Predicted Epitopes of *E. coli.* that Cross-react with Human antigens as potential mediators of autoimmunity

A DISSERTATION

SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

Submitted by:

Shalu Garg

(Roll No. 2K19/MSCBIO/02)

Under the supervision of

Dr. ASMITA DAS

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

MAY, 2021

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, hereby certify that the work is which is presented in Major Project –II entitled "**Predicted Epitopes of *E. coli.* that Cross-react with Human antigens as potential mediators of autoimmunity**" in fulfilment of the requirement for the award of the degree of Master of Science (M.Sc.) in Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own, carried out during a period from 7th Jan 2021 to 28th May 2021, under the supervision of Dr. Asmita Das.

The matter presented in this thesis has not been submitted by me for the award of any degree of this or any Institute. The work has been communicated in SCI expanded journal.

Title of paper: HLA Supertype impacts susceptibility to autoimmune diseases

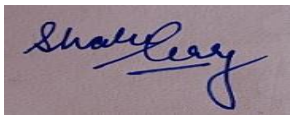
Author name: Shalu Garg (Author), Dr. Asmita Das (Co-Author)

Name of journal: Indian Journal of Biotechnology

Status of paper: Communicated

Date of paper communication: 28/05/2021

Shalu Garg
(2K19/MSCBIO/02)



DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

SUPERVISOR CERTIFICATE

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the student is correct.

Place: Delhi

Date: 30.05.2021

Dr. Asmita Das

SUPERVISOR

Assistant Professor

Department Of Biotechnology

Delhi Technological University

ACKNOWLEDGEMENT

It is my privilege to express my profound sense of gratitude and indebtedness to my mentor Dr. Asmita Das, Assistant Professor in the Department of Biotechnology, Delhi Technological University, for her valuable guidance and consistent encouragement during the progress of the project work. The dissertation would not be completed within a short period without her insightful suggestions and support. I am equally grateful and wish to express my wholehearted thanks to respected lab seniors Mr. Sunil and Ms. Ritu Dhankas for their kind support and help in the course of my research work. On a personal note, I wish to express my gratitude and affection to my family for their constant love and support.

I would like to thank Mr. Piyush Kumar and all my friends for their boundless support, which motivated me to complete the project work in the given time.

Shalu Garg

ABSTRACT

For several decades many efforts are being made to understand autoimmunity. Through research, it has become apparent that immune system has evolved multiple mechanisms to control self-reactivity. The human system contains microbiota, and their interactions help to regulate immune system. However, any defects in these mechanisms can break down immune tolerance. Generally, a significant portion is occupied by bacteria, but some viruses, protozoa as well as fungi also reside as the normal microflora. Microbial dysbiosis initiates dysregulation of immune system on its own via many processes, including T helper cell skewing, epitope spread, bystander activation, cross-reactive, and dual T cell receptor (TCR) stimulation act on susceptibility, initiation, and disease propagation. The human body contains both commensals and pathogenic bacteria. This work aims to study various commensal bacteria linked with autoimmune diseases. Under this provision, we will focus on the immunological mechanisms that give rise to autoimmune diseases. Further, we can predict potential epitopes in bacteria's to find similarity with human sequences and perform molecular docking of bacterial peptides with MHC class II and TCR. Through this study, new preventive and therapeutic strategies can be developed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SERIAL NO.	ABBREVIATIONS	EXPLAINATION
1	AID	Autoimmune Diseases
2	APC	Antigen Presenting Cell
3	TCR	T- Cell Receptor
4	MHC	Major Histocompatibility Complex
5	TH cell	T-Helper cell
6	CD	Cluster of Differentiation
7	MAMPs	Microbe-Associated Molecular Patterns
8	LPS	Lipopolysaccharide
9	NOD	Nucleotide binding Oligomerization Domain
10	TLR	Toll Like Receptor
11	IgA	Immunoglobulin A
12	ACE	Atomic Contact Energy
13	HB	Hydrogen Bonds
14	VdW	Van der Waals

CHAPTER 1

MICROBIOTA AND AUTOIMMUNITY

1.1 ESCHERICHIA GENERA HISTORY

In 1885, E. coli was discovered by a German scientist named Odor Escherichia. He discovered it in the excreta of healthy individuals and hence named it Bacterium coli due to their presence in the colon of humans. Based on the bacterial shape and its motility, it was placed in the genera, according to the early prokaryotic classification. After further studies, Ernst Haeckel placed E.coli in the Monera kingdom, and this classification was named Ernst Haeckel's bacterial classification. In 1895, E.coli was not considered in the Monera kingdom; instead, it was reclassified in the genus Escherichia, which was named after its discoverer, and this was classified by the scientist named Migula. Formally this genus was belonging to the bacterial group known as "coliforms," these come under the Enterobacteriaceae family; hence they are called "the enterics[1]."

1.2 SALIENT FEATURES OF ESCHERICHIA COLI

Escherichia coli is a Gram-negative and facultative anaerobic bacteria. It is rod-shaped. Its diameter is about 0.25 to 1.0 micrometers, and its length is 2.0 micrometers[2]. Strains that contain flagella are called motile strains because they have the ability to move. In the majority, flagellar strains have a peritrichous arrangement that means multiple flagella may be randomly distributed over the cell surface[3]. Bacteria required optimum conditions for their growth, such as temperature, pH, salinity. E.coli. required 37°C (98.6°F) specific temperature for their multiplication[4], but there are many laboratory strains, which shows their growth at 49°C (120°F) of temperature. For their multiplication, a large number of substrates are needed, such as oxygen, hydrogen, different types of acid, including pyruvic as well as formic acid. They utilize these by involving the oxidation-reduction reactions[5].

1.3 ESCHERICHIA COLI PHENOTYPIC DIVERSITY

Escherichia coli encompass a large inhabitant of bacteria that denotes a very high level of phenotypic as well as genetic diversity. Reclassification of the taxonomy is needed after focusing on the different isolates of E.coli in a large number[6]. E.coli is considered one of the diverse species in the groups of bacteria. The genomic sequence of different strains of E.coli is also quite variable. Only 20% of genomic sequence is shared between the strains hence called as they have only 20% sequence similarity[7].

From an evolutionary perspective, members of Shigella genus such as S. flexneri, S. sonnei, and S. dysenteriae are considered E.coli strains. This process is known as taxa in disguise. Likewise, K-12 strain of E.coli majorly helpful in the field of recombinant DNA technology; hence they are called warrant classification[8].

A subgroup of species that contains some different and unique characteristics from the other members, thus differentiating it from the others, is known as strain. We can find these small characteristic variations only by looking at molecular level. This will tell us about the genes which are responsible for the functioning as well as life cycle of the bacterium. For instance, one strain may attain the resistance towards the antimicrobial agents while others remain sensitive[9].

Serotype is a common subdivision system of E.coli based on their cell surface antigens, which was observed through evolutionary relatedness. For instance, O-antigen which are present in LPS layer, flagellin protein, K-antigen as well as in the capsule structure[10]. In this case, O is considered as the serotype. The common strain of laboratory has a mutation which does not allow formation of an O-antigen hence it is non type able[11]. Till now only 190 serogroups are found in the literature.

Natural biological process of horizontal gene transfer, mutation as well as gene duplication are the major phenomenon which is responsible for evolution of new strains of E.coli. Every E.coli strain is taken from either E.coli B or E.coli K strains. There are many strains that acquire characteristics which can be deleterious to human host[12]. Strain which is responsible for various illness, and even death in weak immunity compromised person specially aged or in young ones is O157:H7, which is a more dangerous strain[13]. Most of the labortatory strain MG1655 genome was obtained through the process of horizontal gene trasfer from species of salmonella divergence[14].

1.4 LIFE CYCLE OF ESCHERICHIA COLI

In life cycle of Escherichia coli, one cell divides into two cells and they are called as daughter cells. Hence, they divide by the process of binary fission. These daughter cells are genetically identical to theire parental cells under the normal conditions, this is referred to as local doubling of cells. E. coli take generation time of 20 minutes.Nonetheless, both daughter cells will survive or die, is not necessary, their growth is exponential when surviving cells excceds one on an average[15].

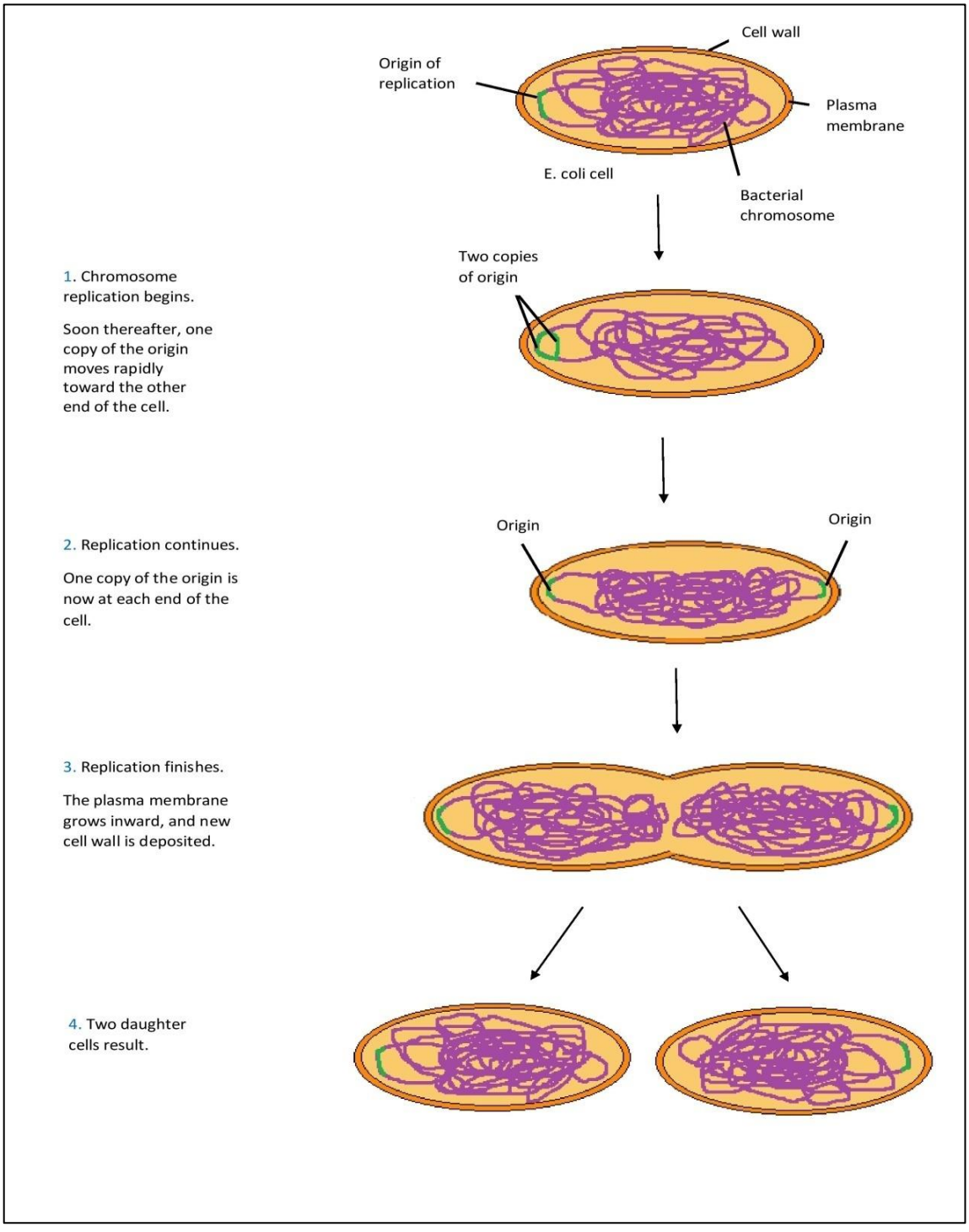


Figure 1.1 Life Cycle of E. coli.

Garg & Das (Manuscript under preparation)

1.5 BASIC CHARACTERISTICS AND LOCALISATION OF NORMAL MICROFLORA

Every living organism interacts with environment, and this interaction is important, but sometimes it proved to be endangered. As epithelial cells are in direct contact with environment, so it represents the most important interaction between organism and environment. About $300m^2$ surface area of humans body is represented by mucosa whereas $2m^2$ by skin. After the delivery of a fetus, or we can say delivery from sterile uterine environment, represents the first contact of macro with microorganisms. Major way for microbial entry is skin, gastrointestinal, urogenital tract as well as respiratory tract. Physiological relationship with bacteria brings settlement of epithelial surface and beneficial for the host as well as for bacteria also, hence, this interaction is known as commensalism. Sometimes this relationship turns into the opportunistic behavior, where endogenous microbes harms their host and now this is called parasitism[16]–[18]. In the major portion, bacteria are the microflora in the body but it some viruses, protozoa as well as fungi also resides as the normal microflora.

In gastrointestinal tract , majority of bacteria are anaerobic in nature. In the gut system, stomach is the major organ which is responsible for blocking the microbes entry from the outer environment to the lower parts of tract because it contains different types of proteolytic enzymes as well as it releases some content which are acidic in nature. Even after that few bacteria still resides there, these bacterias are often attached to gastric epithelia and also present in mucus[19]. Ileum contains a large number of bacteria while duodenum and jejunum have very less number of bacteria's. Vast diversity and highest number of bacteria located in colon, these bacteria's either attached to mucosal layer or found in colon's content[18], [20].

Disturbance in intestinal mucosa can be because of two resons:-first, Microorganisms and their toxins secondly, poor functioning of parts of mucosal immune system. Many inflammatory processes may occur due to pathologically enhances activities of immune system. Disturbances in mucosal barriers functioning and because of change in mechanisms regulating mucosal

immunity various chronic disease may occur. For instance inflammatory diseases (allergies), multi-organ failure as well as autoimmune disease (AIDs). These diseases may develop either in their starting phase or throughout the surface[19].

