M.TECH. (BIOINFORMATICS) SAURABH BISWAS 2022

MUCORMYCOSIS VACCINE DESIGN USING BIOINFORMATIC TOOLS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
BIOINFORMATICS

Submitted by:

SAURABH BISWAS

Under the supervision of:
PROFESSOR YASHA HASIJA

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY,2022

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Saurabh Biswas, Roll No. 2K20/BIO/02 of M.Tech. (Bioinformatics), hereby declare that the project Dissertation titled "Mucormycosis Vaccine Design Using Bioinformatic Tools" which is submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi **Saurabh Biswas**

Date:

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled" Mucormycosis Vaccine Design Using Bioinformatic Tools" which is submitted by Saurabh Biswas, Roll No. 2K20/BIO/02, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirement of the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University of elsewhere.

Place: Delhi	**Professor Pravir Kumar**	**Professor Yasha Hasija**
Date:	Professor and HOD	SUPERVISOR
	Department of Biotechnology	Professor
	Delhi Technological University	Department of Biotechnology
		Delhi Technological University

ACKNOWLEDGEMENT

First and foremost, I would like to thank my research supervisor, Professor Yasha Hasija. Without her support, guidance, and patience throughout the process, this project would have never been accomplished. I would like to thank her very much for her encouragement and understanding.

I am thankful to Professor Pravir Kumar (Head of Department, Department of Biotechnology) for teaching me biotechnological concepts and supporting me with the research facilities without which the research work would not have been completed.

I would like to thank Mr. Raj Kumar Chakraborty for his help and assistance in completing the project.

I am thankful to my parents whose constant support and care kept me inspired and focused on the work. My achievements and accomplishments are because they believed in me. My sincere thanks to my brother, who kept me grounded, reminded me of what is important in life and was always supportive of my decisions.

Place: Delhi **Saurabh Biswas**

Date:

ABSTRACT

Mucormycosis or black fungus is an uncommon and serious disease initiated by a set of molds known as mucormycetes. The molds are present all over the environment. This disease generally occurs in immunocompromised individuals who ingest medicines to decrease their immune system's ability to fight against pathogens or individuals suffering from health issues. This disease usually affects lungs or sinuses after the spores have entered the body. A rise in muormycosis cases was seen with the rise of coronavirus infections in India. Hence, there is a vital requirement for a mucormycosis vaccine due to the presence covid-19 cases.

Vaccines are biological products that can be used to prevent or treat diseases. Vaccine development is a lengthy and expensive process. To decrease the cost and duration of vaccine development, bioinformatic methods can be used. In this project we have used these techniques to design and develop a mucormycosis vaccine.

Epitope screening using computational methods predicts that the CotH1, CotH2, CotH3, Fet3 and Ftr1 proteins can be utilized for constructing vaccines. Molecular docking tests predict that the epitopes (AADVISYRVSTAVWHVSWGD, ADVISYRVSTAVWHVSWGDP, DTQRKTVKDVYLTAAQRVSV, EWHLESGLAVVFVEAPDIAQ, SGASFNFPGDFDSFIPNVES, IWALPEGQFMANRNWFKIRH, SFLVAIAHAASVQFNLIAPS, IGSNYVHSFANVSFGIHGAG, FLVAITHAASIKFNVIAPNA, QNADKKATFEKYLTETVRVL) have low energy minimization score which show sufficient binding with the MHC class II molecules and give stability to the complex. More wet lab tests are further required to use of these peptides as a vaccine against Mucormycosis.

CONTENTS

CANDIDATE'S DECLARATION...i
CERTIFICATE ii
ACKNOWLEDGEMENT iii
ABSTRACT iv
CONTENTS v
LIST OF TABLES... vii
LIST OF FIGURES... viii
LIST OF SYMBOLS, ABBREVIATIONS x
Chapter 1 : INTRODUCTION.. 1
1.1 Symptoms of Mucormycosis 1
1.2 Individuals at risk... 2
1.3 Spread of mucormycosis and its prevention 3
1.4 Important fungal proteins causing mucormycosis 4
1.4.1 CotH proteins 4
1.4.2 Iron uptake proteins 5
1.5 Diagnosis and Treatment 6
Chapter 2 : LITERATURE REVIEW 8
2.1 Reverse Vaccinology.. 8
2.1.1 Filtering tools.. 9
2.1.2 Classifying tools 10
2.2 Immunoinformatics 11
2.2.1 Prediction of B-cell epitopes 12
2.2.2 Prediction of T-cell epitopes 15
2.3 Structural Vaccinology 17
Chapter 3 : METHODOLOGY 19
3.1 Retrieval of sequences and their analysis 19

3.2 Prediction of the epitopes 19
3.3 Population Coverage Estimation 20
3.4 Protein Retrieval and epitope design 21
3.5 Molecular docking study 21
Chapter 4 : RESULTS AND DISCUSSION 23
4.1 Physiological analysis of the proteins 23
4.2 Epitope predictions 23
4.3 Population Coverage Estimation 26
4.4 Analysis of the molecular docking and visualization 27
Chapter 5 : CONCLUSION AND FUTURE SCOPE 39
REFERENCES 40
List of Publications 43

LIST OF TABLES

Table No.	Title	Page no.
2.1	Comparative analysis of different reverse vaccinology tools	11
2.2	Comparative analysis of the B-cell prediction tools	14
2.3	Comparative analysis of the T-cell prediction tools	17
4.1	Physiological properties of the polypeptides	23
4.2	Selection of Epitopes for Ftr1 gene	24
4.3	Selection of Epitopes for Fet3 gene	24
4.4	Selection of epitopes for CotH1 gene	24
4.5	Selection of epitopes for the CotH2 gene	24
4.6	Selection of epitopes for the CotH3 gene	25
4.7	List of selected epitopes	25
4.8	The best binding energy scores for docking of the epitopes with DRB1*07:01	27

LIST OF FIGURES

Figure No.	Title	Page no.
1.1	Symptoms of mucormycosis according to the organ affected.	2
4.1	Population coverage of the epitopes worldwide	26
4.2	Population coverage of the epitopes in India	26
4.3	3D structure of DRB1*07:01(1AQD)	27
4.4	The interactions between the epitope (SGASFNFPGDFDSFIPNVES) and protein visualized in: (a) 2D form and (b) 3D form.	28
4.5	The interactions between the epitope (IWALPEGQFMANRNWFKIRH) and protein visualized in: (a) 2D and (b) 3D form.	29
1.6	The interactions between the epitope (SFLVAIAHAASVQFNLIAPS) and protein visualized in: (a) 2D and (b) 3D form.	30
4.7	The interactions between the epitope (IGSNYVHSFANVSFGIHGAG) and protein visualized in: (a) 2D and (b) 3D form.	31
4.8	The interactions between the epitope (FLVAITHAASIKFNVIAPNA) and protein visualized in: (a) 2D and (b) 3D form.	32
4.9	The interactions between the epitope (QNADKKATFEKYLTETVRVL) and protein visualized in: (a) 2D and (b) 3D form.	33

4.10	The interactions between the epitope (DTQRKTVKDVYLTAAQRVSV) and protein visualized in: (a) 2D and (b) 3D form.	34
4.11	The interactions between the epitope (EWHLESGLAVVFVEAPDIAQ) and protein visualized in: (a) 2D and (b) 3D form.	35
4.12	The interactions between the epitope (AADVISYRVSTAVWHVSWGD) and protein visualized in: (a) 2D and (b) 3D form.	36
4.13	The interactions between the epitope (ADVISYRVSTAVWHVSWGDP) and protein visualized in: (a) 2D and (b) 3D form.	37

LIST OF SYMBOLS, ABBREVIATIONS

ACC	Auto Cross Conversion
ANN	Artificial Neural Networks
APC	Antigen Presenting Cell
CotH	Spore Coat Protein Homolog
Covid-19	Coronavirus Disease 2019
CT	Computed Tomography
CTL	Cytotoxic T Lymphocyte
EBI	European Bioinformatics Institute
ER	Endoplasmic Reticulum
ExPASy	Expert Protein Analysis System
Fet3	Ferrous Transport 3
Fre	Ferric Reductase transmembrane
Ftr1	Fe Transporter 1
GRP78	Glucose-Regulated Protein 78
HLA	Human Leukocyte Antigen
HSP70	Heat Shock Protein 70
IEDB	Immune Epitope Database
MHC	Major Histocompatibility Complex
NBRF	National Biomedical Research Foundation
NERVE	New Enhanced Reverse Vaccinology Environment
PDB	Protein Data Bank
PIR	Protein Information Resource
RV	Reverse Vaccinology

SEPPA	Spatial Epitope Prediction of Protein Antigen
SIB	Swiss Institute of Bioinformatics
SVM	Support Vector Machines
Th cell	T Helper Cell
Treg cell	Regulatory T Cell
TrEMBL	Translated EMBL
UniProt	Universal Protein resource
Vaxign	Vaccine Design
WBC	White Blood Cell

Chapter 1: INTRODUCTION

Mucormycosis or black fungus is an uncommon and serious disease initiated by a set of molds known as mucormycetes (J.L. Hernández, et. al., 2021). The molds are present all over the environment. This disease generally occurs in immunocompromised individuals who ingest medicines to decrease their immune system's ability to fight against pathogens or individuals suffering from health issues. This disease usually affects the lungs or sinuses after the spores have entered the body.

A rise in mucormycosis cases was seen with the rise of coronavirus infections in India. To reduce the inflammation in the lungs due to a coronavirus infection, steroids are used. These steroids can decrease immunity and increase blood sugar in covid-19 patients. Hence, there is a vital requirement for a mucormycosis vaccine due to the presence covid-19 cases.

Vaccines are biological products that can be used to prevent or treat diseases. Vaccine development is a lengthy and expensive process. To decrease the cost and duration of vaccine development, bioinformatic methods can be used. Reverse vaccinology, immunoinformatics, and structural vaccinology techniques can be used to develop vaccines against various organisms. In this project, we have used these techniques to design and develop a mucormycosis vaccine.

1.1 Symptoms of Mucormycosis

The symptoms due to this infection vary based on the location where the infection occurs (G. Petrikkos, et. al., 2012). The fungus usually infects the brain, eyes, sinuses, and nose leading to tissue death, dislocation or swelling of the eye, blurry vision, fever, headache, pain, swelling on the face, and a runny nose (M.M. Roden, et. al., 2005).

Other variants of this disease could affect the skin, intestines, stomach, and lungs. Disseminated mucormycosis affects individuals suffering from different ailments, so it is hard to tell which symptoms may be associated with mucormycosis.

