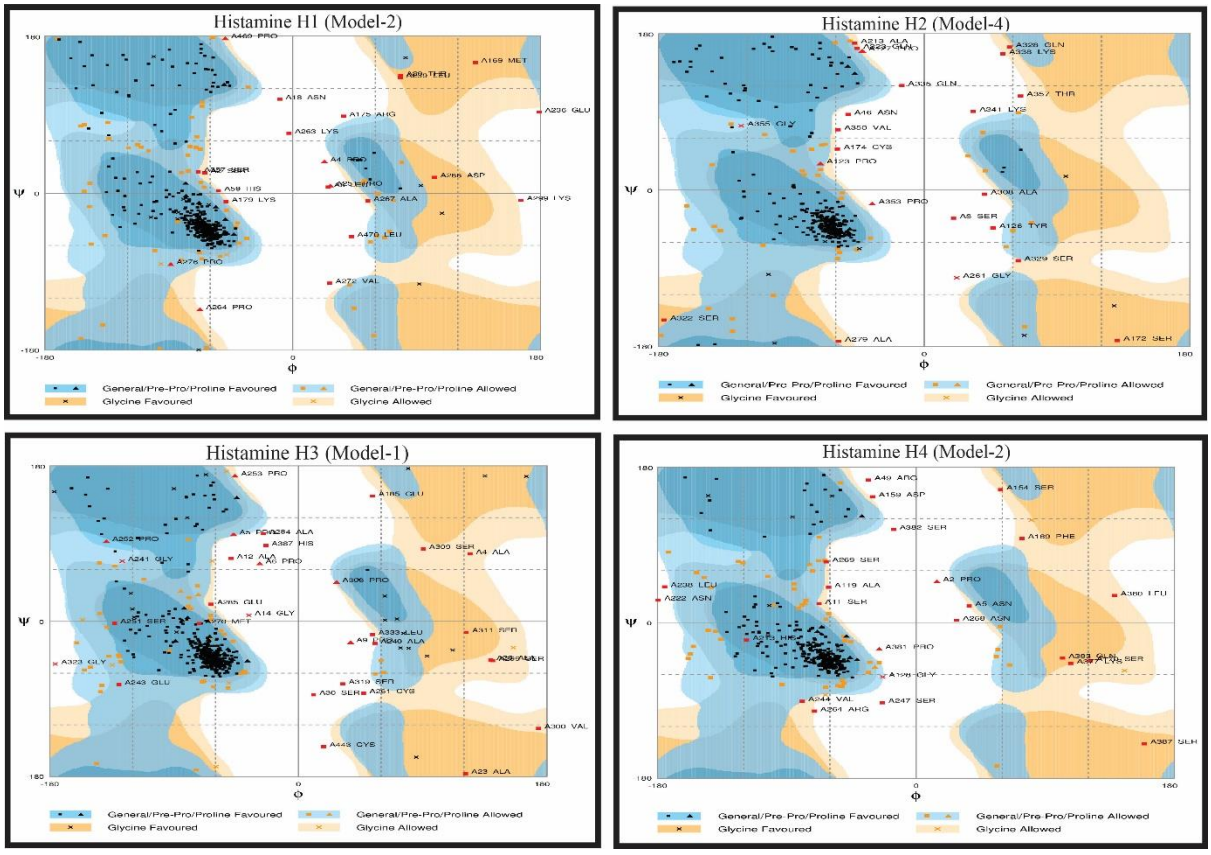

Figure 4.8: RAMPAGE output of best models for Histamine H1, H2, H3 and H4 receptor.

PROCHECK tests stereochemical quality of protein structure by evaluating residue-by-residue geometry and overall structural geometry. It provides amino acid residues distribution on Ramachandran plot divided into four colour coated regions. The regions are Residues in most favored regions, Residues in additional allowed regions, residues in generously allowed regions and residues in disallowed regions. According to PROCHECK standard, a good quality model should posses over 90% amino acid resided in the most favored regions. Table 4.3 suggests that, our predicted models were ranged from

69.9% to 83.7%. The best models form each histamine group suggest that more than 77% amino acid residues were in most favored region for these models (Figure 4.9).

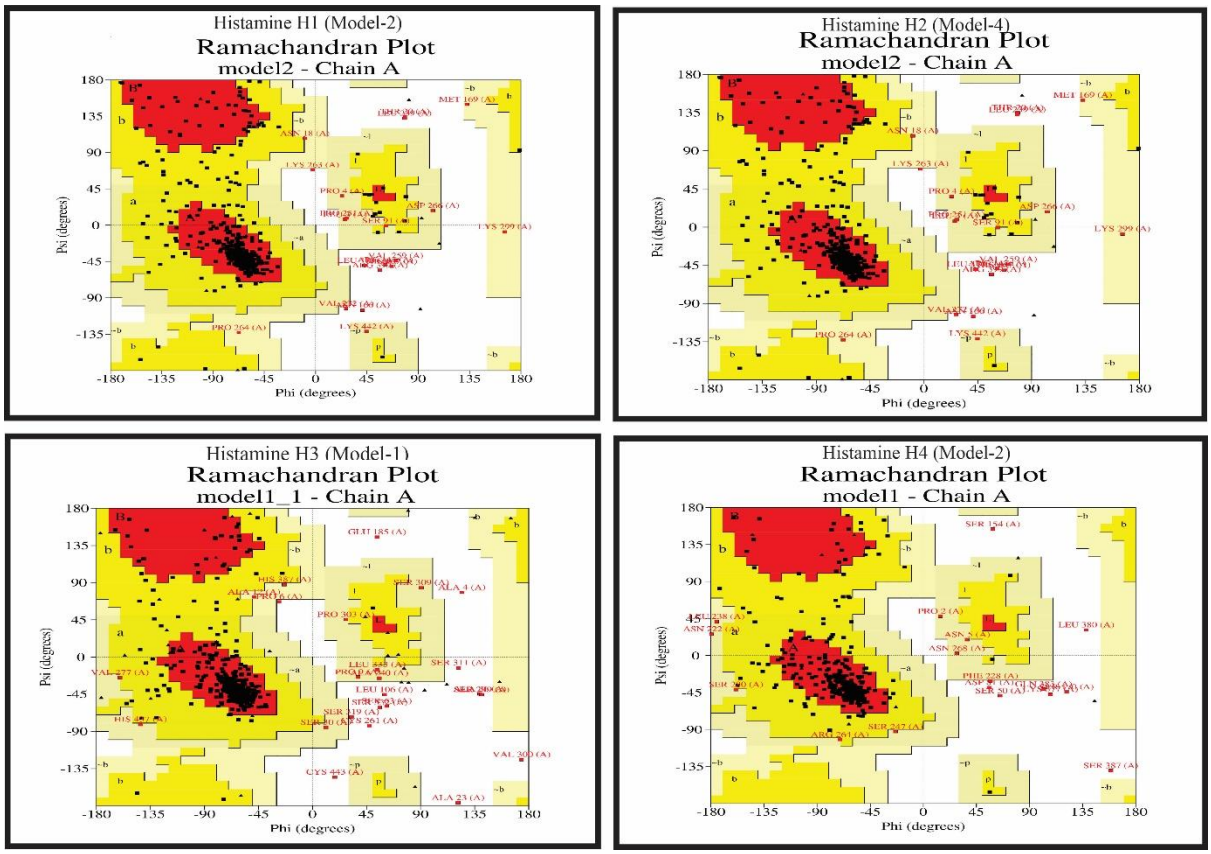

Figure 4.9: PROCHECK analysis result for best models of Histamine H1, H2, H3 and H4 receptor. Color codes are Red color- most favorable regions, yellow color region- allowed region, and pale yellow generously allowed region and white color- disallowed regions.

Analyzing the results of different validation tools, we selected top model for each histamine receptor. From the Table-3, we found that model 2 of histamine H1 receptor, model 4 of histamine H2 receptor, model 1 of histamine H3 receptor and model 2 of histamine H4 receptor are outperformed other models according to the validation scores attained from different tools. Figure 4.10 represents best model for different histamine receptor.

Model 2 of histamine H1 receptor posses ERRAT score of 93.946%, RAMPAGE score (residues in favoured region) 83.5% and RAMACHANDRAN PLOT score (% of Residues in most favoured regions) 77.2%. Similarly, Model 4 of histamine H2 receptor carries ERRAT score of 92.8776%, RAMPAGE score (residues in favoured region) 83.2% and RAMACHANDRAN PLOT score (% of Residues in most

favoured regions) 77.9%. Model 1 of histamine H3 receptor carries ERRAT score of 90.618%, RAMPAGE score (residues in favoured region) 82.6% and RAMACHANDRAN PLOT score (% of Residues in most favoured regions) 82.5%. Lastly, model 2 of histamine H4 receptor possess ERRAT score of 92.147%, RAMPAGE score (residues in favoured region) 87.4% and RAMACHANDRAN PLOT score (% of Residues in most favoured regions) 83.7%. Homology modeling of these receptors has shown heavy load of alpha helix in each receptor's 3D structure. Validation of 3D structure of these receptors also performed to check the quality of our predicted models as well as to select the best model of each histamine receptor.

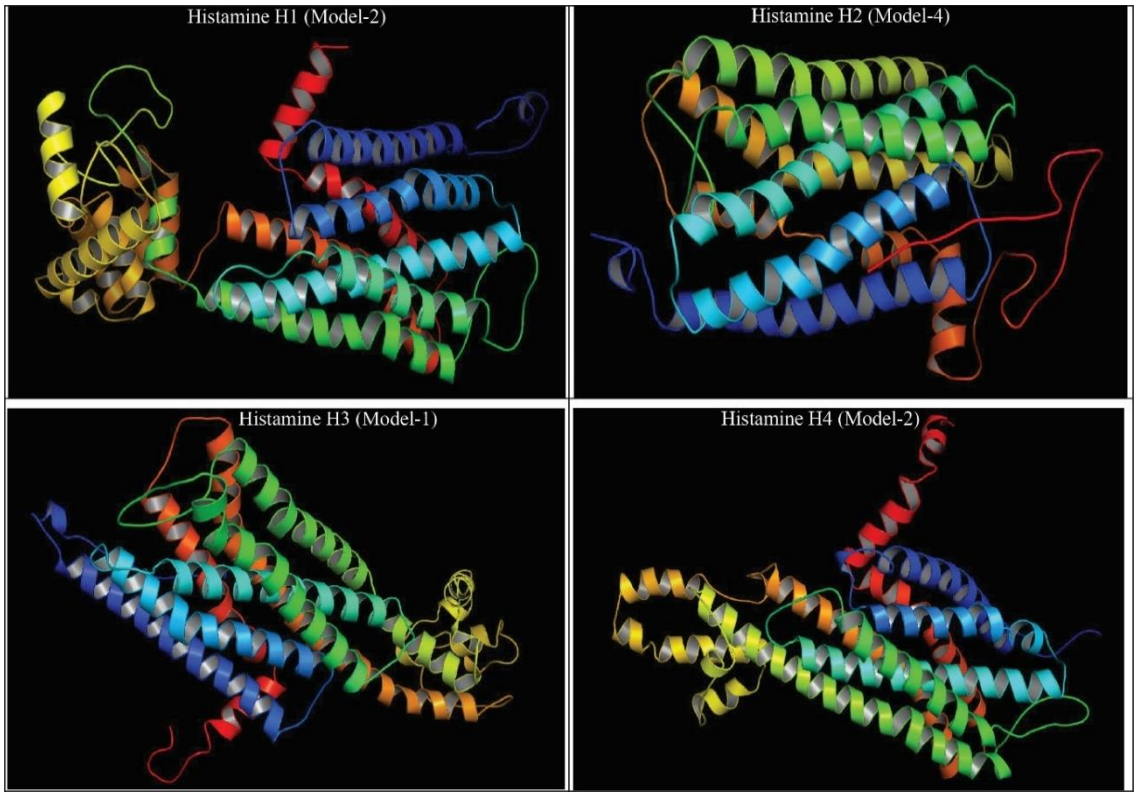

Figure 4.10: Confirmation of best model for Histamine H1, H2, H3 and H4 receptor by different model validation tools such as ERRAT, RAMPAGE & PROCHECK. For histamine H1 receptor the best model was model-2, in case of histamine H2 receptor the best model was model model-4 and for histamine H3 and H4 receptor the best model was model-1 and model-2, respectively.

4.3 Analysis of Vaccine prediction for Api m3 Allergen

4.3.1 Physicochemical Properties of Api m3

Analysis of physicochemical properties of Api m3 protein was conducted by using ExPASy protparam tool. Api m3 consisted of 373 amino acids in its protein sequence and it poses

molecular weight around 44Kda. From the total number of positively & negatively charged residues, it can be concluded that Api m3 is hydrophobic in nature. The instability index and aliphatic index was found 53.69 & 92.76, respectively. Instability index value greater than 40 is regarded as the instability of the protein.

Table 4.4: Analysis of Different Physicochemical Properties of Api m3 Protein

Number of amino acids (Nos.)	373
Molecular weight (Da)	43905.19
Theoretical pI	5.63
Total number of negatively charged residues (Asp + Glu) (Nos.)	51
Total number of positively charged residues (Arg + Lys) (Nos.)	43
Instability index	53.69
Aliphatic index	92.76
Grand average of hydropathicity (GRAVY)	-0.430

So, Api m3 is fairly instable in nature and aliphatic index value suggests relative thermostability of Api m3 protein. It was also found that, the GRAVY value of protein was -0.43. The parameters calculated by ExPASy protparam tool is presented in Table 4.4.

4.3.2 Determination of Secondary Structure

Secondary structural features were calculated by SOPMA tool. Secondary structure of Api m3 dominated by alpha helix (43.97%) and random coil (30.83%). Minority part of the secondary structure poised by extended strand (14.75%) and Beta turn (10.46%).