1.6 INVOLVEMENT OF COMMENSAL BACTERIA AND THEIR COMPONENTS IN THE DEVELOPMENT OF AUTOIMMUNE DISEASES

Cancer as well as heart diseases (cardiovascular diseases) are the major victims in enhancing the death rate of the world. In addition to it, many other chronic disorders are also prevalent in causing drastic morbidity including inflammatory diseases, different types of allergies, arthritis as well as autoimmune diseases[21], [22]. The above-described disorders shows predominant medical problem due to their deleterious impact on the life, also require a kind of medical care which are long-standing.

There are two major deleterious features of autoimmune diseases as well as inflammatory diseases, firstly destruction of tissues, and, secondly impairement of functions which are caused due to immunologically mediated mechanisms. For immunomodulatory mechanisms, bacteria as well as elements such as metabolites are accountable[23]. Recognizable work on neoplastic diseases, autoimmune as well as inflammatory diseases directed at looking environmental agent's infective action, in addition with microbial components[24], [25].

In view of several cases, diminished role of intestine barrier tends to elevated in antibodies aimed against antigen located in intestine lumen. Recent studies showed that presence of autoantibodies in person(s) having no symptoms can have essential predictive value addition for progression of autoimmune as well as inflammatory diseases[26], [27].

Immense efforts have been made to acknowledge the mechanism which leads to destruction of self-tolerance in case of autoimmunity. In genetically predisposed people infection were

observed to begin this action. One of the major hypothesis describing about how infectious agents can lead to autoimmunity based upon cross-reactivity, known as "**molecular mimicry",** resemblance within autoantigenic epitopes and harmless antigens in environment[28], [29]. Infection start build-up of autoimmunity by inefficient stimulation of immune system[30]. Microbial components lead to the activation of APC's which lead to the unusual antigen processing and presentation through their adjuvant activity. These components are superantigens that are noted for the induction of inflammatory reactions.

It was noticed that, in the immunosuppressed individuals, pathogenic microorganisms as well as toxins that attacks the intestine may disturb homeostasis of the intestinal mucosa. Hypo-activity or hyperactivity of immune system affects intestinal mucosa[31], [32]. Studies indicating that inter-individual differences as well as patterns which are specific to disease in composition of the microbiota in humans are notable. Even though, the complexity and differences of gut microbiota between individuals states a surprising factor in the hard work to find out the possible consequence of individual commensal microbes in disease pathogenesis.

The study of early events of any disease become difficult because patients go to clinic after the disease become symptomatic. Animal models that are induced experimentally or developed spontaneously are used to study role of environmental factors along with genetic factors in earlier phase of the development of disease, to interpret mechanism of pathogenesis and mainly for development of new medical approaches such as preventive or therapeutic strategies, Induced experimentally. Spontaneous making models of animal allow to study role of genetic factors as well as environmental factors during initial events of disease build-up, to interpret pathogenic phenomenon and to succeed new preventive as well as therapeutic strategies. However, they sometimes become too unreal to compare it with human disease[33].

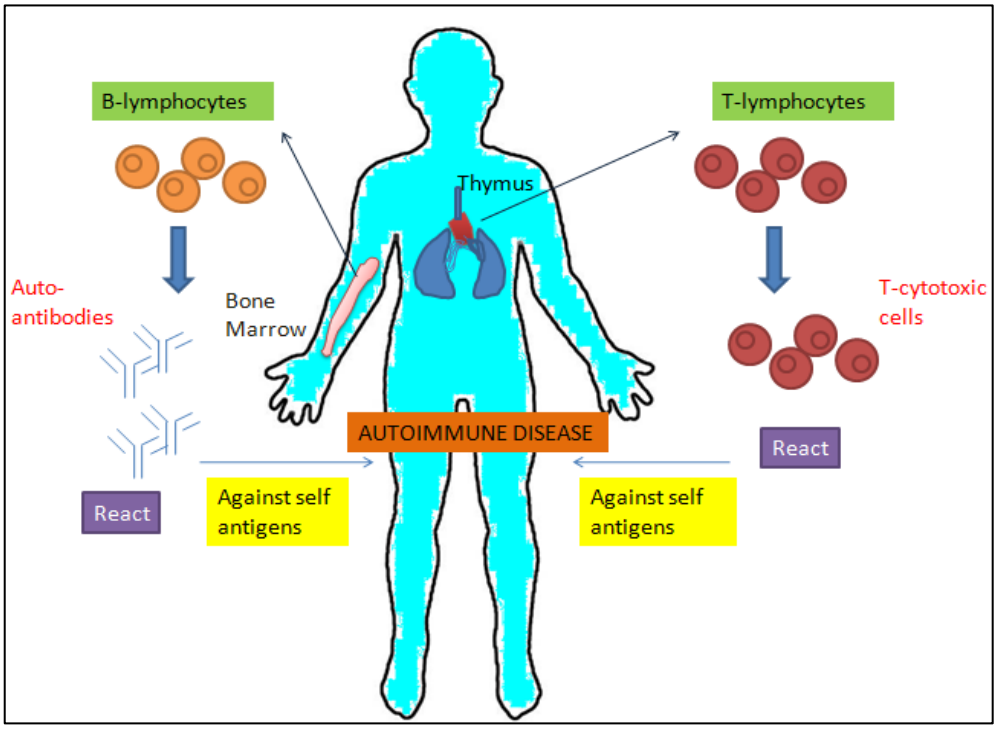


Fig 1.2 Schematic representation of autoimmunity

Garg & Das (Manuscript under preparation)

1.7 IMMUNE TOLERANCE

There are many factors that contribute to dysregulation of immune system and disturbance in peripheral tolerance[34] which can lead to abnormal immune responses towards normal flora and dysbiosis. These factors include epigenetics, environmental as well as genetic factors. These deleterious effects may come up during early childhood or maybe ignited during change in the hormone level after puberty when different diseases manifest which are related to immune system[35], [36]. As per hygiene hypothesis, dysbiosis in microbial community may get transferred to future generation and this may result in autoimmune disease[37], [38]. Epitope spread, bystander activation, cross-reactive as well as dual TCR stimulation are ways by which dysbiosis may initiate immune system dysregulation that results in development of the disease. We will study about immunological processes that give rise to autoimmune diseases, even though they likely to apply for allergic diseases[39].

1.1.1 T helper cell skewing

For the damage of inflammatory tissues two cells are mainly responsible, TH1 and TH17 cells. There is homeostasis between immune response and microbiota and may unbalanced during microbial action or injury of tissue[40], [41]. Markedly, programmed cell death during pathogenic inflammation directed to autoreactive TH17 cell responses[42], [43]. Furthermore, pin-pointed commensals induce mucosal T cell responses that are autoreactive. Additionally, few commensals may induce differentiation of mucosa T cell towards T_C cells, TH1, or TH17 which encourage immune-related diseases[44]. Notably, conditions are different during homeostasis and inflammation.

1.1.2 Bystander activation

Activation of autoreactive T-cells and B-cells in an antigen-independent manner in time of infection by generating inflammatory conditions, this is called bystander activation which can be done via pathogens[34], [45], [46]. Pathobionts encourages pro-inflammatory signals and are found within microbial community. Commensal-derived MAMPs like LPS, peptidoglycan may ignite NOD-like receptors and TLRs which further play role in autoimmunity as well as inflammatory diseases[47]–[51].

1.1.3 Epitope spread

Infections like viral and other cytolytic may cause release of antigen as well as epitope spreading that can be intramolecular or intermolecular. Contribution to this mechanism are also done by microbiota residing in gut that cause tissue damage under inflammatory conditions like infection of viral or change in function of gut barrier. Autoantigens may modify post-translationally which can also cause autoimmunity. It is a alternate way used by commensals to encourage multiple epitopes targeting. [52].

1.1.4 Cross- reactivity

It is usually referred to as molecular mimicry. Cross reactivity may occur due to steady colonization with cross-reactive microflora in host having adequate haplotypes of MHC, which

initiate development of disease[53]–[60]. In human microflora, cross-reactive antigens are provided by orthologues of autoantigens or non-orthologue mimotopes of antigens that initiate T-cells as well as B-cells which are autoreactive in nature[55], [56], [60]. Likewise, cross-reactivity with commensal antigen in human body affects other organs than gut[61]. There is no need of bacterial translocation for non-gut autoimmunity, particularly if abnormal gut homing of immune cells lead to spreading of autoimmune responses[62] but it contribute where gut barrier structural integrity is disturbed.

1.1.5 Dual T cell receptors

Another mechanism of how commensal microflora can activate autoimmunity is through TCR. T- cell receptor which is present on T-lymphocytes may recognize self antigens as well as microbial antigens. In humans, this mechanism is not known till now but it was observed in mice model. In Crohn's diseased patient, IgA-coated adherent-invasive E. coli was found to be associated with spondylarthritis, which encourage TH17 cells that are specific to commensal antigen along with autoantibodies[63]. Investigations shows that T-cells with dual TCRs can also a factor for causing autoimmunity, but this is not known whether this pathobiont give rise to immune cell stimulation by this mechanism or via other mechanisms.

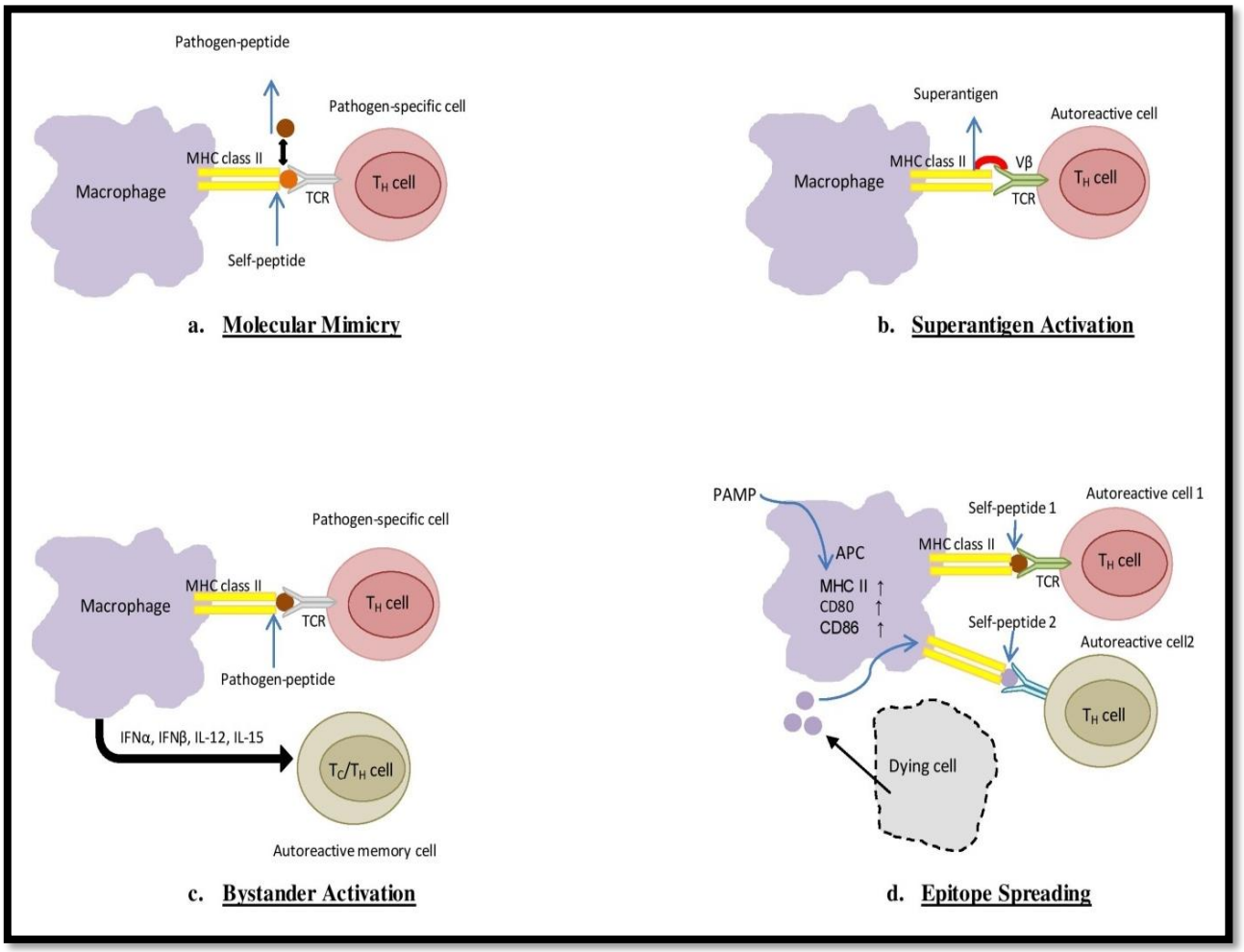


Figure 1.3 Mechanism of normal flora induced autoimmunity

Garg & Das (Manuscript under preparation)

CHAPTER 2

METHODOLOGY

2.1 MATERIALS AND METHODS:

2.1.1 Database used:

i) Immune Epitope Database (IEDB)
ii) Protein Data Bank
iii) Uniprot
iv) National Centre for Biotechnology Information (NCBI)
v) Protein BLAST

2.1.2 Servers used:

i) PEP-FOLD 3
ii) PatchDock
iii) FireDock

2.1.3 Software used:

i) PyMOL

2.2 METHODS

2.2.1 Literature review and sequence retrieval

To determine the names of commensal bacteria, literature review was carried out that may cause autoimmune diseases. Out of all the bacteria, one specific organism was selected for the further study based on their characteristics. After selection, FASTA sequences of the obtained antigen of a particular bacteria was retrieved using UniProt tool.

