1. ABSTRACT

Human mouth flora has large number of microorganisms and periodontal complications leads to Coronary heart disease (CHD). Periodontitis which is caused by consortia of microbes such as Porphyromonas gingivalis, Aggrigatibacter actinomycetemcomitans, Chlamydia pneumoniae, Campylobacter rectus and entry of these microbes in blood vessel initiate a proinflammatory response. Also antigen presenting cells like macrophages and dentritic cells activate the adenosine receptors (A2bAR and A3AR) and atheroma formation, which leads to atherosclerosis. There is twenty five percent hike in the risk of CHD connected with periodontal disease. The nucleotide sequence from Porphyromonas gingivalis strain ATCC33277, which is the main causative agent of periodontitis were blasted against other microbes associated with dental caries and periodontal disease. The protein sequence of Transglutaminase2 (TG2), which forms tight association with fibronectin(FN), was retrieved and the protein binding regions/sites were predicted using PredictProtein tool. The adhesion molecule- Gingipains R(RgpA and RgpB) both showed similarity in binding sites among TG2 and FN. The predicted B cell epitope of RgpA and RgpB using various tools such as ABCpred Prediction Server, Discotope, Bepipred Linear Epitope Prediction, were subjected to threshold structure prediction. Flexibility, antigencity and surface accessibility were predicted. Further T cell epitope stretch was predicted, that was encompassed in B cell epitope stretch which triggers both CMI and humoral immune response. The T cell epitopes were identified using tools like Multipred, NetMHC, CTLpred, PrediVac and TAP binding region and proteosomal cleavage sites are predicted through TAPPred and PAPROC respectively. The population coverage prediction is also done using PrediVac for identifying the prevalent peptides in different geographical location. Ultimately two epitopes were obtained which inhibit arginine gingipains and predicted as a probable vaccine candidate for periodontitis causing atherosclerosis.

2. INTRODUCTION

2.1 Periodontal diseases

Huge and diverse amount of microbial population inhabits in oral cavity of human beings which enters the two vital system of the body that is in respiratory and gastrointestinal tract and functions accordingly. The mouth contains the highly diverse, enormous and complex microbial population. These microorganisms inhabit the different surfaces of the normal human mouth. In biofilms, bacteria gather and acquire on the both hard and soft tissues of mouth. Some bacterial population causes the oral disease like dental caries, gingivitis, periodontitis etc. (Seymour et al.,2007)

The periodontal diseases are a wider conglomerate of clinical bodies in which stimulation of an inflammatory method results in degradation of the element which are involved in attachment, alveolar bone's loss, so, in this case it is not treated there is a loss of tooth. Periodontal disease is one of the most extensively known diseases of the oral cavities and this is responsible for causing tooth loss in young population. Nowadays, there has been huge interest in the connection of periodontal disease to vital systemic diseases, for example- cardiovascular disease.

The consortia of bacteria recognised as periodontitis causing agents are A.actinomycetemcomitans, Porphyromonas gingivalis, Treponema denticola, T. socranskii, P. Intermedia and Campylobacter rectus. (O'Toole G et al.,2000)

2.2 Atherosclerosis

Cardiovascular disease (CVD), which affects the heart and related process, is caused by heart and blood vessels disorders. It includes coronary heart disease (CHD), cerebrovascular disease, and increase in blood pressure, congenital heart disease, rheumatic heart disease peripheral artery disease and heart failure. Atherosclerosis relates to many such conditions.

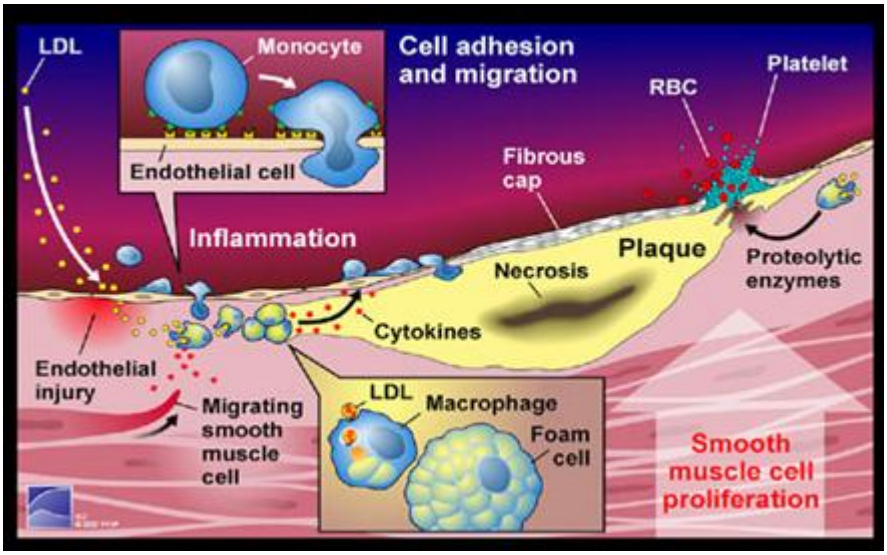


Fig1: Phenomenon showing atherosclerosis progression

Atherosclerosis is a disease in which a person develops a substance called plaque or atheroma. The plaque starts forming in arteries and a layer is visible on their walls. It involves a step by step and slow accumulation of lipids, smooth muscle cells, leucocytes, cholesterol particles, calcium content, and fibrous connective tissue beneath the endothelium surface of the artery, finally forming a build up plaque that comes out into the vessel's lumen and due to this reduces blood flow.

2.3 Relation between periodontal disease and atherosclerosis

Chronic periodontitis and gingivitis which are conditions of oral diseases are ambiguous present worldwide, also it is one of the most prevalent microbial diseases found in human beings.

There is an association between the two disease which been realised by one of the researchers named Chun Xio in the year 2009.

From the research it is shown than bacteria are present in atherosclerotic plaques. According to Ford et al., studies Porphyromonas gingivalis as a causative agent is present in 100% of atherosclerotic plaques and this is shown with the help of real time PCR. Chlamydia pneumonia was found in approximately. And the outcome shows that these microorganisms are not only present in mouth but they also invade blood vessel walls.

Four criteria explaining the association between Periodontitis and atherosclerosis:

1. Some of the common risk factors like smoking, obesity and diabetes are reflected in both the diseases that is atherosclerosis and as well as in periodontitis.

2. According to the theory of auto immunity, immunoglobulins against antigens of bacteria may also contact and react to endothelial protein and in this way can damage the wall of arteries as well as later on begins the arterial lesion.

3. Atherosclerosis and periodontitis both are inflammatory processes. Amount of inflammation in gums elevates by the hike of white blood cells, cells involved in immunological processes and other markers for inflammation in the population suffering with periodontitis.

4. When bacteria which is responsible for the cause of periodontitis get entry inside the blood stream due to injury, cut and tears made by dental procedures, they causes the activation of some immunological cells and produces enzyme which is responsible for causing the blood platelets to become more viscous and blood clots are also formed which in a way ultimately enhances the atheroma formation and hence contribute to the development of atherosclerosis. (Wong et al.,2004)

For adherence and entry into bacterial surface some surface structures are important which include:Fimbrae i.e, binding partner Fibronectin(FN) and Arg-gingipain A and Arg-gingipain B and they form the complex with tranglutaminase of the host epithlial cell.

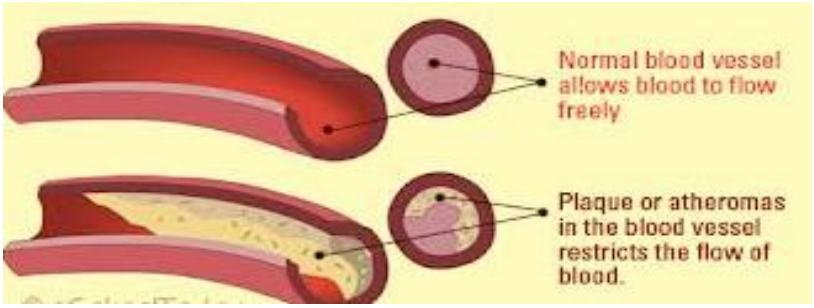


Fig2: Difference in normal and atherosclerotic condition of blood vessel

The connection is indeed biologically feasible. The association among the two diseases have been shown from many epidemiological studies in the past years. According to the studies periodontal disease has 25% chances for chronic heart disease (CHD). With the passage of time the disease become more crucial and increase more risk of CHD.

2.4 Vaccine development

A vaccine is prepared by biological elements either by using killed or attenuated agent which acts against the antigen and when induced inside the host elicit the immune response. These agents are in fact the part of the disease causing organism but it has zero pathogenicity. The agent stimulates the immune system to recognize it as foreign to the bogy and creates a memory so that if the same microorganisms attack the body again, the system encounters it and destroys it. By using bioinformatics tools epitope can be designed for the vaccine and this can reduce the excess time as well as cost or resources for vaccine development.

2.5 Emerging concepts regarding vaccine development

The emerging concepts of development of vaccine for periodontal diseases are:

First one is this that this is the disease which is responsible for causing loss of tooth world-wide. Secondly this disease is caused by large number of microorganisms that is it is polymicrobial. Third one is that this causes very high chances of cardiovascular diseases and one of them is Atherosclerosis. These concepts influence to go for the development of vaccine and to eradicate the disease and help the clinicians in the pathway of the therapeutic studies related to both the diseases.

3. REVIEW OF LITERATURE

3.1 Periodontitis and Atherosclerosis

The micro-organims are present ambiguously in environment and also inside the human body. Mouth has large number of flora. Bacterial species usually exist in ecological balance with each other and with host but some of these in their higher concentration act as antigens and lead to different diseases (McNicol et al.,2010). Forty percent of the human population world-wide suffered from cardiovascular diseases. The cholesterol level in the body is responsible for the development of the disease. The higher concentration of triglycerides and LDL that is low density lipoprotein as well as HDL's less concentration was noted in periodontal patients before treatment. (Mattila et al.,2005)

According to the research, it has been noticed that this oral disease somehow acts as the crucial factor for cardiovascular disease such as atherosclerosis. The association and the relation between oral infections and atherosclerosis have been discussed by many researchers for years. (Niedzielska et al.,2008)

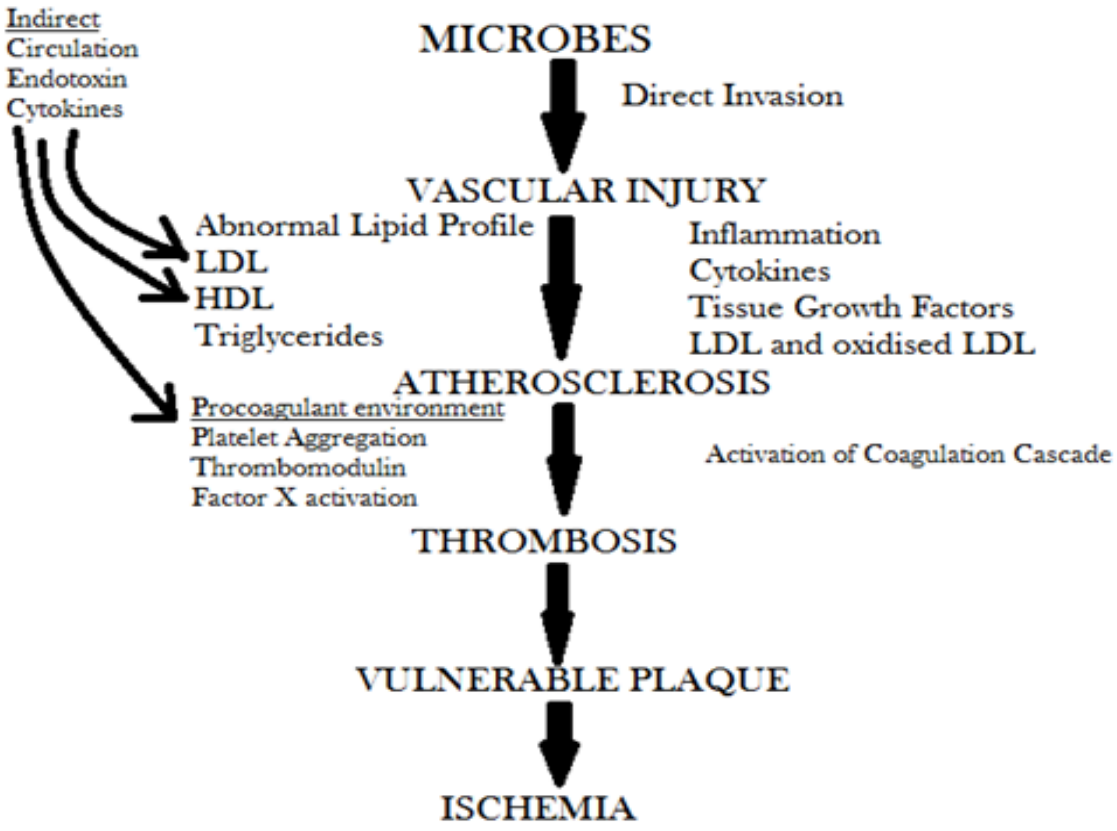


Fig3: Mechanism of infectious agents in Atherosclerosis

It has been proposed that the chronic bacterial infection of periodontitis may promote and modify atherosclerosis by establishing a burden of systemic inflammatory cytokines bacterial antigens, bacterial pathogens and endotoxins which may contribute to the progress of

atherogenetic and thromboembolic events. A study using a non-human primate model has described systemic manifestations of periodontitis, including detection of inflammatory biomarkers, end products of bacteria, and varied concentration level of lipids consistent with an increased atherogenic risk. Studies on human atheromas obtained during endoarterectomy have found multiple periodontal pathogens in the atheromas, including Aggregatibacter actinomycetemcomitans, P. gingivalis, Porphyromonas intermedia and Bacteroids forsythus. Since certain bacteria from dental plaque can cause infective endocarditis and disseminated intravascular coagulation, and have also been localized in human atheromas, these pathogens may promote development of atherosclerosis and trigger coronary thrombosis. The characteristic feature of periodontitis is the loss of tissue bone and due to this the bleeding occurs and the gums get swollen, also there is a movement of tooth which weakens the root and this could lead to complete loss of tooth. Out of twenty five thousand bacterial strains around thousand lies in the oral cavity only that forms the dental plaque. And therefore this process begins with the formation of dental plaque on the surface of tooth. (Axelsson et al.,1981) The consortia of bacteria are recognised as causative microbes in periodontitis are A. actinomycetemcomitans, Porphyromonas gingivalis, Bacteroides forsythus, C.rectus, Treponema denticola, T. socranskii, and P. intermedia. Periodontal diseases may be a risk factor in preterm low birth deliveries, cardiovascular disease, diabetes mellitus, respiratory diseases, and other diseases.

Some of the studies have shown that Porphyromonas gingivalis possess many properties that could lead to CVD as potential mediator of atherosclerosis. There is an emergence of higher platelet activation and aggregation, induction of pro inflammatory mediators and increase in serum lipid level due to P. gingivalis infection. Porphyromonas gingivalis lipo-polysaccharide produces more number of TNF-α and IL-1 by monocytes compared to with E. coli LPS, providing evidence for the potential of this bacterial component to play a critical role in the chronic inflammatory response associated with periodontitis.

3.2 Inflammatory disease

Chronic inflammatory periodontal diseases are found all around the world and are one of the most common chronic infections in human beings. Around ten to fifteen percent population has been observed as affected by advanced for of disease and leads to considerable inflammatory burden. The patients suffering from periodontitis have significantly higher risk of cardiovascular disease (CVD). The systemic inflammation process enhances the level of cytokines and inflammatory mediators. These circulatory cytokines invade the endothelium vessels and destroy it. They subsequently enter the blood stream leading to endothelium dysfunction increase in inflammation and its effects, finally causes atherosclerosis. (Mallika et al.,2007)

3.3 Gingipains produced by P.gingivalis

Actin like adhesions, lipopolysaccharide, hemolysis, capsular polysaccharide, hemagglutinins and fimbriae as well as numerous proteolytic enzymes are different virulent factors produced by Porphyromonas gingivalis.molysins. Gingipains are one of the proteases and they are a

group of cysteine proteases produced by P. gingivalis. Gingipains are of two types one is RgpA and other is RgpB, both can cleave in the arginine residue region. These two are encoded by the gene that is rgpA and rgpB respectively. Gingipains as substrate degrade fibronectin as well as collagen and also damage immunoglobulins, the enzyme protease inhibitors are inactivated and facilitated acquisition of iron. (Gibson et al.,2011)

3.4 TG2 is an important mediator in P.gingivalis infection

Tranglutaminase 2 or TG2 is an enzyme present in the body inside as well as outside the cells and it has a transamidating activity. It is a family of nine proteins of human genome which is evolved from papain group of cysteine proteases. Transglutaminase form the bond with the fibronectin and make a close association with it. This TG2-FN complex interacts with the gingipains cystein proteases and invades into the vessels. P.gingivalis on attachment internalised into human cells; adherence and entry are mediated by bacterial surface structures like gingipain cysteine proteinases and fimbriae. Some surface components of eukaryotic cells are suggested to serve as receptors. Fimbriae as a part of microbial surface include fibronectn or FN its integrin receptors as binding partners. The binding domains of arg-gingipain A and lys-gingipain adhere to epithelial cells (Boisvert et.,2014). TG2 has a high affinity for FN (LeMosy EK, et al.,1992; Gaudry CA, et al.,1999), also fibronectin is a binding surface structure for Porphyromonas gingivalis (Hanazawa et al.,1997; Murakami Y, et al.,1996; Duncan MJ et al.,2008), it has been examined that for P.gingivalis and host interaction TG2 and FN association is necessary.

Epidemiological studies which include the patterns, effects and conditions of disease, have proposed that periodontal diseases are linked with greater risk of Cardiovascular disease (Dzink et al.,1988; Sundqvist et al.,1993; Persson et al.,1994; Houston etal.,1999). It was shown that patients with periodontitis have a nineteen percent greater risk of CVD compared to subjects without periodontitis (Page et al.,2000; Gibson et al.,2004). Moreover, the intima media stiffening is connected with higher risk of acute periodontitis, whereas a systemic antibody response to a periodontal organism is associated with coronary heart disease (Kadowaki et al.,1994). Although observational studies suggest that such an association is not dependent on known confounders like obesity, smoking, hypertension, and diabetes mellitus, the following argument has not been established (Slots et al.,2004).