Table 4.5: Determination of Secondary Structure Data of Api m3

Features of Secondary	Percentage
Alpha helix (Hh)	43.97%
310 helix (Gg)	0.00%
Pi helix (Ii)	0.00%
Beta bridge (Bb)	0.00%
Extended strand (Ee)	14.75%
Beta turn (Tt)	10.46%
Bend region (Ss)	0.00%
Random coil (Cc)	30.83%

The features calculated by SOPMA tool is presented in Table 4.5. The secondary structure plot is shown in Figure 4.11.

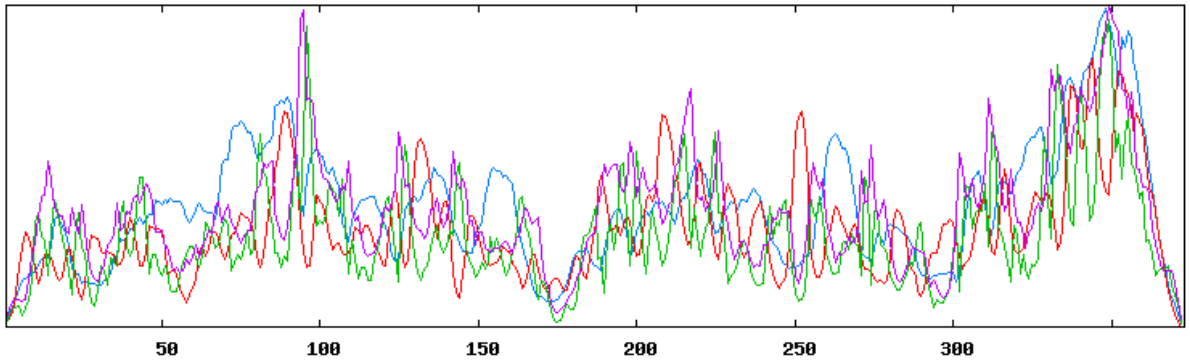

Figure 4.11: Secondary structure plot of api m3 of *Apis mellifera*. Here, helix is indicated by blue, random coils are shown in purple, while extended strands and beta turns are indicated by red and green, respectively.

4.3.3 Prediction of T-cell Epitope

T-cell epitopes were predicted by implying NetCTL 1.2 server which provides combined score based on each MHC class 1 supertypes, proteasomal C terminal cleavage and TAP transport efficiency. The result generated by NetCTL 1.2 server for query sequence Api m3, we selected five epitopes based on the highest combinational score. Selected five epitopes were ITTPWDYYY, SVSALSSFY, FYDRTKMSL, YTEESVSAL and AEQSYGLTL, with the combinational score of 2.6286, 2.5280, 2.3858, 2.2899 and 2.0585, respectively. (Data not shown) Pursuant of IEDB MHC class1 binding prediction tool the previously selected epitopes were found to be acquainted by wide range of MHC class 1 molecule. The tool is principled on Stabilized Matrix Method (SMM) and provides HLA binding affinity of selected epitopes based on IC50 value as an output. There is an inverse relationship between MHC-I molecules. In this study, we set threshold value of IC50 less than 200 nM for the section of MHC-I molecules. With this criterion the selection of MHC-I molecules was ensured with higher affinity to bind with selected epitopes.

IEDB offers can excellent MHC-I processing efficiency tool, which predicts combined score for each of the previously selected epitopes based on three factors, such as proteasomal cleavage efficiency, TAP transport efficiency and MHC-I binding efficiency, accumulatively. Proteasomal cleavage efficiency stands for enzymatic digestion of a protein to form fragmented peptides. These peptide fragments are recognized by MHC-I molecules and simultaneously, forms MHC-Peptide complex. The complexes are then transported to the endoplasmic reticulum. This process rigorously influenced by Transport Associated Proteins (TAP) and it happens before being faced to T-cells on the plasma membrane of the cell.

Table 4.6: Five Most Potential T-cell Epitope Candidates with Interacting MHC-1 Alleles, Total Processing Score and Epitope Conservancy Result.

Epitope	Interacting MHC-1 allele with an affinity, 200 (total score of proteasome score, TAP score, MHC score, processing score and MHC-1 binding)	Epitope conservancy analysis result
ITTPWDYYY	HLA-A*01:01 (1.12)	100%
	HLA-A*29:02 (1.78)	
	HLA-A*30:02 (2.52)	
	HLA-A*68:01 (0.44)	
	HLA-B*35:01 (0.33)	
	HLA-B*57:01 (0.41)	
	HLA-B*58:01 (1.16)	
	HLA-C*16:01 (0.41)	
	HLA-C*03:02 (0.34)	
SVSALSSFY	HLA-A*01:01 (0.23)	100%
	HLA-A*03:01 (0.29)	
	HLA-A*11:01 (0.81)	
	HLA-A*26:01 (0.23)	
	HLA-A*29:02 (1.40)	
	HLA-B*15:01 (0.42)	
	HLA-B*15:02 (0.43)	
	HLA-A*30:02 (1.16)	
	HLA-B*35:01 (0.50)	
	HLA-B*15:25 (0.68)	
	HLA-A*68:01 (0.60)	
	HLA-C*12:03 (0.10)	
FYDRTKMSL	HLA-C*14:02 (0.90)	100%
	HLA-C*07:02(-0.16)	
YTEESVSAL	HLA-A*02:06 (-0.47)	100%
	HLA-B*39:01 (-0.59)	
	HLA-C*02:02 (-0.46)	
	HLA-C*02:09 (-0.46)	
	HLA-C*03:03 (0.57)	
	HLA-C*03:02 (0.20)	
	HLA-C*03:04 (0.57)	
	HLA-C*05:01 (0.01)	
	HLA-C*08:01 (-0.01)	
	HLA-C*08:02 (-0.09)	
	HLA-C*12:02 (-0.33)	
	HLA-C*12:03 (0.24)	
	HLA-C*14:02 (-0.19)	
	HLA-C*15:02 (-0.49)	
	HLA-C*16:01 (0.10)	
AEQSYGLTL	HLA-B*13:01 (-0.27)	100%
	HLA-B*37:01 (-0.40)	
	HLA-B*40:01 (0.96)	
	HLA-B*40:02 (0.85)	

The higher combined score indicates more possibilities of accurate presentation of peptides and provoking a desired immune response. The combinational score attained from IEDB MHC-1 binding analysis and processing tools are presented in Table 4.6.

T-cell mediated immune response solely depends on successful recognition of designed peptides by HLA molecules with higher affinity. From that point, it can be concluded that the peptide recognized by most HLA alleles carry the high potentiality to induce a

significant immune response. From selected epitopes, one epitope out smarted others by the mean of HLA allele interactions. The 9-mer peptide epitope YTEESVSAL showed interactive affinity for 15 MHC class I molecules, which includes HLA-A*02:06, HLA-B*39:01, HLA-C*02:02, HLA-C*02:09, HLA-C*03:03, HLA-C*03:02, HLA-C*03:04, HLA- C*05:01, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02 and HLA-C*16:01 (Table 4.6).

Also, MHC class II T-cell epitopes were found based on HLA-DQA1*01:01/ DQB1*05:01, HLA- DRB1*11:01, HLA-DRB1*03:01 and HLADRB1*15:01 restriction are presented in Table 5. 4. YPKDPYLYYDFYPLE and GGPLLRIFTKHMLDV were found as prominent MHC-II type T-cell epitopes of Api m3 protein (Table 4.7).

Table 4.7: Predicted Two MHC Class II T-cell Epitopes with Their Position, Length & Conservancy Score

Epitope	Interacting MHC-II	Position	Number of Amino	Epitope conservancy
YPKDPYLYYDFYPLE	HLA-DQB1*05:01	23-37	15	100%
GGPLLRIFTKHMLDV	HLA-DRB1*11:01	226-240	15	100%

Systematic selection of T-cell epitopes can be very useful in stimulation of clinical and immunological resistance to allergen in a refined treatment strategy. From that point our predicted epitopes critically evaluated as per current theorems of potential allergen T-cell epitope [127]. Researchers found effective T- cell epitopes these can effectively trigger immunogenic Th2 response in different food and insect allergen [128] – [131].

4.3.4 Prediction of Epitope Conservancy

In present study, five 9 mer T-cell epitopes for MHC class I molecules and two 15 mer T-cell epitopes for MHC class II molecules were predicted. Conservancy of these epitopes is expected as the more conserved epitopes provides better immunization. All seven of our predicted epitopes including five MHC class I and two MHC class II peptides showed 100% conservancy, data is shown in Table 4.6 & Table 4.7, respectively.

4.3.5 Population Coverage Prediction

Suitable epitopes should cover wide range of ethnic population all over the world. For the purpose, we checked our predicted epitopes population coverage through IEDB's population coverage analysis tool. From our predicted MHC class I epitopes, we found that MHC-

I epitope ITTPWDYYY, SVSALSSFY, FYDRTKMSL, YTEESVSAL and AEQSYGLTL covered 9.84%, 9.84%, 3.42%, 62.71% and 4.66% population of the world on an average, respectively (Table 4.8). YTEESVSAL

Table 4.8: Population Coverage of Five Predicted MHC class I T-Cell Epitopes.

Area	ITTPWDYYY			SVSALSSFY			FYDRTKMSL			YTEESVSAL			AEQSYGLTL		
	Class I			Class I			Class I			Class I			Class I		
	coverage[a]	Average _hit[b]	pc 90[c]	coverage[a]	Average _hit[b]	pc 90[c]	coverage[a]	average_ hit[b]	pc 90[c]	coverage[a]	Average _hit[b]	pc 90[c]	coverage[a]	average_ hit[b]	pc 90[c]
Central Africa	2.84%	0.03	0.1	2.84%	0.03	0.1	3.21%	0.03	0.1	56.62%	0.66	0.23	0.53%	0.01	0.1
East Africa	11.24%	0.11	0.11	11.24%	0.11	0.11	1.07%	0.01	0.1	49.22%	0.57	0.2	2.76%	0.03	0.1
East Asia	2.55%	0.03	0.1	2.55%	0.03	0.1	11.28%	0.11	0.11	81.27%	1.2	0.53	0.06%	0	0.1
Europe	25.67%	0.26	0.13	25.67%	0.26	0.13	2.01%	0.02	0.1	69.56%	0.89	0.33	0.18%	0	0.1
North Africa	14.18%	0.14	0.12	14.18%	0.14	0.12	2.48%	0.02	0.1	59.39%	0.71	0.25	1.34%	0.01	0.1
North America	12.72%	0.13	0.11	12.72%	0.13	0.11	3.55%	0.04	0.1	65.58%	0.83	0.29	9.59%	0.1	0.11
Northeast Asia	3.44%	0.03	0.1	3.44%	0.03	0.1	6.91%	0.07	0.11	71.42%	0.94	0.35	21.57%	0.22	0.13
Oceania	11.38%	0.11	0.11	11.38%	0.11	0.11	0.67%	0.01	0.1	57.64%	0.7	0.24	0.05%	0.03	-59
South America	6.03%	0.06	0.11	6.03%	0.06	0.11	1.93%	0.02	0.1	53.75%	0.63	0.22	0.00%	0	0.1
South Asia	13.80%	0.14	0.12	13.80%	0.14	0.12	4.23%	0.04	0.1	59.80%	0.75	0.25	2.60%	0.03	0.1
Southeast Asia	2.10%	0.02	0.1	2.10%	0.02	0.1	4.34%	0.04	0.1	73.65%	1	0.38	10.93%	0.11	0.11
Southwest Asia	14.66%	0.15	0.12	14.66%	0.15	0.12	2.75%	0.03	0.1	47.69%	0.55	0.19	1.65%	0.02	0.1
West Africa	7.31%	0.07	0.11	7.31%	0.07	0.11	0.00%	0	0.1	69.66%	0.86	0.33	0.00%	0	0.1
Average	**9.84**	**0.1**	**0.11**	**9.84**	**0.1**	**0.11**	**3.42**	**0.03**	**0.1**	**62.71**	**0.79**	**0.29**	**3.94**	**0.04**	**-4.4**
Standard deviation	**6.47**	**0.07**	**0.01**	**6.47**	**0.07**	**0.01**	**2.85**	**0.03**	**0**	**9.66**	**0.18**	**0.09**	**6.16**	**0.06**	**15.6**

covered around 62.71% population of the world on an average. This epitope showed 81.27% coverage in East Asia, 71.42% coverage in Northeast Asia, 73.65% coverage in Southeast Asia, 69.66% coverage in West Africa, 59.39% coverage in North Africa, 69.56% coverage in Europe and 65.58% coverage in North America. In case of MHC class II epitopes, we found much lower population coverage comparing to the MHC class I epitope population coverage. YPKDPYLYYDFYPLE and GGPLLRIFTKHMLDV showed average 17.5% and 18.85% world population coverage.

4.3.6 Designing of 3D Structure of T-cell Epitopes

3D structures of T-cell epitopes were predicted for better understanding the molecular actions of these peptides. Designed models are illustrated in Figure 4.12

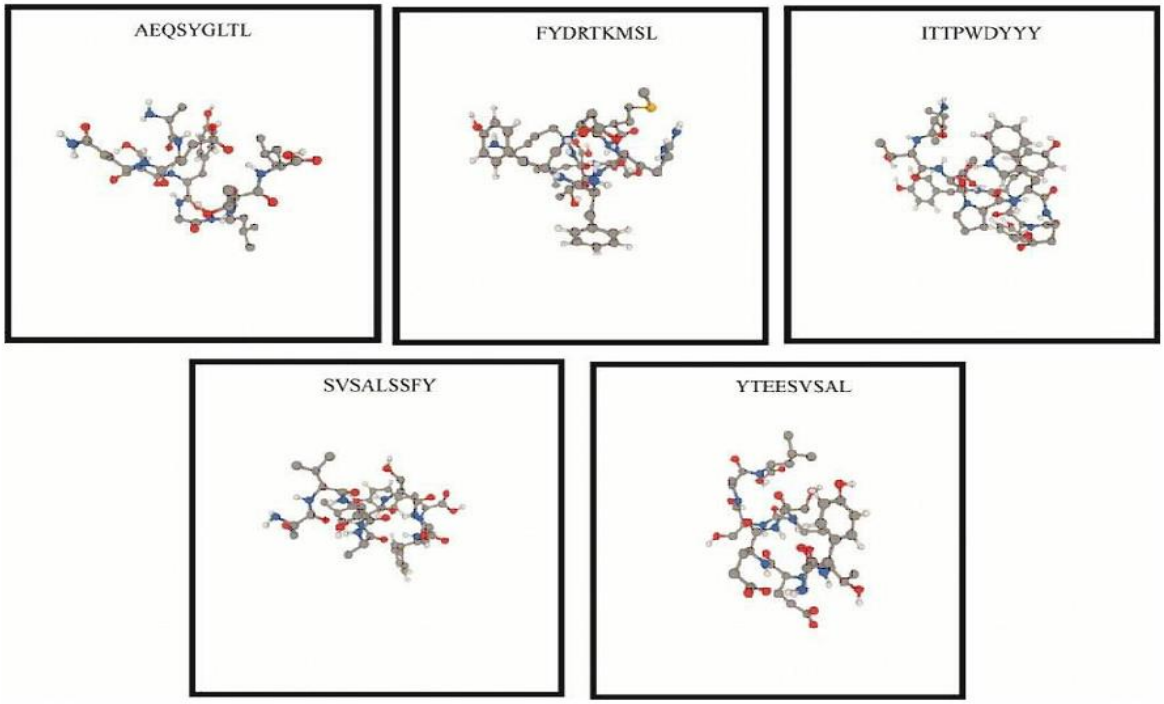

Figure 4.12: Ball & Stick structure of five MHC class I T-cell epitopes.

