

## Prediction of Antigenic MHC Peptide Binders and TAP Binder of COX1 Protein through *In Silico* Approach

Mishra S\* and Gomase VS

Department of Biotechnology, Mewar University, Chittorgarh, India

\*Corresponding author: Sonu Mishra, Department of Biotechnology, Mewar University, Chittorgarh, India, Tel: 9560808369; E-mail: [sonumishra1014@gmail.com](mailto:sonumishra1014@gmail.com); [gomase.viren@gmail.com](mailto:gomase.viren@gmail.com)

Received date: Mar 14, 2016; Accepted date: Mar 30, 2016; Published date: April 4, 2016

Copyright: © 2016 Mishra S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

In the current analysis Cytochrome c oxidase subunit I (CO1 or MT-CO1) protein sequence from GWD has been used to study the MHC binding antigenic peptide, antigenic peptide prediction through different B cell prediction method, protein solvent accessibility, polar and nonpolar residue to analyze the regions which are probably exposed on the protein surface. From the protein the peptide fragment can be used to analyze and specific nonamer can be selected for the rational vaccine designing. In this investigation, PSSM and SVM algorithms are applied for finding of MHC I and MHC II binding peptides. We also predicted the high affinity TAP binding peptides of CO1 protein from GWD, having 205 amino acids which show 197 nonamers. From the opted outcomes we predict that, the possibilities that, the antigenic peptide of cytochrome c oxidase subunit I (mitochondrion) protein might play a major role and could be the most suitable candidate for subunit vaccine development on the bases of the finding that, with single epitope, the immune response can be generated in large population.

**Keywords:** Antigenic peptides; GWD; MHC-binders; TapPred; PSSM; SVM; Nonamers; COX1

### Abbreviations

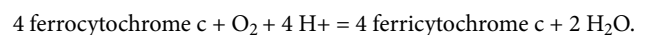
CCOI: Cytochrome C Oxidase Subunit I; MT-COI: Mitochondrial Cytochrome C Oxidase Subunit I; ETC: Electron Transport Chain; mtDNA: Mitochondrial DNA (Deoxyribonucleic Acid); MTCO3: Mitochondrial Cytochrome C Oxidase Subunit I; DCCD: Di-Cyclohexylcarbodiimide; MHC I: Major Histocompatibility Complex-Class I; MHC II: Major Histocompatibility Complex-Class II; PSSM: Position Specific Scoring Matrices; SVM: Support Vector Machine; GWD: Guinea Worm Disease; UniProt: The Universal Protein Resource; NCBI: National Center for Biotechnology Information; TAP: Transporter Associated with Antigen Processing; HPLC: High Performance Liquid Chromatography; GRAVY: Grand Average of Hydropathicity; TapPred: TAPPred is an on-line service for predicting binding affinity of peptides toward the TAP transporter. The prediction of TAP binding peptides is important in order to identify the MHC class-I restricted T cell epitopes. The Prediction is based on cascade SVM, using sequence and properties of the amino acids. The correlation coefficient of 0.88 was obtained by using jack-knife validation test

Rankpep: This server predicts peptide binders to MHCI and MHCII molecules from protein sequence/s or sequence alignments using Position Specific Scoring Matrices (PSSMs). In addition, it predicts those MHCI ligands whose C-terminal end is likely to be the result of proteasomal cleavage

### Introduction

COX1 (Cytochrome c Oxidase I) are commonly known as Mitochondrially encoded Cytochrome c Oxidase I (MT-CO1). This protein encodes of MT-CO1 gene in humans, whereas in other

eukaryotes, this gene is addressed as COX1, CO1 or COI [1]. COX 1 is a primary subunit of cytochrome c oxidase complex. The subunit I of Cytochrome c Oxidase (CO1 or MT-CO1) is one out of three mitochondrial DNA (mtDNA) encoded subunits (MT-CO1, MT-CO2, MT-CO3) of respiratory complex IV. Complex IV is considered to be the third and final enzyme of mitochondrial oxidative phosphorylation of the ETC. In aerobic metabolism, Cytochrome c oxidase (EC 1.9.3.1) is a primal enzyme. Study suggest that in prokaryotes, this enzyme complex is consist of three to four subunits and up to thirteen polypeptides in mammals, but out of which only the catalytic subunit is found in all heme-copper respiratory oxidases. The enzyme complexes modify in heme and copper composition, substrate type and substrate affinity. The different respiratory oxidases allow the cells to tailor-make their respiratory systems followed by a diverse environmental growth circumstances [2]. Catalytic activity of the COX1 protein in oxidative phosphorylation is:



