In silico identification of molecular mimicry between T-cell epitopes of *Borrelia burgdorferi* and the human proteome: Implications for autoimmune response in Lyme disease

By

Tahsin Mahmud Tanim

ID- 18236010

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of B.Sc. in Biotechnology

Department of Mathematics and Natural Sciences

Brac University

December 2022

©2022. Brac University

All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Tahsin Mahmud Tanim
Student ID : 18236010

Approval

The thesis/project titled "[Thesis/Project Title]" submitted by

1. Tahsin Mahmud Tanim (18236010) of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B. Sc in Biotechnology on 15.11.2022.

Examining Committee:

Supervisor:
(Member)

_____

Mohammed Sayem
Lecturer, Department of Mathematics and Natural Science, Brac University

Program Coordinator:
(Member)

_____

Dr. Munima Haque,
Associate Professor, Department of Mathematics and Natural Sciences, Brac University

Departmental Head:
(Chair)

_____

A F M Yusuf Haider
Professor and Chairperson, Department of Mathematics and Natural Sciences
Brac University

Abstract:

Autoimmune disorder is a dysfunctional immune system in a body, which attacks its own cells. *Borrelia burgdorferi*, a pathogenic bacterium, shares an intrinsic relation to Lyme disease in humans causing autoreactive responses that remained a serious health concern for the public Molecular mimicry is one of the primary mechanisms by which pathogen or toxic chemicals induce autoimmunity. In this research, a pipeline of bioinformatics tools was applied for examining the molecular mimicry between human and *B. burgdorferi* protein sequences from a structural perspective. In addition, the potential antigenic peptides of this bacteria similar to the peptides of human were classified and the fundamental molecular mechanisms of autoimmune disorders were analyzed. To initiate the study, proteome databases were retrieved and removed their paralogous sequences. Subsequently, the homologous proteins were identified along with enrichment pathways. Then, the subcellular localization of proteins were predicted followed by the prediction of MHC class II epitopes. Additionally, the 3D structure of the epitopes was modelled for comparing the structure of host and pathogen. The bacterial epitopes DFTIEVERSLRVLDG, LRLKKLIIDIMSNQI and VIAQLLFLESEDSSK that superimposed with the respective host epitopes DFTIEVERALRVLDG, MKLKKQLYNIYAKHT and VIAQLLFLQSESNKK showed 100% binding which were highly predicted cross reactivity. Finally, docking scores of these bacterial epitopes revealed the binding affinity towards the host HLA alleles including DRB1*07:01, DRB1*01:01and DRB4*01:01 that may lead to autoimmune disorders in humans caused by *B. burgdorferi* infection.

Keywords: Autoimmune disease, *Borrelia burgdorferi*, Bioinformatics, Molecular mimicry, Antigen, Infection, MHCII binding Epitopes, Homologous.

Acknowledgement

I thank Allah for guiding me throughout my academic careers, including this particular endeavor, in my work and in my studies. This research would not have been possible without the help of others, who are respectfully recognized here.

I would like to express my gratitude to Professor A F M Yusuf Haider, Chairperson, Department of Mathematics and Natural Sciences for allowing me to finish my undergraduate thesis.

Throughout this process, my supervisor, Mohammed Sayem, Lecturer, Department of Mathematics and Natural Sciences, has shown me a lot of support and dedication. I am immensely grateful to him and owe him a lot of appreciation because he has also provided me with useful recommendations and guidance whenever I needed it.

Contents

List of Tables

List of Figures

List of Acronyms

AD Autoimmune disease

APC Antigen presenting cell

Osp Outer surface protein

MHC Major histocompatibility complex

SCL Subcellular localization

HLA Human leukocyte antigens

LFA-1$_\alpha$ Leukocyte function-associated antigen-1

AITD Autoimmune thyroid disease

HSP Heat shock protein

RA Rheumatoid arthritis

TM Template modelling

RMSD Root mean square deviation

GAPDH Glyceraldehyde 3-phosphate dehydrogenase

SP Substance P

PCNA Proliferating Cell Nuclear Antigen

JIA Juvenile idiopathic arthritis

ELISA Enzyme linked immunosorbent assay

AIH Autoimmune hepatitis

EBV Epstein Barr virus

CMV Cytomegalovirus

NK-1R Neurokinin 1 receptor

TSH-R Thyroid-stimulating hormone receptor

BLAST Basic Local Alignment Search Tool

CD-HIT Cluster Database at High Identity with Tolerance

IFNγ Interferon gamma

IEDB Immunological Epitope Database

Th T helper

IC50 Half maximal inhibitory concentration

PDB Protein Data Bank

Chapter 1

Introduction

Autoimmune disease (AD) takes place, when the body's own immune system recognizes the self-antigen as hetero-antigen that may lead to the destruction of its own healthy cells and tissues. [1] This AD may occur in any specific locations including tissues, organs or even all over the body's parts at once. Globally, around 5% of the population are affected by this disease which boost the infection as well as the mortality rates. [2] The incidence of autoimmune disorders greatly rose over the last century and its causes are almost unknown. This autoimmunity became a severe health problem worldwide, which are mostly chronic and or even fatal [3]. AD rank among the most common illnesses in the United States of America with a prevalence of around 7%. There are eighty types of AD in total and majority of them are lifelong and permanent [4].

Borrelia burgdorferi is a pathogenic bacterial species that belongs to the genus *Borrelia* of the spirochete class. It is one of the etiological agents of the vector-borne disease, Lyme disease, which is frequently observed in the Northern Hemisphere [5]. These bacteria are spread throughout the United States by hard-bodied ticks, such as *Ixodes pacificus* and *Ixodes scapularis*, causing about 300,000 cases per year. A research claimed that an interleukin-2 binding protein inhibited the proliferation of T cell in *Ixodes scapularis* saliva which may facilitate pathogen transmission locally or even systemically by the immune system. [8] The life cycle of these ticks includes egg, larva, nymph, and adult stages which lasts between 2 and 4 years and usually get infected during their larval phase by feeding on the blood of small mammals or even larger ones containing *B. burgdorferi* (**Figure 1**). Spirochaetes use internal periplasmic flagella to enter the skin during a tick bite and go to distal areas like the heart and joints. If infections are not treated, they may result in multisystemic symptoms, including rheumatoid arthritis, heart issues, and neurological disorders [6, 55]. The early inflammatory response is likely to be significantly influenced by the interaction of *Borrelia's* lipoprotein with Toll-like receptor 2 [7]. Some of these lipoproteins, including the outer surface proteins (Osp) A, B, and C, are important immunogens and make up the major portion of the overall quantity

of protein produced by *B. burgdorferi*. [8] Some reports indicate that individuals having Lyme disease (Lyme borreliosis, Lyme neuroborreliosis and Lyme arthritis) for several months may acquire systemic autoimmune disorders including rheumatoid arthritis and spondyloarthritis. However, the mechanism for the development of ADs from Lyme disease is yet unknown. [9,10]

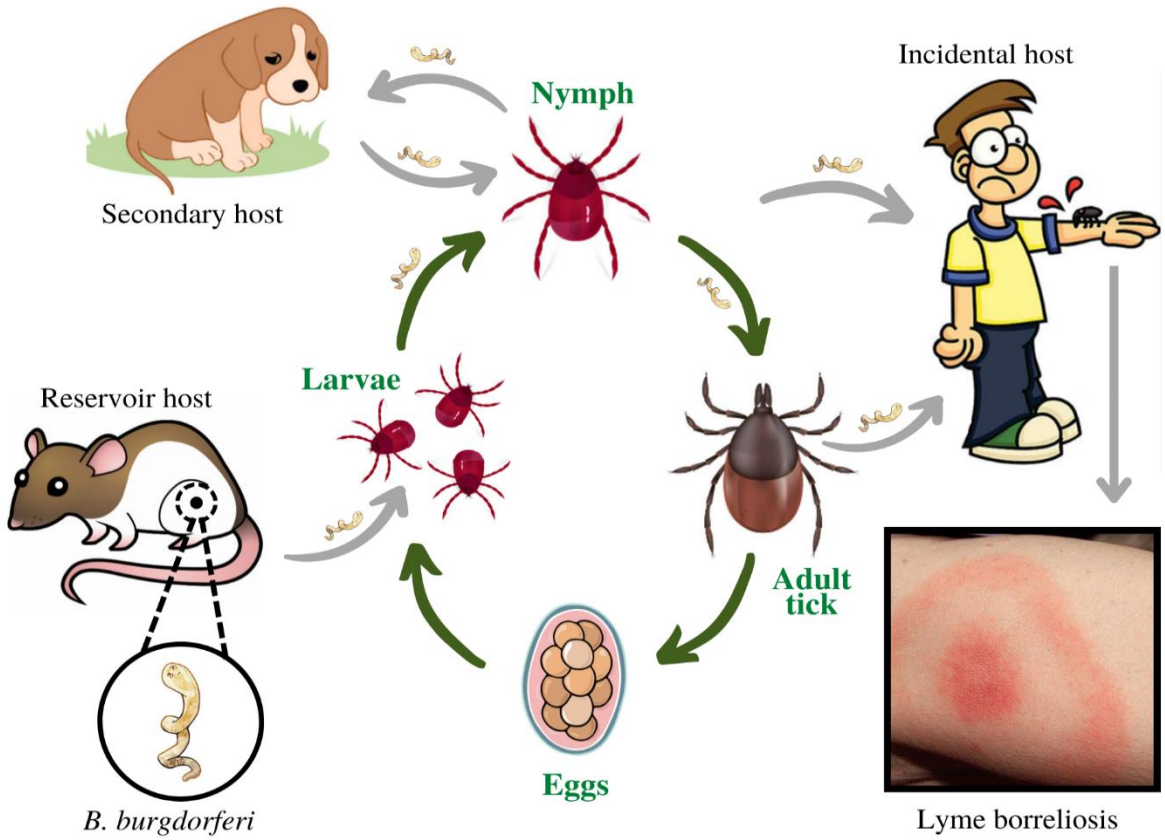


Figure 1: The life cycles of Ixodes scapularis and *Borrelia burgdorferi*. When the larvae hatches from the egg, they feed on the blood containing *B. burgdorferi* of small mammals or birds (reservoir host) and initially get infected. Subsequently, this vector-borne pathogen spread throughout the cycle by a tick's bite causing Lyme borreliosis. The nymph may either transmit bacteria to or get infected from a secondary host but can only transmit in incidental host. However, the bacteria is not passed to the eggs. (Idea taken from various research articles)

Several pathways were identified through which pathogens or toxic chemicals may cause the development and aggravation of autoimmune diseases. In 1964, Damian coined the term "molecular mimicry, is one of the mechanism, in which bacteria and human epitopes have similar structure or sequence characteristics. [1, 30] While performing three-dimensional structural mimicry, it was demonstrated that virulence factors that are exact replicas of host proteins induce an autoimmune response. [12] Autoimmune disorders are characterized by auto

reactive T or B cells that causes chronic tissue inflammation and in many cases, permanent structural and functional damage. Additionally, B-lymphocytes stimulates the production of autoantibodies, also serve as antigen-presenting cells (APCs) and a cytokine source. [14] Molecular mimicry comprises autoimmune disease, as T cell proliferation occurs for bacterial antigen that can self-react with homologous peptides. Moreover, major histocompatibility complex (MHC) polymorphism (especially class II molecules crucial for antigen identification by T cells) are almost associated with every autoimmune disease. [13].

Many studies has shown that CD4+ T cells mediation are linked to autoimmunity. Dendritic cells engulf the pathogens inside the body via endocytosis whose exogenous antigens are processed by the dendritic cells in endosomal compartments before they are presented to CD4+ T cells via MHC class II molecules [57]. Once CD4+ cells exposed to a particular antigen they multiply and differentiate into either Th1 or Th2 cells. This Th1 cell boost the immune system in the body to defend the infection. Additionally, this subset of CD4+ T cell play a role in autoimmunity when they recognize the presented antigenic peptide in the context of an HLA molecule, which serves as the host's antigen presenter, becomes activated if they receive the proper signals [58].