4.3.7 Prediction of B-cell Epitopes

Primarily, linear B-cell epitopes were predicted by using BepiPred, Bcepred, ABCpred and BCPREDS are shown in Table 4.9 - Table 4.12.

Table 4.9: Prediction of B-cell Epitopes of Api m3 of *Apis mellifera* by Using BepiPred.

No	Start	End	Peptide	Length
1	12	30	GDRIPDEKNEMYPKDPYLY	19
2	35	50	PLERGELTNSGKMREY	16
3	61	63	GDF	3
4	66	75	DIYTEESVSA	10
5	78	82	SFYDR	5
6	95	114	YPPNKLQQWNEDLNWQPIAT	20
7	117	119	LRR	3
8	124	125	IF	2
9	140	173	VLESPRGKYEFSKYDKLKKKLEEWTGKNITTPWD	34
10	183	184	AE	2
11	186	200	SYGLTLPSWTNNIFP	15
12	202	205	GELF	4
13	213	220	NITNSTPL	8
14	241	248	VSGTQKKK	8
15	256	261	GHESNI	6
16	273	280	PHVPEYSS	8
17	289	292	IEGT	4
18	303	312	IPSEARELQL	10
19	331	339	VIPSNEELI	9
20	344	354	FVDESANNLSI	11

Table 4.10: Prediction of B-cell Epitopes of Api m3 of *Apis mellifera* by Using Bcepred

No	Start Position	Sequence	End Position
1	24	PKDPYLYYDFYP	35
2	69	TEESVSALSS	78
3	85	MSLQLVLAALYPPN	98
4	111	PIATKYLR	118
5	122	DNIFLPEDCLLFT	134
6	136	ELDRVLE	142
7	172	WDYYYIYHTLVAEQSYGLTL	191
8	205	FDATVFT	211
9	216	NSTPLLKKLYGGPLLRI	232
10	236	HMLDVVSG	243
11	249	RKIYLFS	255
12	260	NIASVLHALQLYYPHVPEYSSSIIME	285
13	291	GTHYVKIVYYLGI	303
14	308	RELQLPGCEVLCPLYKYLQLIENVIP	333
15	336	EELICDK	342
16	353	SIEELDFVKLNLIR	366

Table 4.11: Prediction of B-cell Epitopes of Api m3 of *Apis mellifera* by Using ABCPred

Rank	Sequence	Start position	Score
1	GDRIPDEKNE	12	0.86
2	NEMYPKDPYL	20	0.79
3	TPLLKKLYGG	218	0.76
3	NITTPWDYYY	167	0.76
4	LNWQPIATKY	107	0.75
5	ALSSFYDRTK	75	0.74
5	PLLRIFTKHM	228	0.74
5	DYYYIYHTLV	173	0.74
6	KYDKLKKKLE	152	0.72
7	KLQQWNEDLN	99	0.71
7	IYHTLVAEQS	177	0.71
7	PEDCLLFTIE	127	0.71
8	VIFRHGDRIP	7	0.7
8	LFTIELDRVL	132	0.7
8	YEDNIFLPED	120	0.7
9	RERYGDFLGD	57	0.69
9	EWTGKNITTP	162	0.69
10	KDPYLYYDFY	25	0.68
10	ITNSTPLLKK	214	0.68
10	LFDATVFTYN	204	0.68
10	YGLTLPSWTN	187	0.68

10	AEQSYGLTLP	183	0.68
11	YYDFYPLERG	30	0.67
11	VFTYNITNST	209	0.67
12	KQINVIFRHG	3	0.66
13	LQLVLAALYP	87	0.64
13	YPLERGELTN	34	0.64
13	NIFPRGELFD	197	0.64
14	GKMREYQLGQ	45	0.62
14	IATKYLRRYE	112	0.62
15	KKLYGGPLLR	222	0.61
15	RVLESPRGKY	139	0.61
16	DIYTEESVSA	66	0.6
17	RTKMSLQLVL	82	0.59
18	LYPPNKLQQW	94	0.57
19	KYEFSKYDKL	147	0.54
20	EYQLGQFLRE	49	0.52

Table 4.12: Prediction of B-cell Epitopes of Api m3 of *Apis mellifera* by Using BCPREDS

Position	Epitope	Score
10	RHGDRIPDEKNE	0.975
270	LYYPHVPEYSSS	0.943
210	FTYNITNSTPLL	0.927
66	DIYTEESVSALS	0.863
144	PRGKYEFSKYDK	0.862
96	PPNKLQQWNEDL	0.827
53	GQFLRERYGDFL	0.765
163	WTGKNITTPWDY	0.749
195	TNNIFPRGELFD	0.691
334	SNEELICDKRFV	0.665
246	KKKRKIYLFSGH	0.621
313	PGCEVLCPLYKY	0.616
27	PYLYYDFYPLER	0.363

4.3.8 Validation of Predicted B-cell Epitopes

Validation of B-Cell epitopes were preformed based on the data from Table 4.9 to Table 4.12. Potential B-cell epitopes poise certain features which are essential for effectual recognition by B-cells [132]. These characteristics comprise hydrophilicity, antigenicity, surface accessibility, flexibility and beta turn prediction. Api m3 protein sequence was analyzed through IEDB tools in order to select best epitope from Table 4.9 - Table 4.12.

Surface accessibility is an essential feature of B-cell epitopes. Determination of surface accessibility is prerequisite to select potent B cell epitope because it indicates most exposed area on the surface of a protein and likely to trigger B-cell mediated immune response. Also,

hydrophilic regions of a protein are also generally exposed on the surface. The Emini surface accessibility prediction tool and Parker hydrophilicity prediction tools were used to attain data on surface accessibility and hydrophilicity of api m3 protein. In Table 4.13 data of Emini surface accessibility prediction is presented and Figure 4.13 shows hydrophilicity of api m3 protein.

Table 4.13: Emini Surface Accessibility Prediction

No.	Start	End	Peptide	Length
1	14	29	RIPDEKNEMYPKDPYL	16
2	44	51	SGKMREYQ	8
3	79	85	FYDRTKM	7
4	96	108	PPNKLQQWNEDLN	13
5	115	121	KYLRRYE	7
6	143	161	SPRGKYEFSKYDKLKKKLE	19
7	244	250	TQKKKRK	7
8	273	278	PHVPEY	6

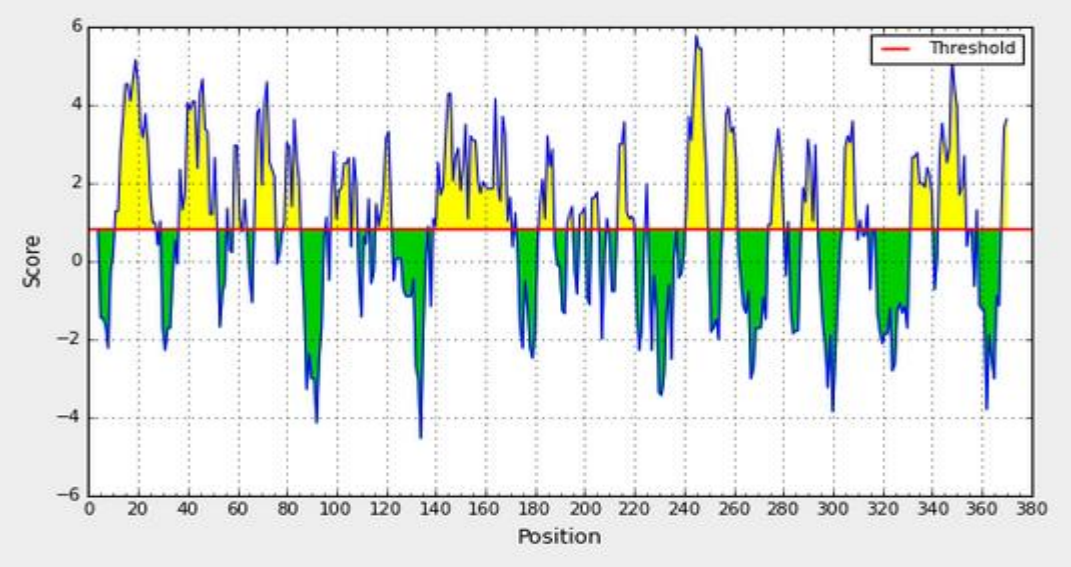

Figure 4.13: Parker hydrophilicity prediction of api m3 protein of *Apis mellifera*. Here, the x-axis and y-axis represents the position and score, respectively. The threshold value is 0.822. In the Graph, hydrophilic regions in the protein are shown in yellow color, and are above the threshold value.

Kolaskar and Tongaonkar antigenicity prediction tool evaluated assessed antigenic propensity value of the protein for B cell epitopes. The Kolaskarand and Tongaonkar antigenicity prediction tool subsequently provided the result, shown in Table 4.14. The average antigenic propensity value of api m3 was calculated 1.034, whereas the maximum value was 1.232 and minimum value of 0.869. We found, potent antigenic region was from 261-281.

Table 4.14: Kolaskar and Tongaonkar Antigenicity Prediction

No.	Start	End	Peptide	Length
1	4	10	QINVIFR	7
2	26	34	DPYLYYDFY	9
3	73	79	VSALSSF	7
4	87	98	LQLVLAALYPPN	12
5	127	135	PEDCLLFTI	9
6	137	143	LDRVLES	7
7	174	185	YYYIYHTLVAEQ	12
8	206	212	DATVFTY	7
9	226	233	GGPLLRIF	8
10	237	243	MLDVVSG	7
11	261	281	IASVLHALQLYYPHVPEYSSS	21
12	294	304	YVKIVYYLGIP	11
13	312	331	LPGCEVLCPLYKYLQLIENV	20
14	359	365	FVKLNLI	7

Another feature of B-cell epitope is the beta turns. Theoretically, beta turn regions of a protein are mostly hydrophilic and functionally surface accessible. We performed, Chou and Fasman Beta turn prediction to find out potential antigenic region of a protein, as these two factors are positively correlated. We found regions from 13-37 & 273-280 were potential beta turn regions (Figure 4.14) and 14-29 & 273-278 was most accessible region on the surface (Table 4.13). These regions were the best candidates to from B-cell epitopes.

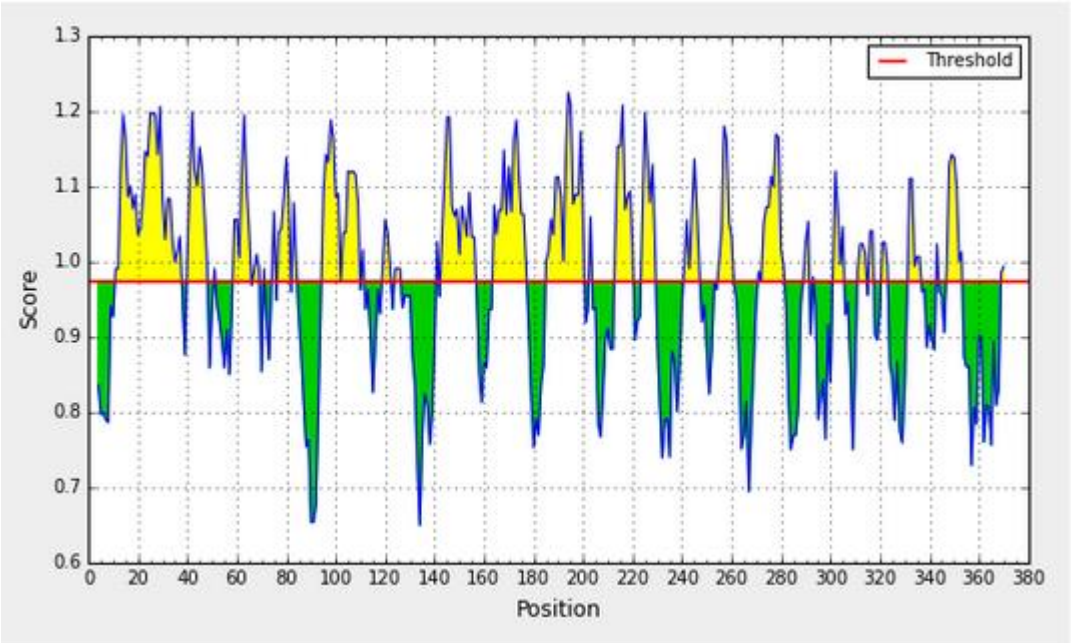

Figure 4.14: Chou and Fasman Beta turn prediction of api m3 protein of *Apis mellifera*. Here, the x-axis and y- axis represents the position and score, respectively. The threshold value is 0.975. In the Graph, regions having beta turns in the protein are shown in yellow color, and are above the threshold value.