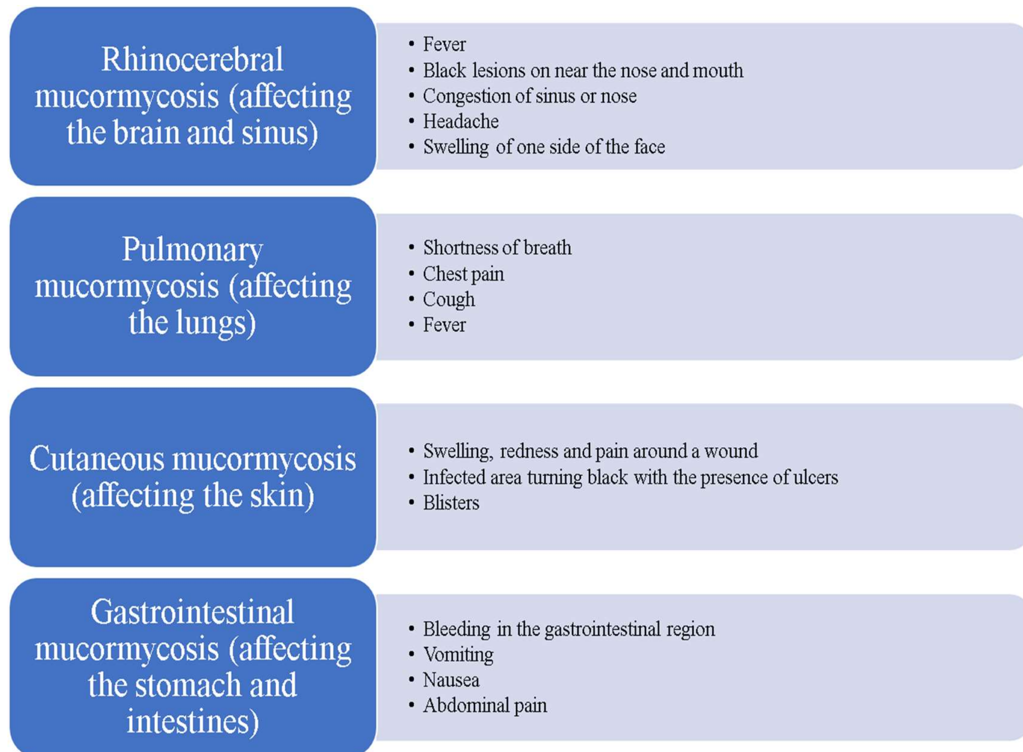

Figure 1.1 Symptoms of mucormycosis according to the organ affected.

1.2 Individuals at risk

Although mucormycosis is very uncommon, it is commonly found amongst immunocompromised individuals. The risk of this infection can increase due to (G. Reid, et. al., 2020):

- Low weight during birth and prematurity
- Injuries on the skin due to wounds, burns, or surgeries
- High quantities of iron inside the body
- Use of drugs

- Use of corticosteroids for extended periods
- Low quantity of WBCs (Neutropenia)
- Transplant of stem cells
- Transplant of organ
- Cancer
- Diabetes

1.3 Spread of mucormycosis and its prevention

The infection spreads through spores via inhalation, consumption of contaminated food, or the contamination of open wounds (B. Spellberg, et. al., 2005). The fungus is generally found in animal manure, decomposing food, and soil, but they generally don't affect people. The disease is not transmitted among people.

It is hard to not inhale the spores of fungi causing the infection as they are commonly found in the environment. No vaccine is currently available to prevent mucormycosis. For people having weak immunity, there are a few methods to decrease the probability of suffering from this disease:

- Antifungal medication: If an individual has a high risk of being infected with mucormycosis, then doctors can suggest medications to prevent the infection.
- Take protection from the environment:
    - Do not get perform activities where contact with dust or soil is required, such as gardening. In situations where this is not possible, cover yourself with gloves and full-sleeved tops, and clean injured skin with soap and water after contact with dust and soil.
    - Stay away from buildings damaged by hurricanes and floods.
    - Stay away from areas containing lots of dust such as construction sites. If these locations cannot be avoided, then wear face masks at the location.

1.4 Important fungal proteins causing mucormycosis

Many different species of fungi can cause mucormycosis infection. The fungi are of the order of Mucorales and are known as mucormycetes. Mucor and Rhizopus are well-known types of fungi that can cause infection.

Proteins associated with mucormycosis pathogenesis used in this project are: CotH proteins(CotH1,CotH2,CotH3) and Iron uptake proteins(Fet3,Ftr1). The properties of these proteins are described below:

1.4.1 CotH proteins

These are protein kinases that are parts of the spores of various pathogens. Studies have shown that these polypeptides are not found in the spores of non-invasive microorganisms, but are a common property of Mucorales strains causing mucormycosis.

The CotH gene copy number is linked to the pathogenicity of a mucormycosis-causing species. Therefore, the common mucormycosis-causing pathogens from the Rhizopus genus contain 6-7 copies, less commonly infecting genera have 3-7 copies, and rarely infecting genera have 1-2 copies. The main causal agent of mucormycosis is from the Rhizopus genus. *Rhizopus delemar* contains 8 CotH polypeptides with 6 sequences present on the fungal coat.

The CotH contains a motif that is essential for interactions among human endothelial cells and the spores, important for the invasion of the fungi. The expression of CotH cells during the host cell and *R.delemar* spores shows their contribution to the infection. Therefore, CotH polypeptides having the ability to harm and attach to host cells, are conserved in Mucorales species. The tissue invasion by the pathogens due to the CotH polypeptides depends upon the interactions between glucose-regulated protein

78(GRP78) and the CotH proteins (CotH3, CotH2, and CotH1) on the epithelial cell surface. The CotH protein having the highest binding affinity with GRP78 is CotH3, next is CotH2.

The heat-shock protein, GRP78 belonging to the HSP70 family, is present inside the ER of the human endothelial cells. During a stressful condition, GRP78 cells are overexpressed and are released into the cell surface, where they interact with CotH proteins.

Through in-vivo models, the role of GRP78 and CotH in mucormycosis pathogenesis was studied. Diabetic ketoacidosis led to greater disease susceptibility in mice under study. GRP78 overexpression due to the glucose generated, and higher concentrations of iron. A decrease in CotH expression decreases the disease severity, even if the GRP78 expression is higher than normal. Antifungal and anti-CotH3 antibody treatment in mice helped in achieving 100 percent mice survival against different species of Mucorales. Additionally, through PCR using CotH genes, mucormycosis infection can be detected during the initial phases of infection.

1.4.2 Iron uptake proteins

The mechanism for iron uptake in Mucorales has been studied in detail. Three proteins associated with iron uptake are Ftr1(Iron permease), Fet3(ferroxidase), and Fre (iron reductase). For the accurate functioning of iron uptake, ferroxidase and permease are essential proteins. Low iron concentration leads to the expression of the iron uptake system in *Lichtheimia corymbifera*, *M.circinelloides*, and *Rhizopus delemar*. During *M.circinelloides* infection in the lungs, ferroxidase genes were found to be overexpressed. The removal of these genes leads to a decrease in the infection, depicting that they are an essential virulence factor in the infection. The removal of *R.delemar*'s ftr1 gene was found to produce growth defects when there was a low availability of iron,

and also decreased the virulence of disease in the mouse. Additionally, the essential role of the Ftr1 gene was confirmed when iron starvation in *R. delemar* lead to an apoptotic process dependent on metacaspase. The observations show the importance of permease protein for iron uptake during infection and survival of the fungi in low iron availability conditions.

1.5 Diagnosis and Treatment

For diagnosing mucormycosis, clinicians can go through a person's test reports, physical examinations, symptoms, and medical history. Doctors predicting sinus or lung infections may take fluid samples from the respiratory tract and send them into labs for analysis.

Diagnosis methods include culture, biopsy and imaging can assist in assessing the severity of the disease (M.E. Grossman, et. al., 2011). In a biopsy, some samples of tissues are checked for evidence of infection in a culture or through a microscope. Imaging tests like CT scans of the suspected location of infection may also be needed. A decisive diagnosis for the infection needs a positive result from a culture or histopathological indication. Samples taken from sterile locations in the body give a strong indication of an invading infection as compared to colonization. Non-sterile location culture can be useful in infected patients that are clinically similar to the fungal infection.

Prognosis and correct treatment are essential for decreasing the effect of the infection. Mucormycosis is a severe infection requiring treatment with the prescribed antifungal drugs, such as isavuconazole, posaconazole, or amphotericin B. These compounds can be given through oral route (e.g. Posaconazole, isavuconazole) or injected through a vein (e.g. amphotericin B, isavuconazole). The first line of treatment for the infection involves the use of amphotericin B lipid formulations. Usually, during this infection,

surgery is required to remove the infected tissue. The infection progresses quickly and is lethal in around 50% of sinus cases and nearly all the widespread cases.

The first mucormycosis case took place in 1855 and was explained by Friedrich Küchenmeister. This disease has been seen to occur during natural disasters such as the Missouri tornado in 2011 and the Indian Ocean tsunami in 2004 (E. Dannaoui, et. al., 2019). An association between mucormycosis and coronavirus was found during the COVID-19 pandemic. This link is found to be due to the decrease in immunity during the disease and the use of glucocorticoid therapy for treatment(D. Garg, et. al., 2021). A rise in mucormycosis cases has been particularly observed in India(A.K. Singh, et. al., 2021).

For the prevention or treatment of diseases, vaccines are used. It is a very expensive and time-consuming process to develop vaccines. Bioinformatics is a method that can overcome these challenges and assist in developing vaccines. Structural vaccinology, reverse vaccinology, and immunoinformatics is some techniques that can help in developing vaccines against different organisms(K.L. Seib, et. al., 2012). In this project, we will utilize these bioinformatic tools for designing and developing a mucormycosis vaccine.

Chapter 2: LITERATURE REVIEW

The Reverse Vaccinology, Immunoinformatic, and Structural Vaccinology techniques used in this project are explained in this section.

2.1 Reverse Vaccinology

Reverse vaccinology is the process of discovering antigens using genomic data. This method identifies antigens by computationally scanning the complete set of antigens encoded in the form of genomic information. Reverse vaccinology doesn't require the culturing of microorganisms, making it easier to analyze organisms that are hard to culture. The reverse vaccinology process involves in silico comparison of different genomes to detect the conserved antigens in a heterogeneous population and also identify the antigens unique to the pathogen. The time and cost of downstream analysis are reduced by this technique as proteomic and transcriptomic data can additionally be included in the selection.

Reverse vaccinology has led to an evolution in vaccine development by selecting a collection of new vaccine candidates through the scanning of protein sequences in the microorganism. Reverse vaccinology has been utilized efficiently to generate vaccines against different microorganisms. Important properties for the selection of good vaccine candidates include similarity, antigenicity, properties of adhesins, and cellular localization.