A dominant T cell epitope was found in Lyme arthritis patients who were resistant to treatment especially from OspA and hLFA-1 likely to exhibit molecular mimicry. [15]. Additionally, these patients may experience the emergence of autoimmunity within infected joints, especially with the HLA alleles such as DRB1*0401 or DRB1*0101 due to molecular mimicry between a host protein and *B. burgdorferi*'s OspA having an immunogenic epitope of T cell [16]. A region of peptide found in OspA has prevented a well-known peptide ligand from attaching to DRB1*0401. Antigen presenting cell from transgenic mice expressed OspA bearing HLA-DRB1*0401 allele which were effective T cell stimulators for many people with chronic Lyme arthritis. [7]. A possible autoantigen was identified using a homology search from a peptide of human LFA-1$_\alpha$ antigen. Although LFA-1$_\alpha$ partially stimulated OspA reactive T cells, but later on it was predicted that it was unable to represent a significant autoantigen as LFA-1$_\alpha$ failed to bind the DRB1*0101 molecule in an in-vitro assay [17]. A detailed molecular analysis of the epitopes in OspA alone may aid in the search for novel potential autoantigens, even if sequence homology is only one way for locating possible autoantigens.

Objective:

- Our main goal in this analysis was to employ bioinformatics studies for the identification of homologous proteins and their associate epitopes across the proteome datasets from *B. burgdorferi* and human.
- Evaluation of the binding potential of the homologous epitopes with immunogenic receptors for understanding the underlying molecular causes of autoimmunity linked to *Borrelia burgdorferi*.

Chapter 2

Materials and methods:

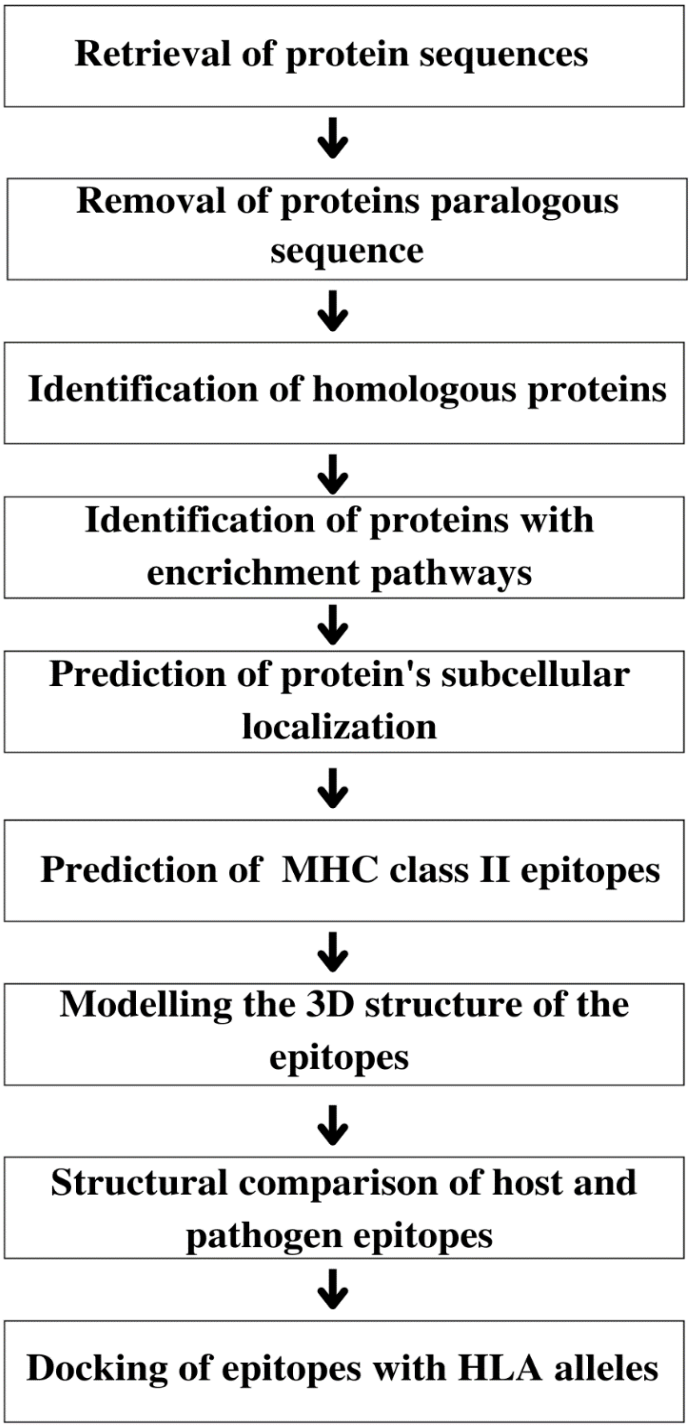


Figure 2: Illustration of the study's methodology in a flow chart

2.1 Retrieval of protein sequences

The study was initiated through the retrieval of proteome of *Homo sapiens* (UP000005640) and *B. burgdorferi* (UP000001807) from Swiss-Prot. Proteome is the entire set of proteins encoded by an organism's genome and Swiss-Prot is the reviewed part of UniProt's proteome database (https://www.uniprot.org) with a very low level of redundancy.

2.2 Identify the non-paralogous protein sequence

Initially the development of CD-HIT (Cluster Database at High Identity with Tolerance) have shown to compare and cluster the sequences of protein with a reduced redundancy for building reference databases. This application is widely used for employing a short-read filtering strategy to cluster protein sequences with lower levels of duplication. The retrieved protein dataset of the *B.burgdorferi* was run into this program (CD-HIT), which had non-redundant sequences. Hence, the main goal was to have a reduced database size by deleting extremely similar or redundant sequences and detect non-paralogous protein sequence of *B. burgdorferi* [19]. With a 90% cutoff, the CD-HIT tool removed the sequences of proteins that are paralogous. Further analysis was performed on the obtained non-paralogous sequences.

2.3 *B. burgdorferi* vs human host protein sequences similarity search

The Basic Local Alignment Search Tool (BLAST) identifies areas of similarity between biological sequences, such as amino acid sequences in proteins or nucleotide sequences in DNA or RNA. In this study, stand-alone BLAST was used on command prompt (cmd) of windows, which is a local installation of the NCBI BLAST. In order to identify host homologous proteins, the protein sequences that are not paralogous to *B. burgdorferi* were BLAST against human (*Homo sapiens*) proteome sequences obtained from the Swiss-Prot. The extraction of homologous proteins were based on E value ($e \leq 10^{-6}$), bit score (100), query coverage (60%) and p-identity (50%). Here, p-identity (or percent identity) is a number that expresses how similar the query and target sequences are. The match is more significant when the percent identity is higher. The percentage of the contig length that matches the NCBI hit is known as query coverage. A low query coverage percentage indicates that just a very small percentage of the contig is aligning. E value is the approximate number of hits that would be

anticipated by chance. E value closer to zero is consider as better results. Bit score is a crucial metric that reveals the statistical importance of an alignment. For instance, the higher the bit score, the closer the two sequences are. Bit scores under 50 are typically regarded as unreliable.

2.4 Analysis of protein's enrichment pathways

Using PathDIP (Pathway Data Integration Portal) (http://ophid.utoronto.ca/pathDIP/), the signaling pathways in humans were analyzed and annotated by combining the physical protein-protein interactions with information from twenty-two core pathway databases to extended pathways. Thus, it predicted the biologically significant protein-pathway relationships [20]. Here, the molecular pathways of the *Homo sapiens* (host) were identified by inserting the Uniprot ID's of the human from blast output and subsequently select the "extended pathway associations" prior searching in the PathDIP analysis. To conduct further research, a cutoff of the p-value $\leq$ to 0.005 and q-value equal to 0.05 was adjusted. Finally, the unique pathway proteins will be taken.

2.5 Computational prediction of bacterial protein subcellular localization

PSORTb (https://www.psort.org/psortb/) is the first subcellular localization (SCL) predictor created to support localizations for a broad spectrum of prokaryotes across multiple subcategories [21]. Studying the subcellular localization of protein molecules is necessary to comprehend their function and structure inside the cell. The prediction of protein SCL of *B. burgdorferi* was done with the help of its online server and gram negative stain was selected. Only the pathway enrichment pathways were analyzed while using this PSORTb.

2.6 Prediction of the epitopes

The Immune Epitope Database (IEDB) (http://tools.iedb.org/mhcii/) is a combined bioinformatics tool used for immune epitope analysis and prediction. The IEDB identifies and characterizes specific epitopes along with its immunogenic receptors as well as other information about immune exposure, induced immune receptors and host organism. Additionally, it covers immune epitope data including antibody, T cell, and MHC binding that

are connected to infectious, allergy, autoimmune, and transplant-related disorders [22]. IEDB was used to predict the epitopes that bind to *B. burgdorferi* and human protein especially the long peptides (13 – 25 amino acids) of the MHC class II molecules. The NN-align-2.2 algorithm along with a threshold of IC50 less than 50 nM predicted the epitopes that binds with a range of class II HLA alleles including HLA-DR, HLA-DQ, and HLA-DP. IC50 (half maximal inhibitory concentration) less than 50 are considered as strong binders.

2.7 Alleles involved in the research

Class II alleles of the human leukocyte antigen were analyzed from IEDB for predicting the binding peptides. Here, the HLA alleles including one DP, one DQ and five DR were employed for the binding of epitopes.

2.8 Generation of 3D structures for HLA and epitope

IPD-IMGT/HLA (https://www.ebi.ac.uk/ipd/imgt/hla/alleles/), an allele query tool, was used to get the protein sequence by searching these HLA alleles. Subsequently, these protein sequences were run in Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/), a free online bioinformatics tool for predicting protein tertiary structures, by selecting the intensive modelling mode. Therefore proteins structures were retrieved from Phyre2 which run in saves (https://saves.mbi.ucla.edu) and the parameters were recorded from the Procheck output for the assessment of protein structure in Ramachandran plot. This Ramachandran plot, in terms of torsion angles, provides a straightforward two-dimensional graphic depiction of all potential protein structures.

The Ressource Parisienne en Bioinformatique Structurale (RPBS) (https://mobyle.rpbs.univ-paris-diderot.fr/) is the result of multiple teams working together for providing unique access to novel structural bioinformatics services. This web-based tool has a variety of functions with the overall purpose of combining the various components of structural bioinformatics [29]. The complex structure and external factors of peptide has a strong regional impact and thus to predict the 3D folding of each peptide, this RPBS structure was used. Therefore, in the program section, the structure option was selected followed by prediction, 3D structure and finally PEP-

FOLD3. In this portion, the epitopes of the mimicry candidates were searched to predict the de novo peptide structure.

2.9 Evaluation of structural similarity of peptides

TM-align (template modelling align) (https://zhanggroup.org/TM-align/) is a technique for comparing protein structures without regard to sequencing. In order to construct the optimum residual alignment based on structural similarity of two proteins with undetermined equivalency, TM-align first generates iterations of heuristic dynamic programming. In addition to the value of TM score, which measures the similarity in structure, providing a structural ideal superposition of two proteins based on the discovered alignment [34]. The predicted structural mimicry candidates of human and bacteria were inserted in this tool for pair-wise structure comparison. The structural similarity of peptides was evaluated using the TM score based on the RMSD (root mean square deviation) value.