Flexible region of a protein aids in successful binding of an antibody to an epitope. Therefore, we predicted flexibility of api m3 by Karplus Schulz flexibility prediction tool. (Figure 4.15)

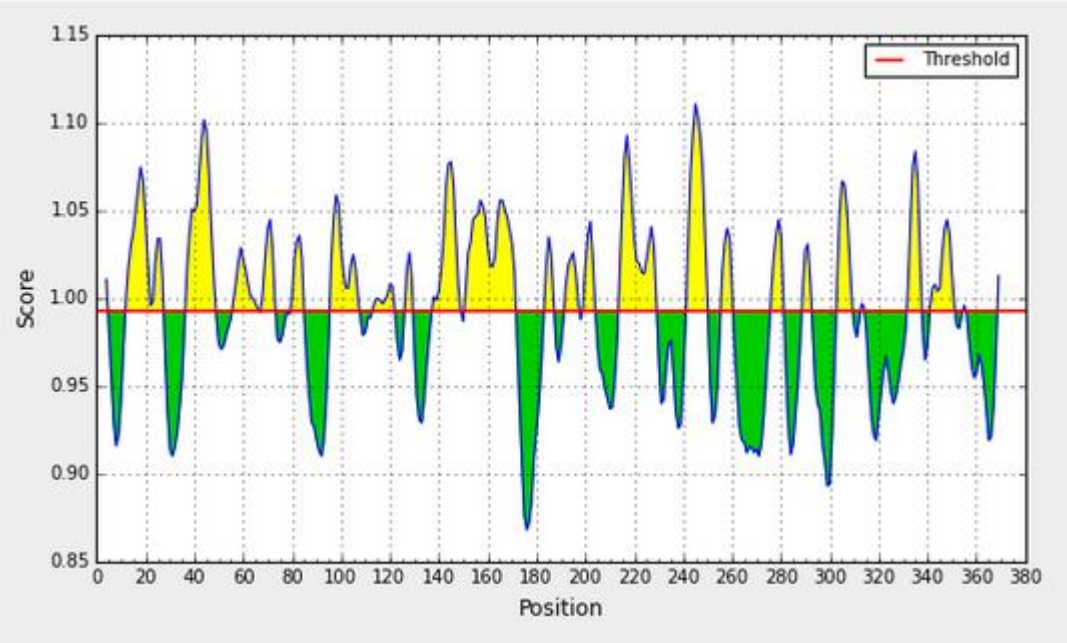

Figure 4.15: Karplus Schulz flexibility prediction of api m3 protein of *Apis mellifera*. Here, the x-axis and y-axis represents the position and score, respectively. The threshold value is 0.992. In the Graph, regions having greater flexibilities in the protein are shown in yellow color, and are above the threshold value.

Considering all the data available from different B-cell prediction tools, we concluded that, the regions from 12-20 (GDRIPDEKN) & 273-281(PHVPEYSSS), two 9 mer peptides could be the most effective B- cell epitopes.

4.3.9 Homology Modeling of Api m3

Homology modeling of Api m3 was performed by SWISS MODEL. The best model selected from SWISS MODEL (Figure 4.16) assessed for general errors by checking the model quality through ERRAT, RAMPAGE & PROCHEK.

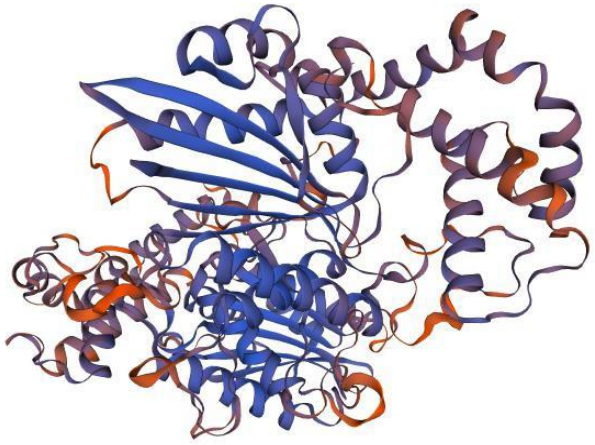

Figure 4.16: 3D model of api m3 protein of *Apis mellifera*.

ERRAT validates models by statistical relation of non-bonded interactions among different atom types based on characteristic atomic interaction [18]. ERRAT evaluates overall quality of a model at 0.01% and 0.05% level of significance and shows result as overall quality factor. Standard high resolution structures generally produces overall quality factor around 95% or higher. Low resolution structures produce overall quality factor around 91%. Our selected model of Api m3 scored 81.505 as overall quality factor (Figure 4.17).

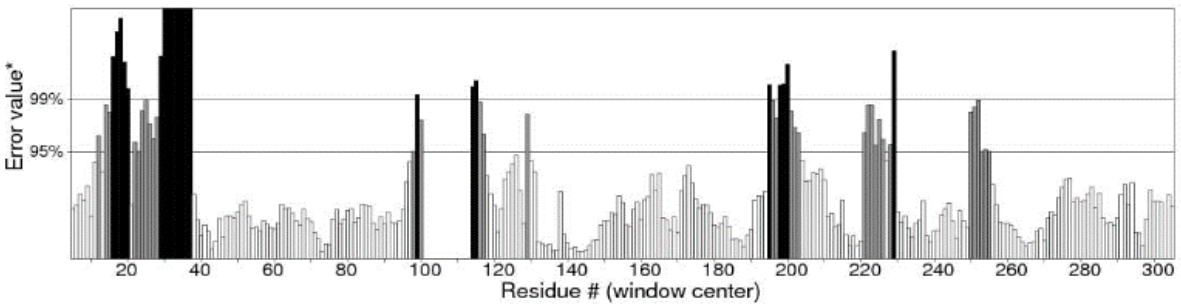

Figure 4.17: ERRAT generated result of api m3 protein of *Apis mellifera*, where 95% indicates rejection limit.

RAMPAGE is another 3D model validation tool, which presents result based on amino acids geometry and deviation. It provides result in three major categories, such as number of residues in favoured region (expected value ~98.0%), number of residues in allowed region (~2.0% expected) and Number of residues in outlier region. Our selected model of Api m3 showed 90.0% amino acid residue resided in the favored regions (Figure 4.18).

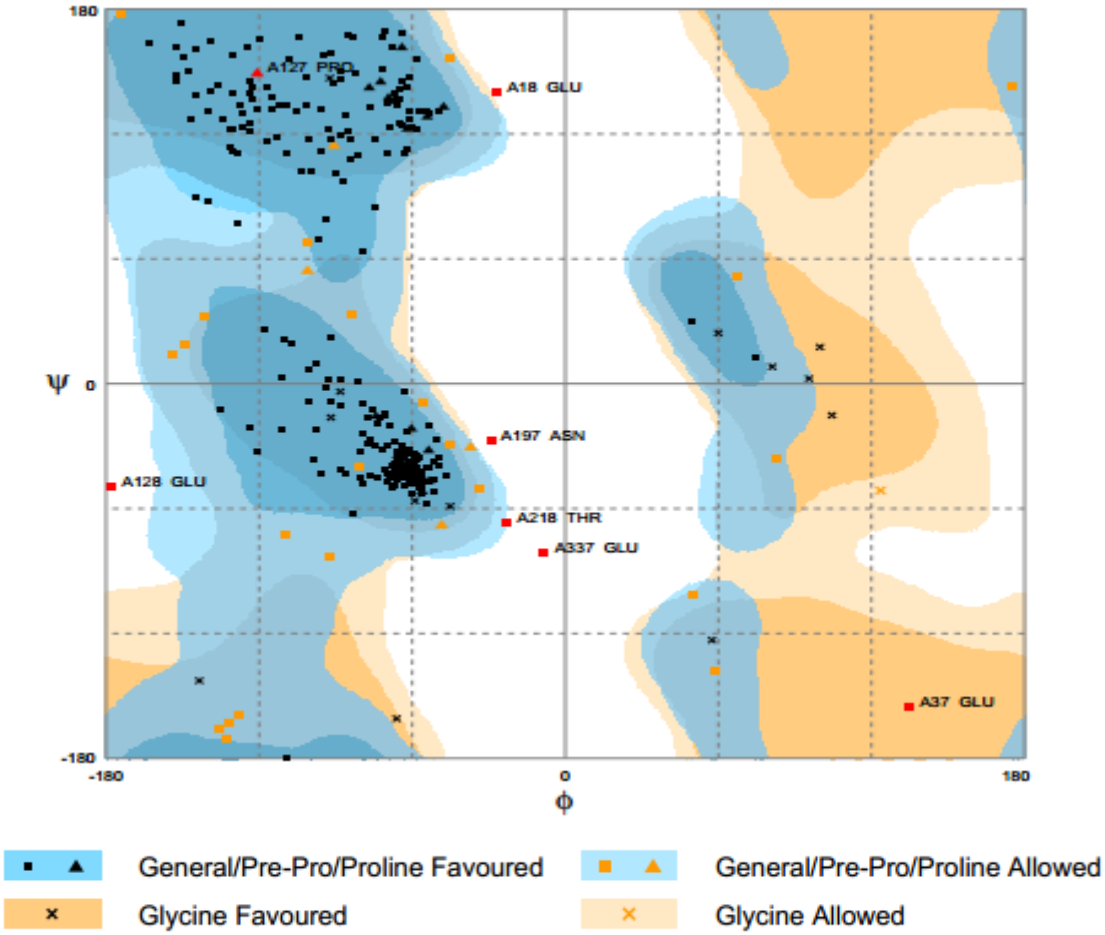

Figure 4.18: Color coded results attained from RAMPAGE. Maximum cluster was observed in favored regions (90.0%)

PROCHECK evaluates stereochemical quality of predicted protein structure by checking residue-by- residue geometry and overall structural geometry. It provides amino acid residues distribution on Ramachandran plot divided into four colour coated regions. Residues in most favored regions colored in red, residues in additional allowed regions colored in yellow, residues in generously allowed regions colored in pale yellow and residues in disallowed regions colored in white (Figure 4.19). According to PROCHECK standard a good quality model should posses over 90% amino acid resided in the most favored regions. Our selected model of Api m3 showed 85.4% amino acid residue resided in the most favored regions.

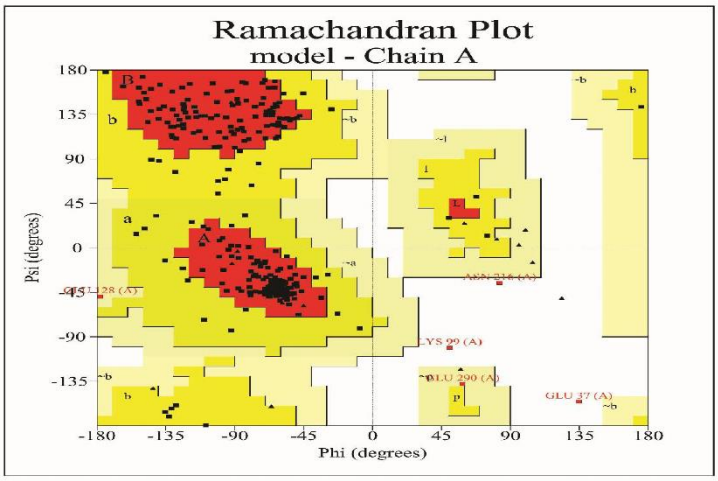

Figure 4.19: PROCHECK analysis result of api m3 protein of *Apis mellifera*, here, Color codes are Red color- most favorable regions, yellow color region- allowed region, and pale yellow generously allowed region and white color- disallowed regions.

After validation of our predicted model, we denoted our proposed T cell and B cell regions onto the model (Figure 4.20, 4.21 and 4.22).

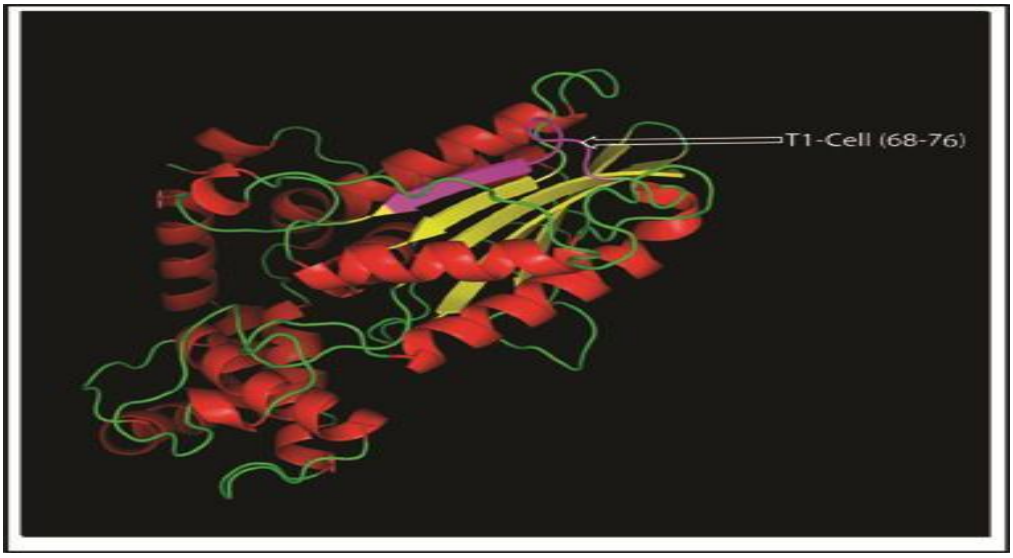

Figure 4.20: MHC class I T-cell epitope is marked in magenta, which is positioned in amino acid residue no. 68-76.