According to the algorithmic technique used, reverse vaccinology can be either classified as classifying or filtering tools.

2.1.1 Filtering tools

These tools select the vaccine candidates by screening the protein sequences using a set of filters until a set of sequences are selected as vaccine candidates. Some filters used include the measurement of molecular weight, subcellular localization, and the probability of being an adhesion protein. When numerical properties are passed through the filters, an apriori cutoff can be used. These tools vary depending upon the number of filters used. Example: Vaxign (Y. He, et. al., 2010), NERVE (S. Vivona, et. al., 2006).

2.1.1.1 Vaxign

Vaxign predicts vaccine candidates based on a set of criteria through the use of peptide sequences and genomic data as input. Some of the features predicted by this tool are binding of the epitope to MHC, adhesion, and antigen sublocation. This tool can be used through a web interface. The two types of Vaxign are Vaxign Query and Dynamic Vaxign. Dynamic implementation and visualization of the output are possible through Dynamic Vaxign, whereas precomputed results are obtained through Vaxign Query.

2.1.1.2 NERVE

NERVE is a tool for the identification of vaccine selection. The tool combines many reliable and popular algorithms for the comparison and analysis of polypeptides. This tool takes the protein sequences of microorganisms as an input and predicts different properties. These properties include sequence conservation in various strains of the same species, sequence similarity with human proteins, protein topology, and subcellular localization. The output is given as a table containing ranked vaccine candidates.

2.1.2 Classifying tools

These tools classify the input protein sequences based on their probability of being a vaccine candidate. A combination of known non-vaccine and vaccine candidates is utilized as a training set. Then the sequence is classified as either a non-vaccine or a vaccine candidate. These tools include Heinson-Bowman (A.I. Heinson, et. al., 2017) and VaxiJen (I.A. Doytchinova, et. al., 2007).

2.1.2.1 Heinson-Bowman

This tool uses the properties of both NERVE and Vaxijen, by using the protein annotations from NERVE and the classification algorithm from Vaxijen. This tool was trained using the SVM algorithm by classifying a collection of 200 protective antigens in bacteria. Ten different properties are used by this tool for classification.

2.1.2.2 VaxiJen

Vaxijen was the first tool developed for predicting protective antigens without the alignment of sequences. It permits antigen classification based on their physicochemical features. This tool uses autocross conversion (ACC) for converting protein sequences into vectors of the same size. To predict antigens in bacteria, ACC conversion is used on a set of bacterial antigens, which are collected through literature. The results are given in the form of targets, polypeptide sequences, probability of the prediction, and classification as either an antigen or a non-antigen (based on a set cutoff).

Table 2.1 Comparative analysis of different reverse vaccinology tools

Tool	Type	Description	Sensitivity
Vaxign	Filtering	Uses a web interface. Has predicted protein features Similar to NERVE.	58%
NERVE	Filtering	Different properties can be predicted using an input pathogenic sequence.	64%
Heinson-Bowman	Classifying	Properties of both NERVE and Vaxijen is used by this tool.	75%
VaxiJen	Classifying	Proteins are converted in vectors of identical length using ACC to make predictions.	76%

2.2 Immunoinformatics

Large quantities of data related to immunological studies have come up as a result of genome sequencing of different organisms. Large quantities of epidemiological and clinical data have also been gathered in medical records and literature. This increase in information provides researchers with an excellent opportunity to understand the pathogenesis of different diseases and the functioning of the immune system. Therefore, this requirement of utilizing the quickly increasing immunological information has led to the creation of a field called immunoinformatics. This field is a link between computer science and immunology. The main aim of this field is to produce vaccines using various features such as the incidence of infectious diseases, antigenic variations, and genetic variations.

Immune system activation occurs through the induction of immune memory and its strength determines the effectiveness of the vaccine. Hence, vaccine efficiency can be obtained through factors stimulating immune memory, antibodies, and memory cell types that are stimulated. The parts of the immune system activated by vaccines are antibodies and T-cells. Vaccines stimulating both the T and B-cell responses have higher efficiency. Although major immune effectors are the B-cells, antibodies and memory cells can be induced by T-cells with a high affinity. The identification of these targets has led to improvements in vaccine design, as the classical methods give low vaccine immunogenicity and are not precise enough for microorganisms that can mutate quickly or have a complex life cycle. Hence, immunoinformatics assists in identifying vaccine candidates for microorganisms and disorders having varying antigens and different modes of treatment.

The epitope is a portion of the antigen that attaches to the antibody. Epitopes can be detected by an immune system, causing an immune response. Identifying these epitopes experimentally is difficult and takes a long time. Hence, bioinformatic tools can be used for detecting these epitopes. The different techniques for identifying these epitopes are:

2.2.1 Prediction of B-cell epitopes

Antibodies are proteins created by B-cells. Receptors of these B-cells can interact with epitopes located on the pathogen surface. After B-cell activation, they differentiate and secrete immunoglobulins in the form of antibodies, which facilitate humoral immunity. The released antibodies can perform different functions like pathogens and toxins neutralization, and tagging so that they are destroyed. A B-cell epitope is the part of the antigen which can bind to an antibody. The epitopes are of two types, based on their structures: continuous epitopes and discontinuous epitopes. Continuous B-cell epitopes have sequential residues, while discontinuous epitopes contain regions that are exposed to the solvent but may not be sequential. Antibodies that can identify continuous

epitopes, can also identify denatured antigens, whereas discontinuous epitopes cannot be identified after denaturation. For predicting the epitopes, sequence and structure-based tools can be utilized. Some of them are described below:

2.2.1.1 SEPPA

This tool was created in 2009 and uses single residue propensity scales and continuous fragment clustering. This tool can predict spatial epitopes of antigens (J. Sun, et. al., 2009). The tool can explain the spatial information on the antigen surface. The 3D properties of epitopes displayed include the spatial clustering coefficient. Additionally, information about the epitopes is also taken from the Protein Data Bank database.

2.2.1.2 DiscoTope

This tool predicts discontinuous epitopes using amino acid data, spatial properties, and exposure of the surface obtained by X-ray crystallography studies (P. H. Andersen, et. al., 2006). This tool was the first to solely focus on discontinuous epitopes. Discotope predictions can give a higher efficacy in epitope predictions and can assist in mapping epitopes for developing diagnostic tools and designing vaccines.

2.2.1.3 ABCPred

The continuous epitopes in an antigen can be obtained using this tool, which is based on ANN (W. Shen, et. al., 2015). This tool helps in identifying the epitopes that can be useful in the diagnosis of diseases, allergy studies, and the selection of vaccine candidates. The results are displayed in the form of a table containing the predicted

polypeptides. The tool uses a database containing information about viruses, bacteria, parasites, and fungi.

2.2.1.4 BCPred

This tool classifies antigens by training an SVM classifier using linear B-cell epitope data (Y. El-Manzalawy, et. al., 2008). As we can compare predictions through different techniques, accuracy is increased. This server provides the operator with a choice to select the technique to be utilized for predictions. Users can choose either Amino Acids Pair Scaling, BCPred, or FBCPred technique.

Table 2.2 Comparative analysis of the B-cell prediction tools

Tool	Type	Description	Sensitivity
SEPPA	Discontinuous	Uses single residue propensity scales and continual segment clustering.	28.9%
DISCOTOPE	Discontinuous	Uses amino acid statistics, spatial knowledge, and surface exposure for the predictions.	93%
ABCPRED	Continuous	ANN is used to predict continuous B cell epitopes.	50.7%
BCPred	Continuous	To predict the epitopes, a kernel-based SVM classifier is used.	99%

2.2.2 Prediction of T-cell epitopes

Antigens have to be bound with MHCs for T-cells to detect these antigens. Hence, an important role is played by the detection of MHC-peptides binding for predicting the T-cell epitopes. T-cell receptors can identify antigens when they are presented to them by APC (Antigen Presenting Cell) bound to MHCs. The antigen's presentation to T-cells takes place with the help of MHC molecules, which can recognize CD4 and CD8 T-cells respectively. CD4 T-cells are converted into Treg(regulatory) or Th(helper) T-cells after recognizing the epitope. The T helper cells can enhance the immune response. Whereas, the CD8 T-cells are converted to CTL cells after recognizing the epitope. For predicting the epitopes, sequence and structure-based tools can be utilized. Some of them are described below:

2.2.2.1 NetMHCIIpan

The tool uses ANN to predict the MHC class II binding peptide molecules. For training, binding affinity data and eluted ligand mass spectrometry were used (K.K. Jensen, et. al., 2018). The output from this model is in the form of a score depicting the peptide presentation probablity by a class II MHC receptor. The % rank score is another output that is given, which can compare the predictions to a collection of random proteins for normalizing the score. This server can predict the sequences of different lengths.

2.2.2.2 NetMHCII2

This server can predict peptide binding with class II MHC of humans and mice through the use of ANN (K.K. Jensen, et. al., 2018). The result is obtained in the form of rank and IC_{50} values are displayed. Indications of weakly and strongly binding peptides are also given as an output.

2.2.2.3 NetMHCpan-4.0

This is a technique trained using both the eluted ligand and binding affinity data (V. Jurtz, et. al., 2017). Benchmarks of these tools show better prediction efficiency in detecting T-cell epitopes, neoantigens, and processed ligands, as compared to other tools.

2.2.2.4 NetMHC4

This tool uses ANN for classification, which is trained using human MHC alleles. Predictions for other animals such as monkeys, cattle, pigs, and mice can also be obtained through this tool (W. Zhao, et. al., 2018).The hidden layers in ANN can stimulate peptide binding with MHC. Additional properties such as the weight of interactions among the residues are present on the nodes.

Table 2.3 Comparative analysis of the T-cell prediction tools

Tool	Type	Description	Sensitivity
NetMHCIIpan	Class II MHC	MHCII – peptide binding is predicted using ANN	32.6%
NetMHCII2	Class II MHC	Predicts binding of epitopes with H2 class II alleles, HLA-DR, HLA-DQ, and HLA-DP	57.7%
NetMHCpan4	Class I MHC	Identifies naturally modified ligands, cancer neoantigens , and epitopes.	83.6%
NetMHC4	Class I MHC	Epitopes for MHC class I molecules can be predicted .	83.6%

2.3 Structural Vaccinology

Structural vaccinology uses the structural features of compounds to determine the best vaccine candidates. The 3D structures of pathogens can give important information about the location of epitopes and their structure. Structural vaccinology can create efficient vaccines in the following ways: (1) In-vivo test for the efficiency and safety of vaccine candidates, (2) Integration of the recreated epitope or antigen into a vaccine platform, (3) Using reverse molecular engineering to remodel the epitope or antigen, (4) Obtaining the structure of the antigen-antibody complex or the antigen. The main use of structural vaccinology is to select and design antigens or vaccines generating immunity against different pathogens.