2.10 Molecular docking

The HDOCK (http://hdock.phys.hust.edu.cn) is a set of bioinformatics tools that is integrated fully giving an instant and reliable docking between two similar proteins. This includes a search in homologous pairs, modeling based on a template, prediction of the structure, macromolecular docking, incorporation of biological information, and job administration [48]. The data of both ligand and receptor molecules was inserted to initiate the tool. HDOCK tool is based on either a hybrid template or a free-template docking method, which allows the insertion of data in the form of Protein Data Bank (PDB) or amino acid sequence that can predict the interaction between the molecules of ligand and receptor. HDOCK was used to dock the three dimensional structures of *B. burgdorferi* peptides against human class II HLA alleles. The techniques make it possible to calculate peptides' binding attraction to human HLA. The HDOCK creates many models from the input structure or homology model based on transformation and binding scores, such as confidence scores, docking scores, and ligand RMSDs. A high negative score will increase the probability of binding models. However, the exact score can only be calibrated through experimental data. Therefore, it should not be interpreted as the actual binding affinity of two molecules. Furthermore, the two molecules will have the highest chance of binding if the value of the confidence score is more than 0.7,

possible to bind when the confidence level is between 0.5 and 0.7, and the binding may not occur if the value of the confidence score is less than 0.5. The ligand RMSD is not a reliable indicator of the precision of the docking process as it only measures the distance of predicated model from a homology-modeled structure or the input structure.

Chapter 3

Results

3.1 Retrieval of *B. burgdorferi*'s proteome data

The Uniprot proteome database yielded 1291 protein sequences from *B. burgdorferi* B31. This strain has the spirochete Lyme disease-causing potential in humans, which was taken into consideration, and further analysis was done on these protein sequences.

3.2 Removal of *B. Burgdorferi*'s paralogous proteins

Linux systems were used to test the C++-coded CD-HIT package. Almost all systems that support C++ can run it because it is offered as an open source package and requires little to no change. By creating homologous sequence clusters and removing extraneous protein sequences, the CD-HIT service decreases the amount of proteins [19]. A representative of each paralogous sequence were chosen through the elimination of the redundant sequences from the dataset of proteome and finally with a threshold of 0.9 using the default parameters got 1142 non-paralogous proteins from a total of 1291 protein sequences.

3.3 Analysis of host-pathogen sequence similarity

The BLAST program was used to show how the conserved regions found in pathogen and human proteome have sequence similarities [23]. To identify the homologous protein's mimicry, the non paralogous protein sequences from *B. burgdorferi* were subjected to BLAST against the human proteome. A total of 263 microbial proteins that shows homology with human proteins were identified. Furthermore, the blast output was filtered based on the E-values less than or equal to 10^{-6} and bit score more than or equal to 100. Additionally, 50 percent identity and 60 percent query coverage showed a strong homology between *B. burgdorferi* and human. Finally, with this approach, 10 mimicry candidates were chosen for further examination to predict the development of Lyme arthritis. (**Table 1**)

Table 1: The following 10 protein candidates showed strong homology between *B. burgdorferi* and human.

Sl no.	ID's of Bacterial Protein	Name of the protein	ID's of Human Protein	Name of the protein
1	sp|P0C922|DNAK_BORBU	Chaperone protein DnaK	sp|P38646|GRP75_HUMAN	Stress-70 protein, mitochondrial
2	sp|O30913|EFG1_BORBU	Elongation factor G 1	sp|Q96RP9|EFGM_HUMAN	Elongation factor G, mitochondrial
3	sp|P0C923|CH60_BORBU	Chaperonin GroEL	sp|P10809|CH60_HUMAN	60 kDa heat shock protein, mitochondrial
4	sp|P50062|EFTU_BORBU	Elongation factor Tu	sp|P49411|EFTU_HUMAN	Elongation factor Tu, mitochondrial
5	sp|O51312|ENO_BORBU	Enolase	sp|P09104|ENOG_HUMAN	Gamma-enolase
6	sp|O30564|NAGB_BORBU	Glucosamine-6-phosphate deaminase	tr|D6R9P4|D6R9P4_HUMAN	Glucosamine-6-phosphate isomerase
7	sp|P46795|G3P_BORBU	Glyceraldehyde-3-phosphate dehydrogenase	sp|P04406|G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase
8	sp|O51214|RF1_BORBU	Peptide chain release factor 1	sp|Q9UGC7|RF1ML_HUMAN	Peptide chain release factor 1-like, mitochondrial
9	sp|O51602|GPMA_BORBU	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	sp|P18669|PGAM1_HUMAN	Phosphoglycerate mutase 1
10	sp|O51556|CLPP1_BORBU	ATP-dependent Clp protease proteolytic subunit 1	sp|Q16740|CLPP_HUMAN	ATP-dependent Clp protease proteolytic subunit, mitochondrial

3.4 Identification of molecular enrichment pathways

PathDIP is a bioinformatics tool that integrate the molecular pathways involved in the interaction of proteins, for identifying substantial correlations between curated pathways and proteins, by reducing proteins number without the annotation of the curated pathways [24]. This PathDIP database was used for examining the 10 human protein sequences that were similar to *B. burgdorferi* proteins. The Uniprot IDs of the human proteome sequences were inserted into the entry box. Here, all the 22 pathway sources (ACSN2, HumanCyc, KEGG, Panther_Pathway, RB-Pathways, BioCarta, INOH, NetPath, PharmGKB, REACTOME and others) were used along with the extended pathway associations whose parameters were set to experimentally detected PPIs and 99% cutoff value. After analyzing these 10 nominated proteins, it was found that 8 were pathway based proteins and the rest 2 were pathway independent proteins. Overall, 437 enrichment pathways were found in PathDIP which was further filtered using p value less than 0.005 and prioritized 238 pathways that might be connected to an autoimmune disorder. Enrichment pathways like apoptosis and NF-κB signaling are associated with autoimmunity. If the NF-κB activation is deregulated, then it can lead to the activation of aberrant T-cell, which is highly linked to autoimmune disease [25]. Apoptosis is a key regulator of self- reactive B cells that develop through mutation as blocking Bcl-2-dependent pathway results in autoimmune disorder [26]. Other enrichment pathways like metabolism, glycolysis, cell signaling were also common in the extended pathways.

3.5 Subcellular localization of proteins

In any biological field, protein subcellular localization prediction is essential because it aims to pinpoint previously undiscovered protein localization sites in a cell. Since conventional protein subcellular localization techniques are labor-intensive as well as requires a lot of time, computational methods are required [27]. For this research, PSORTb was used as it has persisted most accurate predictor of bacterial protein subcellular localization [21]. The 8 pathway based proteins will be inserted into this online based tool to understand its localization sites (**Table 2**). Each localization site was given a score out of 10. The result shows that all of these 8 bacterial proteins were identified as cytoplasmic.

Table 2: Identification of protein subcellular localization along with its corresponding score.

Sl. No	*B. burgdorferi* IDs	Bacterial Protein	Sub-Cellular localization	Localization Score
1	sp|P0C922|DNAK_BORBU	Chaperone protein DnaK	Cytoplasmic	9.97
2	sp|O30913|EFG1_BORBU	Elongation factor G 1	Cytoplasmic	10.00
3	sp|P0C923|CH60_BORBU	Chaperonin GroEL	Cytoplasmic	9.97
4	sp|P50062|EFTU_BORBU	Elongation factor Tu	Cytoplasmic	10.00
5	sp|O51312|ENO_BORBU	Enolase	Cytoplasmic	10.00
6	sp|P46795|G3P_BORBU	Glyceraldehyde-3-phosphate dehydrogenase	Cytoplasmic	10.00
7	sp|O51602|GPMA_BORBU	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	Cytoplasmic	8.96
8	sp|O51556|CLPP1_BORBU	ATP-dependent Clp protease proteolytic subunit 1	Cytoplasmic	9.26

3.6 Peptide binding to MHC class II

The IEDB's objective is to offer access to tools for studying epitopes that have been well described and tested via a standardized web interface. Users can quickly compare and contrast different prediction systems using this uniform interface. So, 8 pathway based proteins were further analyzed using this IEDB search engine. After searching, a large dataset of these 8 proteins were achieved which includes HLA alleles, 15-mer peptides, 9-mer core residues and others. It was mainly used to predict the epitopes of *B. burgdorferi* and humans that binds with class II HLA-DR, HLA-DQ and HLA-DP. NN-align is an innovative artificial neural network-based technique that identifies both the binding core and binding affinity of MHC class II simultaneously. Hence, a threshold of half-maximal inhibitory concentration (IC50) less than 50nM were considered as powerful binders of MHC II in human and *B. burgdorferi* peptides [28]. The HLA binders can be considered as T-cell epitope when IC50 less than 50 nM. HLA molecules found in IEDB search were DPA1*03:01/DPB1*04:02,

DQA1*05:01/DQB1*03:01, DRB1*01:01, DRB1*07:01, DRB1*09:01, DRB3*01:01 and HLA-DRB4*01:01. Peptides that were promiscuous binders of these HLA molecules had similar sequence with human epitopes. The epitopes revealed from these HLA molecules are prone to immune response. The ideal peptide-HLA allele binding relationship was mediated by a 15-mer peptide with 9-mer core residues. A total of 288 bacterial epitopes and 291 human epitopes was found. A threshold of more than 2 gaps in the alignment of BLASTp were discarded when compared with the 9-mer core residues of both organism which resulted 31 bacterial peptide and 33 human peptide (**Table 3**). Finally, these strong binder peptides were given top priority for the structural analysis.

Table 3: The match of the 9-mer core residues, were chosen by ≤2 gaps in the alignment of BLASTp, between host and bacteria. Based on these 11 matching residues a total of 31 bacterial peptide and 33 human peptide were chosen.

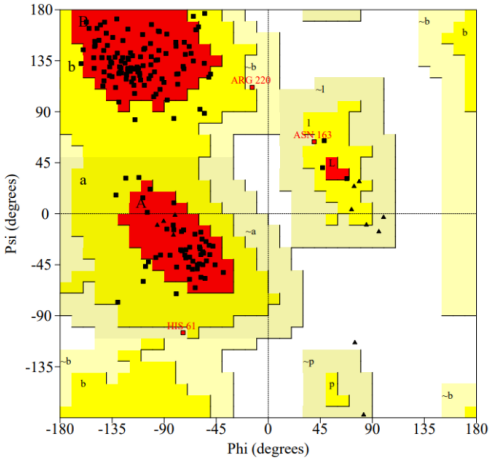
Sl no.	**Bacterial epitope core**	**Matched human sequence**
1	NRILGNFIL	NKLLGQFTL
2	GERLVGQVA	GERLVGMPA
3	ADANGPKHL	MDSSGPKHL
4	KLRNIGISA	KIRNIGISA
5	YVRIYQGIL	YVRSYQGEL
6	FTIEVERSL	FTIEVERAL
7	TTLTAAISI	TTLTAAITK
8	FQRTKPHMN	YVRDKPHVN
9	LRLKKLIID	MKLKKQLYN
10	TVIAQLLFL	LVIAQLLFL
11	VIAQLLFLE	VIAQLLFLQ

3.7 Evaluation of the predicted protein structures

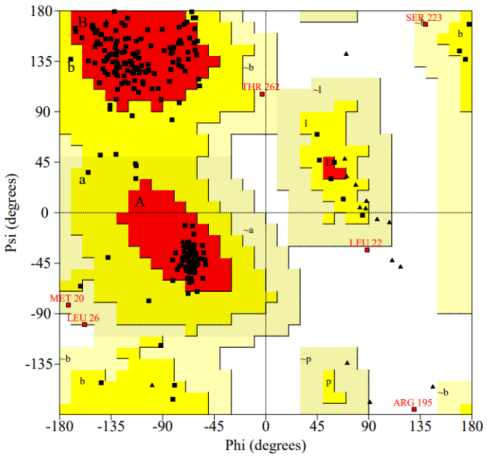
All potential protein structures in terms of torsion angles of amino acids φ (phi) and ψ (psi) are represented graphically in a 2D format by the Ramachandran plot. The PROCHECK evaluates the stereo-chemical quality of a predicted model by analyzing the overall model geometry along with the geometry of each individual residue. Ramachandran plot is produced by PROCHECK tool using input of modeled protein file and predict the quality of the structural protein. A high-quality Ramachandran diagram includes the majority of torsional angles within the allowed region whereas a protein structure with poor quality or low resolution has a significant number of torsional angles in the disallowed region [36][37]. Hence, the parameters were recorded for the assessment of protein structure (**Table 4**).