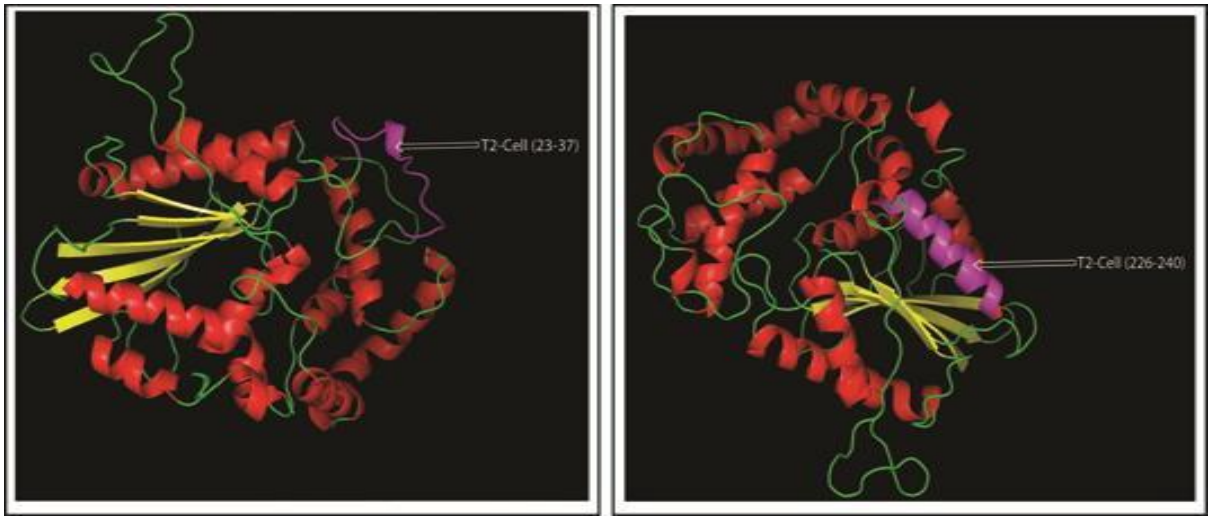

Figure 4.21: Two MHC class I T-cell epitope is marked in magenta, which is positioned amino acid residue no. 23-37 & 226-240.

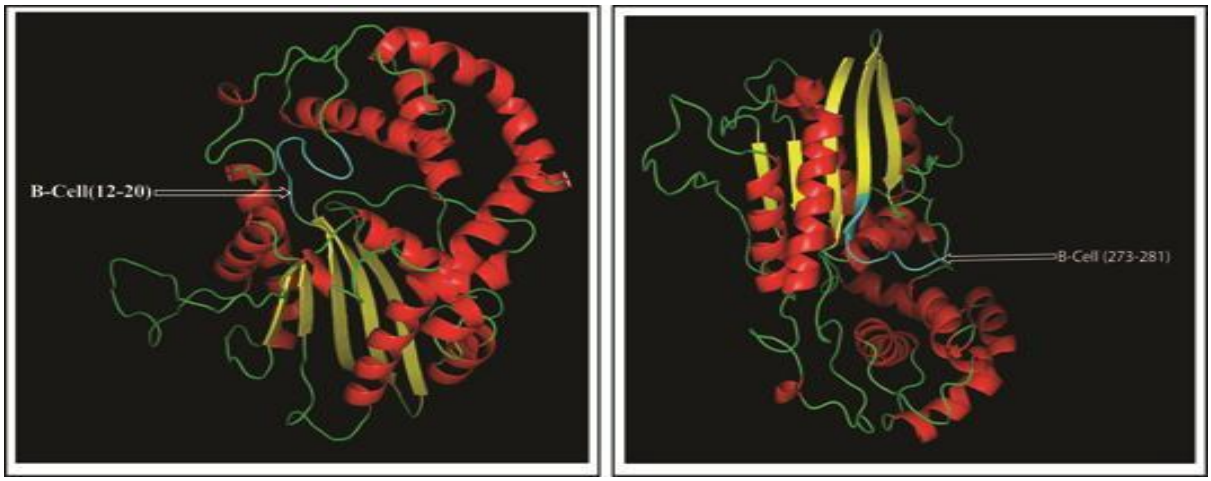

Figure 4.22: B-cell epitopes are highlighted in cyan. First, potential B-cell epitope is region 12-20 (Cyan) and another candidate is 273-281 (Cyan).

4.4 Analysis of B Cell Epitope Prediction Tool

Our proposed algorithm, provided output as a graphical representation for the ease of user. Our concerned parameters for B cell epitope identification were Parameters such as hydrophilicity, accessibility, turns, exposed surface, polarity and antigenic properties of polypeptides chains have been correlated with the location of continuous & discontinuous epitopes. This has led to a search for empirical rules that would allow the position of continuous epitopes to be predicted from certain features of the protein sequence. All prediction calculations are based on propensity scales for each of the 20 amino acids. Each scale consists of 20 values assigned to each of the amino acid residues on the basis of their relative propensity to possess the property described by the scale.

The output of the programme illustrated as graphical representation. On the graphs, the Y-axes depicts for each residue the correspondent score (averaged in the specified window), while the X-axes depicts the residue positions in the sequence.

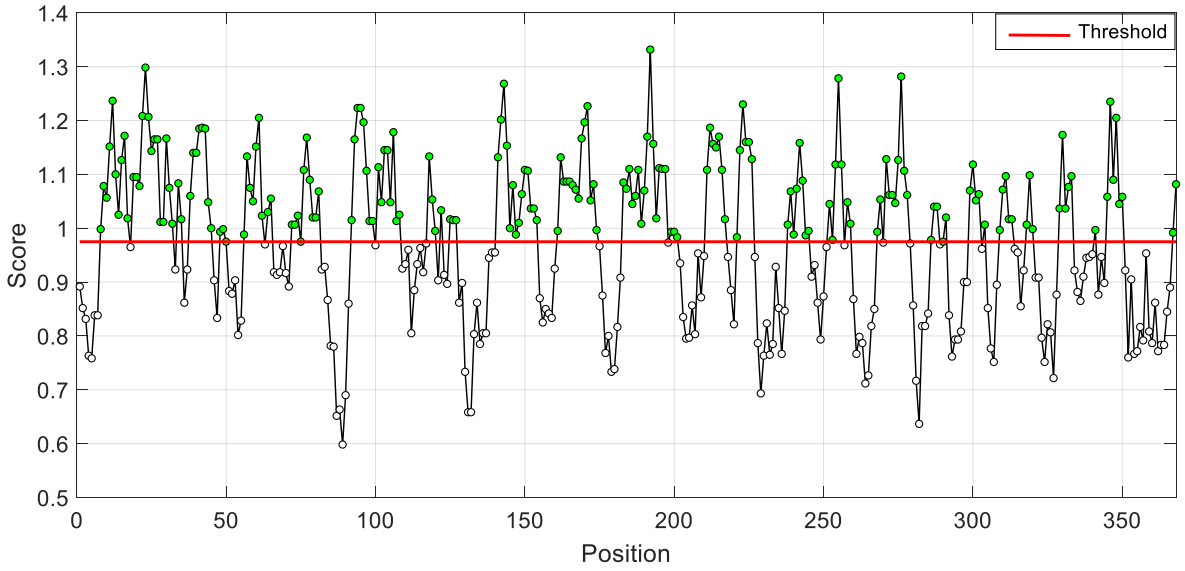

Figure 4.23: Graphical presentation of output for predicting beta turn region in a protein.

The larger score for the residues might be interpreted as that the residue might have a higher probability to be part of epitope (those residues are above the red line threshold on the graphs).

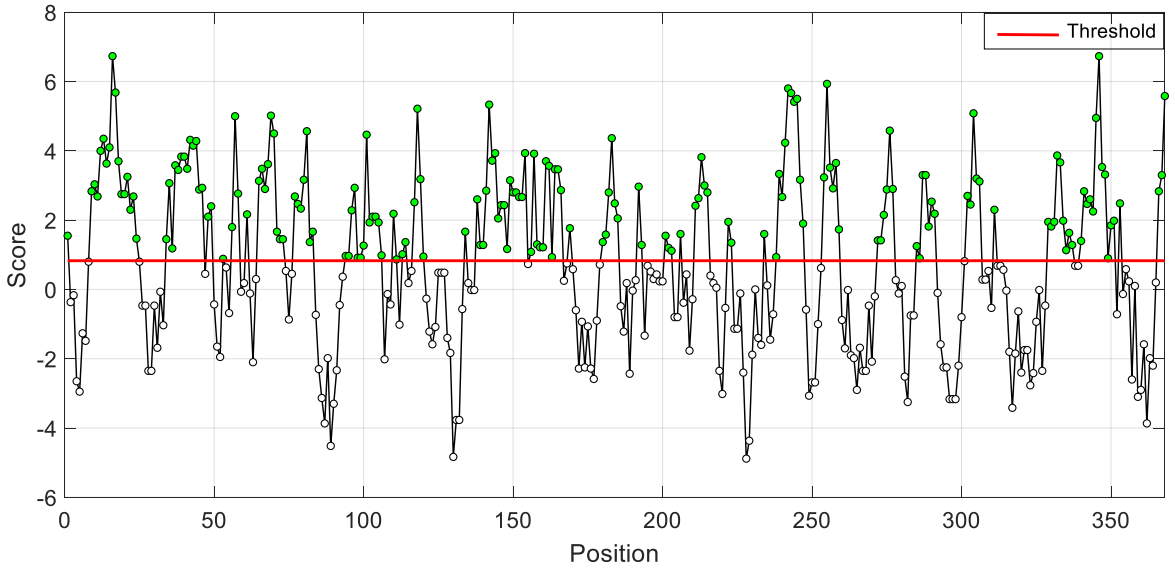

Figure 4.24: Hydropathicity prediction of a protein represented in graphs.

However, the presented methods do not predict the epitopes per se, either linear or discontinuous, they might only guide the researchers to further explore the protein regions on being genuine epitopes.

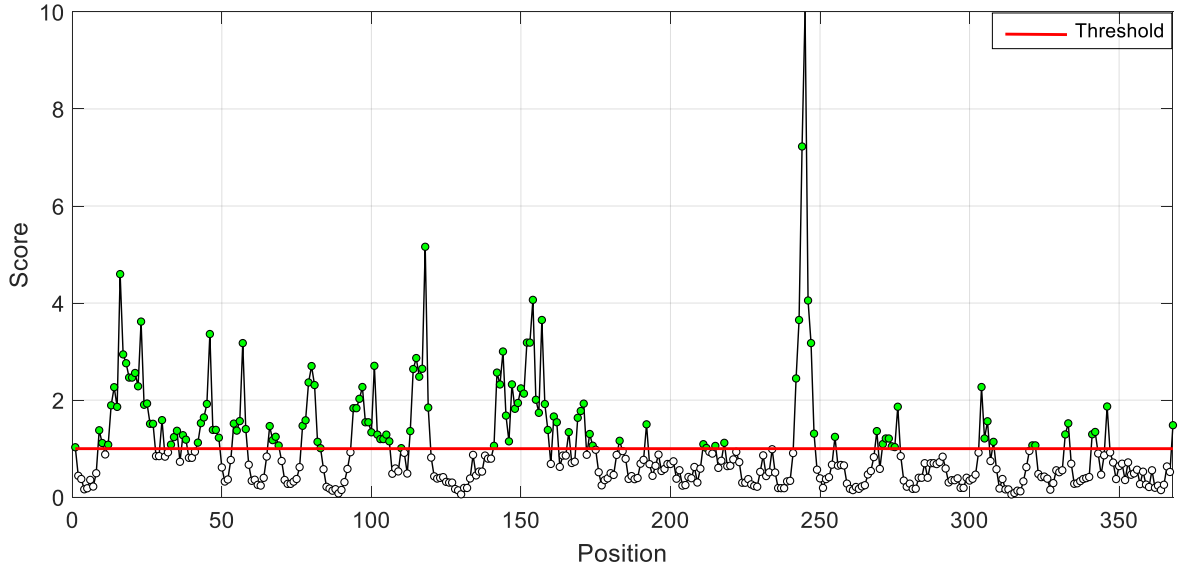

Figure 4.25: Graphical presentation of output for predicting surface accessibility of a protein.

Here, we separately built graphs for beta turn, hydropathicity, surface accessibility and antigenicity. As it was said earlier, higher scores in graphs denote higher probabilities of being a B cell epitope. Finally, all the graphs were superimposed onto each other to identify the region(s) of the protein which is highly antigenic, flexible, situated on the beta turn, hydrophilic and surface accessible, simultaneously. This region will be considered as potential B cell epitope and can be used for further vaccine development.

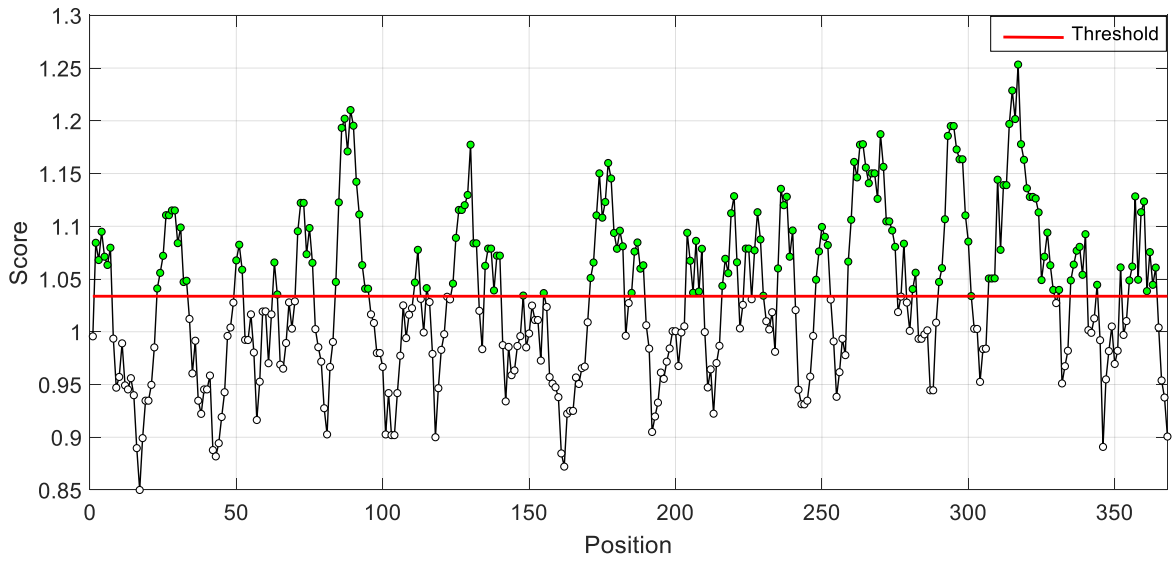

Figure 4.26: Graphical presentation of output for predicting antigenicity of a protein.