2.2.2 B cell and T cell epitope prediction

For B cell and T cell epitope prediction IEDB tool was used. IEDB contains list of epitopes from patent applications and peer-reviewed papers. FASTA sequence retrieved from UniProt is given as input in B cell epitope prediction of linear peptides from protein sequence. As a result, different sized peptides are obtained. Out of all, longest peptide was selected for the prediction of T-cell epitope. We used TepiTool in our study for T cell epitope prediction. Output obtained was then saved for further use in protein BLAST.

2.2.3 Sequence alignment

Most common tool for determining the sequence similarity between two sequence is NCBI BLAST. Protein BLAST compares query sequences to protein database. T- cell epitopes that are obtained in the previous step , now used as predicted peptides .Predicted peptides are compared to protein sequences of humans to determine the similarity between the 2 sequences.

2.2.4 Molecular docking of predicted peptides with MHC Class II

For determining the structures of predicted peptides from the sequence of amino acid given as input PEP-FOLD3 server was used. Amino acid sequence of predicted peptides is used as input to obtain PDB files.

After the PEP-FOLD3 server, PatchDock server was used as an intermediary tool for molecular docking. The result contains potential complexes list that is sorted by shape complementarity criteria. PDB files of both receptor and ligand are loaded onto PatchDock to obtain the desired output.

PatchDock server was followed by FireDock server. It was used to provide scoring of protein-protein docking solutions. Transformation files obtained from PatchDock along with the receptor and ligand PDB files are loaded onto the server. Output contains global energy ranking, attractive vanderwaal, repulsive vanderwaal, ACE, and HB.

2.2.5 Structure Visualization of predicted peptides- MHC class II complex

For the visualization of structures PyMOL software was used. It is a molecular visualization software that help to produce high quality images in three-dimensional. It produces images of small molecules and big macromolecules. Best global energy structure obtained from FireDock was loaded onto PyMOL. The obtained image was then saved in PNG format.

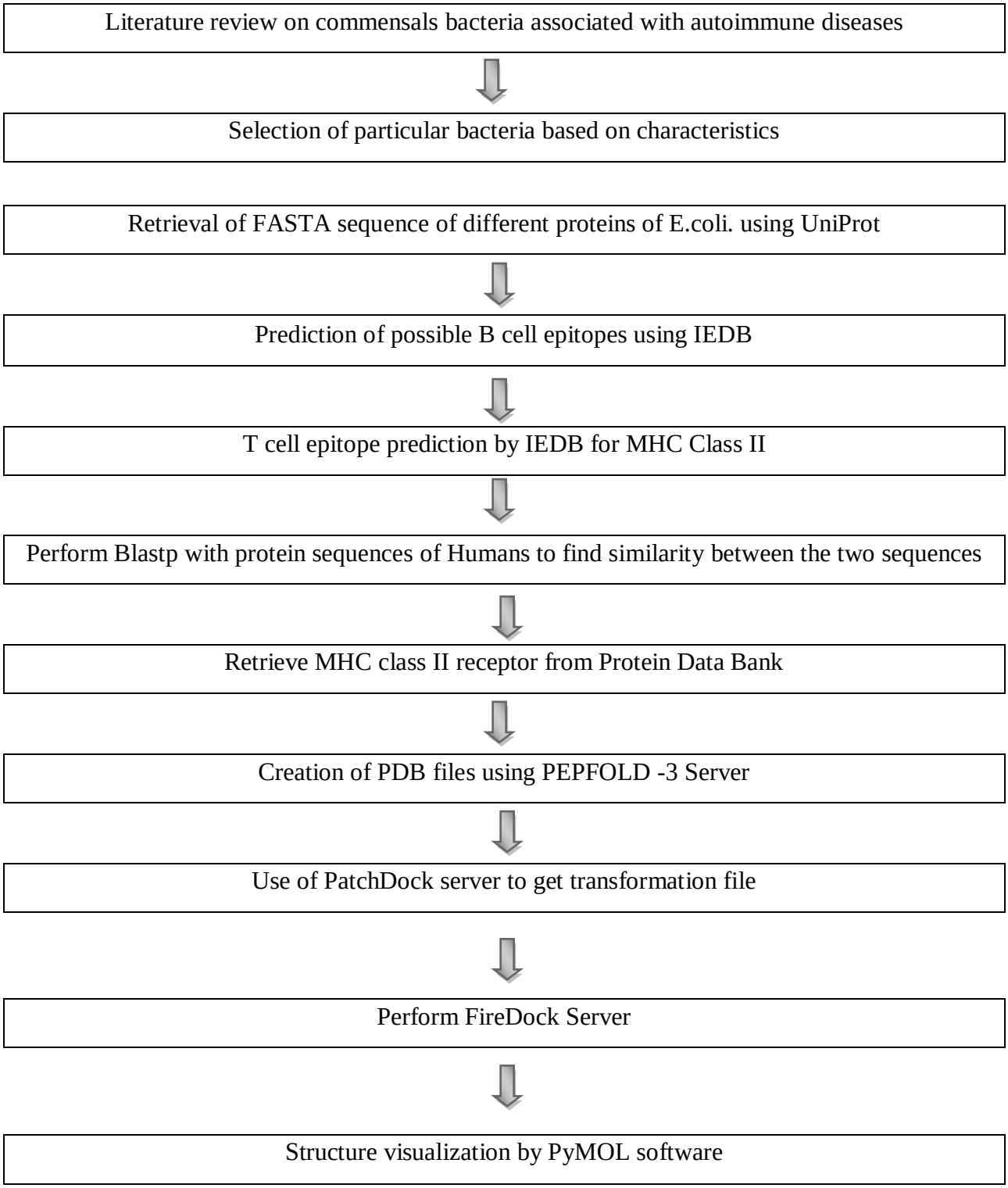
2.2.6 Molecular docking of predicted peptides with TCR

Firstly, TCR sequence and peptide structure was retrieved using PDB followed by molecular docking with predicted peptides. Only five predicted peptides were selected on the basis of their global energy with MHC Class II. These5 peptides were then used for molecular docking with TCR. PatchDock server was used as an intermediary tool for molecular docking. The result contains potential complexes list that are sorted by shape complementarity criteria. PDB files of both receptor and ligand is loaded onto PatchDock to obtain the desired output. This step was followed by FireDock server. It was used to provide scoring of protein-protein docking solutions. Transformation file obtained from PatchDock along with the receptor and ligand PDB files are loaded onto the server. Output contains global energy ranking, attractive vanderwaal, repulsive vanderwaal, ACE and HB.

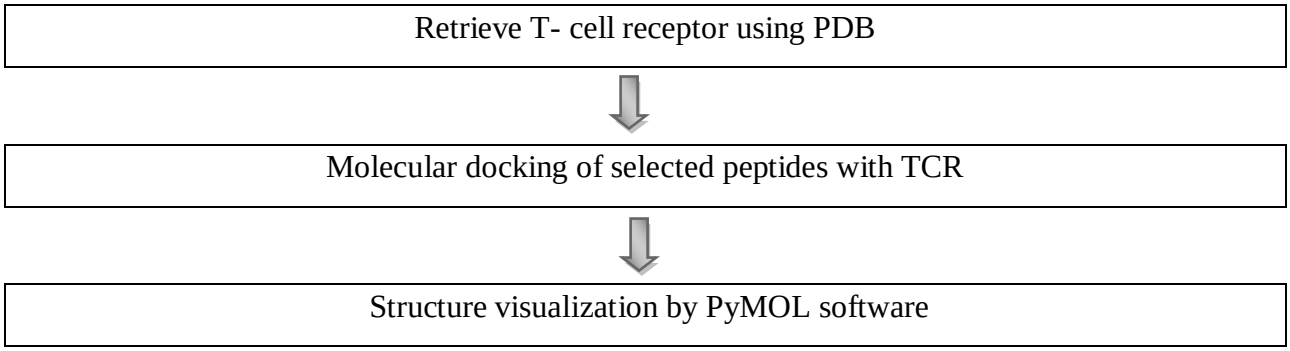
2.2. 7 Structure Visualization of predicted peptides- TCR complex

For the visualization of structures PyMOL software was again used. Best global energy structure obtained from FireDock was loaded onto PyMOL. The obtained image was then saved in PNG format.

2.3 WORKFLOW TO DETERMINE BINDING EFFICIENCY OF PEPTIDES - MHC CLASS II COMPLEX



2.4 WORKFLOW TO DETERMINE BINDING EFFICIENCY OF PEPTIDES- TCR COMPLEX



CHAPTER 3

RESULTS, DISCUSSION, AND CONCLUSION

3.1 B - CELL EPITOPE PREDICTION USING IEDB TOOL

E. coli. has a potential to cross-react with human antigens that may cause autoimmunity by triggering autoreactive B cell activation. Hence we need to check whether B cell epitopes of *E. coli* can activate humoral immunity.

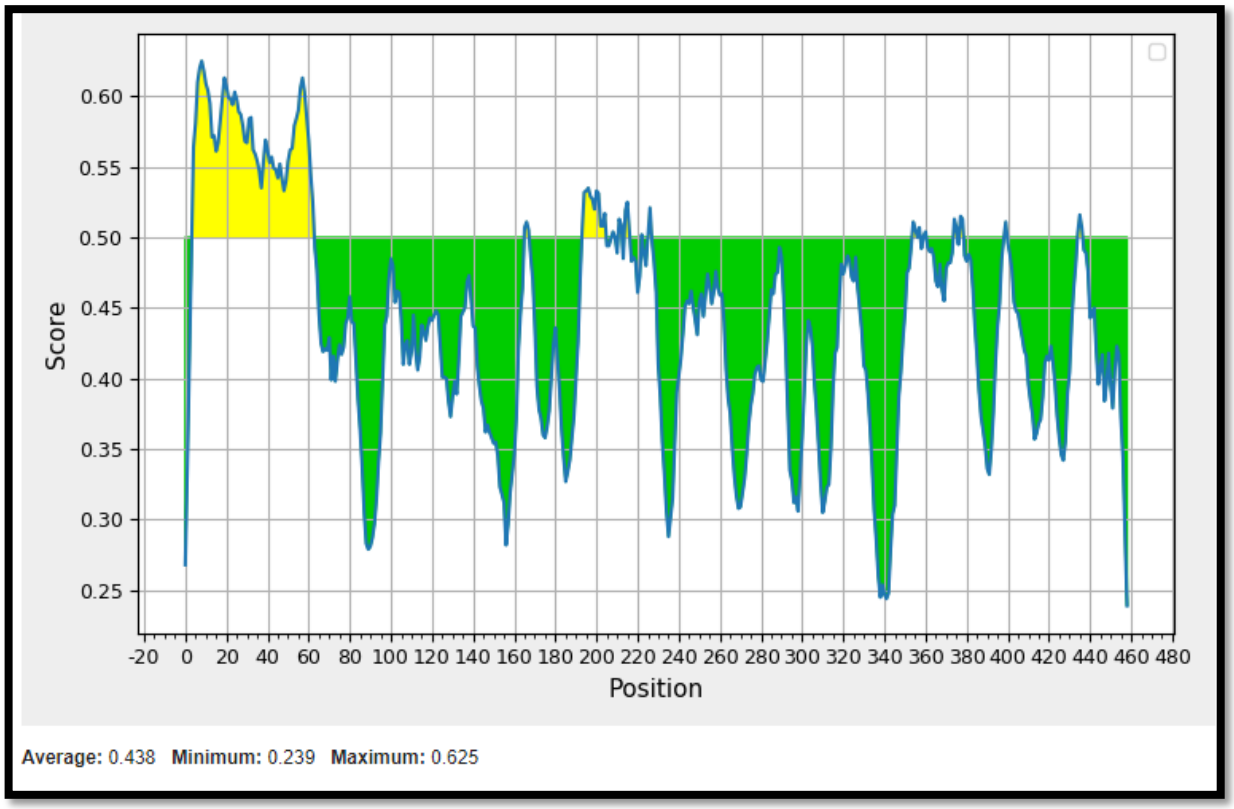


Figure 3.1 Hydropathy plot of E.coli Putrescine aminotransferase protein

No.	Start	End	Peptide	Length
1	5	63	PSSASALACSAHALNLIEKRTLDHEEMKALNREVIEYFKEHVNPGFLEYRKSVTAGGDY	59
2	166	168	SPR	3
3	194	205	STFRKPFMPLLP	12
4	209	210	HV	2
5	212	213	FG	2
6	215	217	IEA	3
7	223	223	N	1
8	226	227	KK	2
9	355	356	LP	2
10	358	358	Q	1
11	360	361	EQ	2
12	375	376	RE	2
13	378	379	PD	2
14	400	400	I	1
15	435	437	IEQ	3