Structural vaccinology is also very helpful in selecting the epitopes on the surface placed near functional positions such as conformational epitopes, catalytic or binding sites. The structural information can be utilized to identify conformational features such as the properties influencing immunogenicity, peptide exposure to the solvent, and the determination of the antigenic locations in various pathogens. At different sequence and structure levels, structural vaccinology can be used. Example- Vaccine candidates against *S.pneumoniae* were determined through epitope predictions and structural modeling.Vaccine failures have been found to occur due to non-recognition by receptors and antibodies caused by protein flexibility.

The protein sequences or 3D structures can be used to predict protein flexibility. Molecular docking techniques can be utilized to select and design the desirable antigens. This method can produce a two-molecule complex (comprising of a protein bound to a protein or a ligand) having the lowest binding energy. To predict the interactions between antibodies or MHC receptors with epitopes, molecular docking techniques can be used. The scoring function obtained by molecular docking can be used to screen probable drug molecules that bind with the required protein. AutoDock Vina is a virtual screening and molecular docking tool (O. Trott, et. al., 2010). AutoDock Vina is quicker and gives a higher accuracy for binding predictions when compared to AutoDock 4.

Chapter 3: METHODOLOGY

3.1 Retrieval of sequences and their analysis

Protein sequences found in mucormycosis causing fungi: Ftr1(I1BRD6), Fet3(A0A068RI08), CotH1(I1BVT3), CotH2(I1C4E4), CotH3(I1CFE1) were taken from the UniProt server. UniProt is an open-access database of polypeptide sequence and functional data, with various entries taken from sequencing projects. It holds a lot of data about the functions of literature-derived polypeptides. The database is maintained by a collection of companies known as the UniProt consortium.

The ProtParam tool was used to identify the molecular weights, isoelectric points, and the length of the sequences. The ProtParam tool is used to calculate different physicochemical features that can be obtained from only the sequence of a polypeptide. The polypeptide may be stored in TrEMBL or Swiss-Prot or it may be a user-entered sequence.

3.2 Prediction of the epitopes

The Ftr1, Fet3, CotH1, CotH2, and CotH3 fungal peptides were examined using BCPred and ABCPred servers to determine the B-cell epitopes. These epitopes were further screened to select the epitopes that can also act as T-cell epitopes.

T-cells having HLA DRB1(A class II MHC receptor) are known to have a contribution to the antifungal mechanism of the body (G. Castellano-González, et. al., 2020). Therefore, the NetMHCII2 tool was used to select the T-cell epitopes. One of the DRB1

alleles present in Indians with the most frequency is the DRBI*07 allele (N.K. Mehra, et. al., 1997). The epitopes binding to this allele were selected.

To determine whether the epitopes are antigenic, the VaxiJen tool was used. Allertop tool was then used to determine whether the peptides were allergenic (I. Dimitrov, et. al., 2013). AllerTop tool represents the properties of amino acids by five descriptors which describe the beta-strand forming tendency, relative amino acid abundance, helix-forming tendency, molecular size, and hydrophobicity of amino acids. The polypeptides are then classified using the KNN algorithm using a large training set of non-allergens and allergens.

To prevent autoimmune reactions, the selected epitopes were used as queries to look for similar protein sequences in SwissProt and UniProtKB servers. Finally, the non-allergenic and antigenic peptides, which do not lead to autoimmune responses were selected.

3.3 Population Coverage Estimation

The utilization of multiple epitopes can lead to coverage by the vaccine. The population coverage worldwide was detected using the IEDB population coverage tool. The population coverage was determined using a set of HLA alleles.

T-cells can identify a complex of an epitope derived from a pathogen and a specific MHC (Major Histocompatibility Complex) compound. A specific epitope will only induce a response in a person expressing an MHC molecule that can bind with a certain epitope. This process is called the MHC restriction. Thousands of different MHC(HLA) alleles can be found in humans as these molecules are very polymorphic. The selection of a variety of peptides having different binding affinities could increase the patient population coverage when peptide vaccines are applied. Calculating the population coverage is very important as different ethnic groups possess different frequencies of

the HLA alleles. Therefore, a careful examination of an ethnically biased region is required to obtain the desired result.

To calculate the population coverage, the IEDB population coverage tool can be used. This tool predicts the portion of people that would respond to a set of epitopes using MHC binding data and genotypic frequencies of HLA. The Allele Frequency database is used to collect the HLA allele genotypic frequencies. This helps the Population Coverage tool in calculating several population coverages parallelly and giving an average coverage as the output.

3.4 Protein Retrieval and epitope design

The structure of DRB1*07:01(1AQD) was collected from the PDB server. DISTILL server was used to design the epitope structures using their similarity to the PDB structures (D. Baú, et. al., 2006). DISTILL is a set of open-access servers that can be used to predict the structural properties of proteins. Multiple query submission is possible from a single interface that can access the servers.

3.5 Molecular docking study

AutoDock Vina tool is used for the molecular docking of the selected epitopes. This tool can produce grid maps and display the results in an easily understandable way. To predict the binding affinities, Negative Gibbs free energy values are utilized. The position of the epitopes with the best binding affinities was selected. Before docking the protein and ligands, water compounds were removed and the energy of protein was minimized to include the hydrogen atoms in the structure.

Finally, the interactions between the protein and epitopes were visualized in a 3D form using Discovery Studio BIOVIA and in a 3D form using LigPlot. Discovery Studio BIOVIA is a comprehensive tool to discover the use of molecules in therapeutics and investigate the chemistry of proteins. LigPlot is a tool that can automatically generate a 2D diagram of the interactions between ligands and proteins.

Chapter 4: RESULTS AND DISCUSSION

4.1 Physiological analysis of the proteins

Table 4.1 displays the physiological properties of the polypeptides. The isoelectric point or pI value is the pH at which the polypeptide has zero charges. The length and molecular weights of the polypeptides are also displayed in the table.

Table 4.1 Physiological properties of the polypeptides

Name	Accession	Length	Molecular Weight	Isoelectric point
CotH1	I1BVT3	609	68759.4	5.29
CotH2	I1C4E4	594	65254.26	4.94
CotH3	I1CFE1	601	65759.18	5.07
Fet3	A0A367JLF7	580	64323.32	4.73
Ftr1	I1BRD6	368	41124.67	5.25

4.2 Epitope predictions

The epitopes that can stimulate both the B and T-cell immunities are known as great vaccine candidates. ABCPred and BCPred servers were utilized to select epitopes from the five polypeptides. These epitopes were used to select the final peptides which can also act as T-cell epitopes through the NetMHCII2 and NetMHCIIpan servers. Two of the highest-scoring epitopes for each polypeptide were selected. The final selected epitopes are antigenic, and non-allergic, and do not cause autoimmune disorders. These antigens are marked (highlighted in green) in the tables below:

Table 4.2 Selection of Epitopes for Ftr1 gene

Sequence	Antigenicity	Allergenicity
NAILGWNNTATYGSIISYCLYW	Non-antigen	Allergen
GGEAADVISYRVSTAVWHVS	Antigen	Allergen
AADVISYRVSTAVWHVSWGD	Antigen	Non-allergen
ADVISYRVSTAVWHVSWGDP	Antigen	Non-allergen
VFIGGVSLGIQGKSIPIAAI	Antigen	Allergen

Table 4.3 Selection of Epitopes for Fet3 gene

Sequence	Antigenicity	Allergenicity
DTQRKTVKDVYLTAAQRVSV	Antigen	Non-allergen
ARVEYWWNITYTTANPDGLF	Non-antigen	Non-allergen
AGMDFTYEFNITQHGSYWLH	Antigen	Allergen
EWHLESGLAVVFVEAPDIAQ	Antigen	Non-allergen
RFRADNPGAWFFHCHIEWHL	Antigen	Allergen

Table 4.4 Selection of epitopes for CotH1 gene

Sequence	Antigenicity	Allergenicity
ILRKMGTYANEANMVRFFIN	Antigen	Allergen
SGASFNFPGDFDSFIPNVES	Antigen	Non-allergen
IWALPEGQFMANRNWFKIRH	Antigen	Non-allergen
KTKITFIGPETINTFEGCTL	Non-antigen	Non-allergen
EQVKTIGKYFDYDQFLRFMV	Antigen	Allergen

Table 4.5 Selection of epitopes for the CotH2 gene

Sequence	Antigenicity	Allergenicity
SFLVAIAHAASVQFNLIAPS	Antigen	Non-allergen
IGSNYVHSFANVSFGIHGAG	Antigen	Non-allergen
QNADKKAKFEQYLTETVRVL	Non-antigen	Non-allergen
FEQYLTETVRVLFNNVTLTN	Non-antigen	Non-allergen
VRVLFNNVTLTNRVLAIHNF	Non-antigen	Allergen

Table 4.6 Selection of epitopes for the CotH3 gene

Sequence	Antigenicity	Allergenicity
FLVAITHAASIKFNVIAPNA	Antigen	Non-allergen
QNADKKATFEKYLTETVRVL	Antigen	Non-allergen
KYLTETVRVLFNNVTLTNRV	Non-antigen	Non-allergen
QSTSGASRSKTAPIVLAISA	Antigen	Allergen
VLFNNVTLTNRVLALHNFLL	Non-antigen	Allergen

The epitopes: AADVISYRVSTAVWHVSWGD, ADVISYRVSTAVWHVSWGDP, DTQRKTVKDVYLTAAQRVSV, EWHLESGLAVVFVEAPDIAQ, SGASFNFPGDFDSFIPNVES, IWALPEGQFMANRNWFKIRH, SFLVAIAHAASVQFNLIAPS, IGSNYVHSFANVSFGIHGAG, FLVAITHAASIKFNVIAPNA, QNADKKATFEKYLTETVRVL were selected.