Figure 4.23 illustrates regions that are most potential beta turn regions. The spikes above the threshold are probable beta turn regions. We also get idea about, most hydrophilic regions in a protein from the Figure 4.24. In this figure, the regions above the red threshold line are hydrophilic regions and the line beneath the threshold are hydrophobic regions. So, our developed algorithm properly classified the hydrophilic regions from the hydrophobic regions. To predict B cell epitopes in a protein, it was prerequisite to identify beta turn regions and hydrophilic regions in a protein. On the other hand, we classified most surface accessible regions in a protein which is another criterion of predicting B cell epitope. We classified the most surface accessible regions by the algorithm, which are the values above the threshold in Figure 4.25. Finally, we classified the most antigenic regions of a protein, as according to the literature B cell epitopes are laid into the antigenic portion of a protein. We determined most antigenic part of a protein which is illustrated in Figure 4.26. The regions above the threshold are determined as the most antigenic regions of a protein.

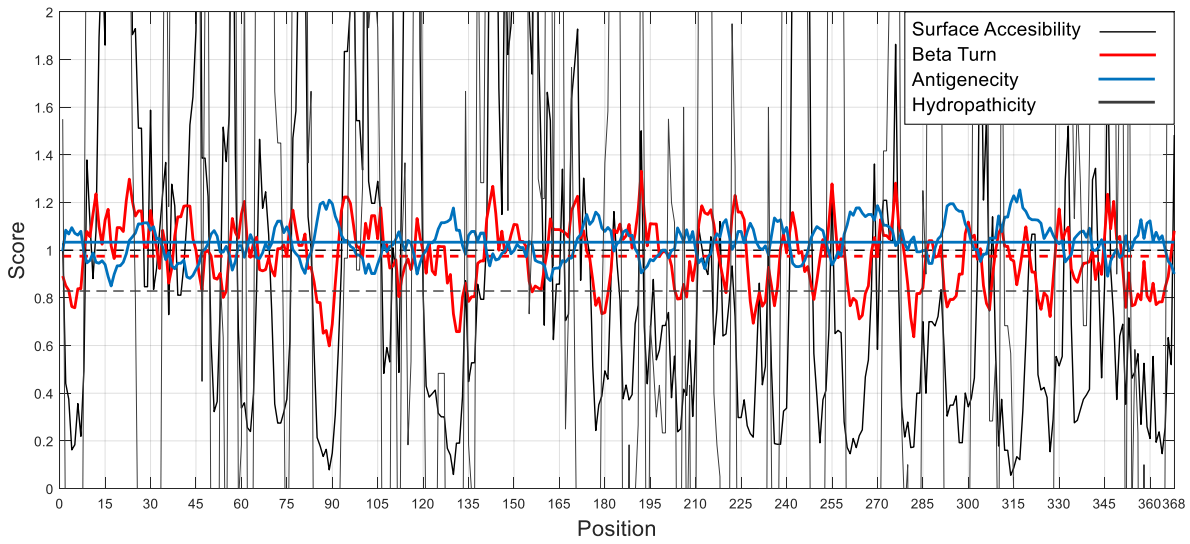

Figure 4.27: Superimposed graphs attained from beta turn prediction, hydropathicity prediction, surface accessibility and antigenicity prediction, for identifying potential B cell epitope.

Lastly, in Figure 6, all four parameters are considered to generate a single graph by superimposing all four graphs and we found that the most probable B cell epitope region is from 272-280 of protein bearing accession AAY57281.1. For attaining better look at the graph provided more zoomed image of the Figure 6, along with the marked predicted B cell epitope is shown in Figure 4.28.

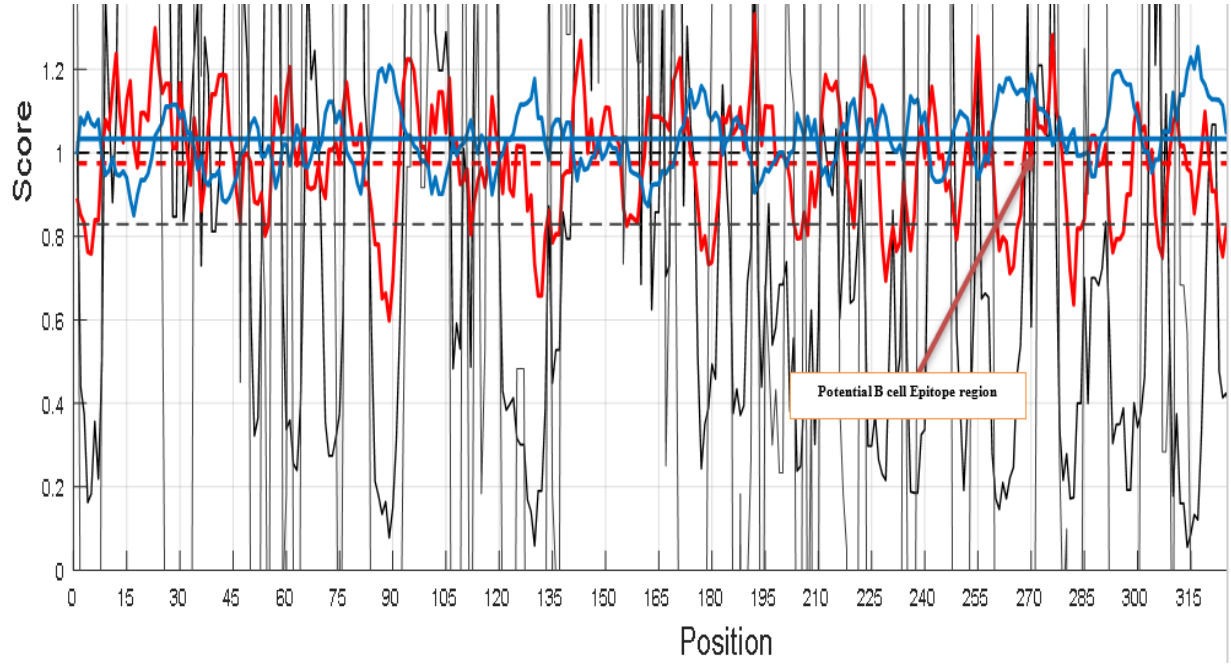

Figure 4.28: Partial zoomed in image for showing potential B cell epitope determined through superimposing.

4.5 Discussions

Histamine plays vital role in molecular mechanism of allergic reactions. Therefore, characterization and homology modeling of Histamine receptors is of great importance. In this study, we focus on the physicochemical properties, motif analysis, transmembrane region prediction and 3D structure analysis of Histamine receptors. Physicochemical properties of these histamine receptors were analyzed and found molecular weight around 55.7 KDa, Theoritical pI 9.33-9.62, instability index 34.93-47.00, aliphatic index (AI) was above 90 and most of the receptors were hydrophobic except histamine H1 receptor. Histamine receptors lacking satisfactory conserved region but region 75-94 was found promising after multiple sequence alignment. Histamine receptors are member of family G protein-coupled rhodopsin-like receptor. We found a profound motif from 84-149 for four histamine receptors with significantly lower E-value. These receptors were seven pass transmembrane protein and we found the gap between transmembrane helix number 5^{th} and $6^{th}$of each histamine receptor except Histamine H2 receptor which can be potential drug target candidate. Finally, 3D models of these receptors were developed through homology modeling and best models were selected by applying different model validation tools.

On the other hand, Api m3 is one of the major allergen in honeybee allergen family. Honeybee allergens can cause severe anaphylaxis, even leads to death. Being an economically important insect, human keeps close contact with it and often becomes life threatening not only in beekeepers but also for mass population. Therefore, vaccination can be a preferred method to counter this issue. Prerequisite for successful vaccine design is to identify effective epitopes. In quest of this, physicochemical properties of Api m3 were analyzed and found relative thermostable nature of api m3 allergen and only 10.46% of the allergen consisted of beta turn region. Five MHC class I T-cell epitopes were identified and through scrutiny of these T-cell epitopes led to find out YTEESVSAL as the best epitope. For MHC class II T-cell epitopes YPKDPYLYYDFYPLE and GGPLLRIFTKHMLDV were found as most prominent T-cell epitopes of api m3 allergen. Linear B-cell epitopes were predicted by using BCPREDS, ABCpred, BepiPred and Bcepred and results were validated by means of hydrophilicity, antigenicity, surface accessibility, flexibility and beta turn region. Current study also revealed that, GDRIPDEKN and PHVPEYSSS, two 9 mer peptides could be the most effective B-cell epitopes of Api m3.

Finally, Correlation between B cell epitope localization and physicochemical properties (e.g., hydrophilicity, solvent accessibility, flexibility, turns, polarity, antigenicity, and surface

exposure) of protein is widely known. In this study, we tried to develop an algorithm-based B cell epitope detection interface. By considering most crucial features of B cell epitopes, we generated specific algorithms and finally compiled the results in a single interface. This study, will efficiently detect B cell epitope in an antigen and also reduce the cost and time required for vaccine development. Much studies need to be carried out on prediction of T cell epitope as well.

4.6 Chapter Summary

Different physiochemical properties of histamines receptors are analysed and discussed in this chapter. Homology models are built for each histamine receptors and checked for their effectivity. B cell and T cell epitopes for api m3 allergen are predicted and analysed. The chapter also shows how effective the predicted epitopes are by analysing epitope conservancy and population coverage. By developing algorithms, B cell epitopes are successfully predicted and validated. This chapter cumulatively answers questions asked in objective part of the study.

Chapter 5

Conclusion

5.1 Outcomes

Histamine receptors are very significant in studying molecular mechanism of allergy. Potent interaction with different components of immune system has evolved histamine receptor as a potent drug target for allergy. A rigorous characterization of different Histamine receptors has been done in this study as well as their 3D models have been developed. Physicochemical properties, instability index and aliphatic index have been examined as well as suitable motifs and potential regions have been suggested for targeted drug binding site. The findings of this study might be helpful in designing more suitable antihistamines and relevant drugs in treatment of allergic diseases. Much study needs to be done in order to analyse, why histamine provokes immune system to perform cascade of allergic reactions and how it could be sensitized without any side effects. On the other hand, Identification of epitopes in api m3 of honeybee is crucial for suitable vaccine assessment and development. The study predicted MHC Class I & Class II T-cell epitopes and B-cell epitopes in api m3 *Apis mellifera*. These peptides should be assessed further for immunoreactivity through *in vivo* studies. The outcome of this study can certainly aid in designing new therapeutic modalities in api m3 venom allergen of *Apis mellifera.* Lastly, it was a challenge to develop efficient B cell epitope prediction methods, which can aid in vaccine development research. Correlation between B cell epitope localization and physicochemical properties (e.g., hydrophilicity, solvent accessibility, flexibility, turns, polarity, antigenicity, and surface exposure) of protein is widely known. In this study, we tried to develop an algorithm-based B cell epitope detection interface. By considering most crucial features of B cell epitopes, we generated specific algorithms and finally compiled the results in a single interface. This study, will efficiently detect B cell epitope in an antigen and also reduce the cost and time required for vaccine development. Much studies need to be carried out on prediction of T cell epitope as well.

5.2 Significance Statement

This study shows potential regions of histamine receptors those can be targeted as drug binding site. 3D models for all four histamine receptors are also proposed. Among these receptors, crystal structure of Histamine H2, Histamine H3 and Histamine H4 receptor yet to be discovered. With 3D models, potential therapeutic peptides can be docked to active

sites for blocking burden amount of histamine by these receptors. Overall, this study might be useful in designing new generations of antihistamines.

Potential T-cell and B-cell epitopes of api m3 allergen are reported. These predicted T-cell and B-cell epitopes of api m3 allergen could help the researchers to test them further for immunoreactivity applying *in vivo* analysis. Still there is no report of T-cell and B-cell epitopes of *Apis mellifera*, this study can be the pioneer in finding effective vaccine against allergens of honeybee.

This research successfully also predicts potential B cell epitope regions from an antigenic protein. The most exciting feature of this study is, it presents results of potential B cell epitopes on a single interface, so that, researchers don't need to search for every feature (e.g., hydrophilicity, antigenicity, beta turn, surface accessibility etc.) separately. Finally, the study can certainly aid in B cell epitope-based vaccine design research.

5.3 Limitations and Future Perspectives

With this study, we developed efficient protocol for modeling Histamine receptors and studied biomolecular features of these receptors. With this protocol, researcher can be able to design homology models for any receptor and proteins those are of causal agent for disease progression. These models will provide us complete idea about the complex structural features of these receptors and according to the acquired knowledge, we can efficiently identify target zones for docking drugs and advance one step closer to fight with the disease associated with the receptor. On the other hand, we have characterized honey bee venom allergen and developed vaccine against it. This vaccine development process reduces huge amount of money, time and labour for designing true vaccines. It reduces cost of over and over wet lab trials and time as well. After assuring, the actual vaccine candidates now we can proceed to develop wet lab vaccine. Also, this method can be applied to design new therapeutics against any allergen protein or antigen protein. Finally, we tried to develop an efficient, user friendly B cell epitope prediction method, with MATLAB algorithms. Our developed interface can be used by the scientists to predicts B cell epitope in an antigen protein. With the same technique, it might be possible to develop T cell epitope prediction interfaces and lastly, with Artificial Neural Networks, if we can train B cell and T cell epitope prediction methods with wet lab data, it might conclude up resulting a novel vaccine predicting tool for future. As an *in silico* study, the outcomes needed wet lab trials and execution to step up in the next stage.

References

[1] H. A. Sampson, A. Muñoz-Furlong, R. L. Campbell, *et al*., "Second symposium on the definition and management of anaphylaxis: summary report—Second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium," The Journal of Allergy and Clinical Immunology, vol. 117, no. 2, pp. 391–397, 2006.