Figure 3.2 Predicted peptides of E.coli Putrescine aminotransferase protein

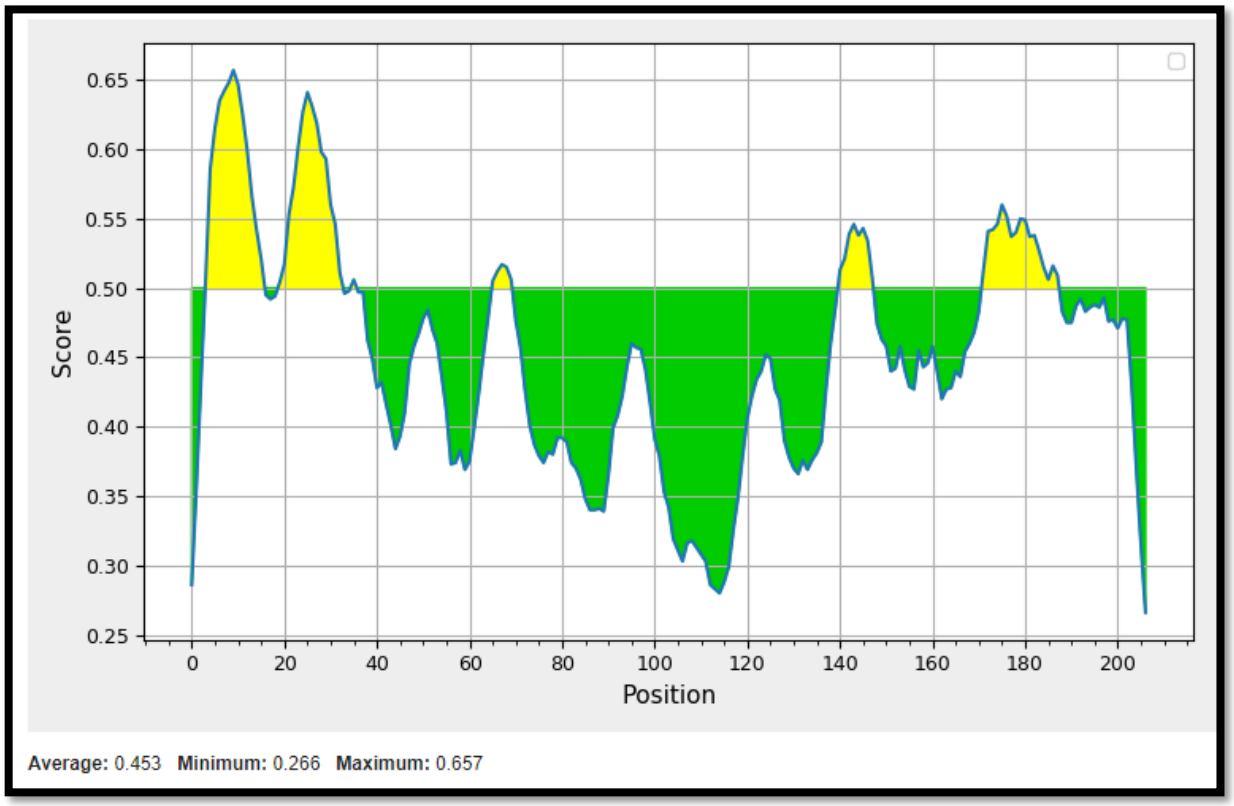


Figure 3.3 Bepipred Linear Epitope Prediction Graph of E.coli ATP-dependent Clp protease proteolytic subunit

No.	Start	End	Peptide	Length
1	4	16	SGERDNFAPHMAL	13
2	20	33	VIEQTSRGERSFDI	14
3	36	36	R	1
4	66	70	AENPE	5
5	141	148	GYQGQATD	8
6	172	188	GQSLEQIERDTERDRFL	17

Figure 3.4 Predicted peptides of E.coli ATP-dependent Clp protease proteolytic subunit

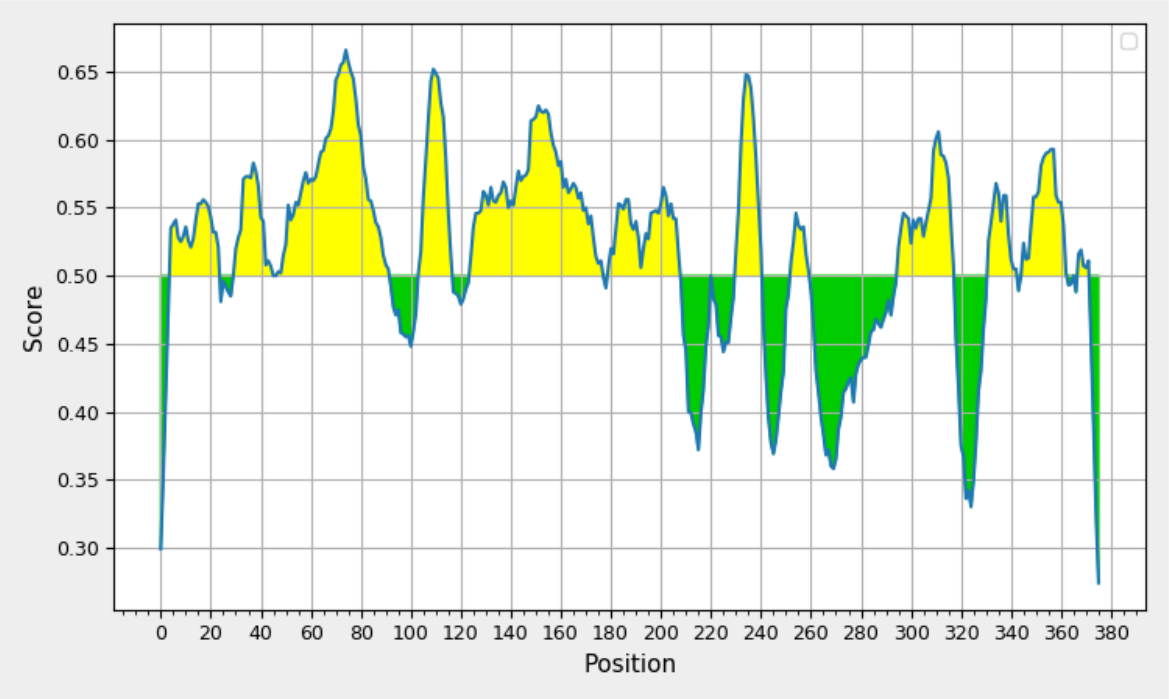
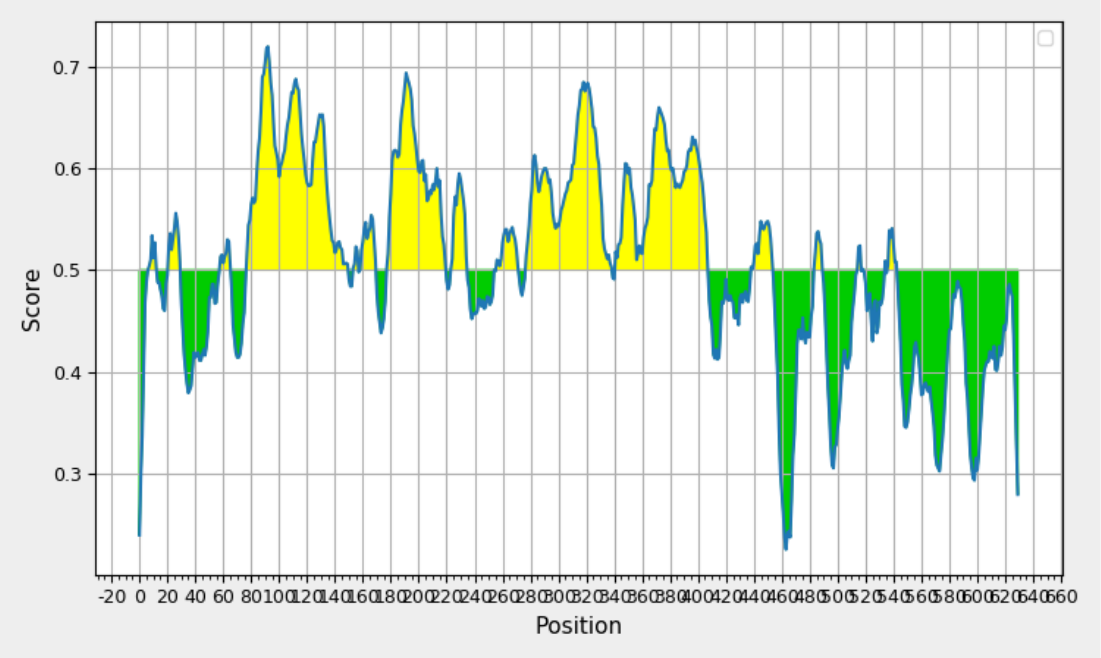


Figure 3.5 Bepipred Linear Epitope Prediction Graph of E.coli Chaperone protein DnaJ

No.	Start	End	Peptide	Length
1	5	24	DYYEILGVSKTAEEREIRKA	20
2	31	46	KYHPDRNQGDKEAEAK	16
3	48	92	KEIKEAYEVLTDSQKRAAYDQYGHAAFEQGGMGGGGFGGGADFSD	45
4	104	117	GGRGRQRAARGADL	14
5	125	177	LEEAVRGVTKEIRIPTLEECDVCHGSGAKPGTQPQTCPTCHGSGQVQMRQGFF	53
6	180	208	QQTCPHCQGRGTLIKDPCNKCHGHGRVER	29
7	221	221	D	1
8	231	241	EGEAGEHGAPA	11
9	253	260	HPIFEREG	8
10	296	318	ETQTGKLFRMRGKGVKSVRGGAQ	23
11	332	343	GLNERQKQLLQE	12
12	345	363	QESFGGPTGEHNSPRSKSF	19
13	368	372	KKFFD	5

Figure 3.6 Predicted peptides of E.coli Chaperone protein DnaJ

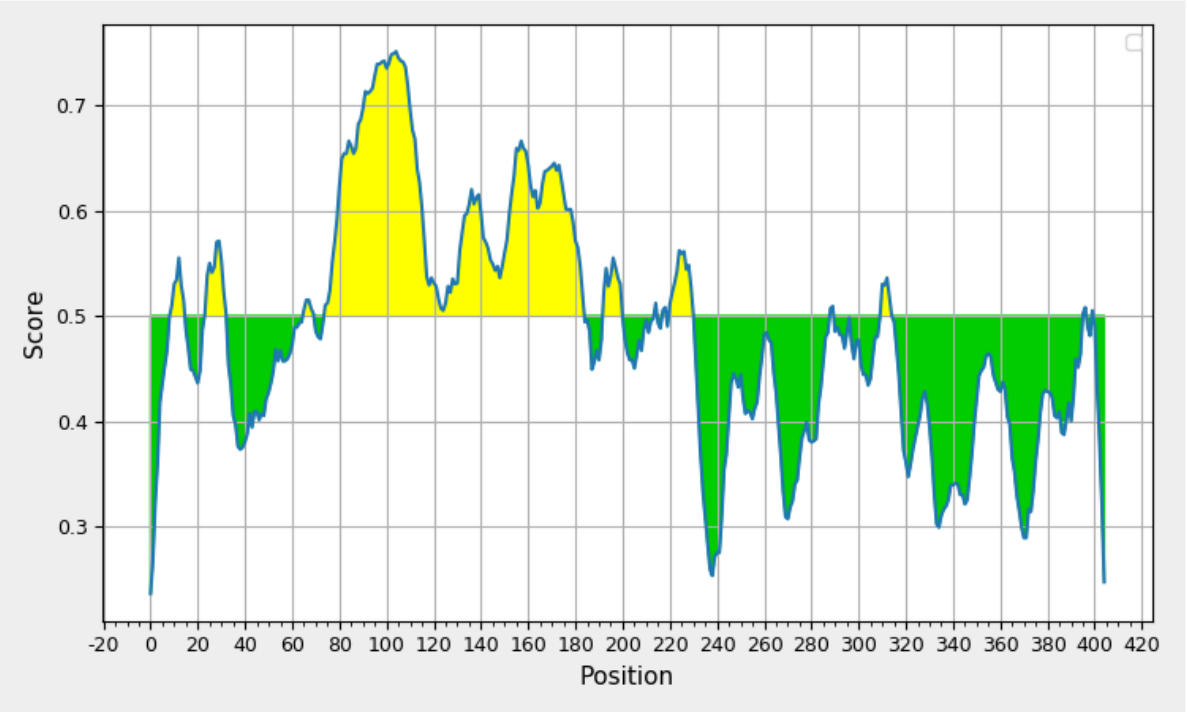


Average: 0.508 **Minimum:** 0.225 **Maximum:** 0.720

Figure 3.7 Bepipred Linear Epitope Prediction Graph of E.coli Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex

No.	Start	End	Peptide	Length
1	7	13	VPDIGAD	7
2	22	30	VKVGDKVEA	9
3	59	66	VSVGDKTQ	8
4	78	150	DGAADAAPAQAEEKKEAAPAAAPAAAAAKDVNVPDIGSDEVEVTEILVKVGDKVEAEQSLITVEGDKASMEVP	73
5	154	157	AGTV	4
6	160	170	IKVNVGDKVST	11
7	178	220	EVAGEAGAAAPAAKQEAAPAAAPAPAAGVKEVNVPDIGGDEVE	43
8	224	235	VMVKVGDKVAAE	12
9	255	255	A	1
10	257	272	VVKELKVNVGDKVKTG	16
11	278	339	FEVEGAAPAAAPAKQEAAAPAPAAKAEAPAAAPAAKAEGKSEFAENDAYVHATPLIRRLARE	62
12	342	408	VNLAKVKGTGRKGRILREDVQAYVKEAIKRAEAAPAATGGGIPGMLPWPKVDFSKFGEIEEVELGRI	67
13	439	454	AFRKQQNEEAAKRKLD	16
14	484	490	DGQRLTL	7
15	516	517	KK	2
16	519	519	I	1
17	535	535	D	1
18	537	543	KLTAGEM	7

Figure 3.8 Predicted peptides of E.coli Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex

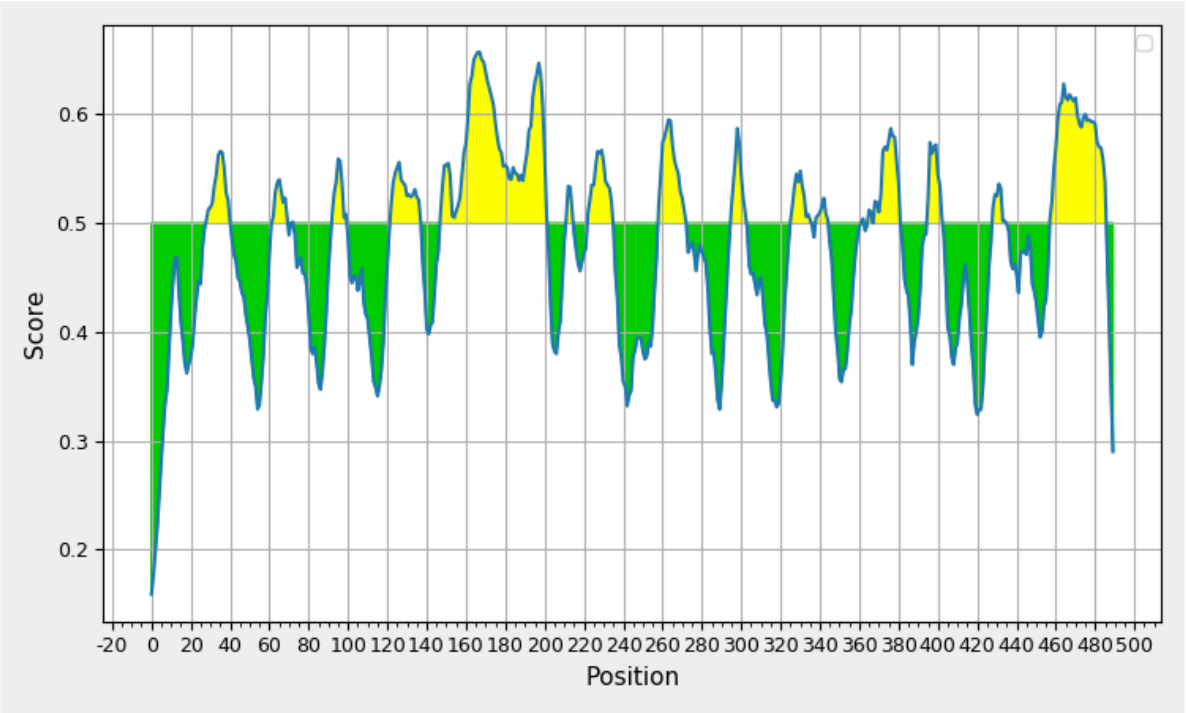


Average: 0.484 **Minimum:** 0.236 **Maximum:** 0.751

Figure 3.9 Bepipred Linear Epitope Prediction Graph of E.coli Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex

No.	Start	End	Peptide	Length
1	10	15	DLPESV	6
2	24	33	HKKPGDAVVR	10
3	66	70	TTVTS	5
4	75	184	GRLREGNSAGKETSAKSEEKASTPAQRQQASLEEQNNDALSPAIRRLLAEHNLDASAIKGTGVGGRLTREDVEKHLAKAPAKESAPAAAAPAAQPALAARSEKRVPMTRL	110
5	193	200	LEAKNSTA	8
6	215	215	D	1
7	218	219	KQ	2
8	221	231	GEAFEKRHGIR	11
9	289	290	TL	2
10	311	315	LTVED	5
11	396	397	ED	2
12	400	400	R	1

Figure 3.10 Predicted peptides of E.coli Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex



Average: 0.478 **Minimum:** 0.159 **Maximum:** 0.657

Figure 3.11 Bepipred Linear Epitope Prediction Graph of E.coli GTPase protein

No.	Start	End	Peptide	Length
1	29	41	LVADFPGLTRDRK	13
2	63	70	TEDGVETR	8
3	72	73	AE	2
4	93	100	AGLMPADE	8
5	123	138	LDPDQAVVDFYSLGLG	16
6	148	202	GRGVLSLLEHVLLPWMEDLAPQEEVDEDAEYWAQFEAEENGEEEEEDDFDPQSLP	55
7	212	215	NVGK	4
8	223	235	LGEERVVVYDMPG	13
9	259	272	GVRKRGKITDAVEK	14
10	296	303	EGISDQDL	8
11	326	336	LSQEVKEQVKE	11
12	339	345	DFRLGFI	7
13	362	363	GN	2
14	366	381	ESVREAYDSSTRRVGT	16
15	396	403	QPPLVRGR	8
16	429	435	DLPDSYK	7
17	458	486	KEGENPYANKRNTLTPTQMRKRKRLMKHI	29

Figure 3.12 Predicted peptides of E.coli GTPase protein

From the B-cell epitope prediction, result was observed in the form of graph and table. In the graph, yellow area which is present are suggested to be a B-cell epitope. In the table, peptides of different length are predicted. Out of all, longest peptide was selected for T-cell epitope prediction. The purpose of T cell prediction is to check whether the longest predicted peptide can also trigger cell-mediated immunity. Hence, T- Cell epitopes were predicted using TepiTool. Output obtained was then saved for further use in protein BLAST.

3.2 SEQUENCE SIMILARITY BETWEEN E. Coli. ANTIGENIC PEPTIDES AND HUMANS USING BLASTP

PEPTIDE SEQUENCE	TARGET	HLA-CLASS II ALLELE	IDENTITY (%)
Putrescine aminotransferase			
KALNREVIEYFKEHV	minichromosome maintenance complex binding protein isoform 3	HLA-DPA1*01:03/DPB1*02:01	100
		HLA-DPA1*02:01/DPB1*01:01	100
		HLA-DPA1*02:01/DPB1*05:01	100
PSSASALACSAHALN	capping protein inhibiting regulator of actin dynamics isoform X6	HLA-DQA1*01:02/DQB1*06:02	100
		HLA-DQA1*05:01/DQB1*03:01	100
NPGFLEYRKSVTAGG	hCG2041195	HLA-DRB1*04:01	72.73
IEYFKEHVNPGFLEY	zinc finger CCHC domain-containing protein 9	HLA-DRB1*04:01	64.29

		HLA-DRB3*02:02	64.29
ATP-dependent Clp protease proteolytic subunit			
SLEQIERDTERDRFL	immunoglobulin heavy chain junction region, partial	HLA-DRB1*03:01	100
Chaperone protein DnaJ			
RGVTKEIRIPTLEEC	stAR-related lipid transfer protein 13 isoform 3	HLA-DQA1*03:01/DQB1*03:02	77.78
		HLA-DQA1*04:01/DQB1*04:02	77.78
CDVCHGSGAKPGTQP	ribonuclease P protein subunit p29	HLA-DQA1*05:01/DQB1*03:01	64.29
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex			
KEAAPAAAPAAAAAK	probable E3 ubiquitin-protein ligase IRF2BPL	HLA-DQA1*01:02/DQB1*06:02	85.71
		HLA-DQA1*03:01/DQB1*03:02	85.71
		HLA-DQA1*04:01/DQB1*04:02	85.71
		HLA-DQA1*05:01/DQB1*	85.71

		03:01	
DGAADAAPAQAEEKK	60S acidic ribosomal protein P2	HLA-DQA1*03:01/DQB1*03:02	76.92
		HLA-DQA1*04:01/DQB1*04:02	76.92
		HLA-DQA1*05:01/DQB1*03:01	76.92
DKVEAEQSLITVEGD	protein AHNAK2 isoform 2	HLA-DQA1*03:01/DQB1*03:02	57.14
		HLA-DQA1*04:01/DQB1*04:02	
DVNVPDIGSDEVEVT	DENN domain-containing protein 11 isoform X1	HLA-DQA1*05:01/DQB1*02:01	76.92
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex			
KESAPAAAAPAAQPA	host cell factor 1 isoform X1	HLA-DQA1*03:01/DQB1*03:02	78.57
		HLA-DQA1*04:01/DQB1*04:02	78.57
		HLA-DQA1*05:01/DQB1*03:01	78.57
AAPAAQPALAARSEK	phosphatidylinositol-4-phosphate 5-kinase-like	HLA-DQA1*05:01/DQB1*	90.91

	1, isoform CRA_a	03:01	
DASAIKGTGVGGRLT	RAN binding protein 17, isoform CRA_g	HLA-DQA1*05:01/DQB1*03:01	88.89
SPAIRRLLAEHNLDA	immunoglobulin heavy chain junction region, partial	HLA-DRB1*01:01	100
		HLA-DRB1*04:01	100
		HLA-DRB1*04:05	100
		HLA-DRB4*01:01	100
GTPase Der			
GRGVLSLLEHVLLPW	transmembrane and coiled-coil domains protein 2 isoform 4	HLA-DPA1*01/DPB1*04:01	100
		HLA-DPA1*02:01/DPB1*01:01	100
		HLA-DPA1*03:01/DPB1*04:02	100
		HLA-DQA1*01:01/DQB1*05:01	100
		HLA-DRB1*01:01	100
EVDEDAEYWAQFEAE	immunoglobulin heavy chain variable region, partial	HLA-DPA1*01:03/DPB1*02:01	77.78
		HLA-DPA1*02:01/DPB1*01:01	77.78

		HLA-DQA1*01:01/DQB1*05:01	77.78
		HLA-DQA1*03:01/DQB1*03:02	77.78
		HLA-DQA1*04:01/DQB1*04:02	77.78
LEHVLLPWMEDLAPQ	pim-2 protooncogene homolog pim-2h	HLA-DPA1*01:03/DPB1*02:01	55
		HLA-DQA1*01:01/DQB1*05:01	55
		HLA-DQA1*05:01/DQB1*02:01	55
WMEDLAPQEEVDEDA	Chain A, Glycogen phosphorylase, muscle form	HLA-DQA1*03:01/DQB1*03:02	57.14
		HLA-DQA1*04:01/DQB1*04:02	57.14
		HLA-DQA1*05:01/DQB1*02:01	57.14
YWAQFEAEENGEEEE	piezo-type mechanosensitive ion channel component 2 isoform 1	HLA-DQA1*03:01/DQB1*03:02	80
		HLA-DQA1*05:01/DQB1*02:01	80

Table 3.1 Blastp results of E.coli antigenic peptide sequence with Human protein sequence

Sequence similarity of all predicted peptides with human protein sequence was observed using blastp. Similarity was found in range of 55% - 100%. These peptides were further used for molecular docking.

3.3 DOCKING TO PREDICT THE BINDING EFFICIENCY OF PREDICTED PEPTIDES TO MHC CLASS II MOLECULE USING FIREDOCK SERVER

Number given to peptides	Peptide sequence	Global energy	Attractive VdW	Repulsive VdW	ACE	HB
1	KALNREVIEYFKEHV	-26.16	-32.26	50.86	4.55	-2.19
2*	PSSASALACSAHALN	-54.59	-27.64	7.36	-3.68	-1.76
3	NPGFLEYRKSVTAGG	-43.49	-32.79	22.37	0.83	-5.89
4	IEYFKEHVNPGFLEY	-41.60	-34.60	32.96	-2.74	-3.53
5	SLEQIERDTERDRFL	-7.15	-4.02	0.39	-1.44	0.00
6	RGVTKEIRIPTLEEC	-47.23	-33.83	10.15	1.38	-0.65
7*	CDVCHGSGAKPGTQP	-51.86	-29.98	17.24	-5.73	-0.47
8	KEAAPAAAPAAAAAK	-43.96	-33.85	16.34	-6.23	-1.64
9	DGAADAAPAQAEEKK	-37.41	-33.64	18.18	6.53	-4.77
10	DKVEAEQSLITVEGD	-14.28	-41.09	90.52	-3.75	-2.53
11	DVNVPDIGSDEVEVT	-42.90	-36.36	28.12	-1.62	-1.32
12*	KESAPAAAAPAAQPA	-56.70	-33.40	18.63	-14.44	-4.36
13	AAPAAQPALAARSEK	-25.80	-36.70	15.73	3.45	-3.15
14	DASAIKGTGVGGRLT	-50.99	-35.59	19.69	-7.53	-3.70
15	SPAIRRLLAEHNLDA	-27.66	-39.70	44.84	-1.42	-4.77
16*	GRGVLSLLEHVLLPW	-71.61	-30.82	19.66	-4.93	-2.47
17	EVDEDAEYWAQFEAE	-0.34	-1.03	0.00	0.42	0.00
18*	LEHVLLPWMEDLAPQ	-82.11	-43.09	41.45	-19.44	-1.37
19	WMEDLAPQEEVDEDA	-12.72	-37.29	43.03	8.27	-4.39
20	YWAQFEAEENGEEEE	0.08	-0.00	0.00	0.00	0.00

Table 3.2 Binding Efficiency of E.coli antigenic peptides- MHC Class II complex

To observe the binding efficiency of peptides with MHC Class II molecule, FireDock server was used. Binding efficiency was observed for all predicted peptides. Out of all, five peptides were selected on the basis of their highest binding energy. These peptides were further used for checking the global energy of peptides with TCR also.

3.4 DOCKING TO PREDICT THE BINDING EFFICIENCY OF PREDICTED PEPTIDES TO TCR USING FIREDOCK SERVER

Peptide sequence	Global energy	Attractive VdW	Repulsive VdW	ACE	HB
2	-23.38	-23.47	24.84	-0.09	-2.34
7	-34.05	-33.97	20.38	-5.17	-4.15
12	-8.36	-8.20	3.22	5.95	-1.69
16	-42.75	-29.89	12.68	-0.61	-1.81
18	-0.41	-1.50	0.00	-1.16	0.00

Table 3.3 Binding efficiency of E.coli antigenic peptides- TCR complex

Highest global energy of Peptide 16 was found when docked with TCR.