3.5 Developing strategies for vaccines

In spite of the fact that there are numbers of cultivable microorganisms which are identified in the subgingival habitant, researchers have cut down them to some putative number of periodontal pathogens down to six or seven, P. gingivalis, Campylobacter rectus Treponema denticola and Tannerella forsythia, Aggregatibacter actinomycetemcomitans, Prevotella intermedia, and Fusobacterium nucleatum, and these are widely also dominantly cultivated in sites demonstrating disease activity. (Socransky et al.,) Scientists proposed the Red complex, namely P. gingivalis, Campylobacter rectus and T. denticola, as the predominant disease-associated organisms. Further F. nucleatum, Porphyromonas intermedia, Dialister

pneumosintes, Campylobacter rectus, Eubacterium nodatum and, were also added as possible periodontal pathogens. (Parra et al.,1996; Contreras et al.,1999; Kubar et al.,2004)

The traditional and novel risk factor for atherosclerosis has led to the determination of some components of atherosclerosis. The pathways of the innate immune response set a trend for an adaptive immune response specifically an autoimmune response which conciliates the advancement of atherosclerosis. In this way atherosclerosis as a disease is a candidate to be treated by the vaccine in classical sense of eliciting the immune responses in the body. However, vaccine development renders accuracy in specifying the characteristic of the desired immune response a useful tool when inscribing a disease as complicated as atherosclerosis with a manifold of inflammatory and autoimmune components.

3.6 Immunization against porphyromonas gingivalis

P. gingivalis has been indiscriminated as a vital periodontal pathogen in human periodontitis. In this context, it has discovered a diverse way of survival strategies enabling it to evade host defence mechanisms. Bacterial cell have some virulent components which include capsular polysaccharide (CPS), cysteine proteases, fimbriae, lipopolysaccharide, and outer membrane vesicles (Sundqvist et al.,1993).

Gingipain, the term adopted for P. gingivalis specific cysteine proteases, represents one of the major pathogenic virulence factors for this organism. It consists of two components: gingipain R (RgpA and RgpB) that cleaves proteins at arginine residues, and gingipain K (porphypain 2, Kgp) that cleaves proteins at lysine residues. Therefore, it has drawn considerable interest as a candidate target antigen for periodontal vaccine development.

Hemaglutinin domain which adhere essentially to erythrocytes as well as catalytic domain of RgpA, RgpB and Kgp plays an important role in invasion of defence mechanism of host system by degrading immunoglobulins and complement proteins and by modulating the functions of neutrophils. By the stimulation of these findings, an active immunization program using purified P. gingivalis cysteine protease (porphypain-2) has been carried out, and this raised the response of immunoglobulin G which resulted in a significantly suppressed P. gingivalis-induced bone loss.

With the experiment performed on the animal model it has been realised that antibodies produced against RgpA hemaglutinin domain specific antibodies could prevent periodontal disease. Later on the extended work suggests that, immunization with the RgpA-RgpB adhesin complexes of P. gingivalis protected against periodontal loss of bone by eliciting a high titer of serum IgG2a response in the rat. This approach seems to open a new venue for further trials to pursue.

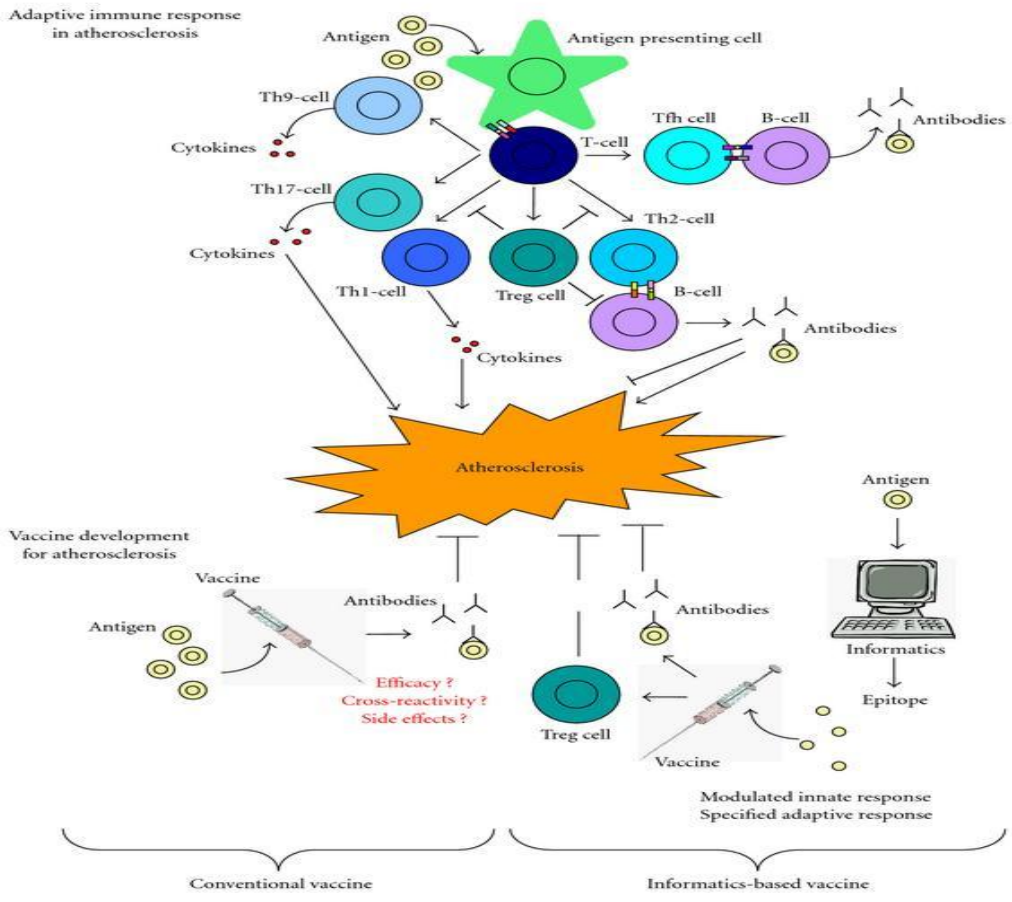


Figure4: Approaches for vaccine development.

3.7 Designing vaccine against periodontitis: Polymicrobial infection

The studies for the immunisation against periodontitis have targeted specifically single pathogenic species in most of the cases. Although, there is a sequence homology which is being shared among varied potential candidate antigenic determinants with other periodontopathic bacteria. These antigens may include phosphorylcholine (Gmur et al.,1999), CPS (Laine et al.,1997), and heat-shock protein (HSP) (Hinode et al.,1998; Maeda et al.,2000).

4. MATERIALS AND METHODOLOGY

4.1 Microorganisms involved

From the literature microorganisms involved in both periodontitis and atherosclerosis are obtained.

1. Porphyromonas gingivalis (strains ATCC33277, W83, W50, TDC60)
2. Aggregatibacter actinomycetemcomitans
3. Chlamydia pneumonia (strains TW183, AR39)
4. Campylobacter rectus (strains RM3267, PW1492, ATCC33238)

4.2 Retrieval of nucleotide sequences from NCBI

The nucleotide sequence of each strain is retrieved from NCBI website (http://www.ncbi.nlm.nih.gov/). Sequence similarity was found out through BLAST among Porphyromonas gingivalis strains. Because P.gingivalis is the main causative agent of atherosclerosis.

4.3 Phylogenetic analysis

A phylogenetic tree or evolutionary tree is a pictorial representation showing branches or "tree" like structure which shows the evolutionary relationships among various species or other entities and the differences and similarities in genetic or physical characters is interpreted through their phylogeny.

Lineages among different sequences or genomic data are typically visualized as parts of a phylogenetic tree.

4.4 Retrieval of protein sequences NCBI

Protein sequences of RgpA and RgpB are retrieved from NCBI and are downloaded in fasta format.

4.5 Similarity searches using BLAST

This program helps in finding regions of local similarity among biological sequences. It searches similarity by comparing the query sequence with the database or by comparing two or more sequences for nucleotide or protein depending upon the sequence and also calculates the values and scores of matches. The similarity search not only finds out the functional and evolutionary relationships between sequences but also identifies the members or sequences of gene families.

4.6 Conserved domains NCBI

It is one of the applications available on NCBI. Te protein sequence query is submitted to get the conserved domains in the sequence. This is completely based on reverse position specific BLAST which results into a position specific matrix scores and generates the conserved domain regions.

It shows the data set containing the sequence alignments as well as profiles which represents the conserved domains in molecular evolution. In the MMDB database the three dimensional structure of proteins and their related information is also present.

4.7 Structure prediction using PHYRE2

As the structure of Transglutaminase is present in bound form so its 3D structure is predicted using Phyre2.

Protein Homology/analogY Recognition Engine V 2.0 versionof Phyre2 is used.

Phyre2 website: sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

(Protein Homology/AnalogY Recognition Engine) is an online service for protein structure prediction and easily as well as freely available for non-commercial use. Phyre is the widely used method for the 3D structure prediction of protein and have many citations. From the primary or linear structure of protein it predicts the 3D structure and also gives the information of secondary structures. Like other remote homology recognition techniques (protein threading), it is able to regularly create reliable protein models.

4.8 Binding sites predicted PredictProtein

The binding sites of Transglutaminase2 and FN are predicted through predict protein tool of protein-protein interaction given on ExPASY.

PredictProtein started out by predicting secondary structure and returning families of related proteins. Also provide structural annotations, by including predictions of non-regular secondary structure and disordered regions, disulphidebridges and inter-residue contacts, and finally by covering trans-membrane beta barrel structures. Annotation of subcellular localization(LocTree, LocTree2, NLSpred), identifying protein-protein interaction sites(ISIS), and protein-DNA binding sites(DISIS) are assisted by prediction of protein function.

4.9 Visualisation by PYMOL

The predicted structure of Transglutaminase2 and gingipains are visualised through PYMOL and binding sites so predicted are also visualised.

PyMOL is a user friendly, open-source, molecular visualization tool which produces high-quality 3D images of biological molecules, for example protein. This tool has wide application in the field of structural biology. The “Py” prefix in the software's name is because of the fact that it is based on the Python programming language.

4.10 B cell epitope prediction by: ABCpred Prediction Server, Discotope, Bepipred Linear Epitope Prediction

(i)ABCpred Prediction Server

The ABCpred server is used to predict B cell epitopes and uses the artificial neural network approach. This server is based on recurrent neural network that is machine based technique which uses fixed length patterns. The server and data sets are available at imtech.res.in/raghava/abcpred/

(ii)Bepipred

BepiPred 1.0 server uses two approaches for prediction of linear B-cell epitopes these are propensity score and hidden markov model and their combination predicts the appropriate location. The protein sequence in FASTA format is uploaded and score threshold of epitope assignment is fixed. The server and data sets are available at cbs.dtu.dk/services/BepiPred/

(iii)DiscoTope

DiscoTope server helps in predicting discontinuous B cell epitopes from three dimensional structures of proteins. In this method surface accessibility is calculated which is estimated in terms of contact numbers and also a novel epitope propensity score of amino acid is calculated. Final scores are calculated after combining both the scores that is the contact numbers and propensity scores of residues in spatial proximity. The server and data sets are available at cbs.dtu.dk/services/DiscoTope/

4.11 Antigenecity, Flexibility, Surface Accessebility Prediction

(i)Antigenecity – Kolaskar and Tongaonkar antigenecity

To predict antigenic determinants on proteins, a semi-empirical method was developed, using physicochemical properties of amino acid residues and occurrence of their frequencies in experimentally known segmental epitopes. This method is much more efficient than any other method analysed through the application based on high number of proteins, as it can predict antigenic determinants with about 75% accuracy.

(ii)Flexibility – Karplus and Schulz Flexibility Prediction

In this method, the known structured are created, according to the known temperature β factors of the α-carbons of 31 proteins which administers the flexibility of segments of these proteins based on mobility. The calculation based on a flexibility scale of these proteins in some way matched to the traditional calculations, the difference is only in the position of the amino acid in the centre of six window size length, as well as for the prediction inspite of one scale, three scales of flexibility are present.

(iii)Surface Accessebility - Emini Surface Accessibility prediction

The calculation for surface accessibility prediction is completely based on the multiplication rather than the addition within the window on accessibility scale. The accessibility profile was obtained using the formula $Sn= (\prod^{6}_{i=1} n+4+i) (0.37)^{-6}$ where Sn is surface probability, dn is the fractional surface probability value and i varies from one to six. A hexapeptide sequence with Sn value greater than 1.0 indicates an increased probability to be found on the surface.

4.12 T cell epitope prediction by: Multipred, NetMHC, CTLpred, PrediVac

(i)Multipred

MULTIPRED2 is a computational method used for screening of peptide binding to multiple alleles which belongs to Class I and Class II MHC of HLA supertypes also to alleles belonging to ones genotype. NetMHCpan and NetMHCIIpan are used for this approach. Binding predictions of 1076 alleles which belongs to 26 different HLA supertypes can be predicted through MULTIPRED2. Built upon the previous developments- MULTIPRED and PEPVAC, MULTIPRED2 is used for mapping of indiscriminating/unselective approach of T-cell epitopes and also the regions of high concentration of these targets which are known as T-cell epitope hotspots. The server and data sets are available at cvc.dfci.harvard.edu/multipred2/

(ii)NetMHC

NetMHC 3.4 server used the approach of artificial neural network for identifying peptides which binds to a number of various HLA alleles. ANNs have been experimented for around 78 different Human MHC (HLA) alleles which represents all twelve HLA A and B Supertypes. The values of predicted peptides are represented by nM IC50 values. For the length from eight to fourteen alleles are predicted using ANNs which is trained with 9mer peptides. This gives the outcomes as Strong and weak binding peptides. The server and this tool is available at cbs.dtu.dk/services/NetMHC-3.4/

(iii)CTLpred

For predicting the T-cell epitopes one of the approach is through CTLPred which predicts and present the information or patterns regarding T-cell epitopes. It typically tells the designing of subunit vaccine. The method is based on the machine learning approach and this technique uses two artificial neural network and support vector machine for the prediction of epitopes. The combination and the consensus of the two methods are used to get better results. The server and data sets are available at imtech.res.in/raghava/ctlpred/

(iv)PrediVac

Predivac is a method to predict HLA classII peptide binding, this is on the basis of SDR concept that is specificity-determining residue. This method integrates the prediction of CD4+ T-cell epitopes and the prediction on the basis of population coverage to obtain the best results. This is an optimal tool to get epitope-based vaccine design in the terms of genetically heterogeneous human population. The server and data sets are available at predivac.biosci.uq.edu.au/

4.13 Prediction of TAP regions by: TAPpred

TAPPred is an on-line service for predicting binding affinities of peptides in context with the TAP transporter. For identification of MHC class I T cell epitopes, prediction of TAP binding peptides is crucial. The Prediction is based on cascade SVM, in which sequence and

properties of the amino acids are used. The jack-knife validation test help to validate 0.88 value of correlation coefficient. The server and data sets are available at imtech.res.in/raghava/tappred/

4.14 Proteosomal site prediction using PAProc

PAProC is a prediction tool which is used to predict cleavages by human as well as yeast proteosomes, and this is based on an experimental cleavage data. The server and data sets are available at www.paproc.de/

4.15 Vaccine stretch prediction according to geographical area by: PrediVac-Population coverage prediction

T-cell epitopes are predicted in proteins or antigens, not for specific MHC class II alleles but also for a large set of HLA class II proteins that are present in a given human population or target population, which is linked to a particular geographic region, country or some ethnicity, on the basis of allelic frequencies. The population coverage is calculated by the combination of two methods: simple search and optimized search. The server and data sets are available at predivac.biosci.uq.edu.au/cgi-bin/population.py

5. RESULTS

5.1 Retrieval of Accession number and GI number of nucleotide sequences of ten microorganisms from NCBI

NCBI's GENOME database (http://www.ncbi.nlm.nih.gov/genome/) was searched with each microorganism name. The genome information was downloaded. The retrieved information included name of the bacterial species, strain, accession number, GI number etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	NAME OF BACTERIAL SPECIES					STRAIN		ACCESSION NUMBER			GI NUMBER		
2													
3	Porphyromonas gingivalis					ATCC33277		Accession:	AP009380.1		GI:	188593544	
4													
5						W83		Accession:	AE015924.1		GI:	34398108	
6													
7						W50		Accession:	NZ_AJZS01000098.1		GI:	419971567	
8													
9						TDC60		Accession:	AP012203.1		GI:	33802964	
10													
11													
12	Aggregatibacter actinomycetemcomitans							Accession:	M27399.1		GI:	141832	
13													
14													
15	Chlamydia pneumoniae					TW183		Accession:	NC_005043.1		GI:	33241335	
16													
17						AR39		Accession:	KC512913.1		GI:	459926476	
18													
19	Campylobacter rectus					RM3267		Accession:	ACFU01000001.1		GI:	222880176	
20													
21						PW1492		Accession:	HQ890331.1		GI:	321530461	
22													
23						ATCC33238		Accession:	EU119866.1		GI:	159080915	

Table1: All the information of retrieved bacterial strains.

5.2 Retrieval of nucleotide sequences from NCBI and performing BLAST.

NCBI Blast (http://www.ncbi.nlm.nih.gov/) Blast search was performed with an E- value threshold of 10^{-6} to find the similarity searches of selected P.gingivalis strains.