Table 4.7 List of selected epitopes

Epitope	Protein
AADVISYRVSTAVWHVSWGD	FTR1
ADVISYRVSTAVWHVSWGDP	FTR1
DTQRKTVKDVYLTAAQRVSV	Fet3
EWHLESGLAVVFVEAPDIAQ	Fet3
SGASFNFPGDFDSFIPNVES	CotH1
IWALPEGQFMANRNWFKIRH	CotH1
SFLVAIAHAASVQFNLIAPS	CotH2
IGSNYVHSFANVSFGIHGAG	CotH2
FLVAITHAASIKFNVIAPNA	CotH3
QNADKKATFEKYLTETVRVL	CotH3

4.3 Population Coverage Estimation

An estimated population coverage of 94.88% was calculated for people living worldwide using a set of most frequently occurring epitopes. A population coverage of 93.04% was found for people living in India. These estimations were obtained through the IEDB population coverage prediction tool.

population/area	Class II		
	coverage[a]	average_hit[b]	pc90[c]
World	94.88%	17.35	11.52
Average	94.88	17.35	11.52
Standard deviation	0.0	0.0	0.0

[a] projected population coverage
[b] average number of epitope hits / HLA combinations recognized by the population
[c] minimum number of epitope hits / HLA combinations recognized by 90% of the population

Figure 4.1 Population coverage of the epitopes worldwide

population/area	Class II		
	coverage[a]	average_hit[b]	pc90[c]
India	93.04%	15.84	10.8
Average	93.04	15.84	10.8
Standard deviation	0.0	0.0	0.0

[a] projected population coverage
[b] average number of epitope hits / HLA combinations recognized by the population
[c] minimum number of epitope hits / HLA combinations recognized by 90% of the population

Figure 4.2 Population coverage of the epitopes in India

4.4 Analysis of the molecular docking and visualization

Molecular docking is an essential part of the drug modeling process. Docking is a new technique through which the ligands can bind to the active site or pockets in the receptor molecule. Ten epitope sequences were successfully docked using the Auto dock vina tool. Strong interactions were seen among the epitope(ligand) and protein molecules. The epitopes displayed binding energy in the range of -7.2 to -13.5 kcal/ mol. The highest binding affinity was seen between the ftr1 epitopes and the protein. The number of H-bonds and protein-ligand binding was then visualized. IWALPEGQFMANRNWFKIRH had the highest number of H-bonds with the protein(4).

Table 4.8 The best binding energy scores for docking of the epitopes with DRB1*07:01

Compound	**Binding Energy (kJ mol^{-1})**
SGASFNFPGDFDSFIPNVES	-8.8
IWALPEGQFMANRNWFKIRH	-10.7
SFLVAIAHAASVQFNLIAPS	-7.8
IGSNYVHSFANVSFGIHGAG	-7.2
FLVAITHAASIKFNVIAPNA	-7.9
QNADKKATFEKYLTETVRVL	-9.6
DTQRKTVKDVYLTAAQRVSV	-11.9
EWHLESGLAVVFVEAPDIAQ	-11.7
AADVISYRVSTAVWHVSWGD	-12.8
ADVISYRVSTAVWHVSWGDP	-13.5

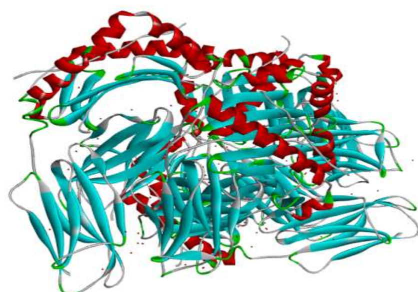

Figure 4.3 3D structure of DRB1*07:01(1AQD)

Visualization of epitope docking with DRB1*07:01:

i) SGASFNFPGDFDSFIPNVES

The epitope has hydrophobic interactions with 25 amino acids of the protein (GLU 134, VAL 42, GLU 46, MET 36, VAL 132, ASP 152, GLU 21, LYS 38, HIS 33, ASP 35, VAL 104, GLY 131, TYR 150, GLY 20, ALA 37, ARG 44, LEU 45, VAL 117, VAL 136, PHE 137, LYS 147, PRO 152, SER 19, TRP 43, THR 129, SER 133). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.4.

a) b)

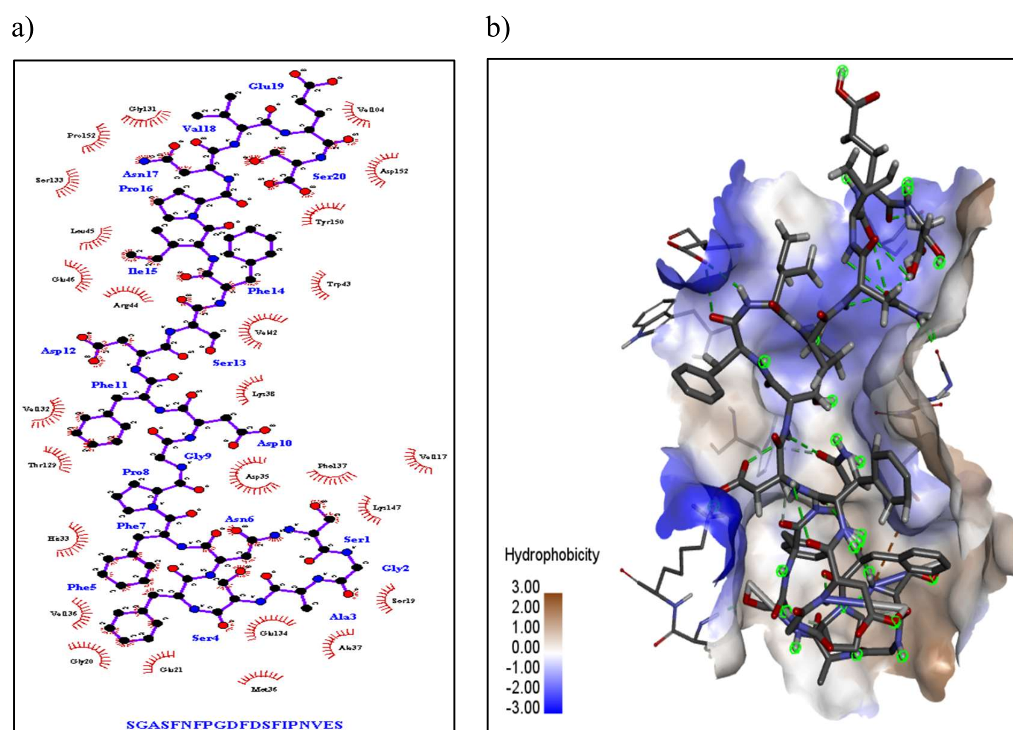

Figure 4.4 The interactions between the epitope (SGASFNFPGDFDSFIPNVES) and protein visualized in: (a) 2D form and (b) 3D form.

ii) IWALPEGQFMANRNWFKIRH

The epitope has hydrophobic interactions with 17 amino acids of the protein (THR 90, ASP 162, VAL 97, TYR 102, SER 104, MET 160, GLU 88, VAL 91, PRO 96, TRP

178, LEU 158, VAL 116, THR 145, VAL 89, SER 95, PHE 180, ASP 181, GLN 156) and forms hydrogen bonds with 4 amino acids of the protein (HIS 177, LYS 176, THR 93, GLU 179). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.5.

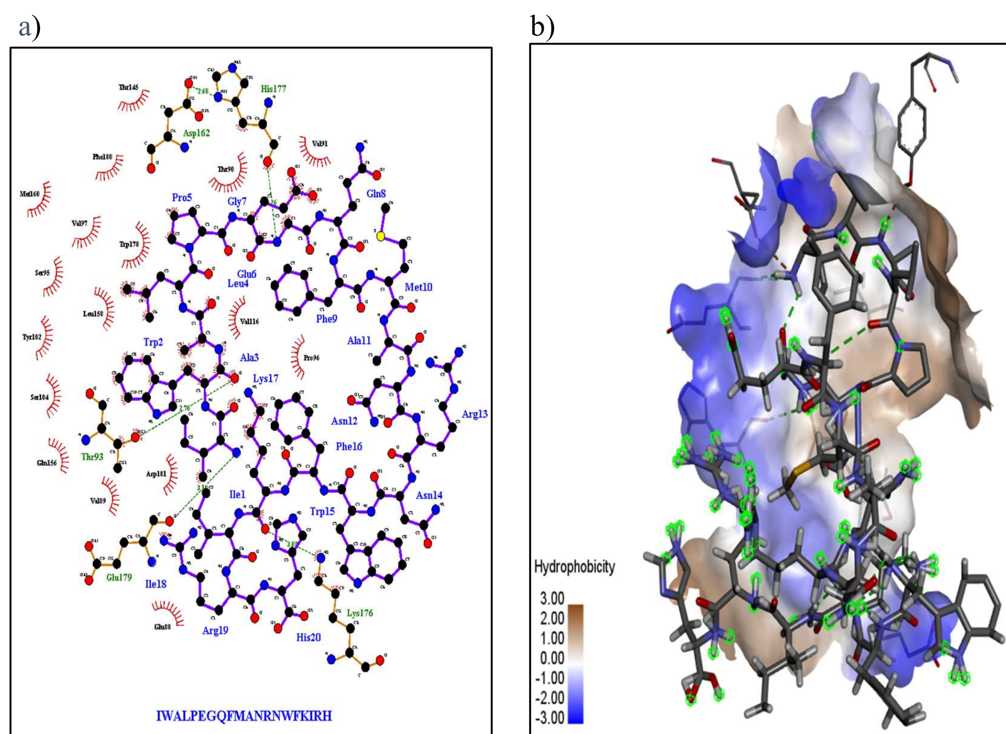

Figure 4.5 The interactions between the epitope (IWALPEGQFMANRNWFKIRH) and protein visualized in: (a) 2D and (b) 3D form.

iii) SFLVAIAHAASVQFNLIAPS

The epitope has hydrophobic interactions with 30 amino acids of the protein(ASP 27,LEU 92,GLU 141,TRP 178,GLU 179,ILE 148,VAL 6,VAL 97,PHE 108,ARG 140,GLN 149,PRO 96,ARG 146,LEU 147,LEU 158,GLU 3,VAL 91,THR 93,ASP 181,LYS 176,THR 145,PHE 155,MET 160,GLU 4,ILE 8,PRO 139,HIS 177,PHE 180,VAL 116,SER 118,GLN 156) and forms hydrogen bonds with 2 amino acids of the protein(TYR 102,THR 90).The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.6.