COX1 involved in oxidative phosphorylation pathway, which is division of Energy metabolism. In order to identify the animal species or closely related species analysis, this COX1 gene is used as the barcode, due to its fastest rate of mutation and sequence conserved specificity among con-specifics. This COX1 gene is frequently practiced as a DNA barcode to identify animal species. In contradiction, raised by the skeptics suggest that, the MT-CO1 sequence differences are available in too minute to be detected among nearly related species, more than 2% sequence divergency is generally detected between such organisms [3]. MTCO1 is encoded by the guanine-rich heavy (H) strand of the mtDNA and situated between nucleotide pairs (nps) 5904 and 7444 [4,5]. It is maternally inherited along with the mtDNA [6,7]. The predicted molecular weight (MW) of MTCO1 is 57 kD [4,5]. However, its apparent MW on SDS-polyacrylamide gels (PAGE) is somewhat less. Using Tris-glycine buffer it runs at 39.5 kD [8-10], whereas results in urea-phosphate test,

it gives an apparent MW of 45 kD [11,12]. Investigations suggest that, this protein is extremely expressed in the cytoplasm of colonic crypts (intestinal glands) of the human colon (large intestine). However, with age in the human it is frequently gets lost in colonic crypts and also it is often absent in field defects that give rise to colon cancers or in portions of colon cancers [13]. CCOI is coded for by the mitochondrial chromosome. The occurrence of the chromosome in the mitochondria in multiple copy forms, which varies between two and six / per mitochondrion [14-16]. The generation of new type of mitochondria occurs due to random segregation of chromosome during the mitochondrial fission, if a mutation takes place in CCOI in one chromosome of a mitochondrion. The occurrence of mitochondria per cell is 100 - 700, depending upon the cell type [15,16]. In rats, the average half-life of mitochondria depends on cell type, and found between 9 and 24 days [17], in mice about 2 days [18], whereas, in human it varies from days to week depending on cell type. The inadequacy of CCOI in a mitochondrion leads to lower reactive oxygen production (and less oxidative damage) and this provides an exclusive vantage in competition with other mitochondria within the same cell to bring forth homoplasmy for CCOI-deficiency [13]. The phylogenetic analysis study conducted by Nguai et al. [19] on the cytochrome c oxidase subunit 1 (cox 1) sequence of *A. ceylanicum* from positive human and animal fecal samples suggest that, considerable level of genetic variation within the cox 1 sequence of *A. ceylanicum* might be a potential haplotype-linked divergences in zoonotic, epidemiological and pathobiological characteristics, a hypothesis which still needs a further future in depth investigation [19]. In another study, the analyses of multiple sequence alignments of mitochondrial 16S rDNA (ribosomal DNA) and cox 1 of *Trichuris skrjabini* revealed high homology with those of *Trichinella* species and this was the first time when the mitochondrial DNA gene sequences of one species of trichurid nematode have been cited [20]. MT-CO1 may act as agent in the pathogenesis of acquired idiopathic sideroblastic anemia, this sickness is characterized by inadequate formation of heme and excessive aggregation of iron in mitochondria. The overloaded iron in mitochondrial may be attributable to mutations of mitochondrial DNA, because impairing the reduction of ferric iron to ferrous iron can cause respiratory chain dysfunction. Whereas, study suggest that the reduced form of iron is substantive in the last step of mitochondrial heme biosynthesis. Insufficiency of COX drives a clinically heterogeneous variety of neuromuscular and non-neuromuscular disorders in childhood and adulthood. The mutation in the COX1 results in the several other variety of the disorders such as Leber hereditary optic neuropathy (LHON) (primary mitochondrial DNA mutations affecting the respiratory chain complexes), mitochondrial complex IV (MTC4D) deficiency, recurrent myoglobinuria mitochondrial (RM-MT), Deafness, sensorineural, mitochondrial disorder, colorectal cancer. Considering the COX1 importance, we have taken this protein for investigation of its antigenicity role, its solvent accessibility property, polar and nonpolar residue analysis. By this time we all are aware that, the regions that are likely exposed on the surface of proteins could be the potentially antigenic that allows potential drug targets to identify active sites against infection as well as for designing and development of effective drug to treat infections. Cytochrome c Oxidase subunit I (mitochondrion) comprised of 205 amino acid residues obtained from *Dracunculus medinensis* for the study of MHC class I and II binding peptide, antigenicity, Solvent accessibility, polar and nonpolar residue to analyze the regions that are likely exposed on the surface of proteins. A little dragon from Medina (*D. medinensis*) is the only species of *Dracunculus* genus [21-24] which causes dracunculiasis in humans,