[2] J. E. Tintinalli, "Emergency Medicine: A Comprehensive Study Guide," New York: McGraw-Hill Companies, 2nd Edition, pp. 177–182, 2010.

[3] FAO/WHO, Codex alimentarius commission: report of the twenty-sixth session of the codex committee on food labeling, Ottawa, Canada: ALINORM 99/22, 1998.

[4] K. D. Jackson, L. D. Howie, L. J. Akinbami, "Trends in allergic conditions among children: United States, 1997–2011," National Center for Health Statistics data brief, Vol. 121, 2013.

[5] M. D. Giuseppe, *et al*., *Nelson Biology*, Toronto: Thomson Canada, Vol. 12, pp. 473, 2003.

[6] S. W. Ludman and R.J. Boyle, "Stinging insect allergy: current perspectives on venom immunotherapy," Journal of Asthma and Allergy, vol. 8, pp. 75–86, 2015.

[7] D. Wiwatwongwana , N. Jariyapan , A. Wiwatwongwana, "Eyelid bee sting with late migration onto the cornea after primary removal: the mystery of the bee stinger," Arch Ophthalmol, Vol. 130, no. 3, pp. 392–3

[8] T. Shimamura *et al*., "Structure of the human histamine H1 receptor complex with doxepin," Nature, vol. 475, no. 7354, pp. 65–70, 2011.

[9] S. J. Hill, C. R. Ganellin, H. Timmerman, J. C. Schwartz, N. P. Shankley, J.M. Young, W. Schunack, R. Levi, H. L. Haas, "International Union of Pharmacology XIII. Classification of histamine receptors," Pharmacological Reviews, vol. 49, no. 3, pp. 253–78, 1997.

[10] M.B. Passani, J.S. Lin, A. Hancock, S. Crochet, P Blandina, "The histamine H3 receptor as a novel therapeutic target for cognitive and sleep disorders," Trends in Pharmacological Sciences, vol. 25, no 12, pp. 618–25, 2004.

[11] G. E. Morris, "Epitope mapping," Methods in Molecular Biology, vol. 295, pp. 255–268, 2005.

[12] J. Huang, B. Ru, and P. Dai, "Bioinformatics resources and tools for phage display," Molecules, vol.16, no.1, pp.694–709,2011.

[13] V. Moreau, C. Granier, S. Villard, D. Laune, and F. Molina, "Discontinuous epitope prediction based on mimotope analysis," Bioinformatics, vol.22, no.9 ,pp.1088–1095, 2006.

[14] R. E. West, A. Zweig, N. Y. Shih, M. I. Siegel, R. W. Egan, M. A. Clark, "Identification of two H3-histamine receptor subtypes," Molecular Pharmacology, vol. 38**,** no. 5, pp. 610–613, 1990.

[15] M. Rapanelli, "The Magnificent Two: Histamine and the H3 Receptor as Key Modulators of Striatal Circuitry," Progress in Neuro-Psychopharmacology and Biological Psychiatry, vol. 73, pp. 36–40, 2017.

[16] B. Sadek, S. Ali, S. Adel, J. Fakhreya, and Holger Stark, "Histamine H3 Receptor as a Potential Target for Cognitive Symptoms in Neuropsychiatric Diseases," Behavioural Brain Research, vol. 312, pp. 415–430, 2016.

[17] S. Attoub, L. Moizo, I Sobhani, J. P. Laigneau, M. J. Lewin, A. Bado, "The H3 receptor is involved in cholecystokinin inhibition of food intake in rats," Life Sciences, Vol. 69, no. 4, pp. 469–478, 2001.

[18] K. Takeshita, K. B. Bacon, and F. Gantner, "Critical role of L-selectin and histamine H4 receptor in zymosan-induced neutrophil recruitment from the bone marrow: comparison with carrageenan," Journal of Pharmacology and Experimental Therapeutics, vol. 310, no. 1, pp. 272–80, 2004.

[19] A. Salem, S. Rozov, A. Al-Samadi, *et al*., "Histamine metabolism and transport are deranged in human keratinocytes in oral lichen planus," British Journal of Dermatology, vol. 176, no. 5, pp. 1213-1223, 2016.

[20] T. Hunt, "Protein sequence motifs involved in recognition and targeting: a new series," Trends Biochem. Sci., vol. 15, no. 8, pp. 305–309, 1990.

[21] J. A. Wells and C. L. McClendon, "Reaching for high-hanging fruit in drug discovery at protein-protein interfaces," Nature, vol. 450, no. 7172, pp. 1001–1009, 2007.

[22] L. T. Vassilev, B. T. Vu, B. Graves, *et al.*, "In vivo activation of the p53 pathway by small-molecule antagonists of MDM2," Science, vol. 303, no. 5659, pp. 844–848, 2004.

[23] S. L. Goodman, G. Hölzemann, G. A. Sulyok, H. Kessler, "Nanomolar small molecule inhibitors for alphav(beta)6, alphav(beta)5, and alphav(beta)3 integrins," Journal of Medical Chemistry, vol. 45, no. 5, pp. 1045–51, 2002.

[24] L. Oliveira-Ferrer, J. Hauschild, W. Fiedler, C. Bokemeyer, J. Nippgen, I. Celik, G. Schuch, "Cilengitide induces cellular detachment and apoptosis in endothelial and glioma cells mediated by inhibition of FAK/src/AKT pathway," Journal of Experimental and Clinical Cancer Research, vol. 27, no. 1, pp. 86, 2008.

[25] B. Gril, M. Vidal, F. Assayag, M. F. Poupon, W. Q. Liu, C. Garbay, "Grb2-SH3 ligand inhibits the growth of HER2+ cancer cells and has antitumor effects in human cancer xenografts alone and in combination with docetaxel," International Journal of Cancer, vol. 121, no. 2, pp. 407–415, 2007.

[26] S. M. Feller and M. Lewitzky, "Potential disease targets for drugs that disrupt protein-protein interactions of Grb2 and Crk family adaptors," Curr. Pharm. Des., vol. 12, no. 5, pp. 529–548, 2006.

[27] *J. F. Manor, S. Esther, A. T. Isaiah,* "Environment Polarity in Proteins Mapped Noninvasively by FTIR Spectroscopy," *The Journal of Physical Chemistry Letters, vol. 3, no. 7, pp. 939–944, 2012.*

[28] *S. R. Goodman,* Medical cell biology, *Academic Press, pp. 37–43, 2008.*

[29] X. *Jin,* Essential bioinformatics, *Cambridge University Press, pp. 208–210, 2006.*

[30] *M. S. Almén, K. J. Nordström, R. Fredriksson, H. B. Schiöth,* "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin," *BMC Biology, vol. 7, pp. 50, 2009.*

[31] *C. Chothia, A. M. Lesk, "*The relation between the divergence of sequence and structure in proteins," *EMBO J., Vol. 5, no. 4, pp. 823–826, 1986.*

[32] *S. Kaczanowski, P. Zielenkiewicz, "*Why similar protein sequences encode similar three-dimensional structures?," *Theoretical Chemistry Accounts, vol. 125, no. 3–6, pp. 643–650, 2010.*

[33] *M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, "Comparative protein structure modeling of genes and genomes," Annu Rev Biophys Biomol Struct., vol. 29, pp. 291–325, 2000.*

[34] S. Y. Chung, S. Subbiah, "A structural explanation for the twilight zone of protein sequence homology," Structure, vol. 4, pp. 1123–1127, 1996.

[35] *A. R. Williamson, "Creating a structural genomics consortium," Nature Structural Biology, vol. 7, no. (S1(11s)), pp. 953, 2000.*

[36] *C. Venclovas, M. Margelevičius, "Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment," Proteins,* vol. *61*, no. *S7, pp. 99–105, 2005.*

[37] S. E. Michael, *"The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae: Apis),"* Journal of Hymenoptera Research, *vol. 8, pp. 165–196, 1996.*

[38] *J. Chen, S. M. Guan, W. Sun, H. Fu,* "Melittin, the Major Pain-Producing Substance of Bee Venom," *Neuroscience Bulletin, vol. 32, no. 3, pp. 265–72, 2016.*

[39] *S. Ramanadham, T Ali, J. W. Ashley, R. N. Bone, W. D. Hancock, X. Lei, "*Calcium-independent phospholipases A2 and their roles in biological processes and diseases," *Journal of Lipid Research, vol. 56, no. 9, pp. 1643–1668, 2015.*

[40] J. H. *Park, B. K. Yim, J. H. Lee, T. H. Kim, "*Risk Associated with Bee Venom Therapy: A Systematic Review and Meta-Analysis," *PLOS One, vol. 10, no. 5, pp. e0126971, 2015.*

[41] *J. A Lee, M. J. Son, J. Choi, J. H. Jun, J. I. Kim, M. S. Lee,* "Bee venom acupuncture for rheumatoid arthritis: A systematic review of randomised clinical trials," *BMJ Open, vol. 4, no. 11, pp. e006140, 2014.*

[42] R. J. *Boyle, M. Elremeli, J. Hockenhull, M. G. Cherry, M. K. Bulsara, M. Daniels, J. N. Elberink, "*Venom immunotherapy for preventing allergic reactions to insect stings," *Cochrane Database of Systematic Reviews, vol. 10, pp. 620-629, 2012.*

[43] "Wasp and Bee venom," *Drugs.com, 2018*

[44] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, "Molecular Biology of the Cell," Garland Science: New York, NY, pp. 1367, 2002.

[45] *S. McClory, T. Hughes, A. G. Freud, E. L. Briercheck, C. Martin, A. J. Trimboli, J. Yu, X. Zhang, G. Leone, G. Nuovo, M. A. Caligiuri,* "Evidence for a stepwise program of extrathymic T cell development within the human tonsil," *The Journal of Clinical Investigation, vol. 122, no. 4, pp. 1403–1415, 2012.*

[46] *P. Vantourout, A. Hayday,* "Six-of-the-best: unique contributions of γδ T cells to immunology," *Nature Reviews of Immunology, vol. 13, no. 2, pp. 88–100, 2013.*

[47] *T. Mallevaey, J. Fontaine, L. Breuilh, C. Paget, A. Castro-Keller, C. Vendeville, M. Capron, M. Leite-de-Moraes, F. Trottein, C. Faveeuw,* "Invariant and noninvariant natural killer T cells exert opposite regulatory functions on the immune response during murine schistosomiasis," *Infection and Immunity, vol. 75, no. 5, pp. 2171–2180, 2007.*

[48] R. J. *Napier, E. J. Adams, M. C. Gold, D. M. Lewinsohn,* "The Role of Mucosal Associated Invariant T Cells in Antimicrobial Immunity," *Frontiers in Immunology, vol. 6, pp. 344, 2015.*

[49] M. C. *Gold and D. M. Lewinsohn,* "Mucosal associated invariant T cells and the immune response to infection," *Microbes and Infection, vol. 13, no. 8–9, pp. 742–748, 2011.*

[50] *Alberts, "Molecular Biology of the Cell," New York: Garland Science. p. 1401, 2002.*

[51] E. *Davidson, J. D. Benjamin,* "A High-throughput Shotgun Mutagenesis Approach to Mapping B-cell Antibody Epitopes," *Immunology*, vol. *143, no. 1, pp. 13–20,* 2014.

[52] E. *Koren and A. S. De-Groot,* "Clinical validation of the "*in silico*" prediction of immunogenicity of a human recombinant therapeutic protein," *Clinical Immunology, vol. 124, no. 1, pp. 26–32, 2007.*

[53] A. S. *De Groot, W. Martin,* "Reducing risk, improving outcomes: Bioengineering less immunogenic protein therapeutics," *Clinical Immunology*, vol. *131, no. 2, pp. 189–201, 2009.*

[54] *J. Walker, R. Ralph, "Molecular bio-methods handbook," Humana Press, pp. 467, 2008.* ISBN 1-60327-374-3.

[55] *N. Biologicals,"*OLLAS Epitope Tag," *Novus Biologicals*, 2011.

[56] B. *Zakeri, "*Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin*," Proceedings of the National Academy of Sciences, vol. 109, no. 12, pp. 690–697, 2012.*

[57] *E. B. Correia, T. J. Bates, J. R. Loomis, G. Baneyx, C. Carrico, G. J. Jardine, P. Rupert, C. Correnti, O. Kalyuzhniy,* "Proof of principle for epitope-focused vaccine design," *Nature, vol. 507,* no. *7491, pp. 201–206, 2014.*

[58] *P.S. McBurney, E. J. Sunshine, S. Gabriel, P. J. Huynh, F. W. Sutton, H. D. Fuller, L. N. Haigwood, B. W. Messer, "*Evaluation of protection induced by a dengue virus serotype 2 envelope domain III protein scaffold/DNA vaccine in non-human primates," *Vaccine, vol. 34, no. 30, pp. 3500–3507, 2016.*

[59] W. E. Paul, Fundamental Immunology, Lippincott Williams & Wilkins, 2012.

[60] B. Sun and Y. Zhang, "Overview of orchestration of CD4+ T cell subsets in immune responses," Advances in Experimental Medicine and Biology, vol. 841, pp. 1–13, 2014.

[61] M. H. Van Regenmortel, "What is a B-cell epitope?," Methods in Molecular Biology, vol. 524, pp. 3–20, 2009.

[62] J. Ponomarenko and M. V. Regenmortel, "B-cell epitope prediction," Structural Bioinformatics, pp. 849–879, John Wiley & Sons, Inc, 2009.

[63] T. A. Ahmad, A. E. Eweida, and L. H. El-Sayed, "T-cell epitope mapping for the design of powerful vaccines," Vaccine Reports, vol. 6, pp. 13–22, 2016.

[64] *K. Murphy, Janeway's Immunobiology (8th ed.). New York: Garland Science, 2016.* ISBN 9780815342434.