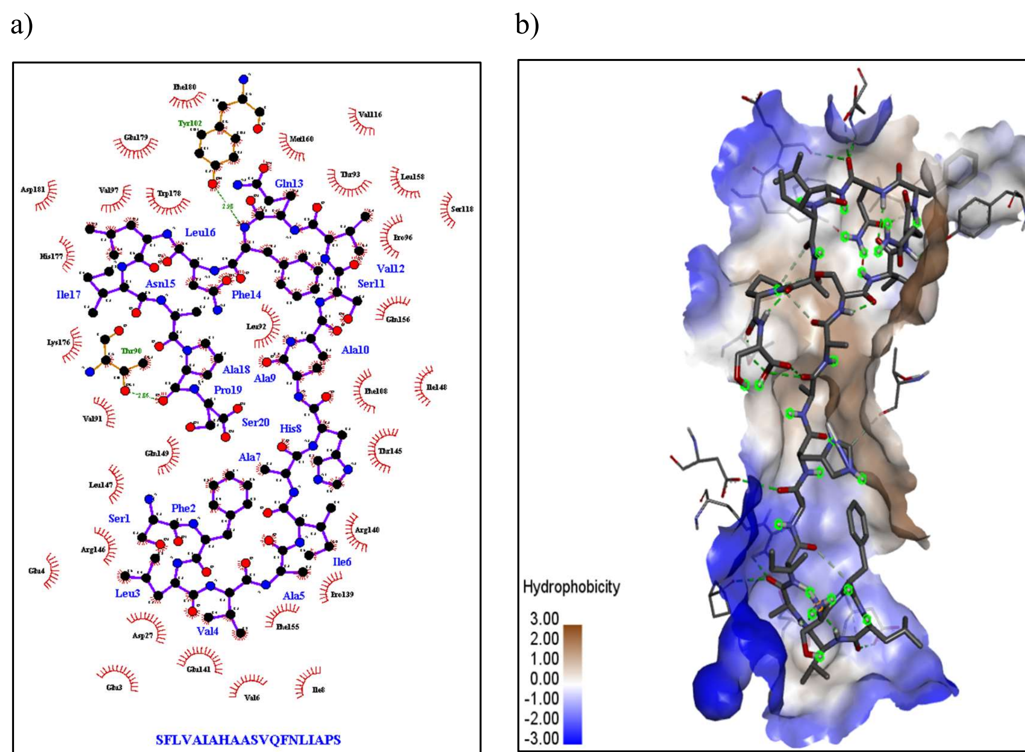

Figure 2.6 The interactions between the epitope (SFLVAIAHAASVQFNLIAPS) and protein visualized in: (a) 2D and (b)3D form.

iv) IGSNYVHSFANVSFGIHGAG

The epitope has hydrophobic interactions with 25 amino acids of the protein (PHE 108, ILE 148, HIS 177, ARG 140, THR 145, ASP 27, THR 90, LEU 92, GLY 146, LEU 147, LEU 158, GLU 141, LYS 176, ASP 25, PRO 139, PHE 155, ILE 8, THR 93, SER 95, ARG 146, GLN 156, VAL 6, ASP 110, GLU 179, ASP 181, MET 160) and forms hydrogen bonds with 4 amino acids of the protein (TRP 178, VAL 91, GLN 149, GLU 3). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.7.

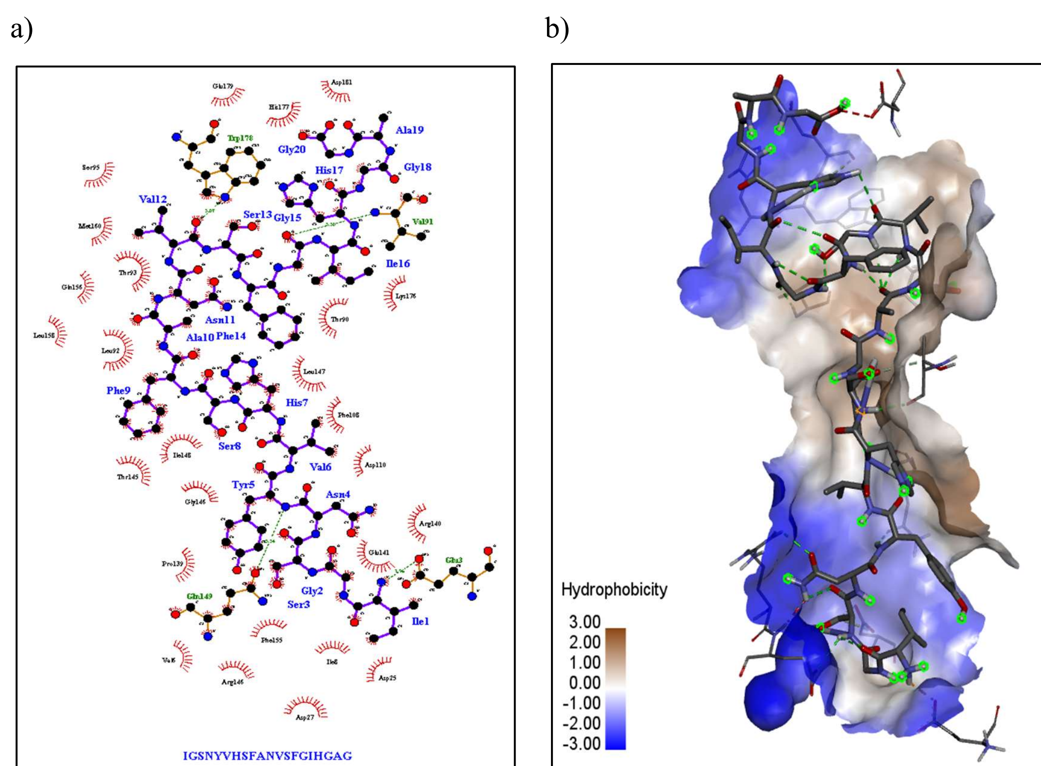

Figure 4.7 The interactions between the epitope (IGSNYVHSFANVSFGIHGAG) and protein visualized in: (a) 2D and (b) 3D form.

v) FLVAITHAASIKFNVIAPNA

The epitope has hydrophobic interactions with 35 amino acids of the protein(LEU 147, GLU 88,GLN 149,THR 93,ASP 27,VAL 6,VAL 97,ILE 148,GLU 128,ARG 146,TRP 178,ILE 127,GLU 3,VAL 91,PHE 108,PRO 139,LYS 176,SER 126,THR 157,ILE 8,LEU 92,SER 95,ARG 140,HIS 177,PHE 180,HIS 16,THR 145,GLY 146,LEU 158,ASP 25,GLY 28,VAL 89,PRO 124,GLY 125,PHE 155,GLN 156) and forms hydrogen bonds with 2 amino acids of the protein(GLU 141, THR 90).The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.8.

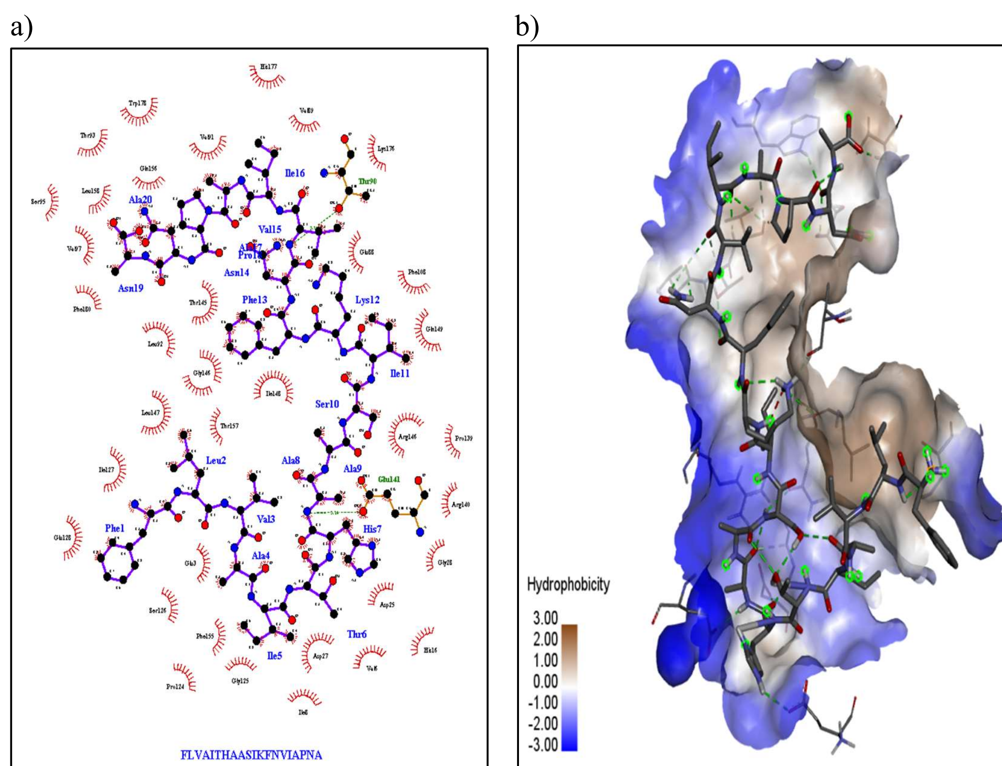

Figure 4.8 The interactions between the epitope (FLVAITHAASIKFNVIAPNA) and protein visualized in: (a) 2D and (b) 3D form.

vi) QNADKKATFEKYLTETVRVL

The epitope has hydrophobic interactions with 16 amino acids of the protein (ASP 27, GLU 128, LEU 147, SER 126, LYS 139, PHE 155, GLY 125, ILE 127, ALA 140, VAL 142, GLU 3, GLY 28, SER 144, GLN 149, GLU 4, PRO 124, GLY 146). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.9.

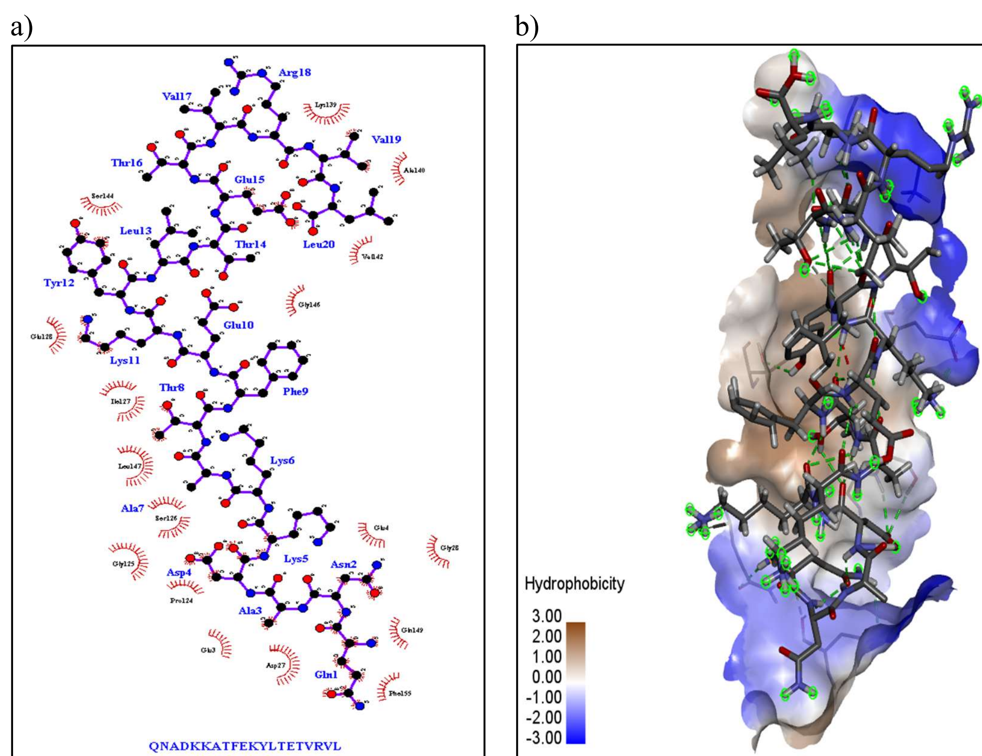

Figure 4.9 The interactions between the epitope (QNADKKATFEKYLTETVRVL) and protein visualized in: (a) 2D and (b) 3D form.

vii) DTQRKTVKDVYLTAAQRVSV

The epitope has hydrophobic interactions with 14 amino acids of the protein (VAL 42, GLU 134, ASN 118, VAL 119, LYS 38, GLU 40, VAL 117, PRO 16, TRP 168, ARG 4, ASP 17, ASP 35, ALA 37, VAL 132, VAL 136). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.10.