commonly well known as "Guinea Worm Disease (GWD)". The other *Dracunculus* species generally resides in the internal tissues and body cavities of non-human mammals and reptiles (snake and turtles) [25]. This parasite follows a very unique life wheels, comprised of six developmental steps with the longest incubation period of about one and half years approximately. This is also the one of the most neglected tropical parasites which has got clinical grandness and needs to be eradicated completely [26]. Once this parasite approaches the maturation stage, these worms copulate and a millions of eggs is formed in uterus of adult female, whereas male worm dies after copulation. Once the incubation period is over the larvae emerge out from the blister once the blister burst out (predominantly localized in the lower extremities (80% - 90%) in most of the reported cases) when an infected individual comes in contact with water. The symptoms developed in the infected individual are slight fever, local skin redness, swelling and severe pruritus around the blister. Other symptoms includes are diarrhea, nausea, vomiting and dizziness [27]. Immersing or pouring water over the blister provides pain relief but this is the point when the adult female is exposed to the external environment [28]. During emergence of the limbs in open water sources it recognizes the temperature difference and releases the milky white liquid in the water which contains millions of immature larvae, when larvae released in water are ingested by copepods where they mount twice and become infective larvae within two weeks [29]. The antigen peptides of *D. medinensis* could be most desirable segment for the development of subunit vaccine on the bases of fact that, the immune response can be generated in large population with the single epitope. This approach is generally based on the phenomenon of cross-protection, whereby the individual can be infected with the mild strain and is protected against a more severe strain of pathogen of the same. The resistant transgenic host's phenotype includes of fewer centers of initial infection, following a delay development in symptom with low accumulation. There is the possibilities that the predicted antigenic peptides from *D. medinensis* could contribute a major role in drug formulation (or peptide vaccine) and disease eradication [30] because a single protein subunit can generate sufficient immune response. In this current investigation work, we have applied the *in silico* approach for MHC class I and class II binding antigenic peptide identification. MHC molecules are cell surface protein which binds to peptides derived from host or antigenic proteins and present them to cell surface for realization by T-cells. T cell recognition is a significant mechanism of the adaptive immune system by which the host distinguishes and responds to foreign antigens [30,31]. The two forms of MHC molecule are extremely polymorphic. MHC class I molecules present peptides from proteins synthesized within the cell, whereas, MHC class II molecule present peptides derived from endocytosed extracellular proteins. MHC molecules take active part in host immune reactions and their contribution in immune response to almost all antigens and it give impacts on specific sites. MHCI binds to some of the peptide fragments generated after proteolytic cleavage of antigen [32]. Identification of MHC-binding peptides and T-cell epitopes helps to improve our understanding of specificity of immune responses [33-36]. Antigenic peptides are most suitable for peptide or synthetic vaccine development.

## Methodology

### Database searching

The cytochrome c oxidase subunit I (mitochondrion) protein sequence of *Dracunculus medinensis* was recovered from

www.ncbi.nlm.nih.gov, UniProt databases which is the essential important step for further investigation [37,38].

### Prediction of the physico-chemical properties of the protein

The physico-chemical properties like molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient [39-41], estimated half-life [42,43], instability index [44], aliphatic index [45] and Grand Average of Hydropathicity (GRAVY) [46] were analyzed by the ProtParam (<http://www.expasy.org/>). Effect of temperature on protein solubility and denaturation and this physicochemical property prediction will help to understand the effect of pH on protein solubility and protein isoelectric point, interaction between protein and water molecules and hydrogen bonds, protein-ligand binding affinity, function of the protein. These property analysis plays an important role and taken into the consideration in the drug development or designing.