Query sequence: Nucleotide sequence of Porphyromonas gingivalis ATCC33277

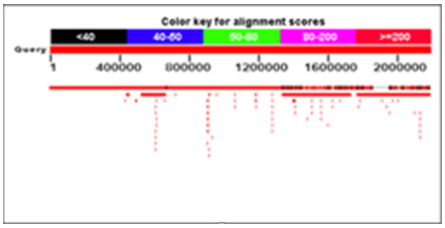


Fig5: BLAST results of Porphyromonas gingivalis ATCC33277

Porphyromonas gingivalis ATCC 33277 DNA, complete genome	1.226e+06	8.375e+06	94%	0.0	99%	AP009380.1
Porphyromonas gingivalis TDC60 DNA, complete genome	7.810e+05	6.653e+06	97%	0.0	100%	AP012203.1
Porphyromonas gingivalis W83, complete genome	7.253e+05	6.540e+06	93%	0.0	100%	AE015924.1
Porphyromonas gingivalis strain HG66 genome	2.587e+05	8.450e+06	96%	0.0	99%	CP007756.1
Porphyromonas gingivalis DNA, CRISPR37, strain: 33277	31462	1.378e+05	2%	0.0	100%	AB757234.1
Porphyromonas gingivalis hemagglutinin A (hagA) gene, complete cds	19509	1.605e+05	1%	0.0	99%	U41807.1
Porphyromonas gingivalis hemagglutinin/protease (hagE) gene, complete cds	15946	78852	1%	0.0	99%	AF026946.1
Porphyromonas gingivalis DNA, CRISPR30, strain: 33277	14552	2.977e+05	0%	0.0	100%	AB757108.1
P.gingivalis prpR1 gene	11891	65470	1%	0.0	99%	X82680.1
Porphyromonas gingivalis Arg-gingipain-1 proteinase gene, complete cds	11635	65928	1%	0.0	99%	U15282.1

Fig6: The similarity searches against Porphyromonas gingivalis ATCC33277

This gives 14 results which have the threshold e-value 10^{-6} and identity greater than or equal to 99%.

Query sequence: Nucleotide sequence of Porphyromonas gingivalis W83

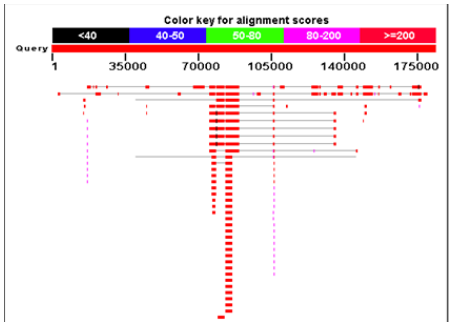


Fig7: BLAST results of Porphyromonas gingivalis W83

Query sequence: Nucleotide sequence of Porphyromonas gingivalis TDC60

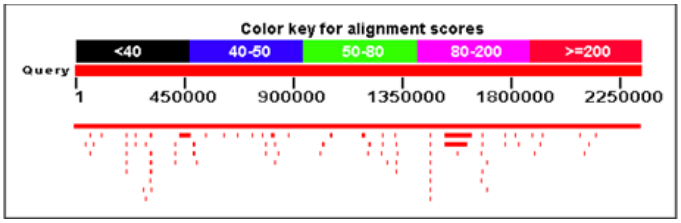


Fig8: BLAST results of Porphyromonas gingivalis TDC60

5.3 Phylogenetic Analysis

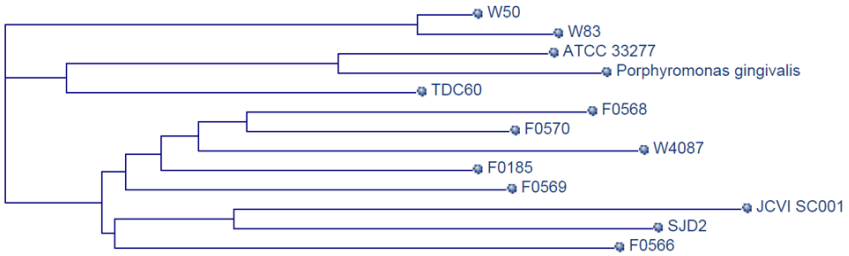


Fig9: Phylogenetic Analysis of different strains of Porphyromonas gingivalis

The phylogenetic analysis suggests that W83 and W50 strain were evolved from same ancestor as their internal node is common and ATCC332277 and TDC60 were evolved from same ancestror. F0568, F0570, W4087, F0185, F0185 strains and JCVI SC001, SJD2, F0566 strains were evolved from same ancestors. The ATCC 33277 and W83 have long branches, therefore they have evolved more over a period of time than W50 and TDC60 strains respectively as shown in figure 9.

5.4 Transglutaminase2 (TG2) in human epithelial cells acts as an adhesion molecule for P.gingivalis associating factors.

Conserved Domains of TG2 are obtained from the "conserved domains" tool given on NCBI site. This shows four conserved domain hits and regions.

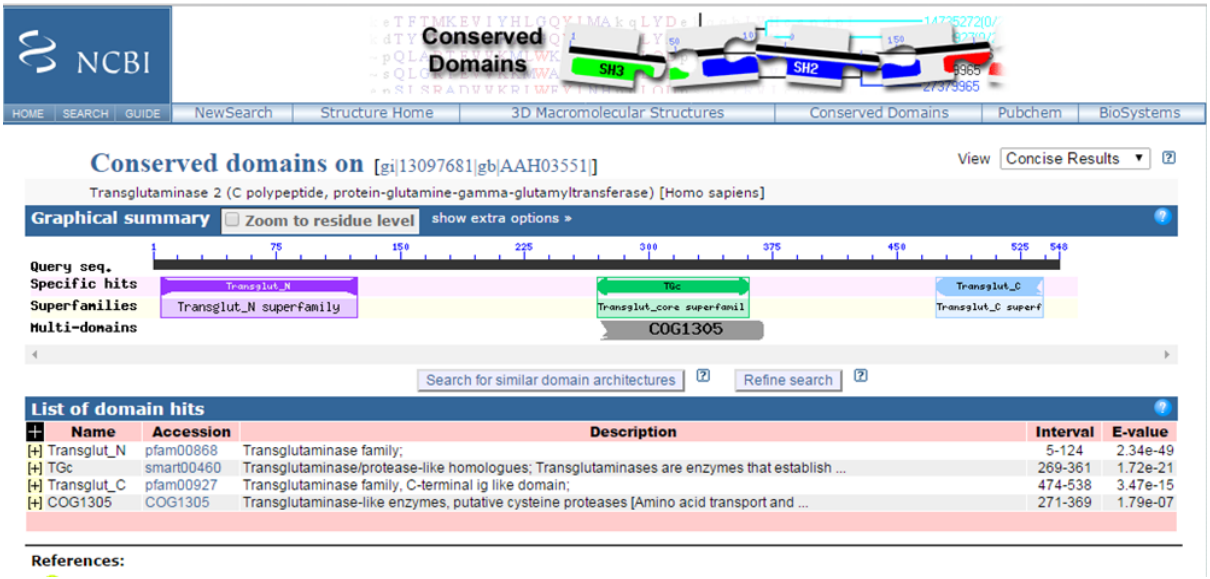


Fig10: Conserved Domains of TG2

This gives the Cd length, Bit Score and E-value for each hit. Transglut_N showed the maximum Bit score i.e, 167.39.

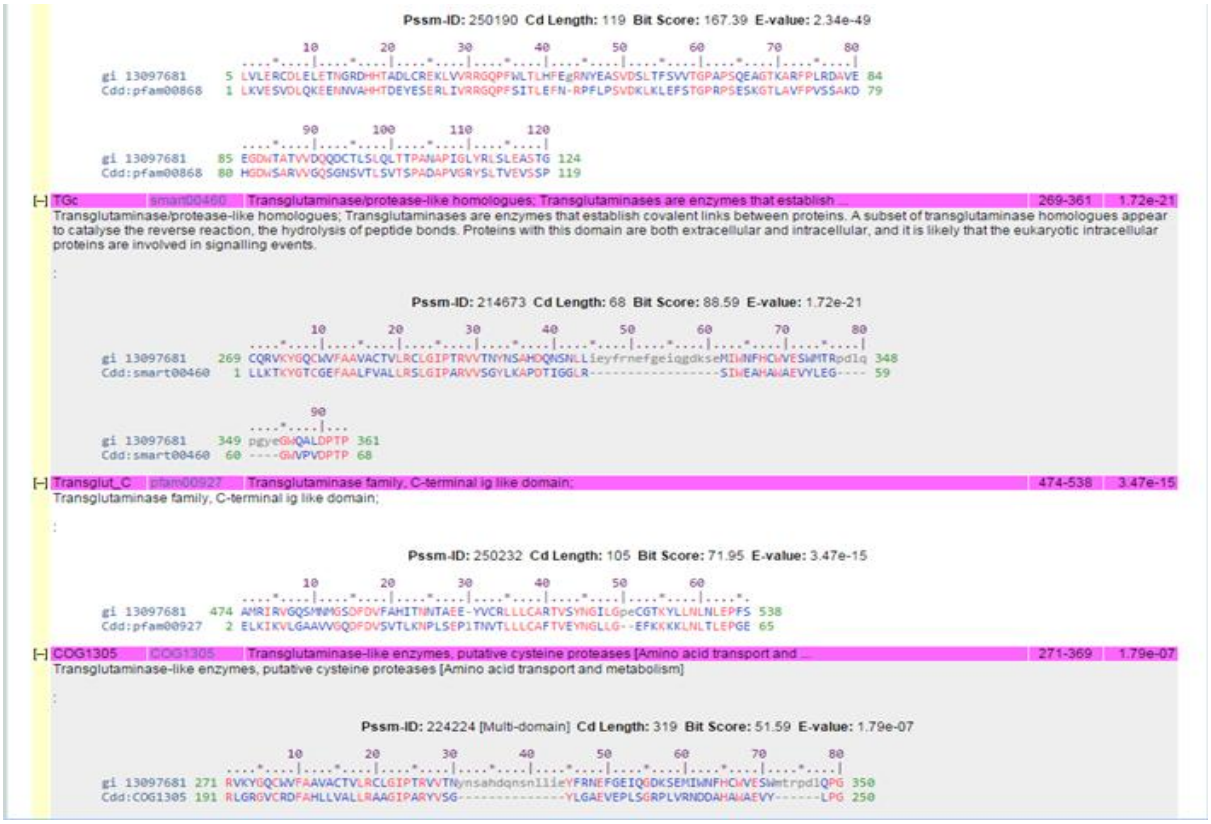


Fig11: Conserved domains with query sequence and four other hits.

5.5 Prediction of 3D structure of TG2 using PHYRE2 tool.

The protein sequence is obtained from NCBI and exported into PHYRE2 which involved six steps:

i) Template used to build model
ii) Secondary structure and disorder prediction (by PSI-pred program)
iii) Domain analysis
iv) Alignment
v) Loop modelling
vi) Generating final results

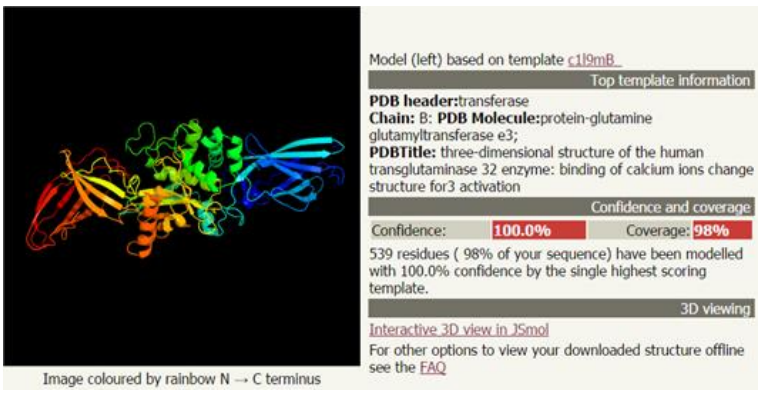


Fig12: Phyre2 result-predicted 3D structure.

The structure shows the information about template, confidence and coverage etc.

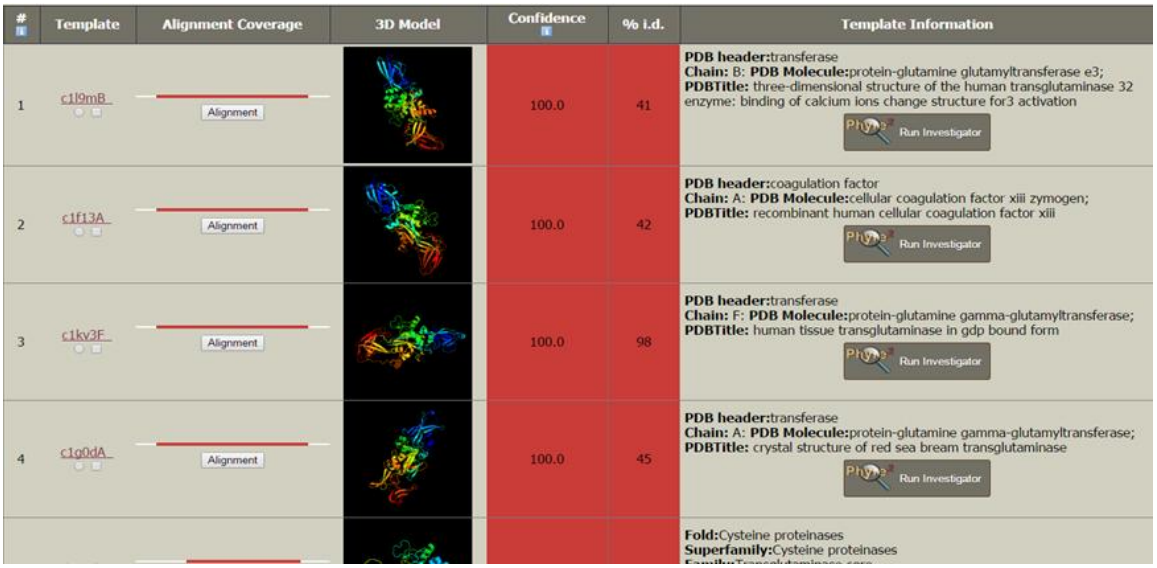


Fig13: Results of Phyre2

Four structures are predicted which are having confidence 100% and percentage identity more than 90% and there 3D models are obtained

5.6 Prediction of protein binding regions/binding sites of TG2 using PredictProtein tool.

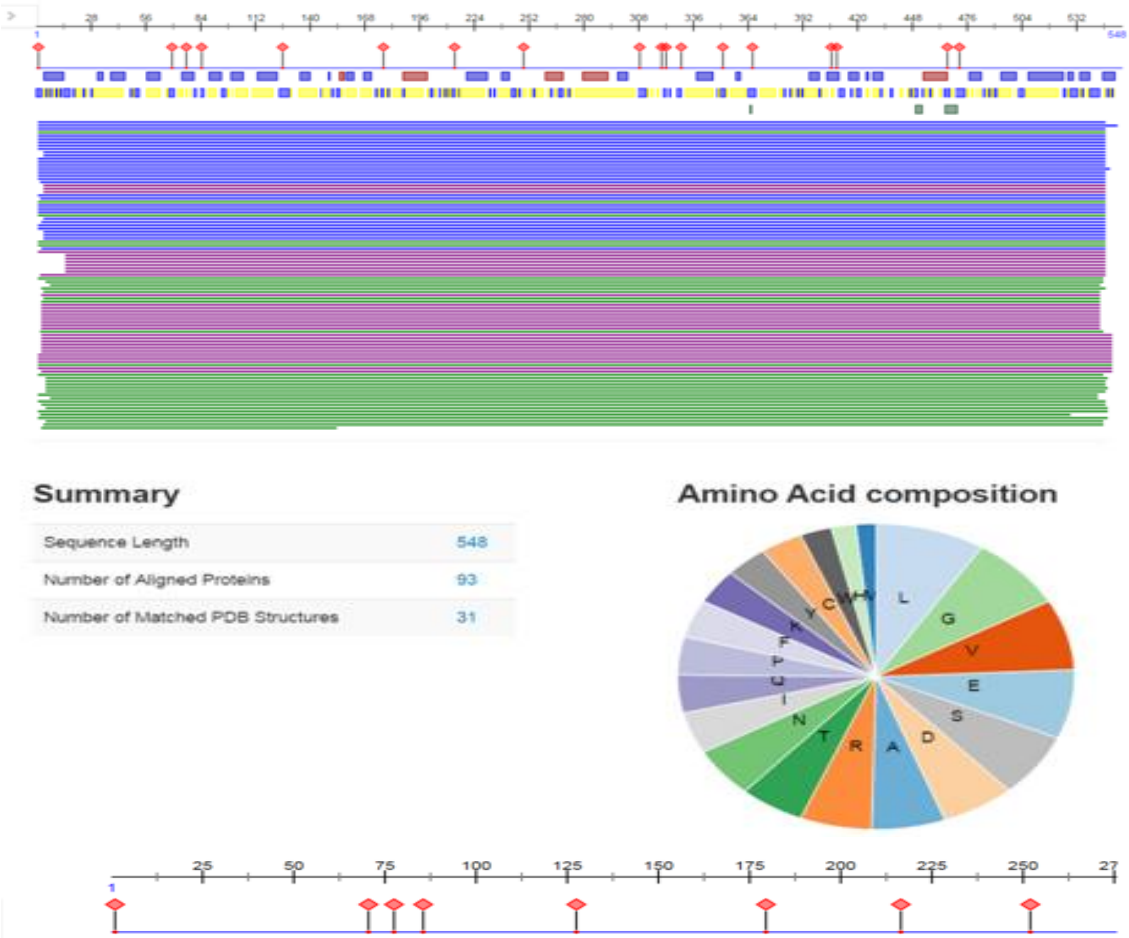


Fig14: Binding sites of TG2 protein. PredictProtein has aligned 93 proteins and 31 matched structures were obtained. Lysine is the most prevalent amino acid according to amino acid composition shown above. 17 binding regions were obtained which are shown in red.

The results are obtained with seventeen protein binding regions including:

Binding	Region	Length size	Binding	Region	Length size
69	70	Length 2	319	320	Length 2
76	76	Length 1	327	327	Length 1
84	86	Length 3	348	358	Length 5
125	126	Length 2	363	366	Length 4
176	176	Length 1	403	403	Length 1
212	212	Length 1	406	407	Length 2
247	247	Length 1	462	464	Length 3
306	308	Length 3	469	469	Length 2
317	317	Length 1			

Table2: Binding regions and their length

Fig15: Visualisation of 3D predicted structure of TG2 using PYMOL

The white spherical balls are showing the predicted protein binding regions of TG2.