Table 4: Ramachandran plot's parameters which were recorded for protein assessment.

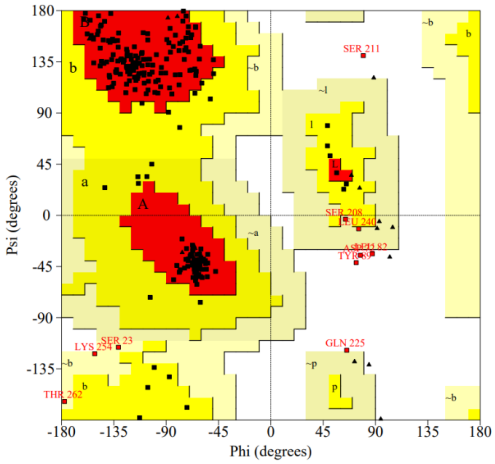
Allele (size)	**Residues in most favored regions (%)**	**Residues in additional allowed regions**	**Residues in generously allowed regions**	**Residues in disallowed regions**
HLA-DRB3*01:01	87.50%	10.70%	1.80%	0.00%
HLA-DRB1*01:01	82.90%	14.50%	1.80%	0.90%
HLA-DRB4*01:01	86.10%	9.60%	2.20%	2.20%
HLA-DPA1*03:01	81.90%	11.90%	2.20%	4.00%
HLA-DRB1*07:01	88.10%	10.20%	0.90%	0.90%
HLA-DRB1*09:01	87.70%	9.70%	1.30%	1.30%
HLA-DPB1*04:02	81.50%	16.10%	1.80%	0.60%
HLA-DQA1*05:01	86.70%	10.60%	1.80%	0.90%
HLA-DQB1*03:01	84.80%	8.20%	4.80%	2.20%



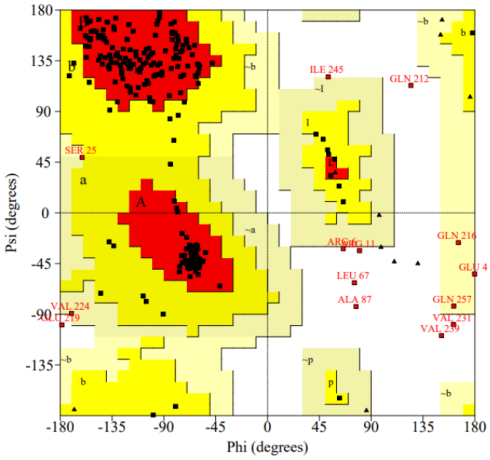
1. HLA-DRB3*01:01



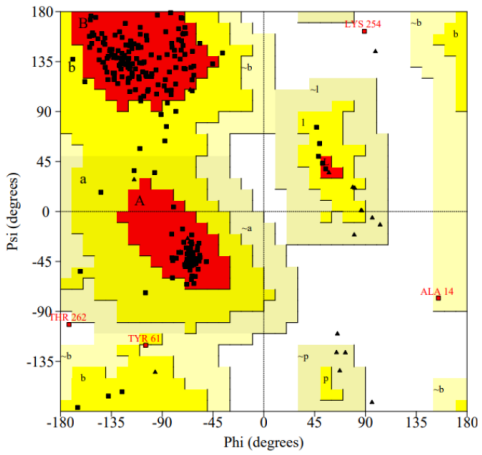
2. HLA-DRB1*01:01



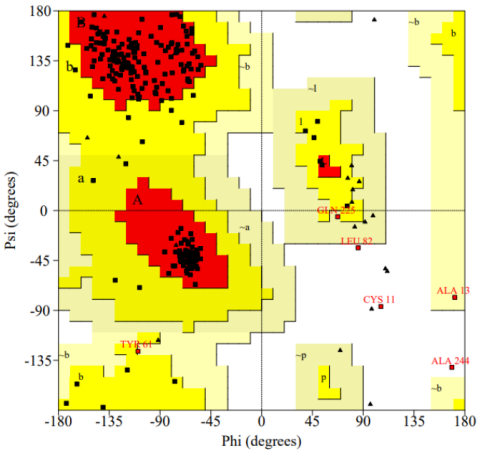
3. HLA-DRB4*01:01



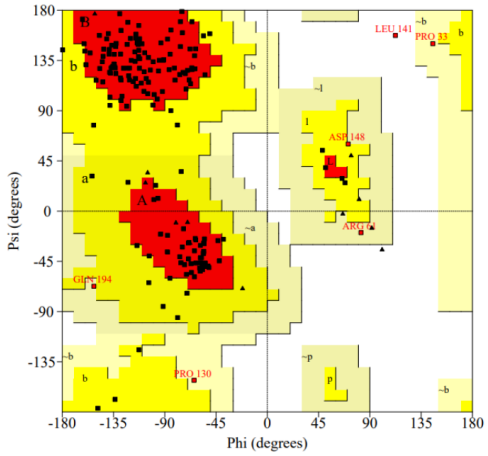
4. HLA-DPA1*03:01



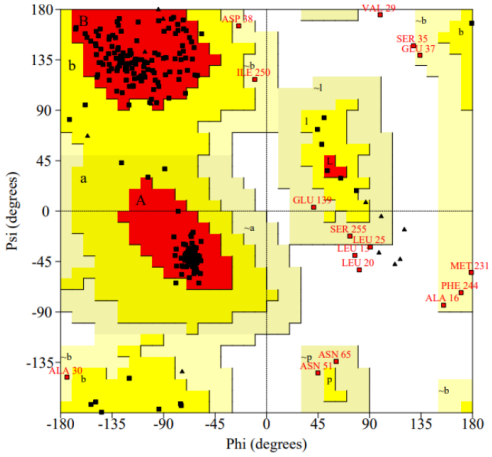
5. HLA-DRB1*07:01



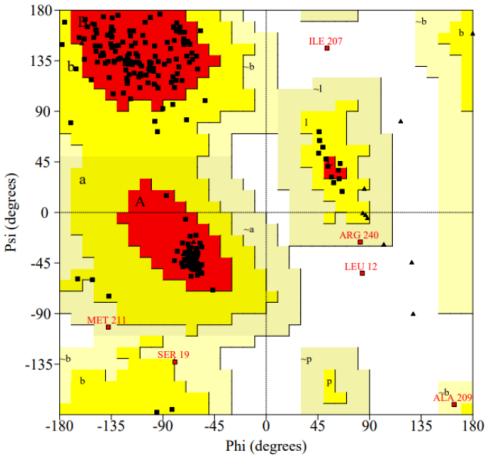
6. HLA-DRB1*09:01



7. HLA-DPB1*04:02



8. HLA-DQB1*03:01



9. HLA-DQA1*05:01

Figure 3: Ramachandran plot showing the all regions of the residues through X-ray crystallography. Black square represents the protein residues where the red is the most favored region, bright yellow is additional allowed regions, pale yellow is generously allowed regions and finally white is the disallowed regions.

3.8 Analysis of the peptide structure

RPBS is a web based bioinformatics structural tool, used for the retrieval of protein data bank (PDB) files from the final mimicry candidates of 31 bacterial and 33 human peptides. So the 15mer- peptides of each epitope were inserted into the PEP-FOLD3 of RPBS portal and downloaded the first PDB file of the predicted epitope from the cluster section. This helped to predict the 3D strucute of each 15- mer peptides (**Figure 4**).

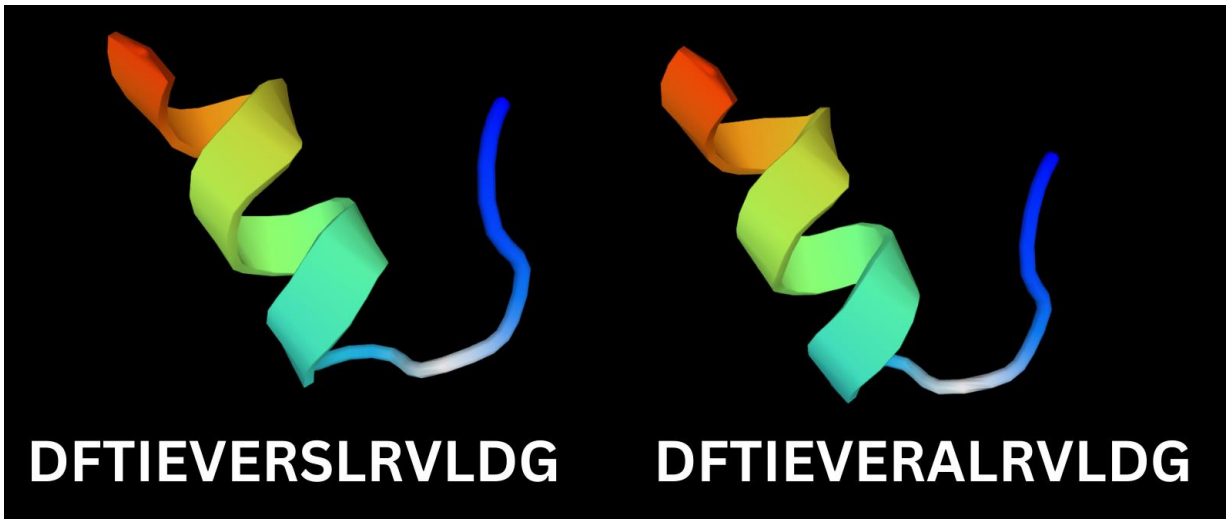
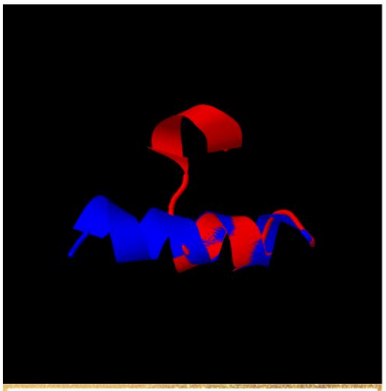


Figure 4: Three-dimensional structure representation of two 15-meric peptides. ("DFTIEVERSLRVLDG" is a bacterial epitope whereas "DFTIEVERALRVLDG" is an epitope of human).

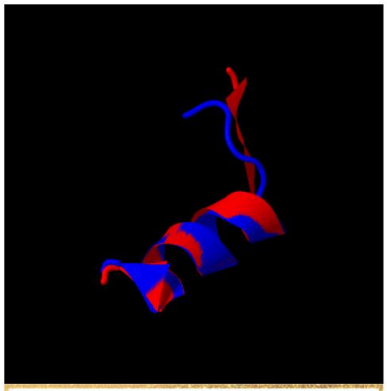
Subsequently, these PDB files of *B. burgdorferi* and human are compared for pair-wise structure in TM-align. The structural alternatives are ranked based on a template modelling (TM) score that considers the RMSD value as well **(Table 5)**. The structural similarity between *B. burgdorferi* and human are of various levels. The compositional similarity between the 3D structures of human and *B. burgdorferi* proteins is determined by the TM-score. The top-ranked bacterial model and the human are compared using the TM score to assess structural similarity.[34] A threshold of less than 0.5 TM value were filtered as higher than 0.5 are assumed to have the same fold. This resulted 15 most strong superimposed structure **(figure 5)** which will be further used for docking. The most often used method for comparing protein structures is to compute RMSD of all identical atom pairs after the optimal superposition of the two structures. The two peptide structures are structurally equivalent if the RMSD is low. Due to their low RMSD values, bacterial and human peptides show structural similarities. These similarities lead to structural mimicry, and these peptides might be helpful for triggering autoimmune diseases.