### Prediction of antigenicity

The prediction antigenicity program, predicts the segments from cytochrome c oxidase subunit I (mitochondrion) protein that are likely to be antigenic by reducing an antibody reaction. In this research work antigenic epitopes of *Dracunculus medinensis*- cytochrome c oxidase subunit I (mitochondrion) are determined by using the various antigenic prediction methods such as Hopp and Woods, Welling, Parker, Bepipred, Kolaskar and Tongaonkar antigenicity methods [47-59].

### Prediction of MHC binding peptide

The Major Histocompatibility Complex (MHC) peptides binding of protein from *Dracunculus medinensis* were predicted using neural networks trained on C terminals of known epitopes. Rank pep tool predicts peptide binders to MHC-I ligands whose C-terminal end is likely to be the result of proteosomal cleavage using Position Specific Scoring Matrices (PSSMs). The sequence similarities are shared by the peptides that bind to a given MHC molecule. Traditionally, the sequence patterns used for the prediction of peptides binding to MHC molecules, such sequence patterns are however, have proven to be too simple, as the complexity of the binding motif cannot be precisely represented by the few residues present in the pattern [60]. RANKPEP uses "Position Specific Scoring Matrices (PSSMs) or profiles" from set of aligned peptides known to bind to a given MHC molecule as the forecaster of MHC-peptide binding and overpower the complexity of the binding motif limitation. Whereas, in the other hand the Support Vector Machine (SVM) based method for prediction of promiscuous MHC class II binding peptides from protein sequence; SVM has been trained on the binary input of single amino acid sequence [61-64].

### Prediction of antigenic peptides by cascade SVM based TAPPred method

We predicted the cascade SVM based several TAP binders which was based on the sequence and the features of amino acids [65]. We found the MHCI binding regions, the binding affinity of cytochrome c oxidase subunit I (mitochondrion).

### Solvent accessible regions

We also investigated the solvent accessible regions of proteins holding highest probability that a given protein region lies on the surface of a protein Surface Accessibility, backbone or chain flexibility via Emimi et al. [66] and Karplus and Schulz [67]. By using different scale we predict the hydrophobic and hydrophilic characteristics of amino acids that are rich in charged and polar residues [68-70] (Figure 1).

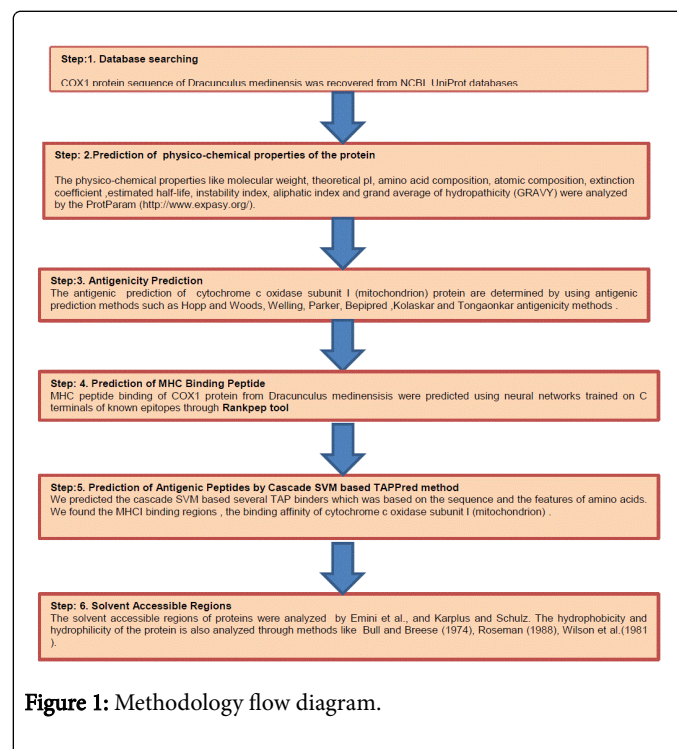


Figure 1: Methodology flow diagram.

### Results and Discussion

*Dracunculus medinensis* protein Cytochrome c oxidase subunit I (mitochondrion), is consist of 205 amino acids long residue with 197 nonamers.