5.7 B-cell epitope prediction using different tools

B cell epitopes were predicted using different tools:

Epitope prediction – ABCpred Prediction Server

Discotope

Bepipred Linear Epitope Prediction

1) Antigenecity – Kolaskar and Tongaonkar antigenecity
2) Flexibility – Karplus and Schulz Flexibility Prediction
3) Surface Accessebility - Emini Surface Accessibility prediction

(i) ABCpred Prediction Server

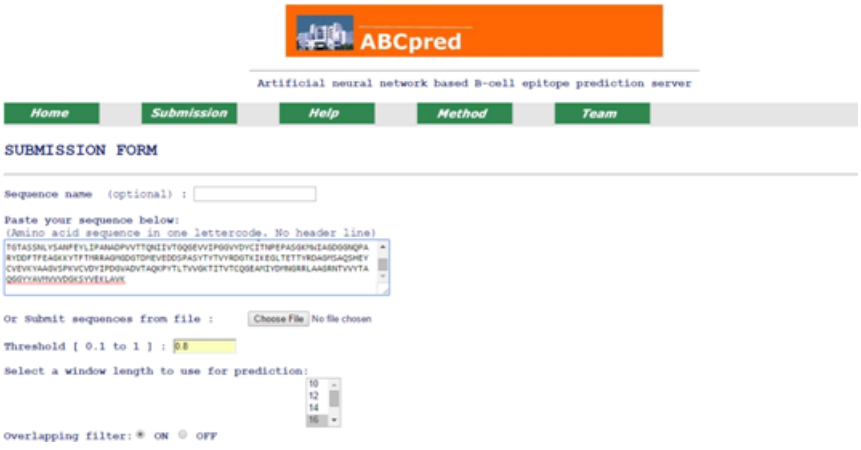


Fig16: Sequence submission for B-cell epitope prediction by ABCpred

ABCpred Prediction Server

INPUT INFORMATION

Sequence name	
Length of the sequence	1703
Number of 14mers from the input sequence	1690
Threshold setting (Default value is 0.5)	0.8

Rank	Sequence	Start position	Score
1	TFEEDGVATGNHEY	1339	0.96
2	GARFGLSTEANGAK	1257	0.93
3	VYRDGTKIKEGLTA	867	0.92
3	INTNGEPNPYQPVS	652	0.92
3	GAIATISANGKMFG	601	0.92
3	IKDFVDWKNQRGLR	248	0.92
4	RGQVVNFAPLQYNP	161	0.91
4	TKYVAFRHFQSTDM	1106	0.91
5	PQSVWIERTVDLPA	1271	0.90
5	AKGVRSPEAIRGRI	1078	0.90
5	YSESFGLGGIGVLT	1006	0.90
6	DVACVNGDFLFSMP	465	0.89
6	SSTHGEAPAEWTTI	1144	0.89
6	DLPAGTKYVAFRHF	1101	0.89
7	LRSGQAEIVLEAHD	715	0.88

Fig17: Predicted B-cell epitope by ABCpred

Using 0.8 and 0.71 threshold, in total 372 peptide stretches is obtained.

Peptide size	Number of peptides obtained
20 mer	80
18 mer	70
16 mer	106
14 mer	83
12 mer	33

Table3: Peptide size and number of peptides obtained by ABCpred

(ii) Discotope

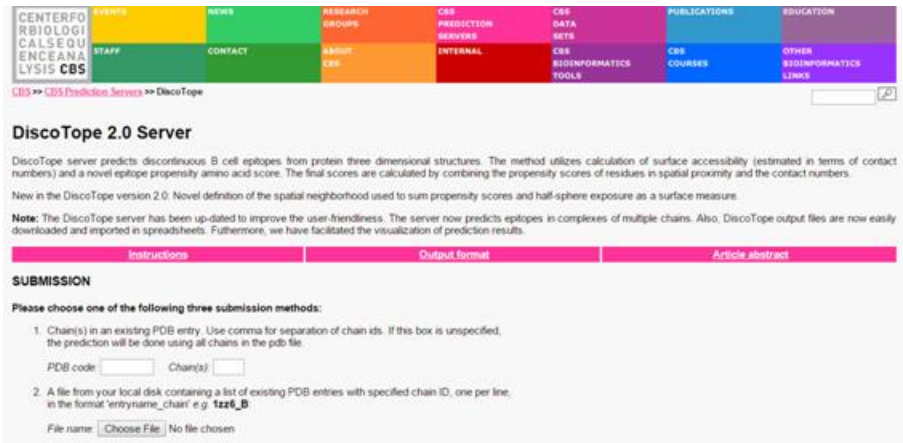


Fig18: Screenshot of DiscoTope 2.0 server

Ten peptides are predicted on the basis of discotope score more than -7.7.

PAS
DADGDGN
PPGGS
NFE
VTAPEAR
FE
SEA
TAPEA
NGIPASWLTIDADGDGN

Table4: Sequences obtained from DiscoTope

(iii) Bepipred Linear Epitope Prediction

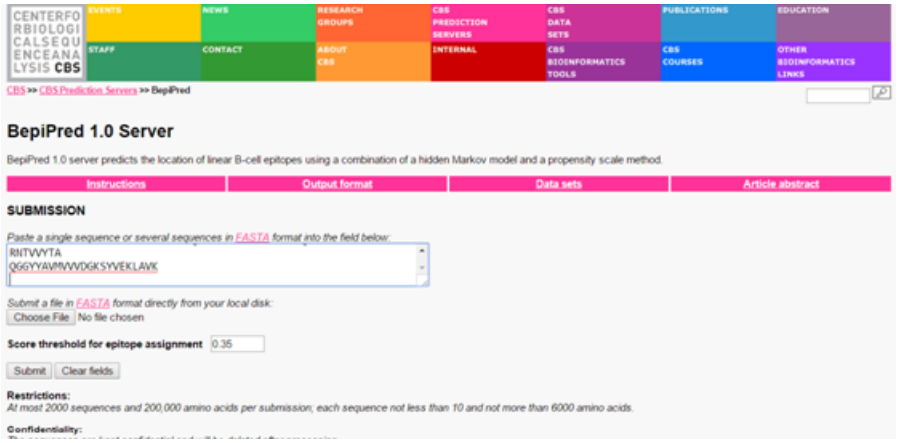


Fig19: Screenshot of BepiPred 1.0 server

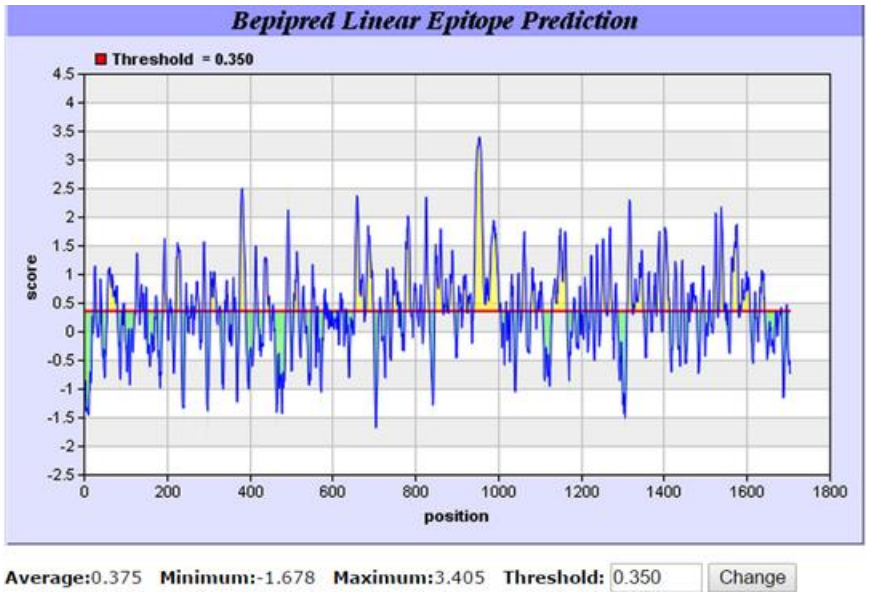


Fig20: Predicted epitopes in yellow peaks by BepiPred

By using 0.35 thresholds value 30 peptides of more than 10mer size are obtained.

Predicted epitopes:

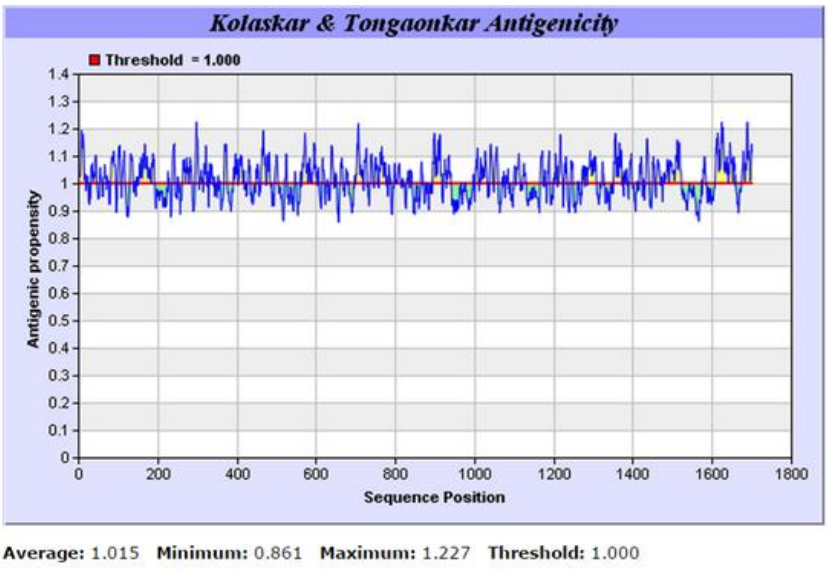
No.	Start Position	End Position	Peptide	Peptide Length
1	23	28	TELGRN	6
2	38	41	QQSV	4
3	56	80	VQTPKGMAQVPTYTEGVNLSEKGMP	25
4	92	92	S	1
5	94	94	T	1
6	96	96	E	1
7	124	133	NEDPKKIPYV	10
8	135	144	GKSYSQNKFF	10
9	148	151	IATL	4
10	172	175	YNPV	4
11	190	199	SETSEQGKNI	10
12	206	209	FAGF	4
13	211	211	D	1
14	221	233	EPGRYTPVEEKQN	13
15	244	247	YEGD	4
16	249	249	K	1

Fig21: Predicted epitope stretches obtained from BepiPred

5.8 Antigenecity, Flexibility and Surface Accessebility prediction

1) Antigenecity – Kolaskar and Tongaonkar antigenecity
2) Flexibility – Karplus and Schulz Flexibility Prediction
3) Surface Accessebility - Emini Surface Accessibility prediction

(1)Antigenecity: Kolaskar and Tongaonkar antigenecity



Average: 1.015 **Minimum:** 0.861 **Maximum:** 1.227 **Threshold:** 1.000

Fig22: Predicted epitopes in yellow peaks by Kolaskar and Tngaonkar antigenecity

By using 1.00 thresholds value 20 peptides of more than 10mer size are obtained.

(2)Flexibility – Karplus and Schulz Flexibility Prediction

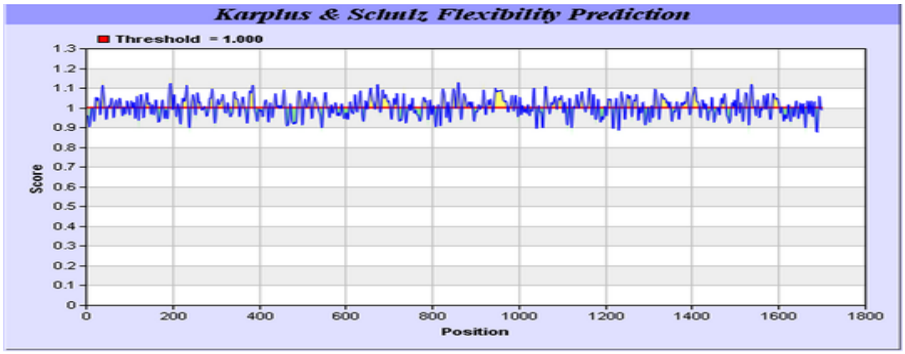


Fig23: Predicted epitopes in yellow peaks by Karplus and Schulz Flexibility

Average: 1.006 Minimum: 0.879 Maximum: 1.130 Download data to file

Position	Residue	Peptide start position	Peptide end position	Peptide	Score
859	G	856	862	SEGGGSD	1.130 (maximum)
860	G	857	863	EGGGSDY	1.126
194	E	191	197	ETSEQGK	1.125
195	Q	192	198	TSEQGKN	1.123
1537	N	1534	1540	DGGNQPA	1.120
673	G	670	676	TTQGQKV	1.119
193	S	190	196	SETSEQG	1.118
1536	G	1533	1539	GDGGNQP	1.117
672	Q	669	675	ATTQGQK	1.116
384	G	381	387	ADNGESD	1.115
858	G	855	861	ISEGGGS	1.114
38	Q	35	41	ESTQQSV	1.114
232	Q	229	235	EEKQNGR	1.114
385	E	382	388	DNGESDI	1.113
1061	G	1058	1064	SSTGNDA	1.109
1406	G	1403	1409	KTEGSRE	1.107
37	T	34	40	LESTQQS	1.107
1538	Q	1535	1541	GGNQPAR	1.105

Fig24: Resulted peptides and their scores obtained from Karplus and Schulz Flexibility

(3)Surface Accessebility - Emini Surface Accessibility prediction

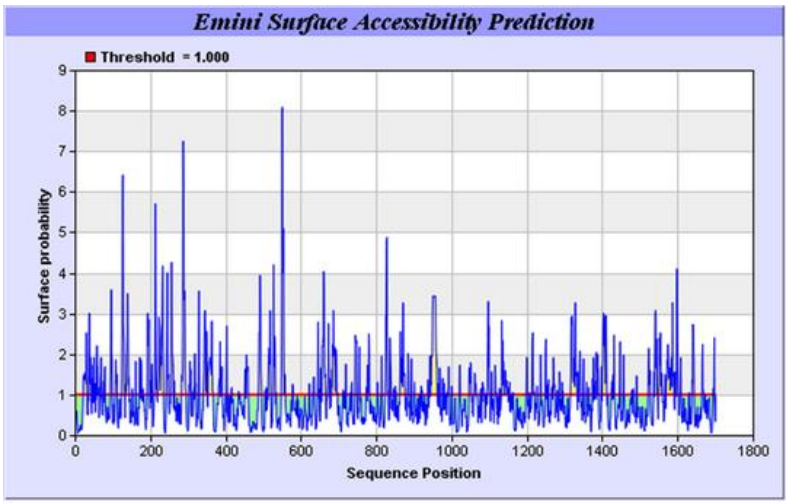


Fig25: Predicted epitopes in yellow peaks by Emini Surface Accessibility

Predicted peptides:

No.	Start Position	End Position	Peptide	Peptide Length
1	21	30	QQTELGRNPN	10
2	92	97	SDTREM	6
3	123	130	RNEDPKKI	8
4	133	142	VYGKSYSQNK	10
5	191	196	ETSEQG	6
6	210	216	EDTYKRM	7
7	219	233	NYEPGRYTPVEEKQN	15
8	242	247	KKYEGD	6
9	254	260	WKNQRGL	7
10	282	293	FVKQEYEKEGND	12
11	340	349	ESKEDLKTQI	10
12	352	365	TIHYERNITTEDKW	14
13	453	458	QLTNSN	6
14	486	494	RAQKDGKPT	9
15	512	518	MRGQDEM	7
16	525	531	KHPNNIK	7
17	547	556	EKYKKDGEKM	10

Fig26: Resulted peptides and their lengths obtained by Emini Surface Accessibility

By using 1.00 thresholds value 10 peptides of more than 10mer size are obtained.

NYEPGRYTPVEEKQN
FVKQEYEKEGND
TIHYERNITTEDKW
WDAPSTKTNATTNTAR
SDYTYTVYRDGTKI
PNGTPNPNPNPNPNPNPG
ANGKRADFTETFESSTHG
SPTPTDYTYTVYRDGTKI
PSAKKTEGSREV
MEVEDDSPASYTYTVYRDGTKI

Table5: Selected peptides of length more than 10mer obtained from Emini Surface Accessibility

Here only surface accessible peptides were selected since they can be recognized as antigenic determinants for the prediction of vaccine candidate.