3.5 3-D STRUCTURES OF PREDICTED PEPTIDES TO MHC CLASS II MOLECULE USING PyMOL SOFTWARE

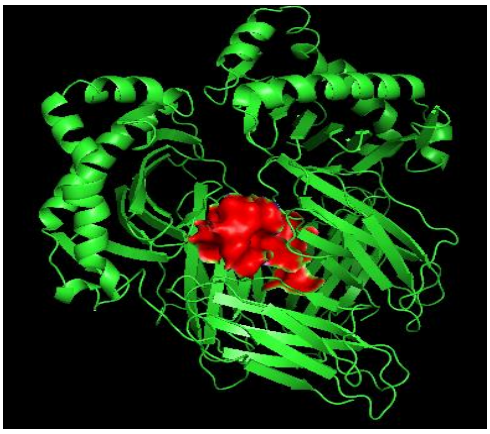


Figure 3.13

MHC Class II-Peptide 1 Complex

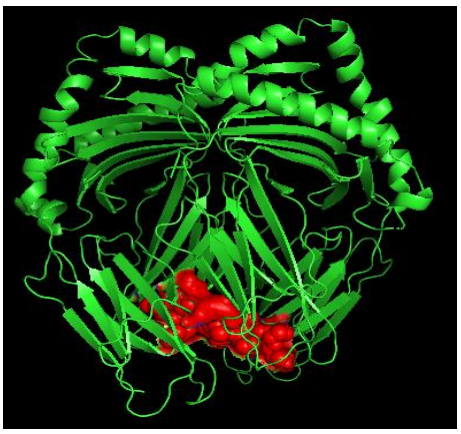


Figure 3.14

MHC Class II-Peptide 2 Complex

Figure 3.15

MHC Class II-Peptide 3 Complex

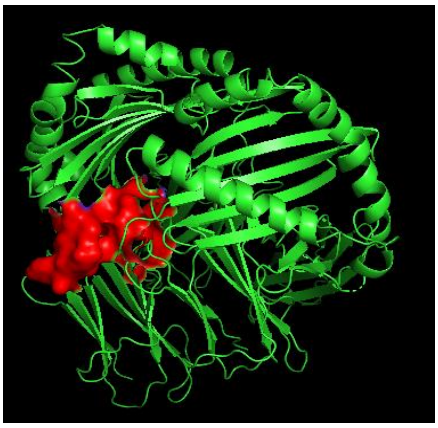


Figure 3.16

MHC Class II-Peptide 4 Complex

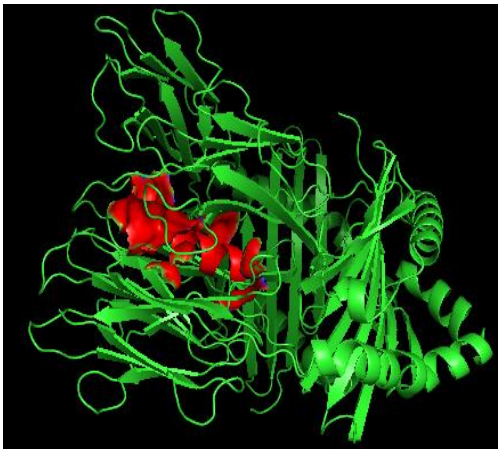


Fig 3.17

MHC Class II-Peptide 5 Complex

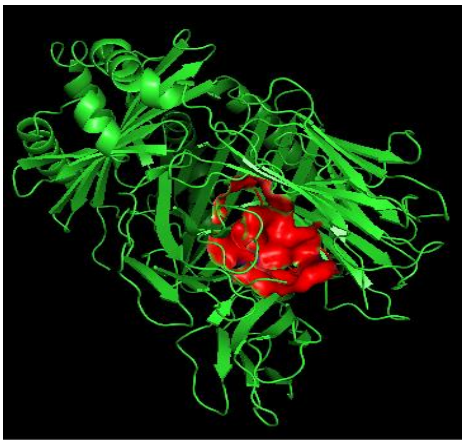


Fig 3.18

MHC Class II-Peptide 6 Complex

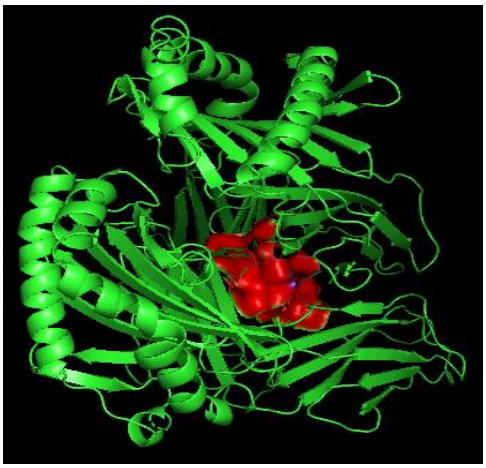


Fig 3.19

MHC Class II-Peptide 7 Complex

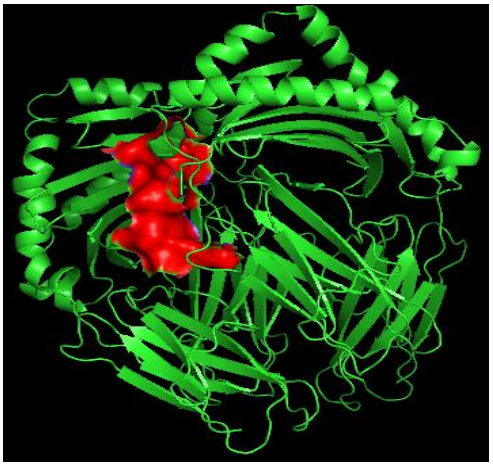


Fig 3.20

MHC Class II-Peptide 8 Complex

Fig 3.21

MHC Class II-Peptide 9 Complex

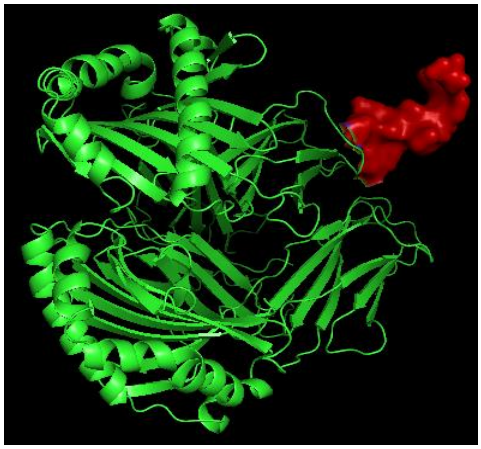


Fig 3.22

MHC Class II-Peptide 10 Complex

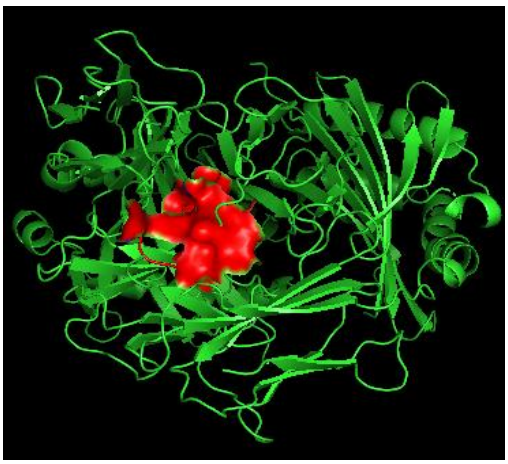


Fig 3.23

MHC Class II-Peptide 11 Complex

Fig 3.24

MHC Class II-Peptide 12 Complex

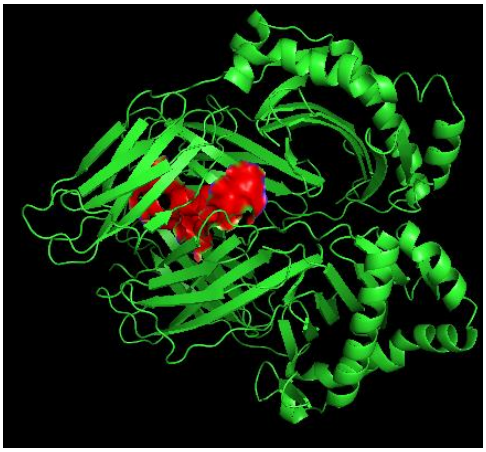


Fig 3.25

MHC Class II-Peptide 13 Complex

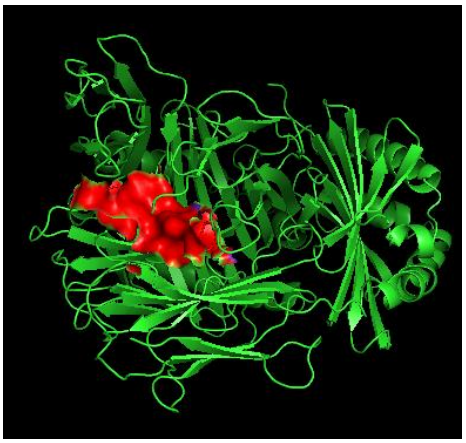


Fig 3.26

MHC Class II-Peptide 14 Complex

Fig 3.27

MHC Class II-Peptide 15 Complex

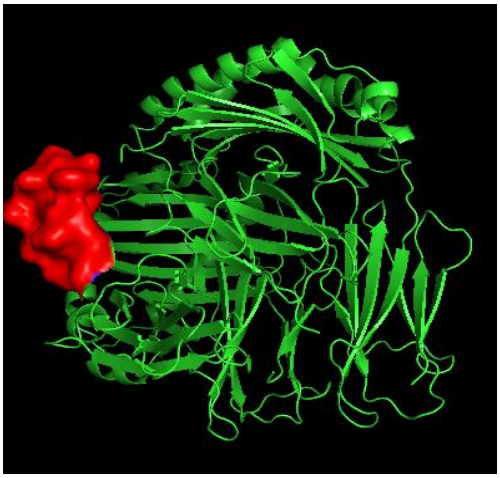


Fig 3.28

MHC Class II-Peptide 16 Complex

Fig 3.29

MHC Class II-Peptide 17 Complex

Fig 3.30

MHC Class II-Peptide 18 Complex

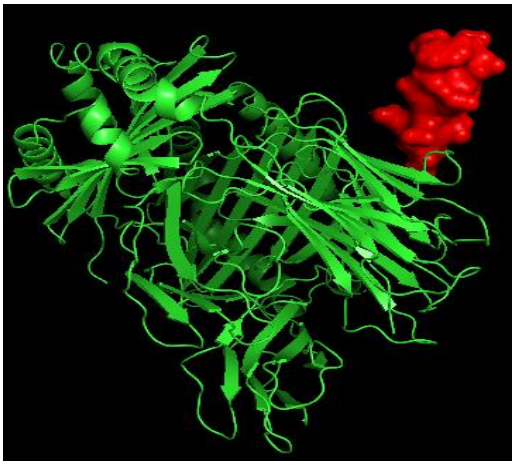


Fig 3.31

MHC Class II-Peptide 19 Complex

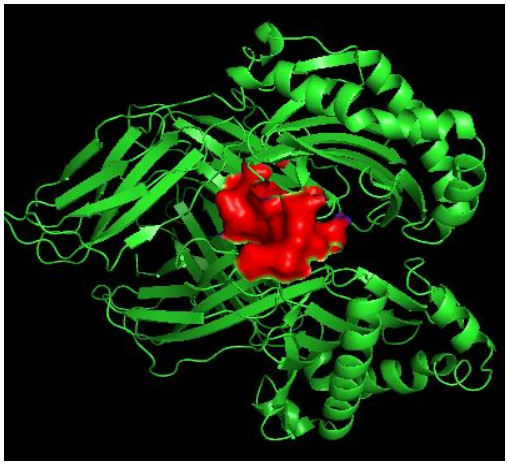


Fig 3.32

MHC Class II-Peptide 20 Complex

3.6 3-D STRUCTURES OF PREDICTED PEPTIDES TO TCR MOLECULE USING PyMOL SOFTWARE

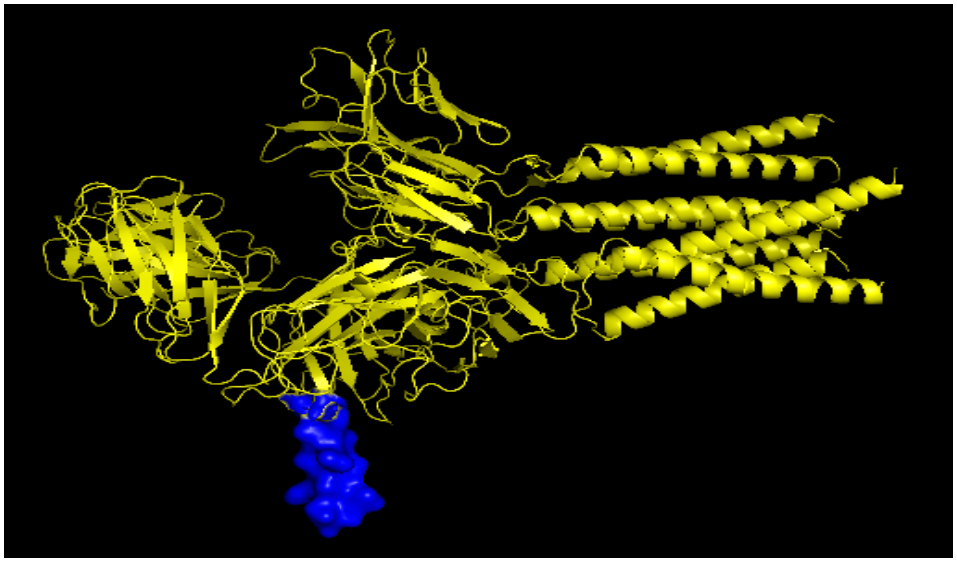


Figure 3.33 TCR- Peptide 2 Complex

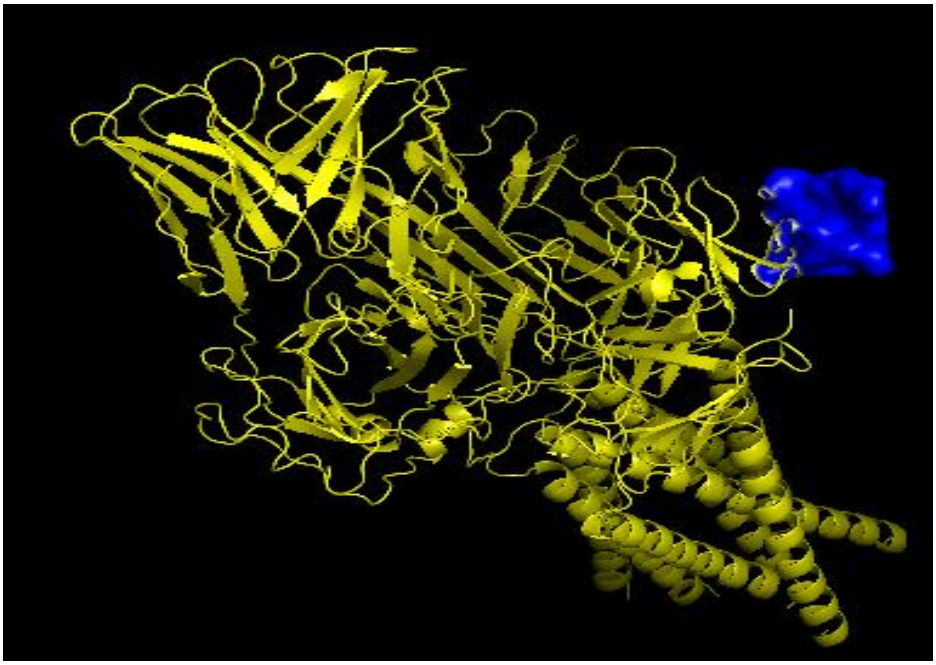


Figure 3.34 TCR- Peptide 7 Complex

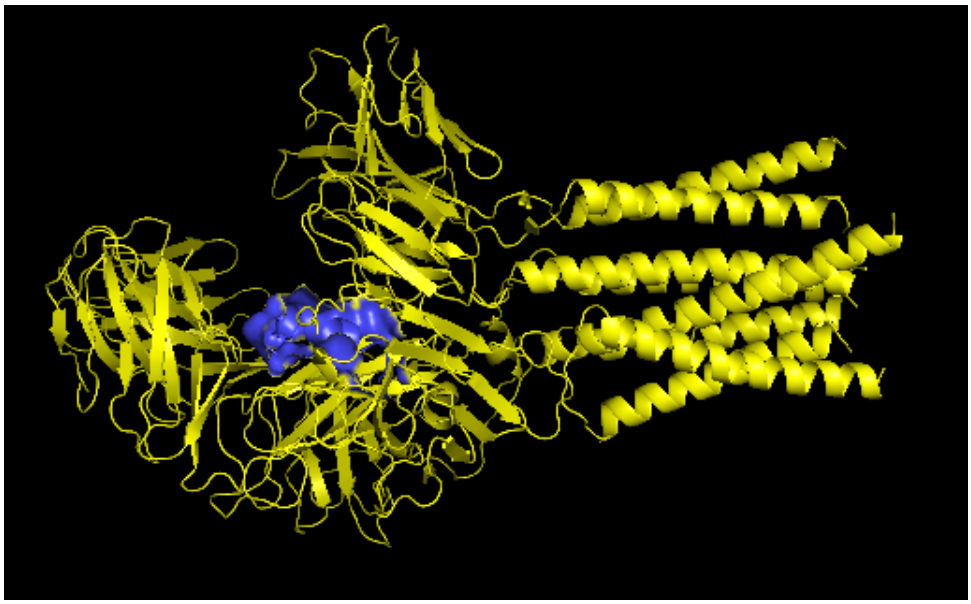


Figure 3.35 TCR- Peptide 12 Complex

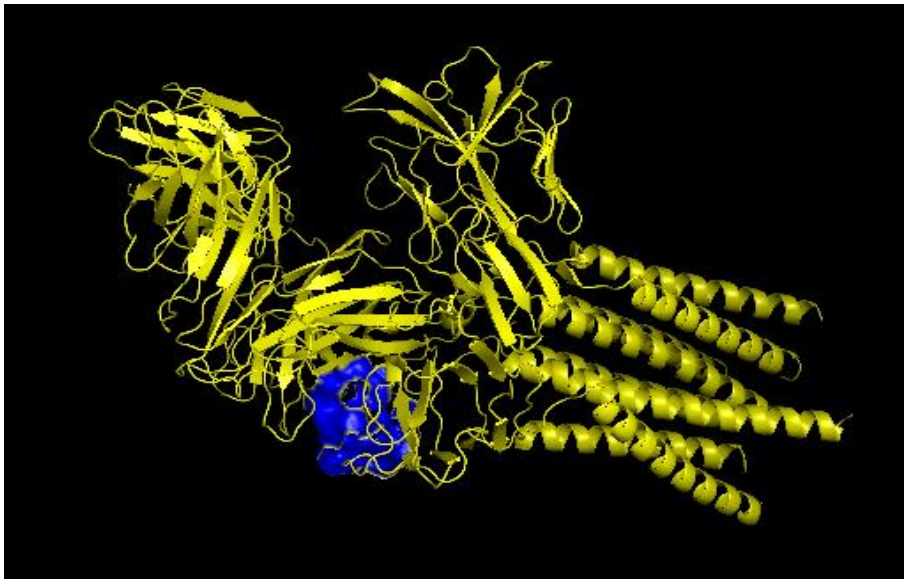


Figure 3.36 TCR- Peptide 16 Complex

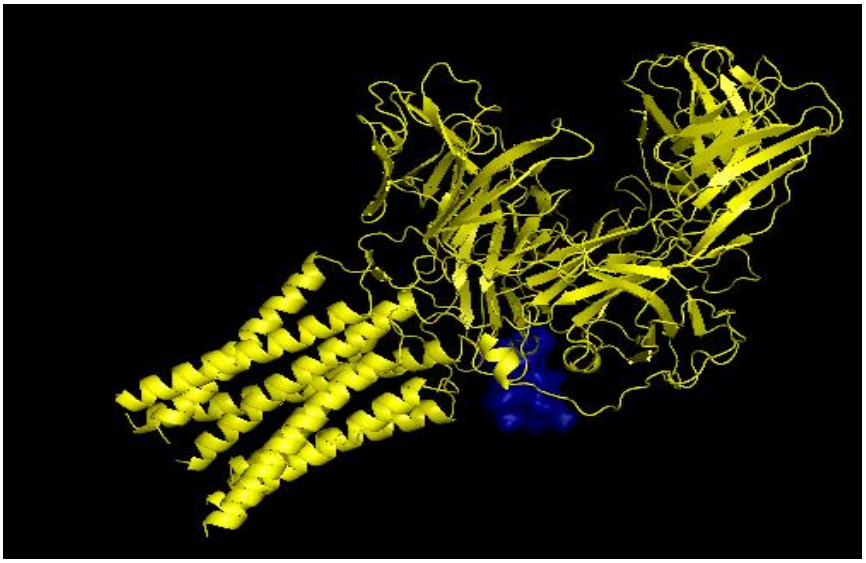


Figure 3.37 TCR- Peptide 18 Complex

Using PyMOL software the 3D structures were visualized for the respective peptides with TCR.

3.7 DISCUSSION

In this study, an E. coli bacterium is taken into consideration which may cause autoimmunity under some conditions. The reasons to study the E.coli includes ubiquitous presence in human ecosystem and it being commonly associated in the microflora. Hence we studied if there are peptides derived from E coli that can crossreact with human peptides such that it may lead to autoimmunity. Autoimmunity is only diagnosed when it gets symptomatic. Applying in-silico approach we can perform pre-symptomatic diagnosis of autoimmune disorders. In contrast to other approaches, epitope prediction study is highly specific. However, for this intervention to be effective, the antigen(s) driving the autoimmune response must be known. To investigate this kind of opportunistic behavior of commensal bacteria, percent similarity between the bacterial antigenic peptides and human peptide sequence was done, which can be responsible for stimulating the autoimmune reactions and hence cause autoimmunity. Firstly, B cell epitopes and T cell epitopes are predicted for a particular bacterium i.e. E. coli. Longest peptide is selected from B cell epitope prediction so that it not show any ambiguity and this is carried out by IEDB followed by T cell epitope prediction using same tool.

Secondly, Protein BLAST is carried out to check similarity with the sequence of Homo sapiens. Molecular docking is performed to identify global energies (binding affinity) for different peptides. This is done to check binding of peptides with both MHC class I and TCR. Since MHC has a broader stringency we need to check global energy for TCR also because our main aim is whether these peptides can be presented to T cells and how well it can bind with TCR to initiate T cell activation and cause an immune response. Five peptides with the highest global energy with MHC class II are selected including PSSASALACSAHALN, CDVCHGSGAKPGTQP, KESAPAAAAPAAQPA, GRGVLSLLEHVLLPW and LEHVLLPWMEDLAPQ. These five peptides are now used to check their global energy with TCR also. Peptide with the highest global energy is GRGVLSLLEHVLLPW i.e., 42.75kcal/mol. PyMOL is used to visualize and obtain 3D structures of peptide-MHC class II and peptide-TCR complex. The peptide with the best global energy with TCR can now used to check its population coverage. This can be done to check the effectiveness of the peptide in Indian population and whether or not it will be well presented by Indian population.