Table 5: *B. burgdorferi* and the human host's MHC class II binding peptides (":" denotes the distance between residue pairs).

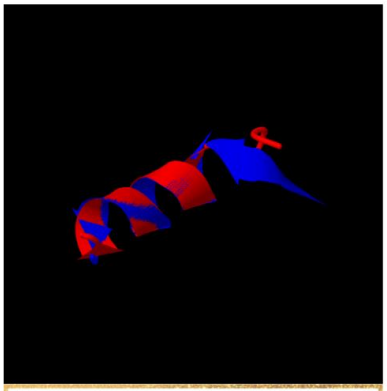
Sl No.	Bacterial Peptide	Human Peptide	Structural Mimicry	RMSD values	TM value
1	KGERLVGQVAKNQMV	GERLVGMPAKRQAVT	KGERLVGQVAKNQMV ::::::::: GERLVGMPAKRQAVT	0.18	0.53497
2	GQLTYVRIYQGILKK	GQLTYVRSYQGELKK	GQLTYVRIYQGILKK .:::::::::::::: GQLTYVRSYQGELKK	1.66	0.6027
3	QLTYVRIYQGILKKG	FGQLTYVRSYQGELK	QLTYVRIYQGILKKG ::::::::::::: FGQLTYVRSYQGELK	1.27	0.55829
4	DFTIEVERSLRVLDG	DFTIEVERALRVLDG	DFTIEVERSLRVLDG ::::::::::::::: DFTIEVERALRVLDG	0.3	0.79365
5	FTIEVERSLRVLDGA	FTIEVERALRVLDGA	FTIEVERSLRVLDGA .::::::::::::::. FTIEVERALRVLDGA	1.74	0.66147
6	HVDFTIEVERSLRVL	HVDFTIEVERALRVL	HVDFTIEVERSLRVL .:::::::::::: HVDFTIEVERALRVL	1.5	0.60421
7	VDFTIEVERSLRVLD	VDFTIEVERALRVLD	VDFTIEVERSLRVLD .:::::::::::::: VDFTIEVERALRVLD	1.12	0.53377
8	HVDHGKTTLTAAISI	HGKTTLTAAITKILA	HVDHGKTTLTAAISI .::::::::: HGKTTLTAAITKILA	1.73	0.50264
9	DHGKTTLTAAISIYC	DHGKTTLTAAITKIL	DHGKTTLTAAISIYC :::::::::::: DHGKTTLTAAITKIL	0.91	0.64929
10	HGKTTLTAAISIYCS	HGKTTLTAAITKILA	HGKTTLTAAISIYCS .:::::::::::: HGKTTLTAAITKILA	1.71	0.56414
11	VDHGKTTLTAAISIY	HGKTTLTAAITKILA	VDHG-KTTLTAAISIY .:::::::::: HGKTTLTAAITKILA	1.37	0.55772
12	LRLKKLIIDIMSNQI	MKLKKQLYNIYAKHT	LRLKKLIIDIMSNQI ::::::::::::::: MKLKKQLYNIYAKHT	0.27	0.80108
13	TVIAQLLFLESEDSS	LVIAQLLFLQSESNK	TVIAQLLFLESEDSS ::::::::::::.. LVIAQLLFLQSESNK	1.97	0.6788
14	VIAQLLFLESEDSSK	VIAQLLFLQSESNKK	VIAQLLFLESEDSSK ::::::::::::::: VIAQLLFLQSESNKK	0.37	0.69614
15	IAQLLFLESEDSSKD	IAQLLFLQSESNKKP	IAQLLFLESEDSSKD ::::::::: IAQLLFLQSESNKKP	0.98	0.5189



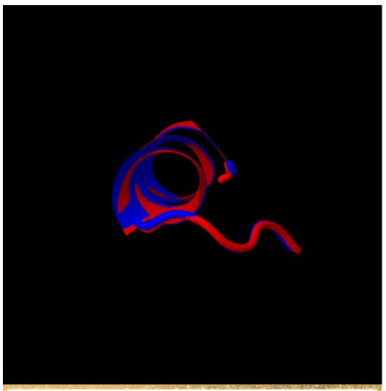
1 "KGERLV------GQVAKNQMV
::::::::
------GERLVGMPAKRQAVT"



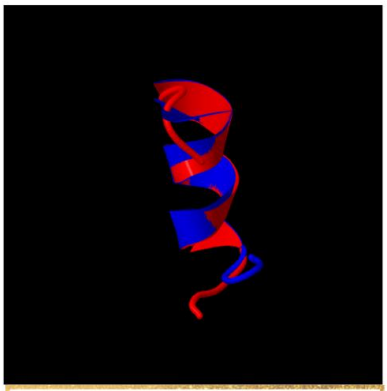
2 "GQLTYVRIYQGILKK
.:::::::::::::::
GQLTYVRSYQGELKK"



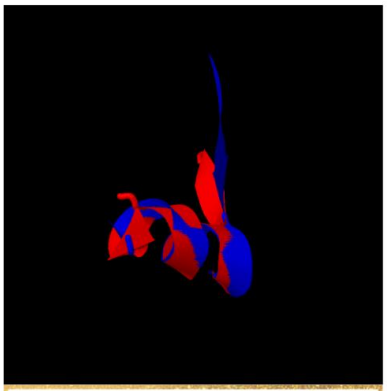
3 "QLTYVRIYQGILKKG-
:::::::::::::::
-FGQLTYVRSYQGELK"



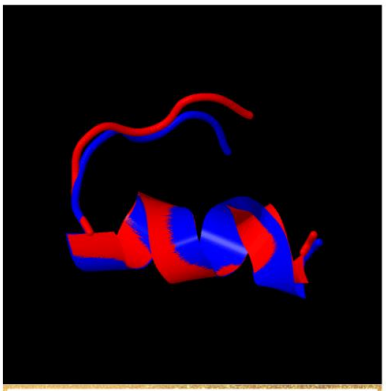
4 "DFTIEVERSLRVLDG
:::::::::::::::
DFTIEVERALRVLDG"



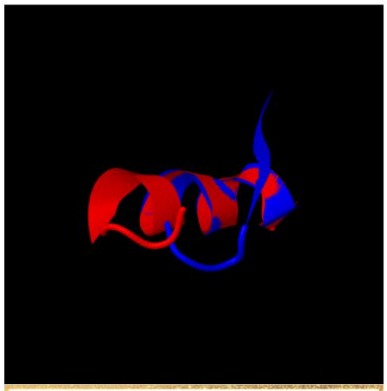
5 "FTIEVERSLRVLDGA
.:::::::::::::::.
FTIEVERALRVLDGA"



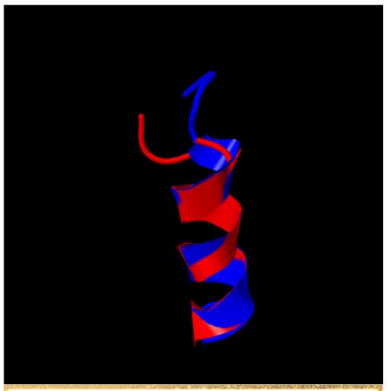
6 "HVD--FTIEVERSLRVL-
.::::::::::::
---HVDFTIEVERALRVL"



7 "V-DFTIEVERSLRVLD
.:::::::::::::
-VDFTIEVERALRVLD"



8 "HVDHG-----KTTLTAAISI
::::::::::
-----HGKTTLTAAITKILA"



9 "DHG--KTTLTAAISIYC
: ::::::::::::
D--HGKTTLTAAITKIL"

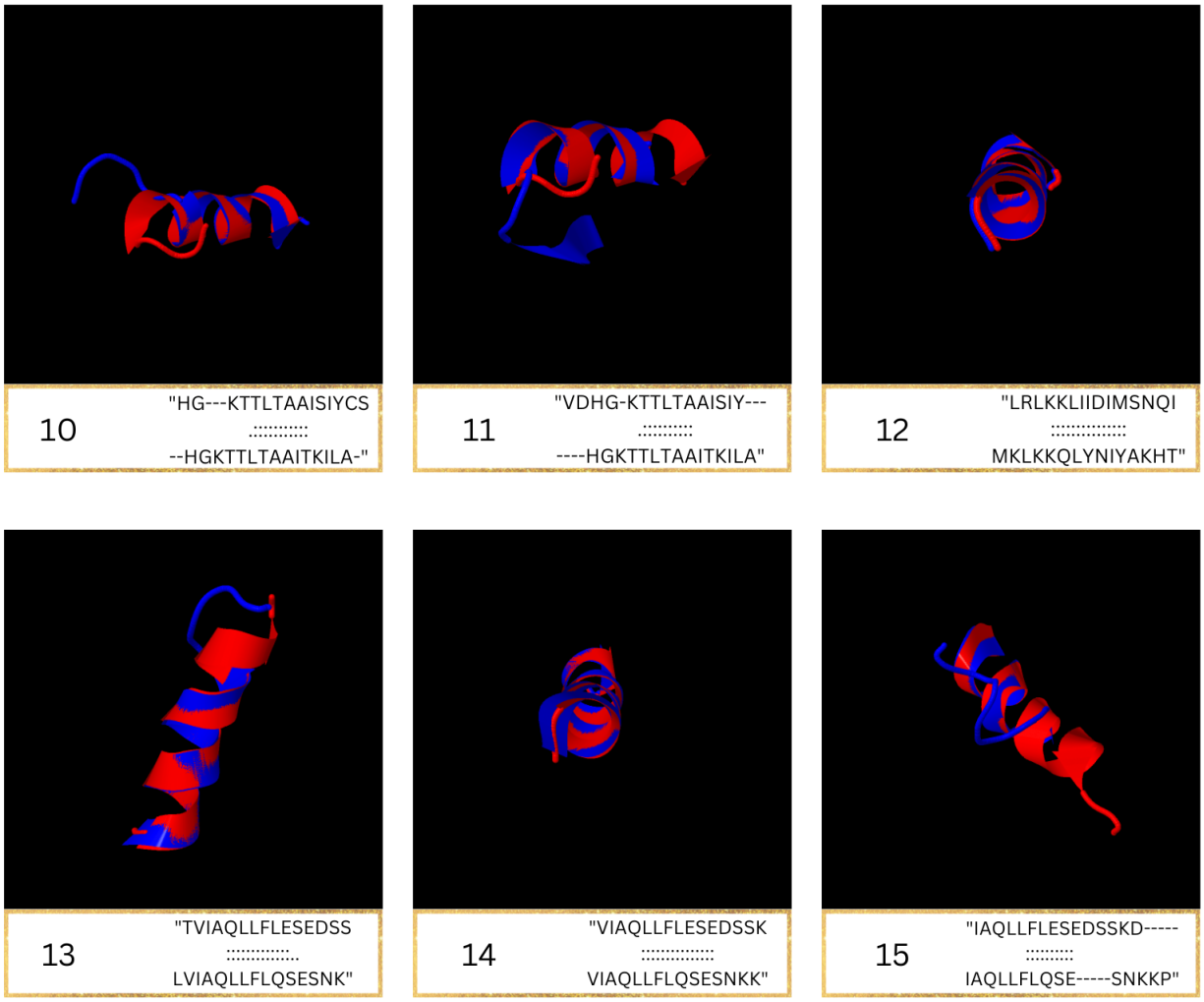


Figure 5: Superposition of two epitopes (bacterial in blue and human in red)

3.9 Molecular docking between bacterial epitope and HLA allele

The HDOCK tool uses a hybrid docking approach to predict the binding complexes between two molecules (receptor and ligand). The 15 final candidates of the bacterial epitopes were docked against human class II HLA alleles. The Docking Candidates produced by the HDOCK method were thoroughly tested. The top 10 models are redirected which aims to find the complex, relatively close solutions based on confidence score, docking scores, and ligand RMSDs (**Table 6, Figure 6)** Out of these, the most negative docking score choose the best model following the confidence score above 0.5 . Subsequently, the most effective peptides, HLA alleles can be chosen. For all life science fields, this technique is simple to implement. All peptides obtained the highest docking scores and most favorable molecular interactions in

various HLA receptors, according to the docking results. Studies using docking scores revealed a correspondence between the binding affinities of peptides to HLA alleles.