5.9 T-cell epitope prediction using different tools

T cell epitopes were predicted using different tools for:

1) Stretches for class I MHC – Multipred, NetMHC
2) Stretches for class II MHC - Multipred, NetMHC, CTLpred

(i)NetMHC

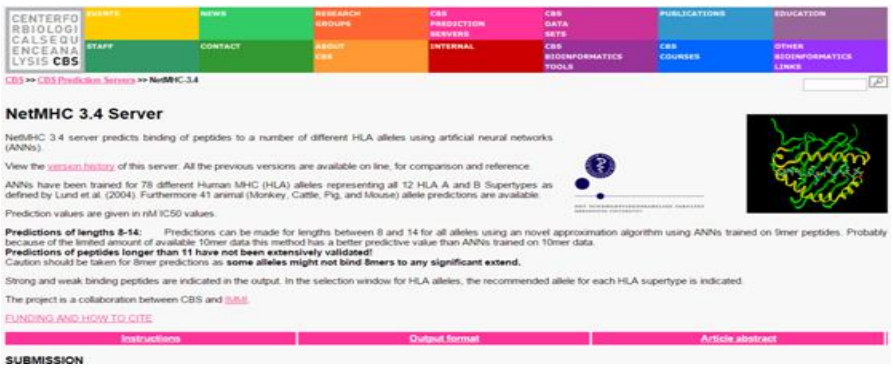
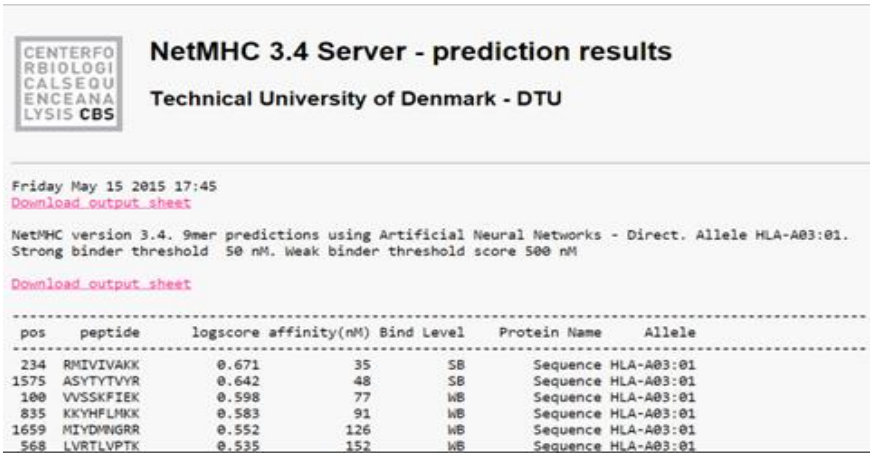


Fig27: Screenshot of NetMHC 3.4 Server



NetMHC 3.4 Server - prediction results

Technical University of Denmark - DTU

Friday May 15 2015 17:45
Download output sheet

NetMHC version 3.4. 9mer predictions using Artificial Neural Networks - Direct. Allele HLA-A03:01.
Strong binder threshold 50 nM. Weak binder threshold score 500 nM

Download output sheet

pos	peptide	logscore	affinity(nM)	Bind Level	Protein Name	Allele
234	RMIVIVAKK	0.671	35	SB	Sequence	HLA-A03:01
1575	ASYTYTVYR	0.642	48	SB	Sequence	HLA-A03:01
100	VVSSKFIEK	0.598	77	WB	Sequence	HLA-A03:01
835	KKYHFLMKK	0.583	91	WB	Sequence	HLA-A03:01
1659	MIYDMNGRR	0.552	126	WB	Sequence	HLA-A03:01
568	LVRTLVPTK	0.535	152	WB	Sequence	HLA-A03:01

Fig28: Predicted peptides and their score by NetMHC

Affinity threshold for Strong binding (SB) peptides = 50.00

Weak binding (WB) peptides = 500.00

Rank threshold for Strong binding (SB) peptides = 0.50

Weak binding (WB) peptides = 2.00

Strong binding peptides are obtained for different HLA molecules:

HLA molecule	No. of High Binders	No. of Weak Binders
A*01:01	9	27
A*02:01	6	26
A*03:01	3	27
A*24:02	3	12
A*26:01	6	23
B*07:02	9	21
B*08:01	2	11
B*27:05	2	16
B*39:01	3	18
B*40:01	5	21
B*58:01	5	24
B*15:01	5	24
Total	58	250

Table6: Strong and weak binders among HLA molecules obtained from NetMHC

According to affinity binding of peptides 58 high binders and 250 weak binders were obtained with the threshold value as 50nM and 500nM respectively. 58 high binder peptides were selected for MHC class I with the affinity range of 35nM to 48nM.

(ii)Multipred

MULTIPRED2 is a computational system used for screening of peptide binding to multiple alleles which belongs Class I and Class II of Human leucocyte antigen (HLA) supertypes also to alleles belonging to ones genotype.

Class I HLA supertypes: A1, A2, A3, A24, A26, B7, B8, B27, B44, B58, B62, C1, C4

Class II HLA supertypes: DR1, DR3, DR4, DR6, DR7, DR8, DR9, DR11, DR12, DR13, DR14, DR15, DR16

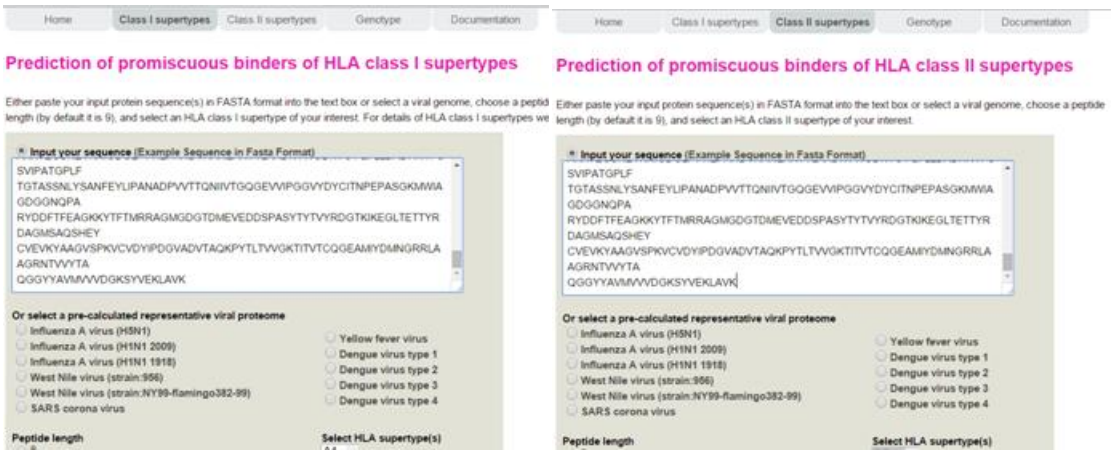


Fig29: Input sequence data for predicting HLA class I and II binders by Multipred2

Prediction was performed on 64 alleles of A2 supertype.
Prediction result for sequence "GI_188595513_DB" for A2

Position	Peptide	Percentage*	IC50 (nM) A*02:01	A*02:02	A*02:03	A*02:04	A*02:05	A*02:06	A*02:07	A*02:09	A*02:11	A*02:12	A*02
4-14	FVSIALCSSLL	75.00%	245.70	24.55	58.36	2307.17	50.38	103.10	7548.17	245.70	96.06	186.11	476.
8-18	ALCSSLLGGMA	7.81%	1344.14	745.70	135.31	9394.82	3234.20	1880.60	21663.93	1344.14	253.98	843.61	479.
10-20	CSSLLGGMAFA	4.69%	9248.84	2436.30	2266.91	17506.67	985.78	1901.35	31140.25	9248.84	3888.97	9168.70	8865
13-23	LLGGMAFAQQT	6.25%	1393.62	646.61	979.15	8386.33	4355.53	2420.77	22634.72	1393.62	188.09	945.48	2172
21-31	QQTELGRNPNV	14.06%	4595.61	2744.51	3829.48	20501.33	1150.39	302.23	33667.98	4595.61	651.14	4432.47	7154
32-42	PLLESTQQSVT	4.69%	1085.59	1050.18	791.15	7093.96	5181.24	1822.86	24029.75	1085.59	76.84	612.81	823.
46-56	TFMDNLKFTEV	26.56%	1103.19	459.02	196.62	11810.08	480.87	223.49	21648.93	1103.19	120.38	827.32	751.
55-65	EVQTPKGMAQV	6.25%	22603.07	10599.78	6520.88	32613.92	6637.77	9129.78	39692.81	22603.07	7700.83	19638.50	1610
62-72	MAQVPTYTEGV	31.25%	2268.97	679.98	1125.41	8388.32	288.30	208.95	25549.78	2268.97	284.96	2387.03	5413
81-91	TLPILSRSLAV	68.75%	397.15	116.47	104.62	3553.65	420.99	324.84	4819.70	397.15	50.90	198.20	321.
84-94	ILSRSLAVSDT	6.69%	3624.62	629.45	246.84	15005.49	4387.41	6351.06	28759.51	3624.62	737.39	1676.28	1105
120-130	MIMRNEDPKKI	4.69%	1841.94	728.39	573.06	8171.53	1817.86	1180.45	22246.61	1841.94	230.50	1111.05	1984
139-149	SQNKFFPGEIA	12.50%	4419.60	2018.84	1172.16	13564.64	554.92	336.96	29506.40	4419.60	1414.97	4826.63	3415
147-157	EIATLDDPFIL	6.25%	10111.22	1241.59	7010.53	24420.56	1108.16	3868.71	34056.48	10111.22	4357.76	9083.55	1736
150-160	TLDDPFILRDV	73.44%	169.09	233.04	390.86	2970.34	1364.18	261.83	8388.51	169.09	23.05	93.86	351.
155-165	FILRDVRGQVV	71.88%	207.92	296.07	68.30	2509.39	819.72	105.10	10976.86	207.92	28.66	174.79	187.
165-175	VNFAPLQYNPV	1.56%	7183.15	6444.41	3356.30	11545.01	2975.27	1053.17	29392.93	7183.15	1197.12	9033.97	7281
177-187	KTLRIYTEITV	64.06%	274.82	855.43	896.66	1079.80	746.11	55.31	14126.76	274.82	20.11	494.16	1264
178-188	TLRIYTEITVA	1.56%	6040.69	1140.23	220.26	15612.91	3905.83	6097.07	30211.03	6040.69	1655.03	5234.18	2307

```
380       390       400       410       420       430       440       450       460       470       480       490       500       510       520       530       540       550       560
|         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
EGGPSADNGESDIQHENVIANLLTQYGYTKIIKCYDPGVTPKNIIDAFNGGISLVNYTGHGSETAWGTSHFGTTHVKQLTNSNQLPFIFDVACVNGDFLFSMPCFAEALMRAQKDGKPTGTVAIIASTINQSNASPHRGQDEMNEILCEKHPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVF
A2                                  IIDAFNGGISL A2                    QLTNSNQLPFI A2      FLFSMPCFAEA A2
                                                                      QLPFIFDVACV A2
```

Fig30: Predicted peptides are shown in yellow on the basis of their IC50 value. Predicted weak binders (IC50<500nM) are shown in green and strong binders (IC50< 50nM) are shown in pink.

CLASS I	No. of peptides	CLASS II	No. of peptides
A2	29	DR1	5
B58	12	DR3	1
C1	4	DR6	6
		DR8	7
		DR16	1

Table7: Number of peptides obtained for MHC class I and II

(iii)CTLpred

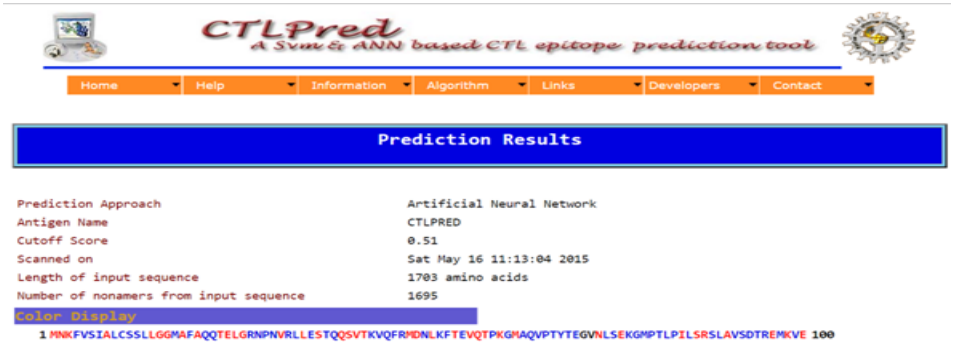


Fig31: Screenshot of CTLPred epitope prediction tool

1	170	LQYNPVTKT	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
2	667	LTATTQGQK	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
3	864	TYTVYRDGT	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
4	998	GYNSNGCVY	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
5	1354	VEVKYTAGV	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
6	1456	HNTFGSVIP	1.000	Epitope	Matrices: New matrices (nHLAPred) Run
7	229	EEKQNGRMI	0.990	Epitope	Matrices: New matrices (nHLAPred) Run
8	294	LTYVLLVGD	0.990	Epitope	Matrices: New matrices (nHLAPred) Run
9	309	KITPGIKSD	0.990	Epitope	Matrices: New matrices (nHLAPred) Run
10	388	IQHENVIAN	0.990	Epitope	Matrices: New matrices (nHLAPred) Run

Fig32: Predicted peptides and their threshold value obtained from CTLPred

Prediction Approach	No. of peptides
Artificial Neural Network	10
Quantitative Matrices	8
Support Vector Matrices	10
Consensus (ANN+SVM)	11
Total	39

Table8: Number of peptides obtained from different T-cell epitope prediction approaches from CTLPred

(iv)PrediVac

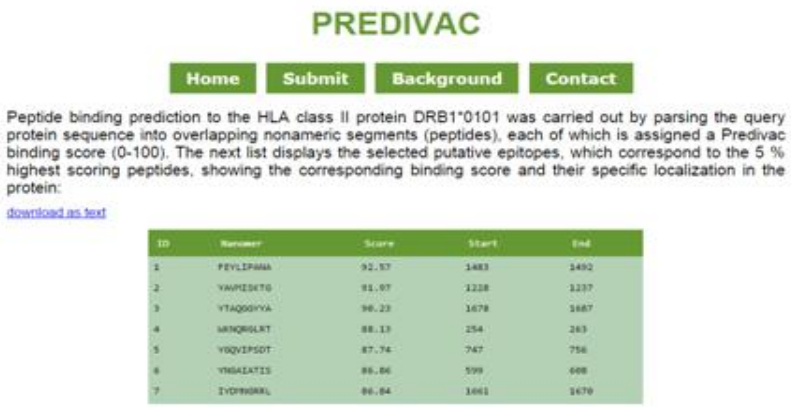


Fig33: Screenshot of PrediVac

55 peptides were predicted to bind to HLA class II (DRB1*0101) which have binding score above 78.

T cell epitopes obtained from all the tools mentioned above when analysed with the predicted B cell epitopes resulted into T cell epitopes which are lying in B cell predicted peptides as:

1	SNLYSANFEYLIPANADPVVITQNIIVTGQGEVVIPGGVYDYCITN
2	WDAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCV
3	APLQYNPVTKTL
4	TGTASSNLYSANFEYLIP
5	AVGQKVTLKWDAPNGT
6	LDDPFILRDVRGQVVNFAPL
7	KFFPGEIATLDDPFILRDVR
8	HEYCVEVKYTAGVSPKECVN
9	TYKRMFMNYEPGRYTP
10	SPASYTYTVYRDGTKI
11	NLSEKGMPTLPILSRSLAVS
12	DPSLLVRTLVPTKMQVTAPA
13	LKTQIDRTIHYERNITTE
14	YLIPANADPVVTTQNIIV

Table9: T-cell epitopes which are lying in B cell predicted peptides

5.10 Predicting vaccine stretches according to geographical area – PrediVac population coverage prediction

Predicting vaccine stretches according to geographical area – PrediVac

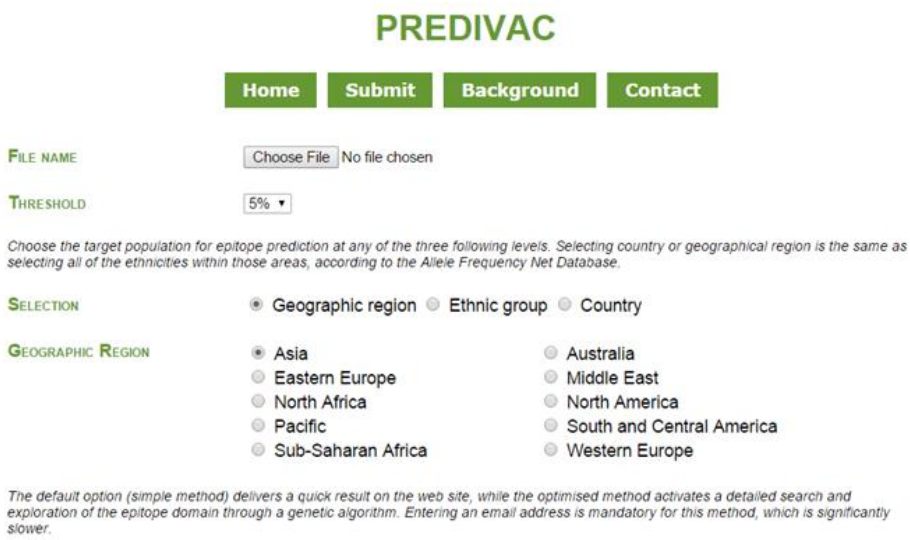


Fig34: Screenshot of Predivac-population coverage prediction tool

Population coverage prediction

Geographical region	Number of peptides
Asia	8
Eastern Europe	4
North Africa	5
Pacific	7
Australia	5
Sub Saharan Africa	6
South and Central America	7
Western Europe	4
Middle east	3
North America	7
Total	56

Table10: Number of peptides obtained in different geographical regions

5.11 Prediction of TAP binding regions by TAPPred

Fig35: Screenshot of TAPPred prediction tool

Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	889	ATGNHEYCV	9.312	High
2	1346	ATGNHEYCV	9.312	High
3	124	NEDPKKIPY	8.490	High
4	1670	AAGRNTVVY	8.423	High
5	1289	VAFRHYNCS	8.307	High
6	212	TYKRMFMNY	8.242	High
7	347	TQIDRTIHY	8.044	High
8	1551	KKYTFTMRR	7.935	High
9	1440	NVWGDNTGY	7.819	High
10	1576	ASYTYTVYR	7.783	High
11	1286	TKYVAFRHY	7.493	High
12	393	VIANLLTQY	7.484	High
13	1258	ARFGLSTEA	7.411	High
14	402	GYTKIIKCY	7.372	High
15	1604	SAQSHEYCV	7.312	High
16	1048	YASEHYAVY	7.182	High
17	467	ACVNGDFLF	7.038	High
18	481	AEALMRAQK	6.913	High
19	1375	TQFNPVKNL	6.840	High

Fig36: Predicted sequence and their score obtained from TAPPred tool

Total 60 peptides are obtained that are having more than six score.

The highlighted stretches are those which contain the TAP binding regions predicted from TAPPred and are present in T cell predicted peptides.

1	SNLYSANFEYLIPANADPVVITQNIIVTGQGEVVIPGGVYDYCITN
2	WDAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCV
3	APLQYNPVTKTL
4	TGTASSNLYSANFEYLIP
5	AVGQKVTLKWDAPNGT
6	LDDPFILRDVRGQVVNFAPL
7	KFFPGEIATLDDPFILRDVR
8	HEYCVEVKYTAGVSPKECVN
9	TYKRMFMNYEPGRYTP
10	SPASYTYTVYRDGTKI
11	NLSEKGMPTLPILSRSLAVS
12	DPSLLVRTLVPTKMQVTAPA
13	LKTQIDRTIHYERNITTE
14	YLIPANADPVVTTQNIIV

Table11: Highlighted regions are shown as peptides containing TAP sites in T cell predicted peptides

5.12 Prediction of proteosomal cleavage sites by PAProc

PAProC is a prediction tool which is used to predict cleavages by human as well as yeast proteosomes, and this is based on an experimental cleavage data.