KDIGSMYLIFGFWSGMVAGLSILIRAE LCKPGFFFGSGQLYNA  
VITSHAIMMIFFMVMP SLIGGFGNWVPLMLGAPDMSFPR LNN  
VSYWLMPVSLMLILSACLVDSSCGT SWTIYPPLSTSGHPGNSVDLA  
IFSLHCSGVSSILGGINFMTTVKNMRSASISLEHLSLFVWTVFVTVF  
LLILTLPLVLAGAITMLLMDRSFNTS

### Prediction of the physicochemical properties of the protein

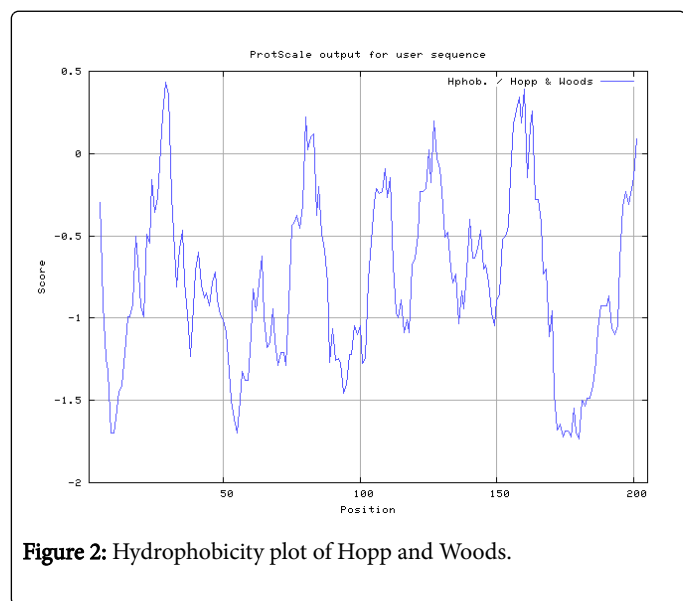
The physico-chemical properties of Cytochrome c oxidase subunit I like molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life instability index, aliphatic index and grand average of hydropathicity (GRAVY) were analyzed by the ProtParam (<http://www.expasy.org/>) and found the instability index is 32.56 which infers that the protein is the stable protein and its grand average hydropathicity predicted is 0.876 (Table 1).

Protein	Gene ID	No. of amino acids	Molecular weight	Theoretical PI	Total no. of -vely charged Residue (Arg + Lys)	Total no. of +vely charged Residue (Arg + Lys)	Total no. of atoms	Extinction coefficient	Instability index	Aliphatic Index	GRAVY
COX1	GI563582560	205	22230.4	7.06	7	7	3144	Extinction coefficients are in units of M-1 cm <sup>-1</sup> , at 280 nm measured in water. Ext. coefficient 33710 Abs 0.1% (= 1 g/l) 1.516, assuming all pairs of Cys residues form cysteines. Ext. coefficient 33460 Abs 0.1% (= 1 g/l) 1.505, assuming all Cys residues are reduced	32.56 (This classifies the protein as stable)	109.8	0.876