Table 6: The binding affinity of bacterial peptide with HLA was revealed by the best molecular docking score.

Sl no.	Receptor	Ligand	Docking Score	Confidence Score	Ligand RMSD
1	HLA-DRB1*01:01	KGERLVGQVAKNQMV	-194.42	0.7086	40.5
2	HLA-DRB1*01:01	GQLTYVRIYQGILKK	-238.36	0.8541	41.28
3	HLA-DRB1*01:01	QLTYVRIYQGILKKG	-208.22	0.7621	36.04
4	HLA-DRB1*01:01	DFTIEVERSLRVLDG	-156.18	0.5309	41.53
5	HLA-DRB1*01:01	FTIEVERSLRVLDGA	-171.26	0.6047	19.49
6	HLA-DRB1*01:01	HVDFTIEVERSLRVL	-191.44	0.6961	41.61
7	HLA-DRB1*01:01	VDFTIEVERSLRVLD	-183.18	0.6601	32.66
8	HLA-DRB1*07:01	HVDHGKTTLTAAISI	-173.04	0.6132	42.93
9	HLA-DRB1*07:01	DHGKTTLTAAISIYC	-210.37	0.7698	44.34
10	HLA-DRB1*07:01	HGKTTLTAAISIYCS	-221.31	0.8063	43.12
11	HLA-DRB1*07:01	VDHGKTTLTAAISIY	-193.87	0.7063	46.98
12	HLA-DRB1*07:01	LRLKKLIIDIMSNQI	-200.34	0.7324	44.43
13	HLA-DRB4*01:01	TVIAQLLFLESEDSS	-168.86	0.5932	43.03
14	HLA-DRB4*01:01	VIAQLLFLESEDSSK	-164.5	0.572	41.02
15	HLA-DRB4*01:01	IAQLLFLESEDSSKD	-157.54	0.5376	40.04

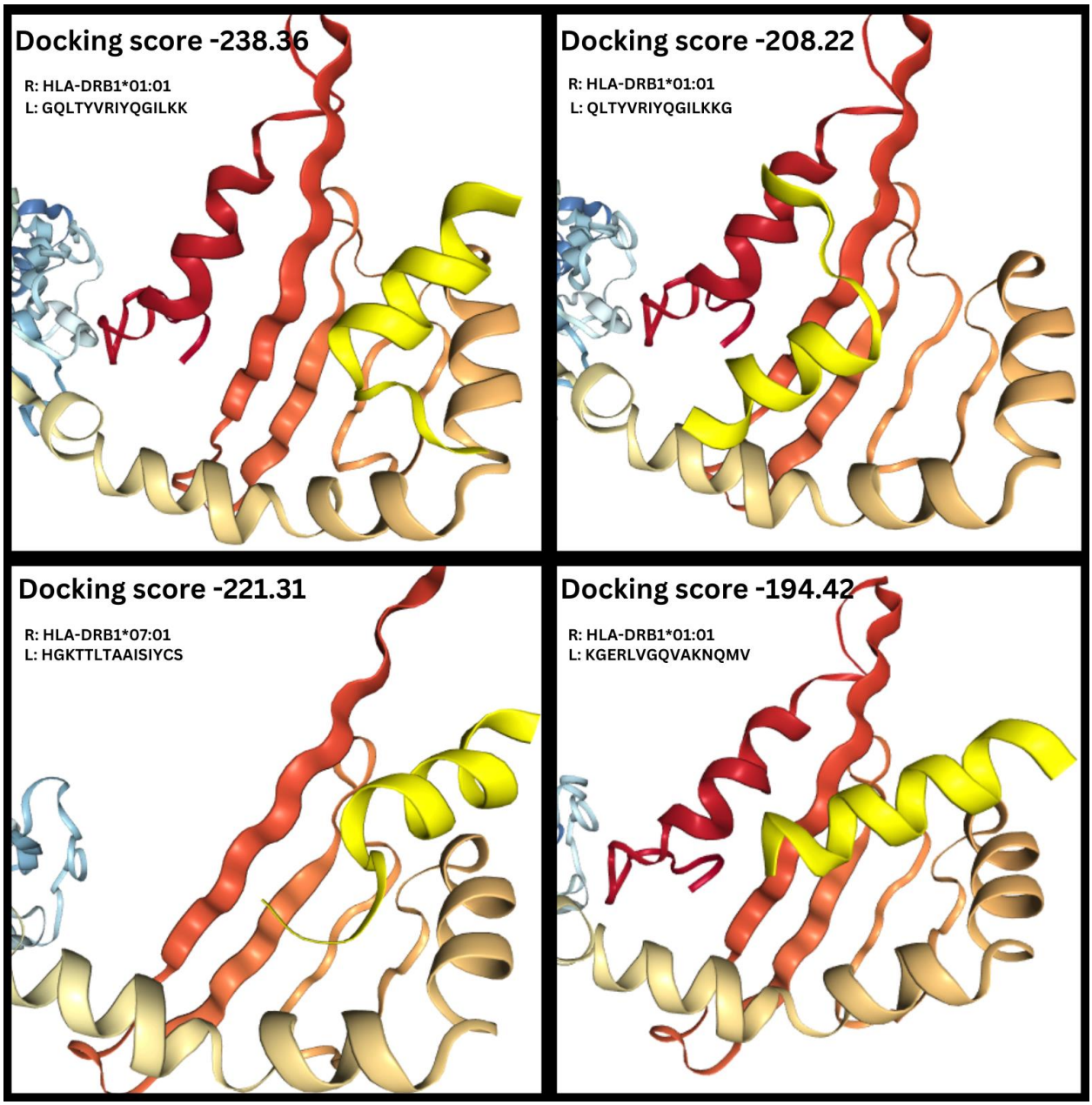


Figure 6: 3D docking of bacterial peptides with human leukocyte antigen. Here, R (Receptor) shown in rainbow and L (ligand) in yellow.

Chapter 4

Discussion

When the immune system interacts with pathogenic antigens, antibodies are produced to fight against infection. However, these antibodies can also cause autoreactive reactions, which are mediated by autoantibodies that react to both pathogenic and human antigens because they share epitopes. Molecular mimicry refers to the possibility of microorganisms that can circumvent the host immune response by using antigens that are similar to those expressed by infectious pathogens and their human hosts. There are now four main criteria used to explain molecular mimicry. These include epitope similarity between host and pathogens; detection of autoreactive T cells or antibodies that cross-react with both human and microbial epitopes; epidemiological evidence correlated autoimmune disease development with exposure to environmental factors or microorganisms; and development of autoimmunity in an animal model after exposure with the proper epitopes [30]. Several autoimmune disorders, including rheumatoid arthritis, Guillain-Barre syndrome, autoimmune thyroid disease, and systemic sclerosis, have been linked to cross-reactivity between pathogens and humans antigens. The identical epitopes between host and pathogen shows the association of cross reactivity with molecular mimicry as it may encourage stimulation of T or even B cells that are autoreactive [32].

The putative mechanism for Lyme neuroborreliosis and Lyme arthritis has been suggested to be molecular mimicry, which has the potential to trigger an autoimmune response. Two proteins that causes cross-reactivity in *B. burgdorferi* and human include heat shock protein 60 (HSP60) and leucocyte function-associated antigen 1 (LFA-1). When humans develop immunity against bacterial HSP60, the autologous HSP60 may results in the adhesion of HSP60-reactive T cells for the early initiation of inflammatory response, where antibodies against HSP60 may cause the atherosclerosis to spread and worsen [33, 53]. Moreover, T cell from a patient having rheumatoid arthritis that is resistant to treatment can identify the LFA-1 residues that have sequence homology with OspA. This suggests that bacteria and the genetic sensitivity of the host are important factors in autoimmunity, which are responsible for mimicry and similarity in proteins. Up to 10% of people have the organ-specific AD known as autoimmune thyroid disease (AITD), which is first driven by T cells and later by B cells. Recent

studies have shown amino acid similarities between human thyroid-stimulating hormone receptor (TSH-R) with the bacterial (*B. burgdorferi*) outer surface protein A, flagellar motor rotation protein A, and DNA recombinase or ATP dependent helicase. These showed a range from 27 to 50% identity and 40 to 75% similarity [33]. Pathogenic role in the emergence of AITD through molecular mimicry is also demonstrated by homologies between *B. burgdorferi* proteins and thyroid autoantigens. These include hTSH-R, hNIS, hTg, and human thyroid peroxidase. To strengthen this notion for various microbiological causes of this AD, it can be determined that a variety of infectious pathogens can trigger immune responses that are reactive to thyroid tissue antigens [31]. In this study, a variety of proteins with similar sequences from humans and *B. burgdorferi* were analyzed. With immunoinformatic techniques, it was discovered that the immunodominant epitopes of these proteins share structural similarities and may possess the ability to bind to human T-cell immune receptors.

PathDIP showed numerous pathways have an excess of similar proteins. The immune system's ability to determine its lineage and function as well as the progression of autoimmune disorders that are governed by metabolic pathways. The most crucial pathways include TNF alpha Signaling, NFkB Signaling Network, MAPK Signaling, Toll-Like receptor(TLR) signaling network, glycolysis, gluconeogenesis, IL-7 signaling and apoptosis. The majority of evidence points to the TLR pathway as a critical factor in the emergence of autoimmune diseases. It is important to consider that ongoing TLR signaling activation or dysregulation fuels the development of autoimmune illness [52].

Heat shock proteins (HSP) are found in the cytoplasm and different organelles where they function as chaperone or even protease. Both bacterial and autologous HSP are highly immunogenic in origin. In addition, immunoreactivity to the Hsp60 chaperone primarily aims to inhibit the immunological response. Several models of mice having autoimmune arthritis that resemble juvenile idiopathic arthritis (JIA) and rheumatoid arthritis (RA) in humans have proven this. In the murine arthritic model, disease regulation was limited to significantly conserved HSP60 derived epitopes used for animal vaccination and thus anergic T cells with immunosuppressive potential were reactive to self-HSP60 [40]. Additionally, it has been demonstrated that EF-Tu binds specifically to substance P (SP), a neuropeptide hormone. Because it triggers the proinflammatory response whenever the binding between SP and

neurokinin 1 receptor (NK-1R) takes place which is associated with many inflammatory illnesses. In mouse models, the hindrance of the activation of NK-1R reduces removal of clearance and raises mortality rates. Infectious disease, autoimmune disorders, psychological problems, cancer, atopic dermatitis, and cell proliferation have all been associated to SP and NK-1R [41]. Patients with autoimmune retinopathy frequently have vision loss in conjunction with autoantibodies against alpha-enolase. Apoptosis of the retinal cells from bipolar and photoreceptors is demonstrated to result from antibodies of anti-recoverin that penetrates the layers of retina correlating to the recoverin cellular site. The inner nuclear layer and retinal ganglion cells were uniquely marked by autoantibodies to alpha-enolase. Scientists found that antibodies were able to penetrate the tissue of retina for targeting the ganglion cells and inner nuclear layers which stimulated the death of cell through an inflammatory mechanism [42]. Glyceraldehyde 3-phosphate dehydrogenase was one of the components of the complexes that specifically reacted with SLE sera but not with sera from individuals with other connective tissue illnesses. Immunofluorescence on Hep-2 cells revealed that the isolated autoantibodies against GAPDH from lupus serum displayed both nuclear speckled and cytoplasmic staining patterns. In addition, an ELISA test showed that 47% of lupus patients had anti-GAPDH autoantibodies. The autoimmune reaction expanded from GAPDH to additional components of PCNA complexes, according to a longitudinal investigation of lupus sera's reactivity to those proteins. In addition to it, a strong positive correlation was found between the elevated serum PCNA levels and the presence of anti-GAPDH antibodies. Thus, GAPDH is crucial for triggering autoimmune reactions against the PCNA complex [43]. Host proteins including Stress-70 protein, Elongation factor G, 60 kDa heat shock protein, Elongation factor Tu, Peptide chain release factor 1-like and ATP-dependent Clp protease proteolytic subunit originated from mitochondria that are homologues to bacterial protein. This shows according to the endosymbiosis theory that mitochondria in a eukaryotic cell evolved from bacterial cells for which most of the homologous peptide belongs from mitochondria [59]. Many studies showed that mitochondria dysfunction induces autoimmune diseases including rheumatoid arthritis, systemic lupus erythematosus and diabetes. Although its mechanism is unknown but showing a strong evidence between autoimmunity and mitochondrial proteins [54].