781	DPSC | S | PTNMI | MDGT | A | SV | NIP
801	AG | TYDF | AI | AA | P | QA | NA | KI | WIA |
821	GQG | PTKED | DYVFE | A | GK | KY | HF
841	L | M | K | KMGSG | DGTEL | T | ISEG | GG
861	S | DYTYT | VY | RDGTKI | K | EG | L | TA
881	TTFE | EDGV | ATGNHEYCV | EVK
901	YT | AGV | S | PKVC | KDVTVEGSNE
921	F | APVQN | L | TGSAVGQKVTLK | W |
941	DAPNG | T | PNPNPNPNPNPG
961	TTT | L | SE | S | F | E | NGIPAS | W | KT | ID
981	A | DGDGH | GW | KP | GNA | PGI | AG | Y | N |
1001	SNG | CVYSESF | G | L | GG | IGVL | TP
1021	DNY | LITPALDL | PNGGKL | TFW |
1041	VC | AQD | ANYA | SEH | YA | VYA | S | ST |
1061	GND | ASNFTNA | LLE | E | TITA | K | G
1081	V | RS | P | EAIR | G | R | IQGTWRQ | KTV
1101	DL | PAGTKY | V | AF | R | HF | Q | ST | DMF |
1121	YI | DLD | EV | EIKA | NGKRA | DFTE |
1141	T | F | E | S | STHGE | APAE | W | TT | I | D | AD
1161	GDGQ | GWLCL | SSGQL | DW | LTAH

Fig37: Screenshot of PAProc and predicted cleavage sites

5.13 Visualisation of predicted epitopes

MNKFVSIALCSSLLGGMAFAQQTELGRNPNVRLLESTQQSVTKVQFRMDNLKFTEVQTPKGMAQVPTYTEGVNLSEKGMPTLPILSRSLAVSDT
REMKVEVVSSKFIEKKNVLIAPSKGMIMRNEDPKKIPYVYGKSYSQNKFFPGEIATLDDPFILRDVRGQVVNFAPLQYNPVTKTLRIYTEITVA
VSETSEQGKNILNKKGTFAGFEDTYKRMFMNYEPGRYTPVEEKQNGRMIVIVAKKYEGDIKDFVDWKNQRGLRTEVKVAEDIASPVTANAIQQF
VKQEYEKEGNDLTYVLLVGDHKDIPAKITPGIKSDQVYGQIVGNDHYNEVFIGRFSCESKEDLKTQIDRTIHYERNITTEDKWLGQALCIASAE
GGPSADNGESDIQHENVIANLLTQYGYTKIIKCYDPGVTPKNIIDAFNGGISLVNYTGHGSETAWGTSHFGTTHVKQLTNSNQLPFIFDVACVN
GDFLFSMPCFAEALMRAQKDGKPTGTVAIIASTINQSWASPMRGQDEMNEILCEKHPNNIKRTFGGVTMNGMFAMVEKYKKDGEKMLDTWTVFG
DPSLLVRTLVPTKMQVTAPAQINLTDASVNVSCDYNGAIATISANGKMFGSAVVENGTATINLTGLTNESTLTLTVVGYNKETVIKTINTNGEP
NPYQPVSNLTATTQGQKVTLKWDAPSTKTNATTNTARSVDGIRELVLLSVSDAPELLRSGQAEIVLEAHDVWNDGSGYQILLDADHDQYGQVIP
SDTHTLWPNCSVPANLFAPFEYTVPENADPSCSPTNMIMDGTASVNIPAGTYDFAIAAPQANAKIWIAGQGPTKEDDYVFEAGKKYHFLMKKMG
SGDGTELTISEGGGSDYTYTVYRDGTKIKEGLTATTFEEDGVATGNHEYCVEVKYTAGVSPKVCKDVTVEGSNEFAPVQNLTGSAVGQKVTLKW
DAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCVYSESFGLGGIGVLTPDNYLITPALDLPNG
GKLTFWVCAQDANYASEHYAVYASSTGNDASNFTNALLEETITAKGVRSPEAIRGRIQGTWRQKTVDLPAGTKYVAFRHFQSTDMFYIDLDEVE
IKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQGWLCLSSGQLDWLTAHGGTNVVASFSWNGMALNPDNYLISKDVTGATKVKYYYAVNDG
FPGDHYAVMISKTGTNAGDFTVVFEETPNGINKGGARFGLSTEANGAKPQSVWIERTVDLPAGTKYVAFRHYNCSDLNYILLDDIQFTMGGSPT
PTDYTYTVYRDGTKIKEGLTETTFEEDGVATGNHEYCVEVKYTAGVSPKECVNVTINPTQFNPVKNLKAQPDGGDVVLKWEAPSAKKTEGSREV
KRIGDGLFVTIEPANDVRANEAKVVLAADNVWGDNTGYQFLLDADHNTFGSVIPATGPLFTGTASSNLYSANFEYLIPANADPVVTTQNIIVTG
QGEVVIPGGVYDYCITNPEPASGKMWIAGDGGNQPARYDDFTFEAGKKYTFTMRRAGMGDGTDMEVEDDSPASYTYTVYRDGTKIKEGLTETTY
RDAGMSAQSHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQGEAMIYDMNGRRLAAGRNTVVYTAQGGYYAVMVVVD
GKSYVEKLAVK

Fig38: Predicted epitopes : the two yellow coloured are finally predicted epitopes and the red one is rejected as the proteosomal cleavage site was lying in the stretch

Fig39: Input sequence as arginin gingipains in Phyre2

The 3D protein structure is obtained using PHYRE2 and the two predicted peptide regions are shown on them in PYMOL visualisation tool. The structure with 100 percent and 99.4 percent confidence and 79 percent i.d were obtained.

APLQYNPVTKTLRIYT

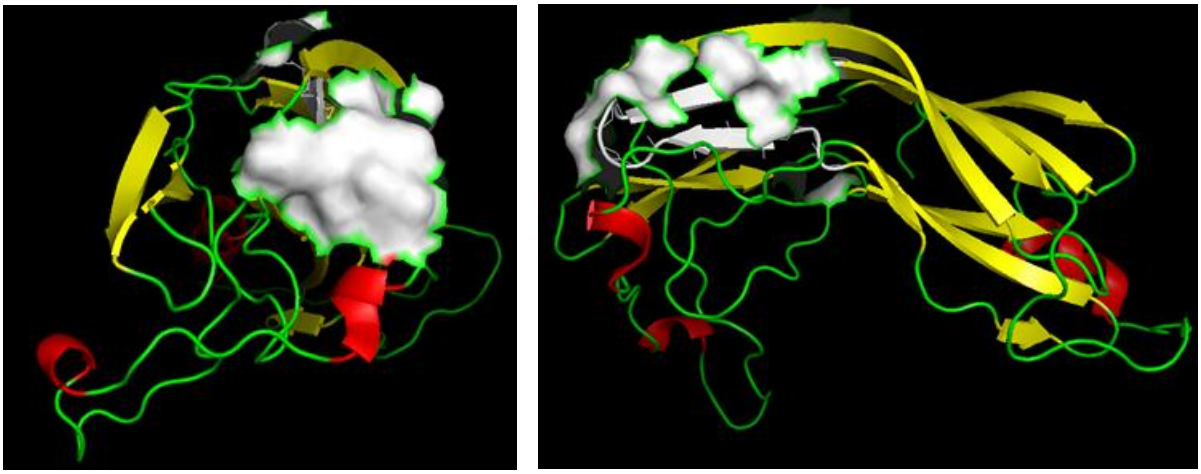


Fig40: APLQYNPVTKTLRIYT (white in color) site on arginin gingipain protein

SNLYSANFEYLIPANA

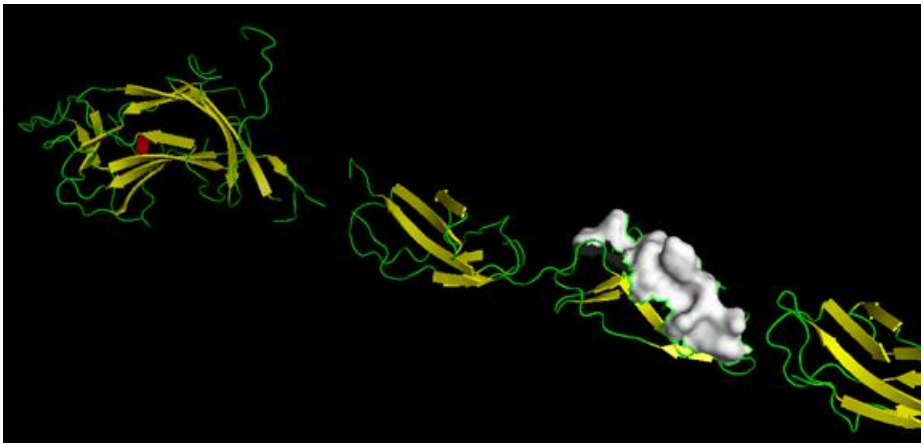


Fig41: SNLYSANFEYLIPANA(white in color) site on arginin gingipain protein

6. CONCLUSION

The given work, directed towards the identification of vaccine candidates for Atherosclerosis caused by the factors of Periodontitis and this has resulted in the prediction of some acceptable epitopes that can be used to elicit immune response against both Atherosclerosis and Periodontitis. WDAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCV, LDDPFILRDVRGQVVNFAPLQYNPVTKTLRIYTEITVAVSET and SNLYSANFEYLIPANADPVVITQNIIVTGQGEVVIPGGVYDYCITN were recognised as B-cell epitopes. These epitope stretches contain MHC class-I binding peptides as NAPGIAGY, LQYNPVT and SNLYS respectively as highlighted above in yellow. Also MHC class-II binding peptides were obtained as IPASWTI, VTKTLRIYT and FEYLIPANA respectively as highlighted above in blue. These peptide stretches bind to most of the MHC class I and II alleles with high affinity and these are antigenic. After finding the TAP binding regions, proteosomal cleavage sites and peptide stretches predicted on the basis of geographical locations, three peptides were predicted, but the proteosomal cleavage sites were lying in IPASWKT|IDADGDGHGW|KPGNAPGIAGY, therefore it is rejected and hence, two probable vaccine candidates were predicted as APLQYNPVTKTLRIYT and SNLYSANFEYLIPANA. So, these are T-cell epitopes derived from B-cell epitopes that may act as vaccine against Periodontitis leading to Atherosclerosis and inhibit arginine gingipains. These insilico results need to be confirmed by experimental validation to develop a good vaccine candidate that can be beneficial for therapeutic purpose.

7. DISCUSSION AND FUTURE PERSPECTIVE

Atherosclerosis is one of the biggest killers of the twenty-first century among CAD. Atherosclerosis is a multi-factorial and a multistep disease which is involved in chronic inflammation at each step, from initiation to the progression, and all the risk factors together contribute to pathogenesis.

The work will help in the future to treat the disease like atherosclerosis which is highly connected with the oral hygiene of the person i.e, the poor oral hygiene lead to periodontitis and further causes the disease-atherosclerosis. For the clinical advantages the importance of inflammation in atherosclerosis gives the outlook to understand the creation of strategies for developing new therapeutics, and also the much better understanding of pathogenesis may help in preparing preventive and therapeutic strategies which can reduce the mortality resulting from CAD.

More deep knowledge of the various pathogenic mechanisms that are involved in atherosclerosis can help in understanding the present knowledge exists regarding CAD epidemic. Its efficient knowledge will surely help clinicians and pharmacists to control and manage the disease in a better way, and Indians are also severely affected by it.

Sub-gingival epithelium region of periodontal disease has shown the presence of Porphyromonas gingivalis. P.gingivalis entering upon tissues adheres to the epithelial cells and it is the first step in colonization and it is important for the development of a disease. The non-covalent interactions of ligands from infecting organism with receptor targets on the surface of host cells through attachment of it to the host cell membrane is mediated. Some of the eukaryotic cell surface receptors which are identified in P. gingivalis–host interactions have extracellular matrix proteins and integrins. Transglutaminase2, the adhesion molecule which is present on host cells form the close association with the arginine gingipains-fibronectin complex and showed binding regions. B cell epitope predictions for Arginine gingipains A and Arginine gingipains B has shown many acceptable peptide stretches.

The advancement in the techniques and the tools available for analysing the sequence data have enabled prediction of probable vaccine candidates from the large number of proteins synthesized by an organism. Attempts are underway to create a vaccine that could prevent Atherosclerosis caused by Periodontitis in humans.

The experimental data has shown that some factors of Periodontitis are responsible for causing Atherosclerosis. Among all the microorganisms studied Porphyromonas gingivalis and its strains are highly responsible for inflammatory action and causing the disease. The phylogenetic analysis among their strains has shown the similarity among themselves. Porphyromonas gingivalis is an important and main causative agent of atherosclerosis specifically ATCC33277 strain of P.gingivalis. All the four strains that are ATCC33277, W83, W50, and TDC60 of P.gingivalis are more than ninety eight percent identical and covering ninety five percent of query length. Arginine gingipains present in Porphyromonas gingivalis form a complex with fibronectin and interacts with the Transglutaminase2 that is present in host cells. Here the two types of Arginine gingivalis (RgpA and RgpB) are the

adhesion molecules and mediates the entry into the host cell and reach to the blood stream which in developmental process causing atherosclerosis.

The B cell epitope predictions of Arginine gingipains A and Arginine gingipains B has shown many acceptable stretches of peptide. Using bioinformatics tools like Bepipred, DiscoTope, ABCpred, IEDB Analysis epitope prediction for Antigenicity, Flexibility, Emini surface accessibility: many peptide stretches were obtained with the length size varying from 3mer to 50mer. Those peptide stretches were obtained which has shown high score for antigenicity, flexibility and surface accessibility. T-cell epitopes for HLA classI and classII were predicted using tools like: Multipred, NetMHC, CTLpred.

In our study we make sure that the selected peptide must contain TAP recognising regions and proteosomal cleavage sites. As in the processing and presentation of peptide, the antigen is processed through proteasome and cleaves proteins into peptides of 15 to 22 amino acid residues and transported through TAP into endoplasmic reticulum to bind to MHC. Hence it is essential to predict proteosomal and Tap recognising regions. TAP processing regions were predicted by TAPpred and Proteosomal cleavage sites were predicted by PAPROC. The T-cell epitopes which present in B-cell epitopes were obtained and 20 such peptide stretches were obtained. With the help of PrediVac tool for population coverage the peptide stretches which can act as possible candidate in different geographical locations: Asia, Eastern Europe, North Africa, Pacific, Sub-Saharan Africa, Australia, Middle East, North America, South and Central America, Western Europe were obtained. The total number of stretches sorted among B-cell epitopes for predicted T-cell epitopes were seven. Out of these seven which are showing their presence in different geographical locations are three. The B-cell epitopes are: WDAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGI AGYNSNGCV, LDDPFILRDVRGQVVNFAPLQYNPVTKTLRIYTEITVAVSET SNLYSANFEYLIPANADPVVITQNIIVTGQGEVVIPGGVYDYCITN. These have overlapping regions as shown:

IPASWKTIDADGDGHGWKPGNAPGIAGY- NAPGIAGY classI, IPASWTI class II, ASWKTIDAD TAP binding regions

APLQYNPVTKTLRIYT- LQYNPVT classI, VTKTLRIYT classII, LQYNPVT TAP binding regions

SNLYSANFEYLIPANA- SNLYS classI, FEYLIPANA classII, ANFEYLIPA TAP binding regions

Different MHC have different propensity to diseases. In our study we have tested peptide binding with different MHC haplotypes that are prevalent in different geographical locations so that vaccine candidate proposed has universal immunogenicity in different geographical populations.

IPASWKTIDADGDGHGWKPGNAPGIAGY is prevalent in North African, Pacific and Sub-Saharan regions where as SNLYSANFEYLIPANA is prevalent in Asian geographical locations. But the IPASWKTIDADGDGHGWKPGNAPGIAGY contains the proteosomal

cleavage sites inside it so not considered. So the later two peptides were predicted as the probable vaccine candidates.

The resultant candidate can act against both atherosclerosis as well as against periodontal disease. Further for the future study other microorganisms causing atherosclerosis can be studied and conserved regions can be find out and a common candidate can be form out of them and which can act as a vaccine candidate for them also.

Additionally, this peptide has been selected in a way that it has the potential to evoke both branches of adaptive immunity.

Such broad range vaccines, if able to clear clinical trials, will be a tool to prevent infectious diseases which have more than one causative agent. The development of such broad range vaccine needs a lot more to be done in the field of immunoinformatics, a field which is in its infantile stage. The successful development of accurate antigen prediction methods will surely cut down the laboratory resources which are required to predict the pathogenic protein that can act as a subunit vaccine.