[65] D. M. *Cooper, "*The early history of B cells," *Nature Reviews Immunology, vol. 15, no. 3, pp. 191–197, 2015.*

[66] *I. Martensson, N. Almqvist, O. Grimsholm, A. Bernardi, "The pre-B cell receptor checkpoint," FEBS Letters,* vol. *584,* no. *12, pp. 2572–2579, 2010.*

[67] *W. T. LeBien, W. Tucker, T. F. Tedder,* "B lymphocytes: how they develop and function," *Blood,* vol. *112, no. 5, pp. 1570–1580, 2008.*

[68] R. *Pelanda, M. R.Torres,* "Central B-Cell Tolerance: Where Selection Begins," *Cold Spring Harbor Perspectives in Biology, vol. 4,* no. *4, pp. a007146, 2014.*

[69] *F. B. Loder, B. Mutschler, J. R. Ray, J. C. Paige, P. Sideras, R. Torres, C. M. Lamers, R. Carsetti,* "B Cell Development in the Spleen Takes Place in Discrete Steps and Is Determined by the Quality of B Cell Receptor–Derived Signals," *The Journal of Experimental Medicine, vol. 190, no. 1, pp. 75–90, 1999.*

[70] B. J. *Chung, M. Silverman, J. G. Monroe,* "Transitional B cells: step by step towards immune competence," *Trends in Immunology*, vol. *24*, no. *6*, *pp. 342–348, 2003.*

[71] *A. Cerutti, C. Montserrat, I. Puga,* "Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes," *Nature Reviews Immunology, vol. 13, no. 2, pp. 118–132, 2013.*

[72] K. Murphy, Janeway's Immunobiology (8th ed.), New York: Garland Science, 2012. ISBN 9780815342434.

[73] L. S. Nutt, D. P. Hodgkin, M. D. Tarlinton, M. L. Corcoran, M. Lynn, "The generation of antibody-secreting plasma cells," Nature Reviews Immunology, vol. 15, no. 3, pp. 160, 2015.

[74] A. *Bortnick, I. Chernova, J. W. Quinn, M. Mugnier, P. M. Cancro, D. Allman,* "Long-Lived Bone Marrow Plasma Cells Are Induced Early in Response to T Cell-Independent or T Cell-Dependent Antigens," *The Journal of Immunology, vol. 188, no. 11, pp. 5389–5396, 2012.*

[75] M. *McHeyzer-Williams, S. Okitsu, N. Wang, L. McHeyzer-Williams,* "Molecular programming of B cell memory," *Nature Reviews Immunology, vol. 12, no. 1, pp. 24–34, 2011.*

[76] *T. Kurosaki, K. Kometani, W. Ise,* "Memory B cells," *Nature Reviews Immunology*, vol. *15*, no. *3*, *pp. 149, 2015.*

[77] *B. Ribourtout and M. Zandecki, "Plasma cell morphology in multiple myeloma and related disorders," Morphologie : Bulletin De l'Association Des Anatomistes*, vol. *99*, no. *325*, *pp. 38–62, 2015.*

[78] *S. Pillai, A. Cariappa, S. T. Moran,* "Marginal Zone B Cells," *Annual Review of Immunology, vol. 23, no. 1, pp. 161–196, 2005.*

[79] N. *Baumgarth,* "The double life of a B-1 cell: self-reactivity selects for protective effector functions," *Nature Reviews Immunology, vol. 11, no. 1, pp. 34–46.*

[80] C. E. *Rosser, C. Mauri,* "Regulatory B Cells: Origin, Phenotype, and Function," *Immunity, vol. 42, no. 4, pp. 607–612, 2015.*

[81] H. Naderi-Manesh, M. Sadeghi, S. Arab, and M. Moosavi, "Prediction of protein surface accessibility with information theory," Proteins, vol. 42, pp. 452–459, 2001.

[82] B. Rost, "Predicting one-dimensional protein structure by profile-based neural networks," Methods Enzymology, vol. 266, pp. 525–539, 1996.

[83] B. Rost, and C. Sander, "Conservation and prediction of solvent accessibility in protein families," Proteins, vol. 20, pp. 216–226, 1994.

[84] Z. Yuan, K. Burrage and J. S. Mattick, "Prediction of protein solvent accessibility using support vector machines," Proteins, vol. 48, pp. 566–570, 2002.

[85] J. Kyte, R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," Journal of Molecular Biology, vol. 157, no. 1, pp.105–32, 1982.

[86] G. W. Canonica, and M. Blaiss, "Antihistaminic, anti-inflammatory, and anti allergic properties of the non sedating second-generation antihistamine desloratadine: a review of the evidence," The World Allergy Organization Journal, vol. 4, no. 2, pp. 47–53, 2011.

[87] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson and D.G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," Mol. Syst. Biol., vol. 7, pp. 539, 2011.

[88] M. Shahid, *et al.,* "Histamine, Histamine Receptors, and their Role in Immunomodulation: An Updated Systematic Review," The Open Immunology Journal, vol.2, pp. 9-41, 2009.

[89] S. Pundir, M.J. Martin and C. O'Donovan, "UniProt Protein Knowledgebase**.** Methods in Molecular Biology," Vol. 1558, pp. 41-55, 2017.

[90] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, "Protein Identification and Analysis Tools on the ExPASy Server' John M. Walker (ed): The Proteomics Protocols Handbook," Humana Press, vol. 8, pp. 571-607, 2005.

[91] K. Sivakumar and S. Balaji, "*In silico* characterization of antifreeze proteins using computational tools and servers," Journal of Chemical Sciences, vol. 119, no. 3, pp. 571–579, 2008.

[92] K. Guruprasad, B. V. B. Reddy and M. W. Pandit, "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence," Protein Engineering, Design and Selection, vol. 4, no. 2, pp. 155–161, 1990.

[93] N. Pradeep, V, K. G. Vidyashree and P. Lakshmi, "*In silico* Characterization of Industrial Important Cellulases using Computational Tools'," Advances in Life Science and Technology, vol, 4, pp. 8–14, 2012.

[94] H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A.P. Cowley and R. Lopez, "Analysis Tool Web Services from the EMBL-EBI," Nucleic acids research, vol. 41, pp. 597-600, 2013.

[95] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G.J. Barton, "Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics," vol. 25, pp. 1189-1191, 2009.

[96] R. D. Finn, *et al.*, "InterPro in 2017 — beyond protein family and domain annotations," Nucleic Acids Research, vol. 45, pp. D190–D199, 2017.

[97] T. L. Bailey, M. Bodén, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W.S. Noble, "MEME SUITE: tools for motif discovery and searching," Nucleic Acids Research, vol. 37, pp. W202-W208, 2009.

[98] S. Moller, M.D.R. Croning and R. Apweiler, "Evaluation of methods for the prediction of membrane spanning regions," Bioinformatics, vol. 17, no. 7, pp. 646-653, 2001.

[99] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, "The I-TASSER Suite: Protein structure and function prediction," Nature Methods, vol.12, pp. 7-8, 2015.

[100] C. Colovos and T. O. Yeates, "Verification of protein structures: Patterns of nonbonded atomic interactions," Protein Science, vol.2, no. 9, pp. 1511–1519, 1993.

[101] R. A. Laskowski, *et al.*, "PROCHECK: a program to check the stereochemical quality of protein structures,"Journal of Applied Crystallography, International Union of Crystallography, vol. 26, pp. 283–291, 1993.

[102] S. C. Lovell, I. W. Davis, W. B. Arendal III, P. I. W. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D.C. Richardson, "Structure Validation by Cα Geometry: ϕ,Ψ and CβDeviation," PROTEINS: Structure, Function, and Genetics, vol. 50, no. 3, pp. 437–450, 2003.

[103] C. Geourjon, G. Deleage, "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments," Comput. Applied Bioscience, vol. 11, no. 6, pp. 681–684, 1995.

[104] C. S. Gill, H.P. Von, "Calculation of protein extinction coefficients from amino acid sequence data," Anal. Biochem., vol. 182, no. 2, pp. 319–32, 1989.

[105] K. Guruprasad, B.V. Reddy, M.W. Pandit, "Correlation between stability of a protein and its dipeptide composition, a novel approach for predicting *in vivo* stability of a protein from its primary sequence," Protein Eng., vol. 4, no. 2, pp. 155–161, 1990.

[106] A. Ikai, "Thermostability and Aliphatic Index of Globular Proteins," Journal of Biochemistry, vol. 88, no. 6, pp.1895–1898, 1980.

[107] V. M. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, *et al.,* "Largescale validation of methods for cytotoxic T-lymphocyte epitope prediction," BMC Bioinform., vol. 8, pp. 424, 2007.

[108] S. Buus, S.L. Lauemoller, P. Worning, C. Kesmir, T.S. Frimurer, *et al.,* "Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach," Tissue Antigens, vol. 62, no. 5, pp. 378–384, 2003.

[109] B. Peters, and A. Sette, "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method," BMC Bioinform., vol. 6, pp. 132, 2005

[110] R. Vita, J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, *et al.,* "The immune epitope database (IEDB) 3.0. Nucleic Acids Research, vol. 43(Database issue), pp. D405–D412, 2015.

[111] H. H. Bui, J. Sidney, W. Li, N. Fusseder, A. Sette, "Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines," BMC Bioinform., vol. 8, no. 1, pp. 361, 2007.

[112] P. Thevenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux, *et al.*, "PEP-FOLD: an updated *de novo* structure prediction server for both linear and disulfide bonded cyclic peptides," Nucleic Acids Res., vol. 40, pp. W288–W293, 2012.

[113] J. Maupetit, P. Derreumaux, P. Tufféry, "PEP-FOLD: an online resource for *de novo* peptide structure prediction," Nucleic Acids Res., vol. 37 (Web Server issue), pp. W498–W503, 2009.

[114] J. Maupetit, P. Derreumaux, P. Tuffery, "A fast and accurate method for large scale de novo peptide structure prediction," J. Comput. Chem., vol. 31, pp. 726–738, 2010.

[115] M. T. Fieser, A. John, H. Tainer, "Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein-helix," Proc. Natl. Acad. Sci., vol. 84, no. 23, pp. 8568–8572, 1987.

[116] W. Zhang, Y. Niu, Y. Xiong, M. Zhao, R. Yu, and J. Liu, "Computational Prediction of Conformational B-Cell Epitopes from Antigen Primary Structures by Ensemble Learning," PLoS ONE, vol. 7, no. 8, pp. e43575, 2012.

[117] S. Saha, and G.P. Raghava, "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network," Proteins, vol. 65, pp. 40-48, 2006.

[118] E. J. Larsen, O. Lund, & M. Nielsen, "Improved method for predicting linear B-cell epitopes," Immunome Research, vol. 2, no. 2, 2006.

[119] S. Saha, and G.P. Raghava, "BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties," In ICARIS, LNCS, vol. 3239, pp. 197–204, 2004.

[120] A. P. Karplus, and G.E. Schulz, "Prediction of chain flexibility in proteins, Naturwissenschaften, vol. 72, pp. 212–213, 1985.

[121] S. A. Kolaskar, and P. C. Tongaonkar, "A semi-empirical method for prediction of anti- genic determinants on protein antigens," FEBS Lett., vol. 276, no. 1–2, pp. 172–174, 1990.

[122] E. A. Emini, J.V. Hughes, D.S. Perlow, and J. Boger, "Induction of hepatitis A virus neutralizing antibody by a virus-specific synthetic peptide," J. Virol., vol. 55, no. 3, pp. 836–839, 1985.

[123] M. J. Parker, D. Guo, R.S. Hodges, "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites," Biochemistry, vol. 25, no. 19, pp. 5425–5432, 1986.

[124] Y. P. Chou, and Fasman G.D., "Prediction of the secondary structure of proteins from their amino acid sequence," Adv. Enzymol. Relat. Areas Mol. Biol., vol. 47, pp. 45–148, 1978.

[125] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling," Bioinformatics, vol. 22, pp. 195-201, 2006.

[126] R. Unger and J.L. Sussman, "The importance of short structural motifs in protein structure analysis," J. Computer-Aided Mol Des, vol.7, no. 4, pp.457-472, 1993.

[127] E. R. O'Hehir, S. R. Prickett, & J. M. Rolland, "T Cell Epitope Peptide Therapy for Allergic Diseases," Current Allergy and Asthma Reports, vol. 16, no. 14, pp. 209-216, 2016.

[128] A. E. Pastorello, *et al.*, "Characterization of the T-cell epitopes of the major peach allergen Pru p 3," Int. Arch. Allergy Immunol., vol. 153, no. 1, pp. 1-12, 2010.

[129] V. Eugene, *et al.*, "Identification and validation of shrimp-tropomyosin specific CD4 T cell epitopes. Human Immunology," vol. 74, no. 12, pp. 1542-1549, 2013.

[130] S. Fanuel, *et al.*, "Analysis of predicted B and T-cell epitopes in Der p 23, allergen from *Dermatophagoides pteronyssinus*," Bioinformation, vol. 13, no. 9, pp. 307-312, 2017.

[131] C. Oseroff, *et al.*, "Analysis of T cell responses to the major allergens from German cockroach: epitope specificity and relationship to IgE production," Journal of Immunology, vol. 189, no. 2, pp. 679–688, 2012.

[132] A. M. Hasan , M. A. Khan, A. Datta, M.H. Mazumder, M.U. Hossain, "A comprehensive immunoinformatics and target site study revealed the corner-stone toward Chikungunya virus treatment," *Mol Immunol.*, vol. 65, no. 1, pp. 189–204, 2015.

[133] J. Janin, S. Wodak, M. Levitt, and B. Maigret, "Conformation of amino acid side-chains in proteins," J. Mol. Biol., vol. 125, no. 357-386, 1978.

List of Publications:

1. N. Zobayer, and A. B. M. A. Hossain, "*In silico* characterization and homology modeling of histamine receptors," Journal of Biological Sciences, vol. 18, pp. 178-191, 2018.
2. N. Zobayer, and A. B. M. A. Hossain, "B and T-cell epitopes based vaccine design in Api m3 allergen of *Apis mellifera*: An immunoinformatics approach," J. Med. Sci., vol. 18, no. 1, pp. 34-47, 2018.
3. N. Zobayer, and A. B. M. A. Hossain, "Algorithmic classifier based B cell epitope prediction. (Submitted in *Bioinformation*)