3.8 CONCLUSION

In this study, we made an attempt to identify the bacterial epitope which may cause autoimmunity in humans. This study will help us to look for commensal bacteria and their similarity with human sequences so that later it could be used as a biomarker for detection of autoimmune diseases. Research in the field considering bacteria should be carried out more in order to find more such pathways that can lead to AID. Accordingly, we require more epidemiological and molecular research in this field to gain knowledge about interaction between bacterial infection and host that may cause autoimmune diseases. Through our study, we find some computational data which will help us in identifying as well as screening of epitopes. For the development of vaccines against epitope, this data may be used. Also, this will provide improved efficacy and safety for vaccine designing. For prevention of autoimmune diseases, this result may provide upgraded regimens. Our findings are fully based on analysis of epitope sequence along with some computational predictions. Nonetheless, in vitro studies as well as in vivo experiment are also needed to prove or to confirm the results, interpreted through in silico study.

REFERENCES

[1] "Bergey's Manual® of Systematic Bacteriology - Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria | George Garrity | Springer." https://www.springer.com/gp/book/9780387241449 (accessed May 24, 2021).

[2] H. E. Kubitschek, "Cell volume increase in Escherichia coli after shifts to richer media," *J. Bacteriol.*, vol. 172, no. 1, pp. 94–101, 1990, doi: 10.1128/jb.172.1.94-101.1990.

[3] N. C. Darnton, L. Turner, S. Rojevsky, and H. C. Berg, "On torque and tumbling in swimming Escherichia coli," *J. Bacteriol.*, vol. 189, no. 5, pp. 1756–1764, Mar. 2007, doi: 10.1128/JB.01501-06.

[4] U. Fotadar, P. Zaveloff, and L. Terracio, "Growth of Escherichia coli at elevated temperatures," *J. Basic Microbiol.*, vol. 45, no. 5, pp. 403–404, 2005, doi: 10.1002/jobm.200410542.

[5] W. J. Ingledew and R. K. Poole, "The respiratory chains of Escherichia coli," *Microbiological Reviews*, vol. 48, no. 3. pp. 222–271, Sep. 01, 1984, doi: 10.1128/mmbr.48.3.222-271.1984.

[6] "Bergey's manual of systematic bacteriology (Book, 1984) [WorldCat.org]." https://www.worldcat.org/title/bergeys-manual-of-systematic-bacteriology/oclc/9042846 (accessed May 24, 2021).

[7] O. Lukjancenko, T. M. Wassenaar, and D. W. Ussery, "Comparison of 61 Sequenced Escherichia coli Genomes," *Microbial Ecology*, vol. 60, no. 4. Springer, pp. 708–720, Nov. 11, 2010, doi: 10.1007/s00248-010-9717-3.

[8] G. N. Schroeder and H. Hilbi, "Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion," *Clinical Microbiology Reviews*, vol. 21, no. 1. American Society for Microbiology (ASM), pp. 134–156, Jan. 2008, doi: 10.1128/CMR.00032-07.

[9] "Thompson A (2007) 'E. coli Thrives in Beach Sands'. Live Science. http://www.livescience.com/health/070604_beach_ecoli.html. Retrieved 2014-07-03."

http://www.sciepub.com/reference/109011 (accessed May 24, 2021).

[10] F. Ørskov and I. Ørskov, “Antibodies against core and O antigen epitopes of Escherichia coli O83 in rabbit antisera and in commercial human immunoglobulin,” *Med. Microbiol. Immunol.*, vol. 180, no. 3, pp. 157–166, May 1991, doi: 10.1007/BF00206120.

[11] R. Stenutz, A. Weintraub, and G. Widmalm, “The structures of Escherichia coli O-polysaccharide antigens,” *FEMS Microbiology Reviews*, vol. 30, no. 3. Oxford Academic, pp. 382–403, May 01, 2006, doi: 10.1111/j.1574-6976.2006.00016.x.

[12] J. P. Nataro and J. B. Kaper, “Diarrheagenic Escherichia coli,” *Clinical Microbiology Reviews*, vol. 11, no. 1. American Society for Microbiology, pp. 142–201, Jan. 01, 1998, doi: 10.1128/cmr.11.1.142.

[13] S. Hudault, J. Guignot, and A. L. Servin, “Escherichia coli strains colonising the gastrointestinal tract protect germfree mice against Salmonella typhimurium infection,” *Gut*, vol. 49, no. 1, pp. 47–55, 2001, doi: 10.1136/gut.49.1.47.

[14] J. G. Lawrence and H. Ochman, “Molecular archaeology of the Escherichia coli genome,” 1998. Accessed: May 24, 2021. [Online]. Available: www.pnas.org.

[15] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, and K. Van’t Riet, “Modeling of the bacterial growth curve,” *Appl. Environ. Microbiol.*, vol. 56, no. 6, pp. 1875–1881, 1990, doi: 10.1128/aem.56.6.1875-1881.1990.