8. ANNEXURE

BEPIPRED
VQTPKGMAQVPTYTEGVNLSEKGMP
EPGRYTPVEEKQN
KDIPAKITPGIKSDQV
AEGGPSADNGESDIQ
GHGSETAWGTSHF
NQSWASPMRGQDE
NTNGEPNPYQPVSNLTATTQGQK
WDAPSTKTNATTNTARSVD
HDQYGQVIPSDTHTL
YTVPENADPSCSPT
GQGPTKEDDYVF
GSGDGTELTISEGGGSDYTYT
ATTFEEDGVATGN
TVEGSNEFAPV
WDAPNGTPNPNPNPNPNPNPGTTTLSESFENGIPASWKTIDADGDGHGWKPGNAPGIAGYNSNGCV
TAKGVRSPEAIRG
GTWRQKTVDLPAG
IKANGKRADFTETFESSTHGEAPAEWTTIDADGDGQ
EETPNGINKGGAR
STEANGAKPQSV
TMGGSPTPTDYTYT
YRDGTKIKEGLTETTFEEDGVATGN
NLKAQPDGGDV
EAPSAKKTEGSREVKR
VIPATGPLFTGTASSNL
WIAGDGGNQPARYDDFTFE
AGMGDGTDMEVEDDSPASYT
YRDGTKIKEGLTETTYRDAGMSAQ
PDGVADVTAQKP

DISCOTOPE

PAS
DADGDGN
PPGGS
NFE
VTAPEAR
FE
SEA
TAPEA
NGIPASWLTIDADGDGN

ABCpred

20 mer	18mer	16mer	14mer
VKYYYAVNDGFPGDHYAVMI	HYAVMISKTGTNAGDFTV	AEDIASPVTANAIQQF	TFEEDGVATGNHEY
PGNAPGIAGYNSNGCVYSES	KQEYEKEGNDLTYVLLVG	KVCVDYIPDGVADVTA	GARFGLSTEANGAK
LDDPFILRDVRGQVVNFAPL	SDQVYGQIVGNDHYNEVF	AGMGDGTDMEVEDDSP	VYRDGTKIKEGLTA
TEVKVAEDIASPVTANAIQQ	YAVMVVVDGKSYVEKLAV	DVTVEGSNEFAPVQNL	INTNGEPNPYQPVS
LDWLTAHGGTNVVASFSWNG	ATGNHEYCVEVKYTAGVS	IKTINTNGEPNPYQPV	GAIATISANGKMFG
VGQKVTLKWDAPNGTPNPNP	KFTEVQTPKGMAQVPTYT	TKIIKCYDPGVTPKNI	IKDFVDWKNQRGLR
QHENVIANLLTQYGYTKIIK	VVLKWEAPSAKKTEGSRE	HKDIPAKITPGIKSDQ	RGQVVNFAPLQYNP
CIASAEGGPSADNGESDIQH	SQNKFFPGEIATLDDPFI	GEVVIPGGVYDYCITN	TKYVAFRHFQSTDM
HEYCVEVKYTAGVSPKECVN	GLTETTYRDAGMSAQSHE	FVTIEPANDVRANEAK	PQSVWIERTVDLPA
QGWLCLSSGQLDWLTAHGGT	KNIIDAFNGGISLVNYTG	NGTPNPNPNPNPNPNP	AKGVRSPEAIRGRI
CAQDANYASEHYAVYASSTG	KVCVDYIPDGVADVTAQK	ASPMRGQDEMNEILCE	YSESFGLGGIGVLT
ETVIKTINTNGEPNPYQPVS	FILRDVRGQVVNFAPLQY	HGSETAWGTSHFGTTH	DVACVNGDFLFSMP
NVVASFSWNGMALNPDNYLI	VVIPGGVYDYCITNPEPA	KEGLTETTYRDAGMSA	SSTHGEAPAEWTTI
RHFQSTDMFYIDLDEVEIKA	ADNVWGDNTGYQFLLDAD	TLKWDAPNGTPNPNPN	DLPAGTKYVAFRHF
YAVYASSTGNDASNFTNALL	INKGGARFGLSTEANGAK	GGSDYTYTVYRDGTKI	LRSGQAEIVLEAHD
KYEGDIKDFVDWKNQRGLRT	ATKVKYYYAVNDGFPGDH	PFEYTVPENADPSCSP	WGTSHFGTTHVKQL
VYRDGTKIKEGLTETTFEED	PASWKTIDADGDGHGWKP	KGMPTLPILSRSLAVS	HYERNITTEDKWLG
NAIQQFVKQEYEKEGNDLTY	NPNPNPNPGTTTLSESFE	PVSNLTATTQGQKVTL	YLISKDVTGATKVK
SCESKEDLKTQIDRTIHYER	MNGMFAMVEKYKKDGEKM	QVVNFAPLQYNPVTKT	VDLPAGTKYVAFRH
DDFTFEAGKKYTFTMRRAGM	TNSNQLPFIFDVACVNGD	PASGKMWIAGDGGNQP	LCLSSGQLDWLTAH
EDGVATGNHEYCVEVKYTAG	DKWLGQALCIASAEGGPS	PGEIATLDDPFILRDV	LPFIFDVACVNGDF
KIWIAGQGPTKEDDYVFEAG	NPEPASGKMWIAGDGGNQ	TPTDYTYTVYRDGTKI	PSADNGESDIQHEN
VNVSCDYNGAIATISANGKM	YLIPANADPVVTTQNIIV	WTTIDADGDGQGWLCL	DKWLGQALCIASAE
GISLVNYTGHGSETAWGTSH	RFGLSTEANGAKPQSVWI	SWKTIDADGDGHGWKP	VFIGRFSCESKEDL
TQYGYTKIIKCYDPGVTPKN	KPGNAPGIAGYNSNGCVY	GPTKEDDYVFEAGKKY	PKECVNVTINPTQF
GDHKDIPAKITPGIKSDQVY	GESDIQHENVIANLLTQY	TVAIIASTINQSWASP	PNPNPNPGTTTLSE
NEAKVVLAADNVWGDNTGYQ	QNIIVTGQGEVVIPGGVY	TELGRNPNVRLLESTQ	EGGGSDYTYTVYRD
DGFPGDHYAVMISKTGTNAG	GKRADFTETFESSTHGEA	YTVYRDGTKIKEGLTE	KMGSGDGTELTISE
GMALNPDNYLISKDVTGATK	DVTVEGSNEFAPVQNLTG	EVKYTAGVSPKECVNV	EAGKKYHFLMKKMG
SNFTNALLEETITAKGVRSP	TGNHEYCVEVKYTAGVSP	YTVYRDGTKIKEGLTE	VPENADPSCSPTNM
WDAPNGTPNPNPNPNPNPNP	DGTKIKEGLTATTFEEDG	TFESSTHGEAPAEWTT	ILLDADHDQYGQVI
NIIDAFNGGISLVNYTGHGS	LWPNCSVPANLFAPFEYT	YTVYRDGTKIKEGLTA	ASPMRGQDEMNEIL
VNFAPLQYNPVTKTLRIYTE	HDQYGQVIPSDTHTLWPN	RGRIQGTWRQKTVDLP	MAFAQQTELGRNPN
AQKPYTLTVVGKTITVTCQG	TLTLTVVGYNKETVIKTI	EETITAKGVRSPEAIR	VVLAADNVWGDNTG
TNESTLTLTVVGYNKETVIK	LLTQYGYTKIIKCYDPGV	NHEYCVEVKYTAGVSP	TYTVYRDGTKIKEG
TAPAQINLTDASVNVSCDYN	TPVEEKQNGRMIVIVAKK	ELTISEGGGSDYTYTV	LCSSLLGGMAFAQQ
DPSLLVRTLVPTKMQVTAPA	DYTYTVYRDGTKIKEGLT	TLKWDAPSTKTNATTN	AEIVLEAHDVWNDG
VDWKNQRGLRTEVKVAEDIA	LKTQIDRTIHYERNITTE	VLLVGDHKDIPAKITP	TQYGYTKIIKCYDP
VVGKTITVTCQGEAMIYDMN	DIPAKITPGIKSDQVYGQ	YEPGRYTPVEEKQNGR	IPAKITPGIKSDQV
DGTKIKEGLTETTYRDAGMS	MGDGTDMEVEDDSPASYT	EAMIYDMNGRRLAAGR	KVAEDIASPVTANA
YAVMISKTGTNAGDFTVVFE	PTQFNPVKNLKAQPDGGD	VEVKYAAGVSPKVCVD	NTGYQFLLDADHNT
NPNPNPGTTTLSESFENGIP	WIERTVDLPAGTKYVAFR	NHEYCVEVKYTAGVSP	GVSPKVCKDVTVEG
YCVEVKYTAGVSPKVCKDVT	DLPAGTKYVAFRHFQSTD	LPAGTKYVAFRHYNCS	LWPNCSVPANLFAP
WPNCSVPANLFAPFEYTVPE	SNEFAPVQNLTGSAVGQK	KGMIMRNEDPKKIPYV	SVDGIRELVLLSVS
TNGEPNPYQPVSNLTATTQG	NTVVYTAQGGYYAVMVVV	YDFAIAAPQANAKIWI	STLTLTVVGYNKET
SAVVENGTATINLTGLTNES	TGTASSNLYSANFEYLIP	SVNIPAGTYDFAIAAP	ETSEQGKNILNKKG
GRMIVIVAKKYEGDIKDFVD	ASSTGNDASNFTNALLEE	HDVWNDGSGYQILLDA	GVATGNHEYCVEVK

VCVDYIPDGVADVTAQKPYT	QGPTKEDDYVFEAGKKYH	PSTKTNATTNTARSVD	NDGFPGDHYAVMIS
TQNIIVTGQGEVVIPGGVYD	LVNYTGHGSETAWGTSHF	GMAQVPTYTEGVNLSE	TWRQKTVDLPAGTK
GSREVKRIGDGLFVTIEPAN	KEGLTETTFEEDGVATGN	KNIIDAFNGGISLVNY	GQGPTKEDDYVFEA
TAGVSPKECVNVTINPTQFN	PSKGMIMRNEDPKKIPYV	KWLGQALCIASAEGGP	EPANDVRANEAKVV
LDEVEIKANGKRADFTETFE	VRSPEAIRGRIQGTWRQK	GGMAFAQQTELGRNPN	NLKAQPDGGDVVLK
LMKKMGSGDGTELTISEGGG	AIAAPQANAKIWIAGQGP	GFPGDHYAVMISKTGT	TKIKEGLTETTFEE
TYDFAIAAPQANAKIWIAGQ	KTINTNGEPNPYQPVSNL	KNVLIAPSKGMIMRNE	HYAVMISKTGTNAG
NLSEKGMPTLPILSRSLAVS	EVEDDSPASYTYTVYRDG	TNMIMDGTASVNIPAG	HYAVYASSTGNDAS
ATTQGQKVTLKWDAPSTKTN	KKIPYVYGKSYSQNKFFP	TQGQKVTLKWDAPSTK	GDGHGWKPGNAPGI
VPTYTEGVNLSEKGMPTLPI	SKDVTGATKVKYYYAVND	WGTSHFGTTHVKQLTN	DYTYTVYRDGTKIK
PGRYTPVEEKQNGRMIVIVA	DWLTAHGGTNVVASFSWN	MIVIVAKKYEGDIKDF	TASVNIPAGTYDFA
SYTYTVYRDGTKIKEGLTET	RHFQSTDMFYIDLDEVEI	YRDAGMSAQSHEYCVE	QIDRTIHYERNITT
TFTMRRAGMGDGTDMEVEDD		SPASYTYTVYRDGTKI	RYTPVEEKQNGRMI
PEPASGKMWIAGDGGNQPAR		DYCITNPEPASGKMWI	INPTQFNPVKNLKA
KKIPYVYGKSYSQNKFFPGE	12mer	AVYASSTGNDASNFTN	HYNCSDLNYILLDD
KGGARFGLSTEANGAKPQSV	KIWIAGQGPTKE	VIPSDTHTLWPNCSVP	STDMFYIDLDEVEI
TPDNYLITPALDLPNGGKLT	QHENVIANLLTQ	VQFRMDNLKFTEVQTP	PGIAGYNSNGCVYS
IVGNDHYNEVFIGRFSCESK	GMAFAQQTELGR	MISKTGTNAGDFTVVF	CSPTNMIMDGTASV
NVRLLESTQQSVTKVQFRMD	QIVGNDHYNEVF	LPAGTKYVAFRHFQST	DAPSTKTNATTNTA
QGKNILNKKGTFAGFEDTYK	KEGNDLTYVLLV	AVGQKVTLKWDAPNGT	KGMAQVPTYTEGVN
KFFPGEIATLDDPFILRDVR	CYDPGVTPKNII	TVVGYNKETVIKTINT	EMNEILCEKHPNNI
GEAPAEWTTIDADGDGQGWL	KRMFMNYEPGRY	ESDIQHENVIANLLTQ	IASAEGGPSADNGE
	APLQYNPVTKTL	YEKEGNDLTYVLLVGD	SDQVYGQIVGNDHY
	VEVVSSKFIEKK	QFVKQEYEKEGNDLTY	GNDLTYVLLVGDHK
	DAPELLRSGQAE	TYKRMFMNYEPGRYTP	LNPDNYLISKDVTG
	PAQINLTDASVN	VAVSETSEQGKNILNK	
	KDGKPTGTVAII	DVTAQKPYTLTVVGKT	
	GRNPNVRLLEST	AGKKYTFTMRRAGMGD	
	REMKVEVVSSKF	PARYDDFTFEAGKKYT	
	TISEGGGSDYTY	DEVEIKANGKRADFTE	
	RELVLLSVSDAP	GVRSPEAIRGRIQGTW	
	MVEKYKKDGEKM	GLGGIGVLTPDNYLIT	
	STQQSVTKVQFR	TNSNQLPFIFDVACVN	
	FVKQEYEKEGND	VNVTINPTQFNPVKNL	
	AQQTELGRNPNV	GGARFGLSTEANGAKP	
	TKVKYYYAVNDG	EDPKKIPYVYGKSYSQ	
	LNPDNYLISKDV	DVTGATKVKYYYAVND	
	STHGEAPAEWTT	SSGQLDWLTAHGGTNV	
	TVEGSNEFAPVQ	PNPGTTTLSESFENGI	
	CSPTNMIMDGTA	EVKYTAGVSPKVCKDV	
	TVPENADPSCSP	TKMQVTAPAQINLTDA	
	LTVVGYNKETVI	IASAEGGPSADNGESD	
	ATINLTGLTNES	EGDIKDFVDWKNQRGL	
	NYTGHGSETAWG	VVLAADNVWGDNTGYQ	
	EDKWLGQALCIA	DGVATGNHEYCVEVKY	
	DRTIHYERNITT	DGEKMLDTWTVFGDPS	
	TGNDASNFTNAL	DGVATGNHEYCVEVKY	
	FWVCAQDANYAS	PQSVWIERTVDLPAGT	

Antigenicity: kolaskar and tongaonkar antigenicity

FVSIALCSSLLGGM
VRLLESTQQSVTKVQ
TLPILSRSLAVS
LDDPFILRDVRGQVVNFAPLQYNPVTKTLRIYTEITVAVSET
RTEVKVAEDIASPVTAN
NEVFIGRFSCE
LPFIFDVACVNGD
LFSMPCFAEA
FGDPSLLVRTLVPTKMQVTAPAQINLTDASVNVSCDYN
VDGIRELVLLSVSDAPELLRS
WPNCSVPANLFAPFEYT
SVNIPAGTYDFAIAAPQ
HEYCVEVKYTAGVSPKVCKDVTVE
GSAVGQKVTLK
GIGVLTPDNYLITPALDL
GTKYVAFRHYNCSDLNYILLDDI
HEYCVEVKYTAGVSPKECVNVT
SNLYSANFEYLIPANADPVVITQNIIVTGQGEVVIPGGVYDYCITN
SHEYCVEVKYAAGVSPKVCVDYIPDGVADVTAQKPYTLTVVGKTITVTCQ

Emini surface Accessibility

NYEPGRYTPVEEKQN
FVKQEYEKEGND
TIHYERNITTEDKW
WDAPSTKTNATTNTAR
SDYTYTVYRDGTKI
PNGTPNPNPNPNPNPNPG
ANGKRADFTETFESSTHG
SPTPTDYTYTVYRDGTKI
PSAKKTEGSREV
MEVEDDSPASYTYTVYRDGTKI

CTLpred

Artificial neural network	Quantitative matrix	Support vector machine	Consensus ANN+SVM
LQYNPVTKT	GRNPNVRLL	GRNPNVRLL	GRNPNVRLL
LTATTQGQK	FRHYNCSDL	NLTDASVNV	STDMFYIDL
TYTVYRDGT	RSVDGIREL	RSVDGIREL	LYSANFEYL
GYNSNGCVY	AVGQKVTLK	STDMFYIDL	KTLRIYTEI
VEVKYTAGV	RYDDFTFEA	RYDDFTFEA	CSDLNYILL
HNTFGSVIP	FRHFQSTDM	RLAAGRNTV	NPNPGTTTL
EEKQNGRMI	GRMIVIVAK	LYSANFEYL	GYNSNGCVY
LTYVLLVGD	FFPGEIATL	SMPCFAEAL	ATKVKYYYA
KITPGIKSD		SFSWNGMAL	TYKRMFMNY
IQHENVIAN		ILRDVRGQV	AVSDTREMK

Predivac : HLA class II (DRB1*0101)

FEYLIPANA	FNPVVKNLKA	YSESFGLGG
YAVMISKTG	LDWLTAHGG	YDFAIAAPQ
YTAQGGYYA	LRDVRGQVV	YTGHGSETA
WKNQRGLRT	MRNEDPKKI	WLTAHGGTN
YGQVIPSDT	FLFSMPCFA	YAVNDGFPG
YNGAIATIS	VTIEPANDV	YIDLDEVI
IYDMNGRRL	YGQIVGNDH	IHYERNITT
WLCLSSGQL	VFEETPNGI	IVTGQGEVV
YEKEGNDLT	VRTLVPTKM	YVFEAGKKY
YAAGVSPKV	YAVMVVVDG	FTGTASSNL
YTAGVSPKV	FAPVQNLTG	VTAQKPYTL
FGVIPATG	LTGLTNEST	MGSGDGTEL
YCITNPEPA	FTEVQTPKG	
WNDGSGYQI	IYTEITVAV	
WKPGNAPGI	VCAQDANYA	
YASSTGNDA	WEAPSAKKT	
MNYEPGRYT	YLIPALDL	
FAIAAPQAN	YVLLVGDHK	
YQPVSNLTA	FPGEIATLD	
VQNLTGSAV	YVLLVGDHK	
FEYTVPENA	FVTIEPAND	
MKKMGSGDG	IANLLTQYG	

Population Coverage: Predivac

Eastern Europe

Peptide	cumulative coverage (%)
YEKEGNDLT	50.90
FNPVKNLKA	53.13
YQPVSNLTA	53.20
YGQIVGNDH	53.24

North Africa

Peptide	cumulative coverage (%)
FNPVKNLKA	92.72
YGQIVGNDH	93.62
ISANGKMFG	97.29
WKPGNAPGI	97.31
YEKEGNDLT	97.32

Pacific

Peptide	cumulative coverage (%)
FNPVKNLKA	88.12
WKPGNAPGI	96.25
MRNEDPKKI	97.33
IPAKITPGI	98.13
MDGTASVNI	98.24
YTAGVSPKE	99.60
FIEKKNVLI	99.64