Binding affinity and specificity are substantially controlled by the binding peptide interaction's 9-mer core region. Similarly, it has been suggested that the peptide residues immediately surrounding the core area established a connection with the molecules of MHC outside of the

binding groove which contribute to the interaction of MHC-peptide. MHC class II molecules present the processed antigens to CD4+ T lymphocytes derived exogenous pathogenic sources and crucial for antigen-specific immune response. In addition, MHC Class II molecules have an open binding groove on both ends, unlikely MHC class I only has one end. Because of this, the peptides length binding to the molecules of class II can vary greatly. Normally amino acids ranges between 13 and 25. Therefore, the ability to locate the proper 9-mer core residues that mediate the binding contact among longer sequences is a requirement for all MHC class II binding prediction methods [50, 56]. Autoimmune diseases are likely caused by MHC class II alleles because the capacity of transmitting peptides through different alleles derived from target cells is numerous to auto-reactive CD4 + T cells. The survival rate of a certain population is likely to be higher, if they have a wide variety of HLA class II alleles. In addition to it, they will have a better immune system to fight against various disease [51]. IEDB-AR resulted the profiles of bacterial epitopes that bind with the peptides of MHC class II peptide to different HLA molecules. Naturally HLA-DR-B1*0401 allele produces a peptide sequence of QKRAA that was found to be identical with *Escherichia coli* heat shock protein such as DnaJ. In addition to it, an elevated frequency of this allele exhibited in individuals having extreme type of Rheumatoid arthritis. It has been demonstrated that this DnaJ QKRAA motif binds bacterial HSP70s. This evidence proves that exposure to enterobacteria causes HLA-DRB1*0401 molecule to express in patients for binding DnaK proteins with QKRAA . Therefore, it causes the proliferation of T cells for reacting with HSP70's as it shares a fixed epitope with human type II collagen [44]. The scientists used an in silico method to discover a homology between thyroid autoantigens and botulinum neurotoxin, which displayed the binding patterns in the regions of HLA-DR3 or HLA-DR7 [45]. Numerous genes in the HLA area were identified for the development of autoimmune hepatitis (AIH). HLA-DQ-B1 has been linked to AIH, but HLA-DR-B1 encoded by the MHC class II molecule was primarily recognized to be the susceptibility gene. Prior viral infections with the infectious agents, specifically CMV, hepatitis virus, varicella zoster virus, and EBV have all been connected to an elevated risk of AIH [30]. Although HLA-DRB1*09:01 is uncommon in Caucasian cultures, Asians have a high prevalence of this HLA-DRB1 allele. Furthermore, HLA-DRB1*09:01 has been linked to a number of autoimmune disorders, including type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus. Therefore, in Japanese and possibly other Asian cultures, HLA-DRB1*09:01 or other genes in tight linkage disequilibrium may contribute to a biological pathway that connects several autoimmune disorders [46]. A borrelial fibronectin binding protein (BBK32), demonstrated a strong DRB binding profile related to refractory arthritis.

Similar to the OspA peptide, the DRB1*0101 molecule and the B5*0101 molecule were mostly bound by the BBK32 peptide. Later, it was identified that the both DRB1*0701 and DRB4*0101 has minimal interaction, that is frequently present in individuals with refractory arthritis, in contrast to the OspA peptide [47].

In this research, three important superposition pairs of *B. burgdorferi* and human host epitopes were identified in Tm-align tool that showed 100% binding based on TM score and RMSD value through a three dimensional modelling. These homologous pairs of bacteria and host include DFTIEVERSLRVLDG & DFTIEVERALRVLDG; LRLKKLIIDIMSNQI & MKLKKQLYNIYAKHT and VIAQLLFLESEDSSK & VIAQLLFLQSESNKK has high possibilities for inducing autoimmune disease. Here, the TM scores of these pairs were above 0.5 which indicates that the structures are roughly in the same fold. In addition, the chosen pairs RMSD values were below 0.4, as a low RMSD value indicates higher structural similarity between 2 epitopes. Furthermore, molecular docking provided information about the powerful binding complex through its docking score and confidence score. HDOCK tool revealed docking pairs, with respect to TM align, that has a highly negative docking score and confidence above 0.7. A highly negative docking score indicates more binding affinity between ligand and the target sequence as more energy released, whereas confidence score above 0.7 considered as a very similar binding between the two molecules. The results showed the docking pairs of the bacterial epitopes LRLKKLIIDIMSNQI, DFTIEVERSLRVLDG and VIAQLLFLESEDSSK with respect to the host alleles HLA-DRB1*07:01, HLA-DRB1*01:01and HLA-DRB4*01:01 have the highest possibility of molecular mimicry in Lyme disease. HLA-DRB1*01:01 have high prevalence with Asian cultures and strongly binds with a Borrelial peptide causing refractory rheumatoid arthritis. Additionally, HLA-DRB1*07:01 and HLA-DRB4*01:01 are common in patients with refractory arthritis. These experimental findings indicate that possibilities of molecular mimicry-based autoimmunity caused by *Borrelia burgdorferi* infection.

Conclusion

This study was done for identifying the three dimensional structural mimicry candidates between *B. burgdorferi* and human that is responsible for inducing autoimmunity. The study was initiated through the retrieval of proteome database and subsequently the bacterial protein sequences were BLAST against the host which revealed 10 protein mimicry candidates of *B. burgdorferi*. These peptides were promiscuous binders of HLA class II molecules and have a high chance to cause Lyme borreliosis. Using all these bioinformatics tools, we can better comprehend the immune system of *Homo sapiens* (humans) and can identify the novel for preventing and treating disease. All in all, improvements of detection techniques in molecular mimicry will present an interesting new area of research that needs to be explored further.

References

[1] Cusick, M.F., Libbey, J.E. & Fujinami, R.S. Molecular Mimicry as a Mechanism of Autoimmune Disease. Clinic Rev Allerg Immunol 42, 102–111 (2012). https://doi.org/10.1007/s12016-011-8294-7

[2] Luan M., Shang Z., Teng Y., Chen X., Zhang M., Lv H., Zhang R. The shared and specific mechanism of four autoimmune diseases. Oncotarget. 2017; 8: 108355-108374. Retrieved from https://doi.org/10.18632/oncotarget.19383

[3] Guan Q., Gao X., Wang J., Sun Y., Shekhar S. "Cytokines in Autoimmune Disease", Mediators of Inflammation, vol. 2017, Article ID 5089815, 2 pages, 2017. https://doi.org/10.1155/2017/5089815

[4] F.W. Miller, L. Alfredsson, K.H. Costenbader, D.L. Kamen, L.M. Nelson, J.M. Norris, A.J. De Roos, The findings from a National Institute of Environmental Health Sciences Expert Panel Workshop include the epidemiology of environmental exposures and human autoimmune diseases, J. Autoimmun. 39 (4) (2012) 259–271. https://doi.org/10.1016/j.jaut.2012.05.002

[5] Katherine A. Wolcott, Gabriele Margos, Volker Fingerle, Noémie S. Becker, Host association of *Borrelia burgdorferi* sensu lato: A review, Ticks and Tick-borne Diseases, Volume 12, Issue 5, 2021, 101766, ISSN 1877-959X, https://doi.org/10.1016/j.ttbdis.2021.101766.

[6] Kurokawa C, Lynn GE, Pedra JHF, Pal U, Narasimhan S, Fikrig E. Interactions between *Borrelia burgdorferi* and ticks. Nat Rev Microbiol. 2020 Oct;18(10):587-600. https://doi.org/10.1038%2Fs41579-020-0400-5

[7] Bolz DD, Weis JJ. Molecular mimicry to *Borrelia burgdorferi*: pathway to autoimmunity? Autoimmunity. 2004 Aug;37(5):387-92. https://doi.org/10.1080/08916930410001713098 . PMID: 15621562.

[8] Guerau-de-Arellano M, Huber BT. Development of autoimmunity in Lyme arthritis. Curr Opin Rheumatol. 2002 Jul;14(4):388-93. https://doi.org/10.1097/00002281-200207000-00009 . PMID: 12118172.

[9] Arvikar SL, Crowley JT, Sulka KB, Steere AC. Autoimmune Arthritides, Rheumatoid Arthritis, Psoriatic Arthritis, or Peripheral Spondyloarthritis Following Lyme Disease. Arthritis Rheumatol. 2017 Jan;69(1):194-202. doi: 10.1002/art.39866. PMID: 27636905; PMCID: PMC5195856.

[10] Pancewicz SA, Rutkowski R, Rutkowski K, Zajkowska JM, Kondrusik M. Zaburzenia immunologiczne w zapaleniu stawów w boreliozie z lyme [Immunopathology of Lyme arthritis]. Pol Merkur Lekarski. 2007 Aug;23(134):141-4. Polish. PMID: 18044347.

[11] Lochhead, R.B., Strle, K., Arvikar, S.L. et al. Lyme arthritis: linking infection, inflammation and autoimmunity. Nat Rev Rheumatol 17, 449–461 (2021). https://doi.org/10.1038/s41584-021-00648-5

[12] Stebbins, C., Galán, J. Structural mimicry in bacterial virulence. Nature 412, 701–705 (2001). https://doi.org/10.1038/35089000

[13] Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet. 2008 Apr 25;4(4):e1000024. https://doi.org/10.1371%2Fjournal.pgen.1000024 . PMID: 18437207; PMCID: PMC2291482.

[14] Izquierdo JH, Cañas CA, Tobón GJ. B lymphocytes in autoimmunity. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, et al., editors. Autoimmunity: From Bench to Bedside [Internet]. Bogota (Colombia): El Rosario University Press; 2013 Jul 18. Chapter 6. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459431/

[15] Steere AC, Gross D, Meyer AL, Huber BT. Autoimmune mechanisms in antibiotic treatment-resistant lyme arthritis. J Autoimmun. 2001 May;16(3):263-8. doi: 10.1006/jaut.2000.0495. PMID: 11334491.

[16] Drouin EE, Glickstein LJ, Steere AC. Molecular characterization of the OspA(161-175) T cell epitope associated with treatment-resistant Lyme arthritis: differences among the three pathogenic species of Borrelia burgdorferi sensu lato. J Autoimmun. 2004 Nov;23(3):281-92. doi: 10.1016/j.jaut.2004.06.005. PMID: 15501399.

[17] Gross DM, Forsthuber T, Tary-Lehmann M, Etling C, Ito K, Nagy ZA, Field JA, Steere AC, Huber BT. Identification of LFA-1 as a candidate autoantigen in treatment-resistant Lyme arthritis. Science. 1998 Jul 31;281(5377):703-6. doi: 10.1126/science.281.5377.703. PMID: 9685265.