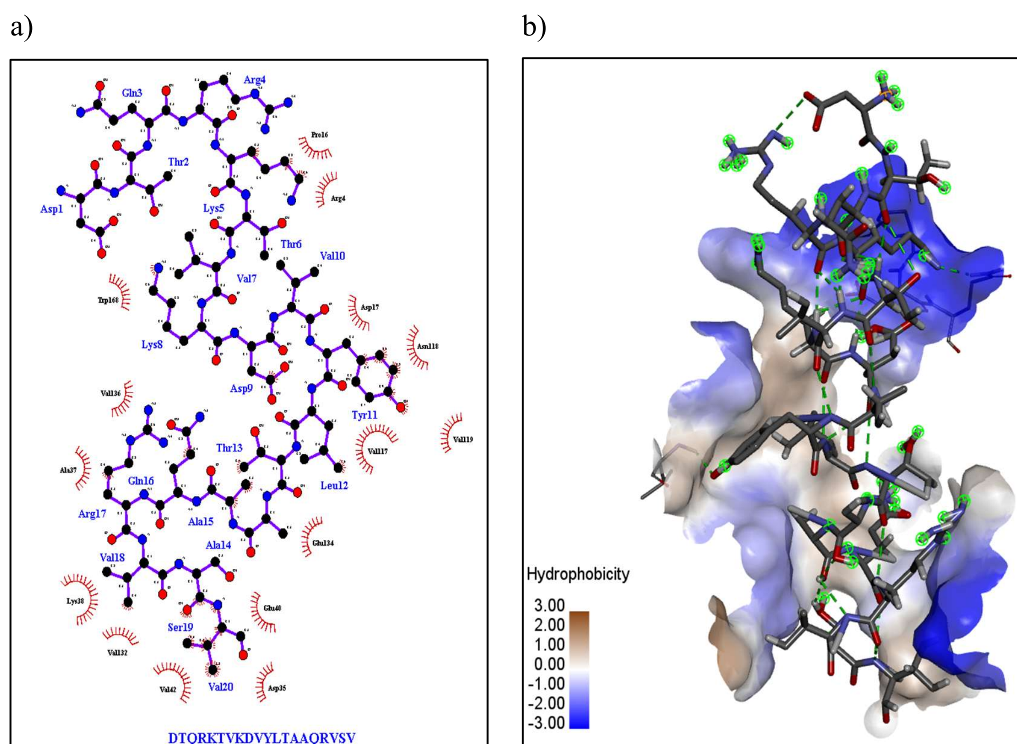

Figure 4.10 The interactions between the epitope (DTQRKTVKDVYLTAAQRVSV) and protein visualized in: (a) 2D and (b) 3D form.

viii) EWHLESGLAVVFVEAPDIAQ

The epitope has hydrophobic interactions with 14 amino acids of the protein(SER 104,GLU 141 ,TRP 178,LYS 111,ASP 142,GLU 179,THR 90,VAL 91,VAL 97,LYS 176,THR 145,VAL 85,LEU 92,LEU 144,ASP 181,GLN 34) and forms hydrogen bonds with 1 amino acids of the protein (ARG 140). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.11.

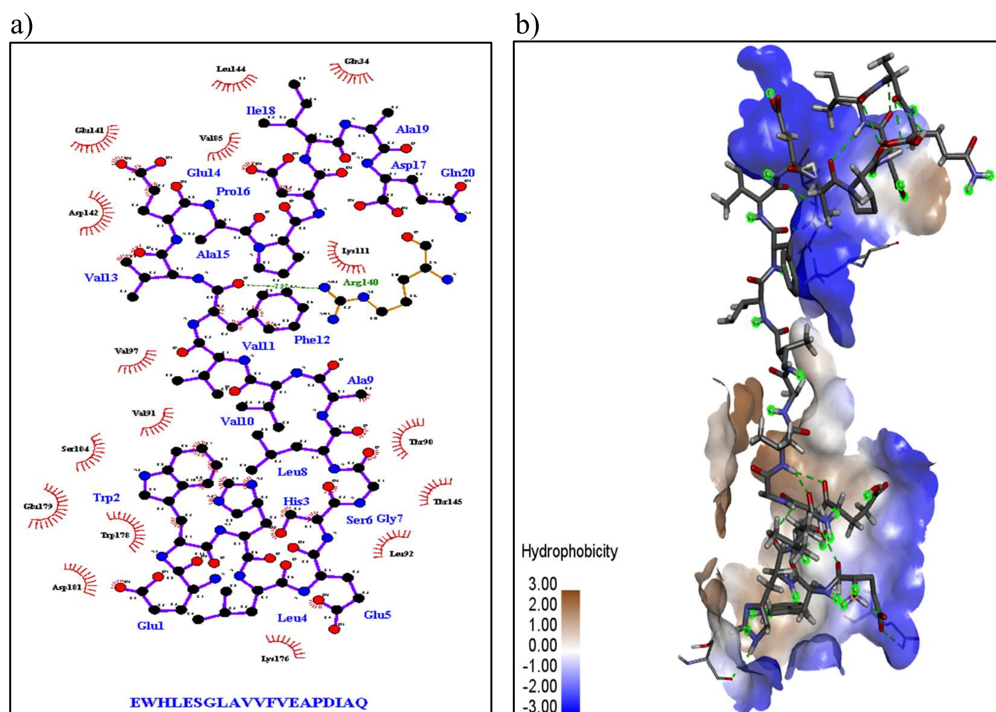

Figure 4.11 The interactions between the epitope (EWHLESGLAVVFVEAPDIAQ) and protein visualized in: (a) 2D and (b) 3D form.

ix) AADVISYRVSTAVWHVSWGD

The epitope has hydrophobic interactions with 15 amino acids of the protein(GLU 46,VAL 117,TRP 43,SER 19,THR 129,ASP 17,VAL 42,ARG 44,LEU 45,VAL 119,VAL 132,ASP 35,ASN 118,SER 133,GLU 134,LYS 147) and forms hydrogen bonds with 1 amino acids of the protein(THR 41).The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.12.

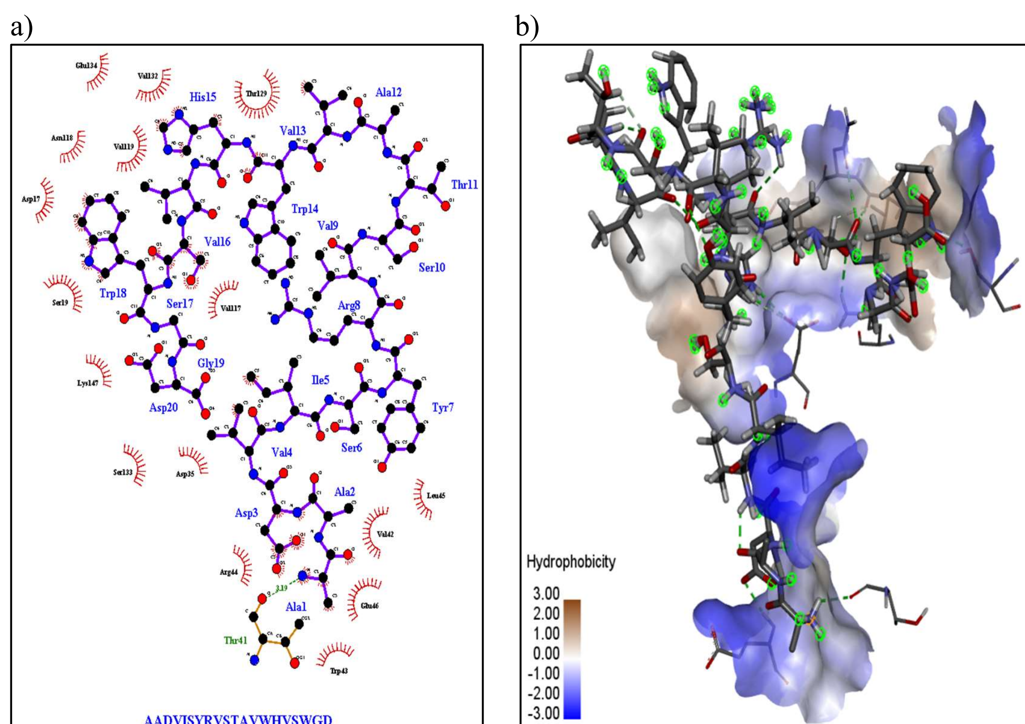

Figure 4.12 The interactions between the epitope (AADVISYRVSTAVWHVSWGD) and protein visualized in: (a) 2D and (b) 3D form.

x) ADVISYRVSTAVWHVSWGDP

The epitope interacts with 16 amino acids of the protein (TRP 178,THR 93,VAL 143,ARG 130,VAL 142,GLU 179,LEU 158,LYS 176,PHE 180,ASP 181,VAL 91,LEU 92,SER 95,HIS 177,TYR 102,GLU 128,LYS 139). The 2D and 3D interactions between the protein and ligand obtained using LigPlot and Discovery Studio respectively are displayed in Fig. 4.13.