Sub Saharan Africa

Peptide	cumulative coverage (%)
FNPVKNLKA	93.58
YEKEGNDLT	97.17
YQPVSNLTA	97.63
WKPGNAPGI	97.68
YGQIVGNDH	97.81
ISANGKMFG	97.91

Australia

Peptide	cumulative coverage (%)
YEKEGNDLT	86.65
FNPVKNLKA	89.08
YQPVSNLTA	89.34
FAPVQNLTG	89.46
YGQIVGNDH	89.74

Middle East

Peptide	cumulative coverage (%)
FNPVKNLKA	97.66
YEKEGNDLT	99.59
YQPVSNLTA	99.68

North America

Peptide	cumulative coverage (%)
FNPVKNLKA	71.27
YEKEGNDLT	74.69
YQPVSNLTA	74.76
FAPVQNLTG	74.78
WKPGNAPGI	74.80
YGQIVGNDH	74.97
FLLDADHNT	74.98

South and Central America

Peptide	cumulative coverage (%)
YEKEGNDLT	85.67
FNPVKNLKA	91.24
YQPVSNLTA	91.55
FAPVQNLTG	92.11
YGQIVGNDH	93.01
YDYCITNPE	93.02
YAVMISKTG	93.03

Western Europe

Peptide	cumulative coverage (%)
FNPVKNLKA	94.82
YEKEGNDLT	97.28
YQPVSNLTA	97.33
YGQIVGNDH	97.44

Ethnicity – India New Delhi Pop2

Peptide	cumulative coverage (%)
YEKEGNDLT	94.44
FNPVKNLKA	99.45
YQPVSNLTA	100.00

Multipred-HLA

A2	B58	C1
FVSIALCSSLL	MAFAQQTEL	FAPLQYNPV
TLPILSRSLAV	KGMAQVPTY	FSMPCFAEA
TLDDPFILRDV	KSYSQNKFF	MIIMDGTASV
FILRDVRGQVV	KTLRIYTEI	YASEHYAVY
KTLRIYTEITV	VTANAIQQF	
WLGQALCIASA	LTNSNQLPF	
IIDAFNGGISL	IASTINQSW	
QLTNSNQLPFI	LTNESTLTL	
QLPFIFDVACN	IVLEAMDVW	
FLFSMCFAEA	VGQKVTLKW	
LTNESTLTLTV	VVLAADNVW	
VLLSVSDAPEL	SSNLYSANFE	
LLSVSDAPELL	YSANFEYLI	
MIMDGTASVNI		
LLEETITAKGV		
FQSTDMFYIDL		
WLTAHGGTNVV		
GMALNPDNYLI		
YLISKDVTGAT		
YILLIDDIQFTM		
YTAGVSPKECV		
VTINPTQFNPV		
FLLDADHNTFG		
NLYSANFEYLI		
YSANFEYLIPA		
YLIPANADPVV		
GMSAQSHEYCV		
AQGGYYAVMVV		
VMVVVDGKSYV		

DR1	DR6	DR8	DR3	DR16
YVLLVGDHK	YVLLVGDHK	VTKTLRIYT	YLITPALDL	KYHFLMKKM
YLIPALDL	FAMVEKYK	RMIVIAKK		
WLCLSSGQL	LVRTLVPTK	MIVIAKKY		
YAVMISKTG	KYHFLMKKM	MFAMVEKYK		
FEYLIPANA	KYVAFRHFQ	FAMVEKYKK		
	KYTFTMRRA	LVRTLVPTK		
	MFAMVEK	KYHFLMKKM		
	YHFLMKKMG	KYHFLMKKMG		
	YVAFRHFQS	KYVAFRHFQ		
		YVAFRHFQS		
		KYTFTMRRA		
		YTFTMRRAG		

NetMHCpan Server : MHC class I prediction

A*01:01	A*02:01	A*03:01	A*24:02	A*26:01
QIDRTIHY	ILSRSLAV	LQYNPVTK	SYSQNKFF	DTYKRMFM
MLDTWTVF	FIFDVACV	RMIVIVAK	KYVAFRHF	ETAWGTSH
ASHEYAVY	KMLDTWTV	GMFAMVEK	YYAVNDGF	NAPGIAGY
FQSTDMFY	KMFGSAVV			ETITAKGV
CSDLNYIL	TLWPNCSV			FTMRRAGM
GTASSNLY	IMDGTASV			YTAQGGYY
ASTYTVY				
MSAQSHEY				
YTAQGGYY				

B*07:02	B*08:01	B*27:05	B*39:01	B*40:01	B*58:01	B*15:01
LPILSRSL	FTMRRAGM	FRMDNLKF	NKFVSIAL	SEKGMPTL	KSYSQNKF	SLLGGMAF
APSKGMIM	MNGRRLAA	GRFSCESK	NQSWASPM	REMKVEVV	ASTINQSW	GMAQVPTY
APLQYNPV			WRQKTVDL	SEQGKNIL	CSVPANLF	FLFSMPCF
SPVTANAI				HENVIANL	LSSGQLDW	ASYTYTVY
KPTGTVAI				GEAMIYDM	MSAQSHEY	MSAQSHEY
IPSDTHTL						
IPASWKTI						
APAEWTTI						
IPANADPV						

9. REFERENCES

- Akimov, SS; Belkin, AM (2001). Cell-surface transglutaminase promotes fibronectin assembly via interaction with the gelatin-binding domain of fibronectin: A role in TGFbeta-dependent matrix deposition. J Cell Sci 114(pt 16):2989–3000.
- Akimov, SS; Krylov, D; Fleischman, LF; Belkin, AM (2000). Tissue transglutaminase is an integrin-binding adhesion coreceptor for fibronectin. J Cell Biol 148(4):825–838.
- Archibald McNicol; Sara, J; Israels (2010). Mechanisms of oral bacteria-induced platelet activation. Canadian Journal of Physiology and Pharmacology. 88(5): 510-524, 10.1139/Y10-029.
- Axelsson, P; Lindhe, J (1981) The significance of maintenance care in the treatment of periodontal disease. J. Clin. Periodontol. 8:281–294.
- Balklava, Z (2002). Analysis of tissue transglutaminase function in the migration of Swiss 3T3 fibroblasts: The active-state conformation of the enzyme does not affect cell motility but is important for its secretion. J Biol Chem 277(19):16567–16575.
- Beck, JD; Eke, P; Heiss, G (2005). Periodontal disease and coronary heart disease: a reappraisal of the exposure. Circulation. vol. 112, no. 1, pp. 19–24.
- Beck, JD; Elter, JR; Heiss, G; Couper, D; Mauriello, SM; Offenbacher, S (2001). Relationship of periodontal disease to carotid artery intima-media wall thickness: the atherosclerosis risk in communities (ARIC) study. Arteriosclerosis, Thrombosis, and Vascular Biology. vol. 21, no. 11, pp. 1816–1822.
- Boisvert, H; Duncan, MJ (2008). Clathrin-dependent entry of a gingipain adhesin peptide and Porphyromonas gingivalis into host cells. Cell Microbiol 10(12):2538–2552.
- Boisvert, H; Lorand, L; Duncan, MJ (2014). Transglutaminase 2 is essential for adherence of Porphyromonas gingivalis to host cells. Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.1402740111
- Bouchard, P; Boutouyrie, P; D'aiuto, F (2010). European workshop in periodontal health and cardiovascular disease consensus document. European Heart Journal Supplement. vol. 12, pp. B13–B22.
- Choi, JI; Schifferle, RE; Yoshimura, F; Kim, BW (1998). Capsular polysaccharide-fimbrial protein conjugate vaccine protects against Porphyromonas gingivalis infection in SCID mice reconstituted with human peripheral blood lymphocytes. Infect Immun. 66:391–393.
- Chun Xiao, SUN; Nathan, RW; Angelov, N; Ririe, C; Chen, JW; Boskovic, DS; Henkin, JM. The Chinese Journal of Dental Research.
- Contreras, A; Umeda, M; Chen, C; Bakker, I; Morrison, JL; Slots, J (1999). Relationship between herpesviruses and adult periodontitis and periodontopathic bacteria. J Periodontol. 70:478–484.
- Curtiss, LK; Tobias, PS (2009). Emerging role of Toll-like receptors in atherosclerosis. J. Lipid Res. 50: S340–S345.
- Dietrich, T; Sharma, P; Walter, C; Weston, P; Beck,J (2013). The epidemiological evidence behind the association between periodontitis and incident atherosclerotic

cardiovascular disease. Journal of Clinical Periodontology. vol. 40, no. 4, supplement pp. S70–S84.

- Dzink, JL; Socransky, SS; Haffajee, AD (1988). The predominant cultivable microbiota of active and inactive lesions of destructive periodontal diseases. J Clin Periodontol. 15:316–323.
- Fesus, L; Piacentini, M (2002). Transglutaminase 2: An enigmatic enzyme with diverse functions. Trends Biochem Sci 27(10):534–539.
- Frank, C; GibsonIII; Genco CA. Prevention of Porphyromonas gingivalis-Induced Oral Bone Loss following Immunization with Gingipain R1
- Gaudry, CA (1999). Cell surface localization of tissue transglutaminase is dependent on a fibronectin-binding site in its N-terminal beta-sandwich domain. J Biol Chem 274(43):30707–30714.
- Gibson, FC; Genco, CA (2001) Prevention of Porphyromonas gingivalis-induced oral bone loss following immunization with gingipain R1. Infect Immun. 69:7959–7963.
- Gibson, FC; Gonzalez, DA; Wong, J; Genco, CA (2004) Porphyromonas gingivalis-specific immunoglobulin G prevents P. gingivalis-elicited oral bone loss in a murine model. Infect Immun. 72:2408–2411.
- Gmur, R; Thurnheer, T; Guggenheim, B (1999). Dominant cross-reactive antibodies generated during the response to a variety of oral bacterial species detect phosphorylcholine. J Dent Res. 78:77–85.
- Hinode, D; Nakamura, R; Grenier, D; Mayrand, D (1998). Cross-reactivity of specific antibodies directed to heat shock proteins from periodontopathogenic bacteria and of human origin. Oral Microbiol Immunol. 13:55–58.
- Holtfreter, B; Empen, K; Glaser, S (2013). "Periodontisis is associated with endothelial dysfunction in a general population: a cross-sectional study". PLoS ONE, vol. 8, no. 12, Article ID e84603.
- Houston, LS; Lukehart, SA; Persson, GR; Page, RC (1999). Function of anti-Porphyromonas gingivalis immunoglobulin classes in immunized Macaca fascicularis. Oral Microbiol Immunol. 14:86–91.
- Imamura, T (2003). The role of gingipains in the pathogenesis of periodontal disease. J Periodontol. 74:111–118.
- Institutet, K (2013). Atherosclerosis and cardiovascular disease. Institute of Environmental Medicine, Unit of Immunology and Chronic Disease, Nobels väg, Stockholm. Sweden Immunity. 171-177.
- Janket, SJ; Baird, AE; Chuang, SK; Jones, JA (2003). Meta-analysis of periodontal disease and risk of coronary heart disease and stroke. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontics. vol. 95, no. 5, pp. 559–569.
- Kadowaki, T; Yoneda, M; Okamoto, K; Maeda, K; Yamamoto, K (1994). Purification and characterization of a novel arginine-specific cysteine proteinase (argingipain) involved in the pathogenesis of periodontal disease from the culture supernatant of Porphyromonas gingivalis. J Biol Chem. 269:21371–21378.

- Kawata, Y; Iwasaka, H; Kitano, S; Hanazawa, S (1997). Porphyromonas gingivalis fimbria- stimulated bone resorption is inhibited through binding of the fimbriae to fibro- nectin. Infect Immun. 65(2):815–817.
- Khader, YS; Albashaireh, ZSM; Alomari, MA (2004). Periodontal diseases and the risk of coronary heart and cerebrovascular diseases: a meta-analysis. Journal of Periodontology. vol. 75, no. 8, pp. 1046–1053.
- Koupenova, M; Johnston-Cox, H; Vezeridis, A; Gavras, H; Yang, D; Zannis, V; Ravid, K (2012). A 2b Adenosine Receptor Regulates Hyperlipidemia and Atherosclerosis. Circulation. 125:354-363.
- Kubar, A; Saygun, I; Yapar, M; Ozdemir, A; Slots, J (2004). Real-time PCR quantification of cytomegalovirus in aggressive periodontitis lesions using TaqMan technology. J Periodontal Res. 39:81–86.
- Kumagai, Y; Yagishita, H; Yajima, A; Okamoto, T; Konishi, K (2005). Molecular mechanism for connective tissue destruction by dipeptidyl aminopeptidase IV produced by the periodontal pathogen Porphyromonas gingivalis. Infect Immun. 73(5):2655–2664.
- Laine, ML; Appelmelk, BJ; van-Winkelhoff, AJ (1997). Prevalence and distribution of six capsular serotypes of Porphyromonas gingivalis in periodontitis patients. J Dent Res. 76:1840–1844.
- LeMosy, EK (1992). Visualization of purified fibronectin-transglutaminase com-plexes. J Biol Chem. 267(11):7880–7885.
- Ling, LJ; Ho, CC; Wu, CY; Chen, YT; Hung, SL (2004). Association between human herpesviruses and the severity of periodontitis. J Periodontol. 75:1479–1485.
- Lockhart, PB; Bolger, AF; Papapanou, PN (2012). Periodontal disease and atherosclerotic vascular disease: does the evidence support an independent association. A scientific statement form the American Heart Association,Circulation. vol. 125, no. 20, pp. 2520–2544.
- Lorand, L (2007) Crosslinks in blood: Transglutaminase and beyond. FASEB J 21(8): 1627–1632.
- Maeda, H; Miyamoto, M; Hongyo, H; Nagai, A; Kurihara, H; Murayama, Y. Heat shock protein 60 (GroEL)
- Mallika, V; Goswami, B; Rajappa M (2007). Atherosclerosis Pathophysiology and the Role of Novel Risk Factors: A Clinicobiochemical Perspective. Sage journals. Angiology. vol. 58 no. 5 513-522.
- Mattila, KJ**;** Pussinen, PJ**;** Paju, S (2005)**.** Dental Infections and Cardiovascular Diseases: A Review. Journal of Periodontology. Vol. 76, No. 11-s, Pages 2085-2088 , DOI 10.1902/jop.76.11-S.2085 (doi:10.1902/jop.2005.76.11-S.2085)
- Moritz, AJ; Cappelli, D; Lantz, MS; Holt, SC; Ebersole, JL (1998). Immunization with Porphyromonas gingivalis cysteine protease: effects on experimental gingivitis and ligature-induced periodontitis in Macaca fascicularis. J Periodontol. 69:686–697.
- Murakami, Y (1996). Porphyromonas gingivalis fimbrillin is one of the fibro- nectin-binding proteins. Infect Immun. 64(7):2571–2576.

- Niedzielska, I; Janic, T; Cierpka, S; Swietochowska, E (2008). The effect of chronic periodontitis on the development of atherosclerosis: review of the literature. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research. 14(7):RA103-6.
- O'Toole, G; Kalpan, HB; Kolter, R (2000). Biofilm formation as microbial development. Annu. Rev. Microbial. 54:49-79.
- Page, RC (2000). Vaccination and periodontitis: myth or reality. J Int Acad Periodontol. 2:31–43.
- Parra, B; Slots, J (1996). Detection of human viruses in periodontal pockets using polymerase chain reaction. Oral Microbiol Immunol. 11:289–293.
- Persson, GR; Engel, D; Whitney, C; Darveau, R; Weinberg, A; Brunsvold, M (1994). Immunization against Porphyromonas gingivalis inhibits progression of experimental periodontitis in nonhuman primates. Infect Immun. 62:1026–1031.
- Radek, JT; Jeong, JM; Murthy, SN; Ingham, KC; Lorand, L (1993). Affinity of human erythrocyte transglutaminase for a 42-kDa gelatin-binding fragment of human plasma fibronectin. Proc Natl Acad Sci USA 90(8):3152–3156.
- Seymour, GY; Ford, PJ; Cullinan, MP; Leishman, S; Yamazaki, K (2007). Faculty of Dentistry, University of Otago, Dunedin, New Zealand, Oral Biology and Pathology, The University of Queensland, Brisbane, Queensland, Australia and Laboratory of Periodontology and Immunology, Department of Oral Health and Welfare, Niigata University Faculty of Dentistry, Niigata, Japan. Journal compilation. European Society of Clinical Microbiology and Infectious Diseases CMI, 13 (Suppl. 4), 3–10.
- Slocum, C; Stephen, R; Coats; Hua, N; Kramer, C; Papadopoulos, G; Ellen, O; Weinberg; Cynthia, V; Gudino; James, A; Hamilton; Richard, P; Darveau; Caroline, A (2014). Distinct Lipid A Moieties Contribute to Pathogen-Induced Site-Specific Vascular Inflammation. PLoS Pathogens, 10 (7): e1004215 DOI: 10.1371/journal.ppat.1004215
- Slots, J (2004). Systemic antibiotics in periodontics. J Periodontol. 75:1553–1565.
- Socransky, SS; Haffajee, AD; Cugini, MA; Smith, C; Kent, RL (1998). Microbial complexes in subgingival plaque. J Clin Periodontol. 25:134–144.
- Sundqvist, G (1993). Pathogenicity and virulence of black-pigmented gram-negative anaerobes. FEMS Immunol Med Microbiol. 6:125–137.
- Travis, J; Pike, R; Imamura, T; Potempa, J (1997). Porphyromonas gingivalis proteinases as virulence factors in the development of periodontitis. J Periodontal Res. 32:120–125.
- Upchurch, HF; Conway, E; Patterson, MK; Maxwell, MD (1991). Localization of cellular transglutaminase on the extracellular matrix after wounding: Characteristics of the matrix bound enzyme. J Cell Physiol 149(3):375–382.
- Wong, J; Attardo, C; Frank, G; Gibson III; Hong, C; Chou, HH; Yumoto, H; Chen, J; Lien, E (2004). Innate Immune Recognition of Invasive Bacteria Accelerates Atherosclerosis in Apolipoprotein E-Deficient Mice. Circulation. 109:2801-2806