[18] Palagi PM, Hernandez P, Walther D, Appel RD. Proteome informatics I: bioinformatics tools for processing experimental data. Proteomics. 2006 Oct;6(20):5435-44. doi: 10.1002/pmic.200600273. PMID: 16991191.

[19] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150-2. doi: 10.1093/bioinformatics/bts565. Epub 2012 Oct 11. PMID: 23060610; PMCID: PMC3516142.

[20] Rahmati S, Abovsky M, Pastrello C, Jurisica I. pathDIP: an annotated resource for known and predicted human gene-pathway associations and pathway enrichment analysis.

Nucleic Acids Res. 2017 Jan 4;45(D1):D419-D426. doi: 10.1093/nar/gkw1082. Epub 2016 Nov 28. PMID: 27899558; PMCID: PMC5210562.

[21] N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, F.S. Brinkman, PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics 26 (13) (2010) 1608–1615. Available from: https://doi.org/10.1093/bioinformatics/btq249

[22] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019 Jan 8;47(D1):D339-D343. doi: 10.1093/nar/gky1006. PMID: 30357391; PMCID: PMC6324067.

[23] T. Bhardwaj, S. Haque, P. Somvanshi, Silico identification of molecular mimics involved in the pathogenesis of Clostridium botulinum ATCC 3502 strain, Microb. Pathog. 121 (2018) 238–244. https://doi.org/10.1016/j.micpath.2018.05.017

[24] Rahmati S, Abovsky M, Pastrello C, Kotlyar M, Lu R, Cumbaa CA, Rahman P, Chandran V, Jurisica I. pathDIP 4: an extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species. Nucleic Acids Res. 2020 Jan 8;48(D1):D479-D488. doi: 10.1093/nar/gkz989. PMID: 31733064; PMCID: PMC7145646.

[25] Liu, T., Zhang, L., Joo, D. et al. NF-κB signaling in inflammation. Sig Transduct Target Ther 2, 17023 (2017). https://doi.org/10.1038/sigtrans.2017.23

[26] Mayer CT, Nieke JP, Gazumyan A, Cipolla M, Wang Q, Oliveira TY, Ramos V, Monette S, Li QZ, Gershwin ME, Kashkar H, Nussenzweig MC. An apoptosis-dependent checkpoint for autoimmunity in memory B and plasma cells. Proc Natl Acad Sci U S A.

2020 Oct 6;117(40):24957-24963. doi: 10.1073/pnas.2015372117. Epub 2020 Sep 22. PMID: 32963096; PMCID: PMC7547173.

[27] Liao Z, Pan G, Sun C, Tang J. Predicting subcellular location of protein with evolution information and sequence-based deep learning. BMC Bioinformatics. 2021 Oct 22;22(Suppl 10):515. doi: 10.1186/s12859-021-04404-0. PMID: 34686152; PMCID: PMC8539821.

[28] Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 10, 296 (2009). https://doi.org/10.1186/1471-2105-10-296

[29] Alland C, Moreews F, Boens D, Carpentier M, Chiusa S, Lonquety M, Renault N, Wong Y, Cantalloube H, Chomilier J, Hochez J, Pothier J, Villoutreix BO, Zagury JF, Tufféry P. RPBS: a web resource for structural bioinformatics. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W44-9. doi: 10.1093/nar/gki477. PMID: 15980507; PMCID: PMC1160237.

[30] Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, Leung PSC, Ansari AA, Gershwin ME, Anaya JM. Molecular mimicry and autoimmunity. J Autoimmun. 2018 Dec;95:100-123. doi: 10.1016/j.jaut.2018.10.012. Epub 2018 Oct 26. PMID: 30509385.

[31] Thaper D, Prabha V. Molecular mimicry: An explanation for autoimmune diseases and infertility. Scand J Immunol. 2018 Jun 23:e12697. doi: 10.1111/sji.12697. Epub ahead of print. PMID: 29935034.

[32] T.G. Benedek, The history of bacteriologic concepts of rheumatic fever and rheumatoid arthritis, Semin. Arthritis Rheum. 36 (No. 2) (2006, October) 109–123. WB Saunders. https://doi.org/10.1016/j.semarthrit.2006.05.001

[33] S. Benvenga, F. Guarneri, M. Vaccaro, L. Santarpia, F. Trimarchi, Homologies between proteins of Borrelia burgdorferi and thyroid autoantigens, Thyroid 14 (2004) 964–966, https://doi.org/10.1089/thy.2004.14.964

[34] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005 Apr 22;33(7):2302-9. doi: 10.1093/nar/gki524. PMID: 15849316; PMCID: PMC1084323.

[35] Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W363-7. doi: 10.1093/nar/gki481. PMID: 15980490; PMCID: PMC1160241.

[36] Kumar, Pranav & Arya, Aditya. (2018). Ramachandran Plot. https://www.researchgate.net/publication/330010893

[37] Shikha Agnihotry, Rajesh Kumar Pathak, Dev Bukhsh Singh, Apoorv Tiwari, Imran Hussain, Chapter 11 - Protein structure prediction, Editor(s): Dev Bukhsh Singh, Rajesh Kumar Pathak, Bioinformatics, Academic Press, 2022, Pages 177-188, ISBN 9780323897754, https://doi.org/10.1016/B978-0-323-89775-4.00023-7

[38] Donald Voet, Judith G. Voet, Charlotte W. Pratt. Fundamentals of Biochemistry, 4th ed.; Ch.6.

[39] Z. Wu, Y. Li, Q. Liu, Y. Liu, L. Chen, H. Zhao, H. Guo, K. Zhu, N. Zhou, T.C. Chai, B. Shi, Pyroptosis engagement and bladder urothelial cell-derived exosomes recruit mast cells and induce barrier dysfunction of bladder urothelium after uropathogenic E. coli infection, Am. J. Physiol. Cell Physiol. 317 (3) (2019) C544–C555.

[40] Tukaj S, Kaminski M. Heat shock proteins in the therapy of autoimmune diseases: too simple to be true? Cell Stress Chaperones. 2019 May;24(3):475-479. doi:

10.1007/s12192-019-01000-3. Epub 2019 May 9. PMID: 31073900; PMCID: PMC6527538.

[41] Harvey KL, Jarocki VM, Charles IG, Djordjevic SP. The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial Pathogenesis. Front Microbiol. 2019 Oct 24;10:2351. doi: 10.3389/fmicb.2019.02351. PMID: 31708880; PMCID: PMC6822514.

[42] Ren G, Adamus G. Cellular targets of anti-alpha-enolase autoantibodies of patients with autoimmune retinopathy. J Autoimmun. 2004 Sep;23(2):161-7. doi: 10.1016/j.jaut.2004.06.003. PMID: 15324934.

[43] Takasaki Y, Kaneda K, Matsushita M, Yamada H, Nawata M, Matsudaira R, Asano M, Mineki R, Shindo N, Hashimoto H. Glyceraldehyde 3-phosphate dehydrogenase is a novel autoantigen leading autoimmune responses to proliferating cell nuclear antigen multiprotein complexes in lupus patients. Int Immunol. 2004 Sep;16(9):1295-304. doi: 10.1093/intimm/dxh131. Epub 2004 Jul 19. PMID: 15262899.

[44] Albani S, Keystone EC, Nelson JL, Ollier WE, La Cava A, Montemayor AC, Weber DA, Montecucco C, Martini A, Carson DA. Positive selection in autoimmunity: abnormal immune responses to a bacterial dnaJ antigenic determinant in patients with early rheumatoid arthritis. Nat Med. 1995 May;1(5):448-52. doi: 10.1038/nm0595-448. PMID: 7585093.

[45] Gregoric, E., Gregoric, J.A., Guarneri, F. *et al.* Injections of *Clostridium botulinum* neurotoxin A may cause thyroid complications in predisposed persons based on molecular mimicry with thyroid autoantigens. *Endocr* 39, 41–47 (2011). https://doi.org/10.1007/s12020-010-9410-9

[46] Tsuchiya, N. Genetics of ANCA-associated vasculitis in Japan: a role for *HLA-DRB1*09:01* haplotype. *Clin Exp Nephrol* 17, 628–630 (2013). https://doi.org/10.1007/s10157-012-0691-6

[47] Drouin EE, Glickstein L, Kwok WW, Nepom GT, Steere AC. Searching for borrelial T cell epitopes associated with antibiotic-refractory Lyme arthritis. Mol Immunol. 2008 Apr;45(8):2323-32. doi: 10.1016/j.molimm.2007.11.010. Epub 2008 Jan 11. Erratum in: Mol Immunol. 2008 Jul;45(12):3508. PMID: 18191206; PMCID: PMC2374920.

[48] Yan, Y., Tao, H., He, J. *et al.* The HDOCK server for integrated protein–protein docking. *Nat Protoc* 15, 1829–1852 (2020). https://doi.org/10.1038/s41596-020-0312-x

[49] Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chem Biol Drug Des. 2019 Jan;93(1):12-20. doi: 10.1111/cbdd.13388. Epub 2018 Oct 8. PMID: 30187647.

[50] Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol. 2008 Apr 4;4(4):e1000048. doi: 10.1371/journal.pcbi.1000048. PMID: 18389056; PMCID: PMC2267221.

[51] C. Bernecker, M. Ostapczuk, S. Vordenb¨ aumen, M. Ehlers, A. Thiel, S. Schinner, H. Willenberg, W.A. Scherbaum, M. Schott, HLA-A2 phenotype may be protective against Graves' disease but not against Hashimoto's thyroiditis in Caucasians, Horm. Metab. Res. 45 (2013) 74–77, 01.

[52] A.M. Hosseini, J. Majidi, B. Baradaran, M. Yousefi, Toll-like receptors in the pathogenesis of autoimmune diseases, Adv. Pharmaceut. Bull. 5 (Suppl 1) (2015) 605.

[53] Duan Y, Tang H, Mitchell-Silbaugh K, Fang X, Han Z, Ouyang K. Heat Shock Protein 60 in Cardiovascular Physiology and Diseases. Front Mol Biosci. 2020 Apr 30;7:73. doi: 10.3389/fmolb.2020.00073. PMID: 32426370; PMCID: PMC7203681

[54] Faas MM, de Vos P. Mitochondrial function in immune cells in health and disease. Biochim Biophys Acta Mol Basis Dis. 2020 Oct 1;1866(10):165845. doi: 10.1016/j.bbadis.2020.165845. Epub 2020 May 28. PMID: 32473386.

[55] Kurokawa, C., Lynn, G.E., Pedra, J.H.F. *et al.* Interactions between *Borrelia burgdorferi* and ticks. *Nat Rev Microbiol* 18, 587–600 (2020). https://doi.org/10.1038/s41579-020-0400-5

[56] Holling TM, Schooten E, van Den Elsen PJ. Function and regulation of MHC class II molecules in T-lymphocytes: of mice and men. Hum Immunol. 2004 Apr;65(4):282-90. doi: 10.1016/j.humimm.2004.01.005. PMID: 15120183.

[57] Mintern, J., Villadangos, J. Antigen-presenting cells look within during influenza infection. *Nat Med* 21, 1123–1125 (2015). https://doi.org/10.1038/nm.3971

[58] Skapenko, A., Leipe, J., Lipsky, P.E. *et al.* The role of the T cell in autoimmune inflammation. *Arthritis Res Ther* 7 (Suppl 2), S4 (2005). https://doi.org/10.1186/ar1703

[59] Archibald JM. Endosymbiosis and Eukaryotic Cell Evolution. Curr Biol. 2015 Oct 5;25(19):R911-21. doi: 10.1016/j.cub.2015.07.055. PMID: 26439354.