[16] “Cellular Microbiology: Bacteria-Host Interactions in Health and Disease | Wiley.” https://www.wiley.com/en-us/Cellular+Microbiology%3A+Bacteria+Host+Interactions+in+Health+and+Disease-p-9780471986812 (accessed May 24, 2021).

[17] R. D. Berg, “Bacterial translocation from the gastrointestinal tract,” *Trends Microbiol.*, vol. 3, no. 4, pp. 149–154, 1995, doi: 10.1016/S0966-842X(00)88906-4.

[18] “Fundamental Immunology.” http://lvts.fr/Pages_html/Encyclopedies/FundamentalImmunology/ramiCOMMAND=applyStylesheet(interface.xsl,CH031_REF.pub)&p_userid=pau&found=.html (accessed May

24, 2021).

[19] H. Tlaskalová-Hogenová *et al.*, "Commensal bacteria (normal microflora), mucosal immunity and chronic inflammatory and autoimmune diseases," *Immunology Letters*, vol. 93, no. 2–3. Immunol Lett, pp. 97–108, May 15, 2004, doi: 10.1016/j.imlet.2004.02.005.

[20] A. C. Ouwehand, "Molecular pathogenesis mechanisms of bacteria," *Trends Immunol.*, vol. 23, no. 10, p. 507, Oct. 2002, doi: 10.1016/s1471-4906(02)02303-7.

[21] J.-F. Bach, "The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases," *N. Engl. J. Med.*, vol. 347, no. 12, pp. 911–920, Sep. 2002, doi: 10.1056/nejmra020100.

[22] F. Bäckhed, "99th Dahlem Conference on Infection, Inflammation and Chronic Inflammatory Disorders: The normal gut microbiota in health and disease," in *Clinical and Experimental Immunology*, Apr. 2010, vol. 160, no. 1, pp. 80–84, doi: 10.1111/j.1365-2249.2010.04123.x.

[23] S. Ehlers and S. H. E. Kaufmann, "99th Dahlem Conference on Infection, Inflammation and Chronic Inflammatory Disorders: Lifestyle changes affecting the host-environment interface: Introduction to and Summary of the 99th Dahlem Conference held in Berlin, June 10-13 2009," in *Clinical and Experimental Immunology*, 2010, vol. 160, no. 1, pp. 10–14, doi: 10.1111/j.1365-2249.2010.04120.x.

[24] P. Youinou, J. O. Pers, M. E. Gershwin, and Y. Shoenfeld, "Geo-epidemiology and autoimmunity," *Journal of Autoimmunity*, vol. 34, no. 3. J Autoimmun, May 2010, doi: 10.1016/j.jaut.2009.12.005.

[25] C. Selmi and M. E. Gershwin, "The role of environmental factors in primary biliary cirrhosis," *Trends in Immunology*, vol. 30, no. 8. Trends Immunol, pp. 415–420, Aug. 2009, doi: 10.1016/j.it.2009.05.006.

[26] "The mosaic of autoimmunity: prediction, autoantibodies, and therapy in autoimmune diseases--2008 - PubMed." https://pubmed.ncbi.nlm.nih.gov/18300564/ (accessed May 24, 2021).

[27] E. Israeli *et al.*, “Anti-Saccharomyces cerevisiae and antineutrophil cytoplasmic antibodies as predictors of inflammatory bowel disease,” *Gut*, vol. 54, no. 9, pp. 1232–1236, Sep. 2005, doi: 10.1136/gut.2004.060228.

[28] F. C. Westall, “Molecular mimicry revisited: Gut bacteria and multiple sclerosis,” *J. Clin. Microbiol.*, vol. 44, no. 6, pp. 2099–2104, Jun. 2006, doi: 10.1128/JCM.02532-05.

[29] M. Blank, O. Barzilai, and Y. Shoenfeld, “Molecular mimicry and auto-immunity,” in *Clinical Reviews in Allergy and Immunology*, 2007, vol. 32, no. 1, pp. 111–118, doi: 10.1007/BF02686087.

[30] W. Van Eden, G. Wick, S. Albani, and I. Cohen, “Stress, heat shock proteins, and autoimmunity: How immune responses to heat shock proteins are to be used for the control of chronic inflammatory diseases,” in *Annals of the New York Academy of Sciences*, 2007, vol. 1113, pp. 217–237, doi: 10.1196/annals.1391.020.

[31] M. J. Barnes and F. Powrie, “Regulatory T Cells Reinforce Intestinal Homeostasis,” *Immunity*, vol. 31, no. 3. Immunity, pp. 401–411, Sep. 18, 2009, doi: 10.1016/j.immuni.2009.08.011.

[32] H. Tlaskalová-Hogenová *et al.*, “Autoimmunity, Immunodeficiency and Mucosal Infections: Chronic Intestinal Inflammation as a Sensitive Indicator of Immunoregulatory Defects in Response to Normal Luminal Microflora,” *Folia Microbiol. (Praha).*, vol. 43, no. 5, pp. 545–550, 1998, doi: 10.1007/BF02820817.

[33] H. Tlaskalová-Hogenová *et al.*, “The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: Contribution of germ-free and gnotobiotic animal models of human diseases,” *Cellular and Molecular Immunology*, vol. 8, no. 2. Cell Mol Immunol, pp. 110–120, Mar. 2011, doi: 10.1038/cmi.2010.67.

[34] A. N. Theofilopoulos, D. H. Kono, and R. Baccala, “The multiple pathways to autoimmunity,” *Nature Immunology*, vol. 18, no. 7. Nature Publishing Group, pp. 716–724, Jun. 20, 2017, doi: 10.1038/ni.3731.

[35] T. Gensollen, S. S. Iyer, D. L. Kasper, and R. S. Blumberg, “How colonization by microbiota in early life shapes the immune system,” *Science*, vol. 352, no. 6285. American Association for the Advancement of Science, pp. 539–544, Apr. 29, 2016, doi: 10.1126/science.aad9378.

[36] A. C. Billi, J. M. Kahlenberg, and J. E. Gudjonsson, “Sex bias in autoimmunity,” *Current opinion in rheumatology*, vol. 31, no. 1. NLM (Medline), pp. 53–61, Jan. 01, 2019, doi: 10.1097/BOR.0000000000000564.

[37] E. Von Mutius and D. Vercelli, “Farm living: Effects on childhood asthma and allergy,” *Nature Reviews Immunology*, vol. 10, no. 12. Nature Publishing Group, pp. 861–868, Nov. 09, 2010, doi: 10.1038/nri2871.

[38] J. F. Bach, “The hygiene hypothesis in autoimmunity: The role of pathogens and commensals,” *Nature Reviews Immunology*, vol. 18, no. 2. Nature Publishing Group, pp. 105–120, Feb. 01, 2018, doi: 10.1038/nri.2017.111.

[39] R. Medzhitov, D. S. Schneider, and M. P. Soares, “Disease tolerance as a defense strategy,” *Science*, vol. 335, no. 6071. American Association for the Advancement of Science, pp. 936–941, Feb. 24, 2012, doi: 10.1126/science.1214935.

[40] T. W. Hand *et al.*, “Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses,” *Science (80-.).*, vol. 337, no. 6101, pp. 1553–1556, Sep. 2012, doi: 10.1126/science.1220961.

[41] O. J. Harrison *et al.*, “Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury,” *Science (80-.).*, vol. 363, no. 6422, Jan. 2019, doi: 10.1126/science.aat6280.

[42] L. Campisi, G. Barbet, Y. Ding, E. Esplugues, R. A. Flavell, and J. M. Blander, “Apoptosis in response to microbial infection induces autoreactive T H17 cells,” *Nat. Immunol.*, vol. 17, no. 9, pp. 1084–1092, Aug. 2016, doi: 10.1038/ni.3512.

[43] M. B. Torchinsky, J. Garaude, A. P. Martin, and J. M. Blander, “Innate immune recognition of infected apoptotic cells directs T H17 cell differentiation,” *Nature*, vol. 458,

no. 7234, pp. 78–82, Mar. 2009, doi: 10.1038/nature07781.

[44] S. Omenetti *et al.*, "The Intestine Harbors Functionally Distinct Homeostatic Tissue-Resident and Inflammatory Th17 Cells," *Immunity*, vol. 51, no. 1, pp. 77-89.e6, Jul. 2019, doi: 10.1016/j.immuni.2019.05.004.

[45] N. S. R. Sanderson *et al.*, "Cocapture of cognate and bystander antigens can activate autoreactive B cells," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 4, pp. 734–739, Jan. 2017, doi: 10.1073/pnas.1614472114.

[46] C. Münz, J. D. Lünemann, M. T. Getts, and S. D. Miller, "Antiviral immune responses: Triggers of or triggered by autoimmunity?," *Nature Reviews Immunology*, vol. 9, no. 4. NIH Public Access, pp. 246–258, Apr. 2009, doi: 10.1038/nri2527.

[47] L. Wen *et al.*, "Innate immunity and intestinal microbiota in the development of Type 1 diabetes," *Nature*, vol. 455, no. 7216, pp. 1109–1113, Oct. 2008, doi: 10.1038/nature07336.

[48] Z. Huang *et al.*, "Antibody neutralization of microbiota-derived circulating peptidoglycan dampens inflammation and ameliorates autoimmunity," *Nature Microbiology*, vol. 4, no. 5. Nature Publishing Group, pp. 766–773, May 01, 2019, doi: 10.1038/s41564-019-0381-1.

[49] T. Vatanen *et al.*, "Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans," *Cell*, vol. 165, no. 4, pp. 842–853, May 2016, doi: 10.1016/j.cell.2016.04.007.

[50] Y. Tsuji, T. Watanabe, M. Kudo, H. Arai, W. Strober, and T. Chiba, "Sensing of commensal organisms by the intracellular sensor nod1 mediates experimental pancreatitis," *Immunity*, vol. 37, no. 2, pp. 326–338, Aug. 2012, doi: 10.1016/j.immuni.2012.05.024.

[51] F. R. C. Costa *et al.*, "Gut microbiota translocation to the pancreatic lymph nodes triggers NOD2 activation and contributes to T1D onset," *J. Exp. Med.*, vol. 213, no. 7, pp. 1223–1239, Jun. 2016, doi: 10.1084/jem.20150744.

[52] M. F. Konig *et al.*, “Aggregatibacter actinomycetemcomitans-induced hypercitrullination links periodontal infection to autoimmunity in rheumatoid arthritis,” *Sci. Transl. Med.*, vol. 8, no. 369, Dec. 2016, doi: 10.1126/scitranslmed.aaj1921.

[53] S. Culina *et al.*, “Islet-reactive CD8+ T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors,” *Sci. Immunol.*, vol. 3, no. 20, 2018, doi: 10.1126/sciimmunol.aao4013.

[54] C. Gil-Cruz *et al.*, “Microbiota-derived peptide mimics drive lethal inflammatory cardiomyopathy,” *Science (80-.).*, vol. 366, no. 6467, pp. 881–886, Nov. 2019, doi: 10.1126/science.aav3487.

[55] W. E. Ruff *et al.*, “Pathogenic Autoreactive T and B Cells Cross-React with Mimotopes Expressed by a Common Human Gut Commensal to Trigger Autoimmunity,” *Cell Host Microbe*, vol. 26, no. 1, pp. 100-113.e8, Jul. 2019, doi: 10.1016/j.chom.2019.05.003.

[56] R. Planas *et al.*, “GDP-l-fucose synthase is a CD4+ T cell-specific autoantigen in DRB3∗02:02 patients with multiple sclerosis,” *Sci. Transl. Med.*, vol. 10, no. 462, Oct. 2018, doi: 10.1126/scitranslmed.aat4301.

[57] R. Hebbandi Nanjundappa *et al.*, “A Gut Microbial Mimic that Hijacks Diabetogenic Autoreactivity to Suppress Colitis,” *Cell*, vol. 171, no. 3, pp. 655-667.e17, Oct. 2017, doi: 10.1016/j.cell.2017.09.022.

[58] N. Tai *et al.*, “Microbial antigen mimics activate diabetogenic CD8 T cells in NOD mice,” *J. Exp. Med.*, vol. 213, no. 10, pp. 2129–2146, Sep. 2016, doi: 10.1084/jem.20160526.

[59] R. Horai *et al.*, “Microbiota-Dependent Activation of an Autoreactive T Cell Receptor Provokes Autoimmunity in an Immunologically Privileged Site,” *Immunity*, vol. 43, no. 2, pp. 343–353, Aug. 2015, doi: 10.1016/j.immuni.2015.07.014.

[60] T. M. Greiling *et al.*, “Commensal orthologs of the human autoantigen Ro60 as triggers of autoimmunity in lupus,” *Sci. Transl. Med.*, vol. 10, no. 434, Mar. 2018, doi: 10.1126/scitranslmed.aan2306.

[61] D. Azzouz *et al.*, “Lupus nephritis is linked to disease-activity associated expansions and

immunity to a gut commensal," *Ann. Rheum. Dis.*, vol. 78, no. 7, pp. 947–956, Jul. 2019, doi: 10.1136/annrheumdis-2018-214856.

[62] C. F. Krebs *et al.*, "Autoimmune Renal Disease Is Exacerbated by S1P-Receptor-1-Dependent Intestinal Th17 Cell Migration to the Kidney," *Immunity*, vol. 45, no. 5, pp. 1078–1092, Nov. 2016, doi: 10.1016/j.immuni.2016.10.020.

[63] M. Viladomiu *et al.*, "IgA-coated E. Coli enriched in Crohn's disease spondyloarthritis promote TH17-dependent inflammation," *Sci. Transl. Med.*, vol. 9, no. 376, Feb. 2017, doi: 10.1126/scitranslmed.aaf9655.