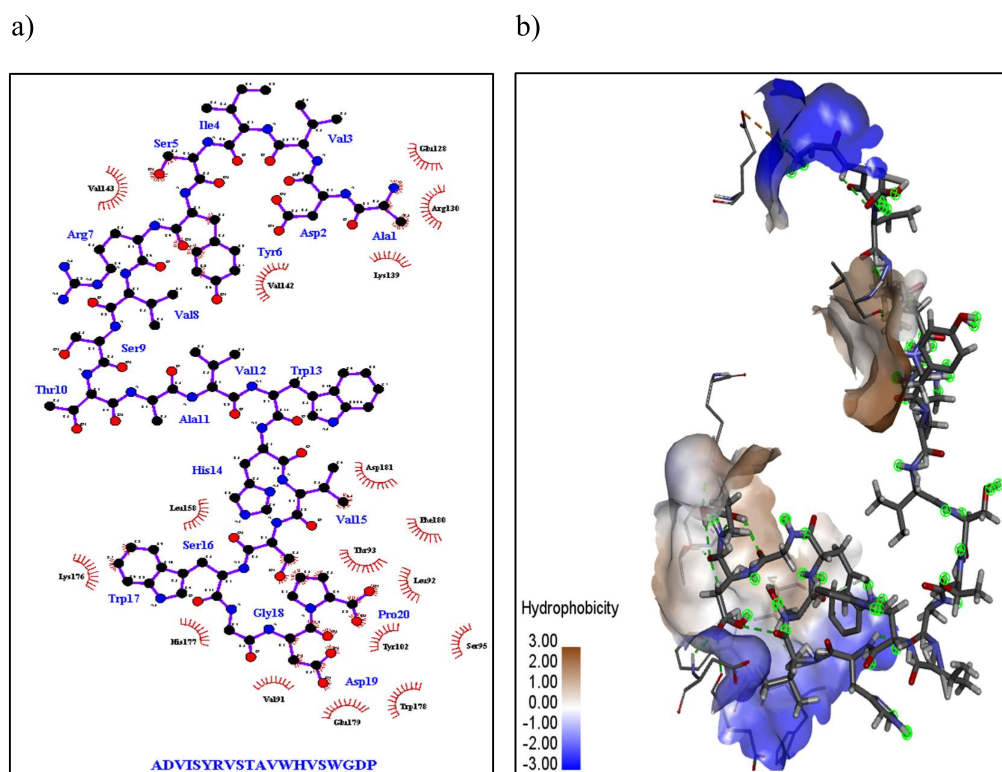

Figure 4.13 The interactions between the epitope (ADVISYRVSTAVWHVSWGDP) and protein visualized in: (a) 2D and (b) 3D form.

In this project, five polypeptide sequences located in mucormycosis-causing organisms were used in predicting ten antigenic sequences. Epitope sequences having the ability to stimulate both T and B cells are known as good vaccine candidates. BCPred and ABCPred tools were utilized in the prediction and selection of B-cell epitopes from the sequences. NetMHCII2 and NetMHCIIpan servers were used to select the sequences which could act as both B and T-cell epitopes. The physiological features of the predicted sequences were computed using the ProtParam tool. The sequence antigenicities were calculated by the VaxiJen tool. Allertop tool was then used to calculate the allergenicity of the sequences. The IEDB population coverage tool was utilized to calculate the population coverage of the sequences. The structure of the epitopes was designed and collected from the DISTILL server and the structure of DRB1*07:01(1AQD) was obtained through the PDB server. AutoDock Vina tool was used to dock the epitope sequences with the protein(1AQD). The dockings were

visualized using Discovery Studio Biovia and the Ligplot tools. Ten different epitope sequences were found in the analysis that could be used to construct a vaccine for mucormycosis.

Chapter 5: CONCLUSION AND FUTURE SCOPE

Epitope screening using computational methods predicts that the CotH1, CotH2, CotH3, Fet3, and Ftr1 proteins can be utilized for constructing vaccines. Molecular docking tests predict that the epitopes (AADVISYRVSTAVWHVSWGD, ADVISYRVSTAVWHVSWGDP, DTQRKTVKDVYLTAAQRVSV, EWHLESGLAVVFVEAPDIAQ, SGASFNFPGDFDSFIPNVES, IWALPEGQFMANRNWFKIRH, SFLVAIAHAASVQFNLIAPS, IGSNYVHSFANVSFGIHGAG, FLVAITHAASIKFNVIAPNA, QNADKKATFEKYLTETVRVL) have low energy minimization score which shows sufficient binding with the MHC class II molecules and gives stability to the complex. More wet lab tests are further required to use these peptides as a vaccine against Mucormycosis. Bioinformatic methods can select vaccine candidates for different diseases, reducing the time taken and the cost of the experiments. A similar technique can be used to select vaccine candidates for different diseases, reducing the cost and time.

REFERENCES

1. A.I. Heinson, Y. Gunawardana, B. Moesker, C.C.H. Denman, E. Vataga, Y. Hall, et al., “Enhancing the biological relevance of machine learning classifiers for reverse vaccinology”, Int. J. Mol. Sci., vol. 18, no. 2, pp. 312, February 2017.
2. A.K. Singh, R. Singh, S.R. Joshi, A. Misra, "Mucormycosis in COVID-19: A systematic review of cases reported worldwide and in India", Diabetes & Metabolic Syndrome, vol. 15, no. 4, pp. 102146, May 2021.
3. B. Spellberg, J. Edwards, A. Ibrahim, "Novel perspectives on mucormycosis: pathophysiology, presentation, and management", Clinical Microbiology Reviews, vol.18, no.3, pp. 556–69, July 2005.
4. D. Baú, A.J. Martin, C. Mooney, et al., “Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins”, BMC Bioinformatics,vol. 7,no. 402, 2006.
5. D. Garg, V. Muthu, I.S. Sehgal, R. Ramachandran, H. Kaur, A. Bhalla, et al., "Coronavirus Disease (Covid-19) Associated Mucormycosis (CAM): Case Report and Systematic Review of Literature",Mycopathologia, vol. **186,** no.2, pp. 289–298, May 2021.
6. E. Dannaoui, M. Lackner, "Special Issue: Mucorales and Mucormycosis", Journal of Fungi, vol. 6, no.1, pp. 6., December 2019.
7. G. Castellano-González, H.M. McGuire, F. Luciani, L.E. Clancy, Z. Li, S. Avdic, B. Hughes, M. Singh, B. Fazekas de St Groth, G. Renga, M. Pariano, M.M. Bellet, L. Romani, D.J. Gottlieb. “Rapidly expanded partially HLA DRB1-matched fungus-specific T cells mediate in vitro and in vivo antifungal activity”, Blood Adv, vol. 4, no.14, pp. 3443-3456, July 2020.
8. G. Petrikkos, A. Skiada, O. Lortholary, E. Roilides, T.J. Walsh, D.P. Kontoyiannis. “Epidemiology and clinical manifestations of mucormycosis”, Clin Infect Dis.,vol. 54,no. 1,pp. S23-34, February 2012.

9. G. Reid, J.P. Lynch, M.C. Fishbein, N.M. Clark, "Mucormycosis", Seminars in Respiratory and Critical Care Medicine, vol. 41, no. 1, pp. 99–114, February 2020.
10. I. Dimitrov, D.R. Flower, I. Doytchinova, “AllerTOP - a server for in silico prediction of allergens”, BMC Bioinformatics vol. 14, no. S4, 2013.
11. I.A. Doytchinova, D.R. Flower, “VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines”, BMC Bioinformatics vol.8, no. 4., January 2007.
12. J. Sun, D. Wu, T. Xu et al., “SEPPA: a computational server for spatial epitope prediction of protein antigens”, Nucleic Acids Research, vol. 37, no. 2, pp. W612–W616, 2009.
13. J.L. Hernández, C.J. Buckley, “Mucormycosis”. Treasure Island, FL: StatPearls Publishing, 2021.
14. K.K. Jensen, M. Andreatta, P. Marcatili, S. Buus, J.A. Greenbaum, Z. Yan, A. Sette, B. Peters, M.Nielsen, “Improved methods for predicting peptide binding affinity to MHC class II molecules”, Immunology,vol. 154,no. 3,pp. 394-406, July 2018.
15. K.L. Seib, X. Zhao, R. Rappuoli, “Developing vaccines in the era of genomics: A decade of reverse vaccinology”, Clinical Microbiology and Infection, vol. 18, no. 5, pp. 109–116, October 2012.
16. M.E. Grossman, L.P. Fox, C. Kovarik, M. Rosenbach, "Subcutaneous and deep mycoses", Cutaneous Manifestations of Infection in the Immunocompromised Host, vol.2, pp. 51–58, November 2011.
17. M.M. Roden, T.E. Zaoutis, W.L. Buchanan, T.A. Knudsen, T.A. Sarkisova, R.L. Schaufele, et al, “Epidemiology and outcome of zygomycosis: a review of 929 reported cases”, Clin Infect Dis., vol. 41, no. 5, pp. 634-53, September 2005.
18. N.K. Mehra, R. Rajalingam, U. Kanga, L. McEnemy, C. Cullen, S. Agarwal, D. Middleton, M. S. Pollack, “Genetic diversity of H LA in the populations of India, Sri Lanka and Iran”, In: Genetic Diversity of HLA: Functional and Medical Implications, eds. Dominique Charron, EDK Publishers, Paris, Vol I, 1997, pp. 314-320.

19. O. Trott, A.J. Olson, “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”, J Comput Chem., vol. 31, no.2, pp. 455-61, January 2010.
20. P. H. Andersen, M. Nielsen, and O. Lund, “Prediction of residues in discontinuous B-cell epitopes using protein 3D structures,” Protein Science, vol. 15, no. 11, pp. 2558–2567, 2006.
21. S. Vivona, F. Bernante, F. Filippini, “NERVE: new enhanced reverse vaccinology environment”, BMC biotechnology, vol. 6, no. 1, pp. 1-8, December 2006.
22. V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, M. Nielsen, “NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data”, J Immunol., vol. 199, no. 9, pp. 3360-3368, November 2017.
23. W. Shen, Y. Cao, L. Cha et al., “Predicting linear B-cell epitopes using amino acid anchoring pair composition,” BioData Mining, vol. 8, no. 1, article no. 14, 2015.
24. W. Zhao, X. Sher, “Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes”, PLoS computational biology,vol. 14,no. 11, e1006457, November 2018.
25. Y. El-Manzalawy, D. Dobbs, V. Honavar, “Predicting linear B-cell epitopes using string kernels”, J Mol Recognit,vol. 21,no. 4,pp. 243-55, July 2008.
26. Y. He, Z. Xiang, H.L. Mobley, “Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development”, Journal of Biomedicine and Biotechnology., vol. 2010, pp. 1-15, October 2010.

List of Publications

Conference Papers

1. S. Biswas, Y. Hasija, “Mucormycosis Vaccine Design using Bioinformatic Tools”, Lecture Notes in Electrical Engineering, vol 853, March 2022.
2. S. Biswas, Y. Hasija, “Machine Learning Methods for Protein Function Prediction”, Proceedings of International Conference on Computational Intelligence and Emerging Power System, December 2021.

Book Chapters

1. S. Biswas, Y. Hasija, “Predicting Depression Through Social Media”, Predictive Analytics of Psychological Disorder on Health, Vol. 128, 2022.
2. S. Biswas, Y. Hasija, “Big Data Analytics in Precision Medicine”, Big Data Analytics for Healthcare, 2022.