**Table 1:** The physicochemical properties of the Cytochrome c oxidase subunit I (mitochondrion).

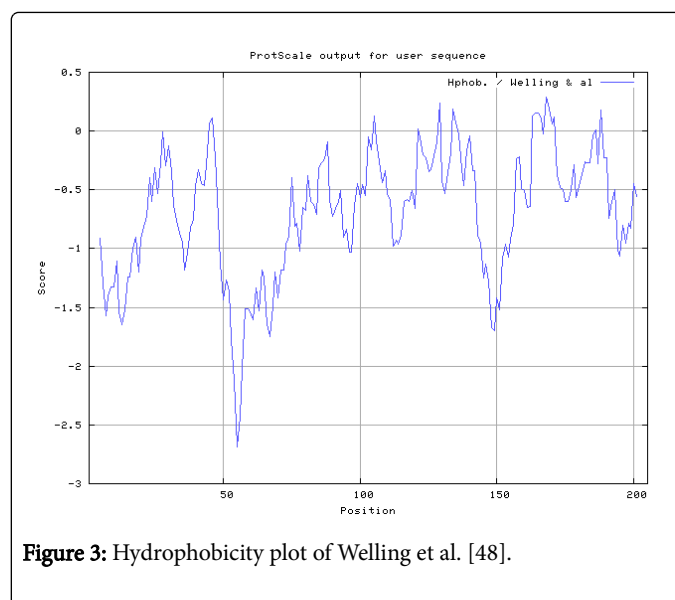
### Prediction of antigenic peptides

In antigenic peptide study, we detected the antigenic determinants by finding the area of greatest local hydrophilicity. In the Hopp-Woods scale Hydrophilicity prediction analysis of the protein found high in position: 29, Score: 0.433 (max), i.e., 23-RAELCKP-32 in a protein sequence, assuming that the antigenic determinants would be displayed on the surface of the protein and thus would be located in hydrophilic regions (Figure 2).



**Figure 2:** Hydrophobicity plot of Hopp and Woods.

In the Welling et al. [48] antigenicity plot gives value as the log of the quotient between percentage in average proteins and percentage in a sample of known antigenic regions, the prediction result data found high in position: 168, Score: 0.289 (max) (177-VTVFLLI-183) (Figure 3).



**Figure 3:** Hydrophobicity plot of Welling et al. [48].

We also study the hydrophobicity plot of HPLC / Parker [49] and the highest peak is obtained in position: 110 (Residue: C) with highest score: 5.971 (107-DSSCGTS-113) (Figure 4).

There are 10 antigenic a determinant sequence is found by Kolaskar and Tongaonkar [50] antigenicity scales (Figures 5a and 5b).

Kolaskar and Tongaonkar antigenicity methods (Figure 6).

Predicted peptides result found, i.e., 4-GSMYLIF-10;17-VGAGLSILIRAE LCKPGF-34;40-QLYNAVITS-48;54-IFFMVMP SLI-63;69-WMVPLML-75;84-RLNNVSYW LMPVSLMLILSACLVDSSCGT-112;114-WTIYPP LST-122;126-PGNSVDLAIFSLHCSGVSSILG-147;161-ASISLEHLSLFVWTVFVTVFLLILTLPVLAGAITM-195 and the predicted antigenic fragments can bind to MHC molecule is the first step in the process of the drug designing or vaccine development.

### Solvent accessible regions

The solvent accessible regions in proteins are also predicted. The different measurement was performed for the prediction of antigenic activity, surface region of peptides. Emini et al. [66] (Figure 7).

Predicts the highest probability i.e. found In position: 201(Residue: S), the sequence is 199-DRSFNT-204 with highest score: 4.77, that a given protein region lies on the surface of a protein and are used to identify antigenic determinants on the surface of proteins. Karplus and Schulz [67] (Figure 8).

The highest score (score: 1.079) is found in the position: 127(Residue: 124) and the sequence is 124-GHPGNSV-130. The other second highest peak has been found in the position: 126(Residue: P) with the sequence 123-SGHPGNS-129 (Score: 1.078). Predict backbone or chain flexibility on the basis of the known temperature B factors of the  $\alpha$ -carbons. The hydrophobicity and hydrophilic characteristics of amino acids is determined by using different scales that are rich in charged and polar residues, i.e., Bull and Breese [68] result high in Position: 125 Score: 0.509 (max) 122-TSGHPGN-128 (Figure 9).

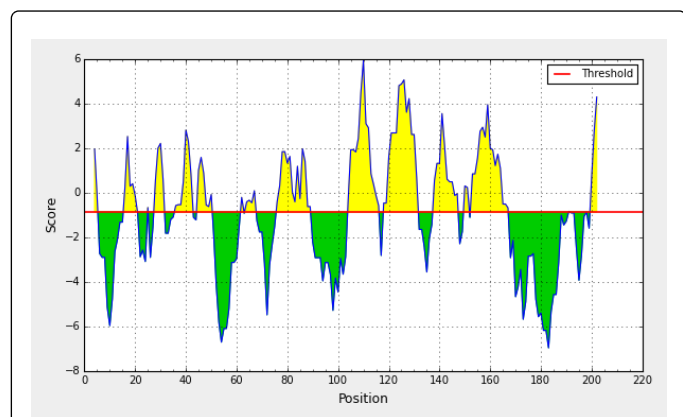


Figure 4: Hydrophobicity plot of HPLC / Parker et al. [49].

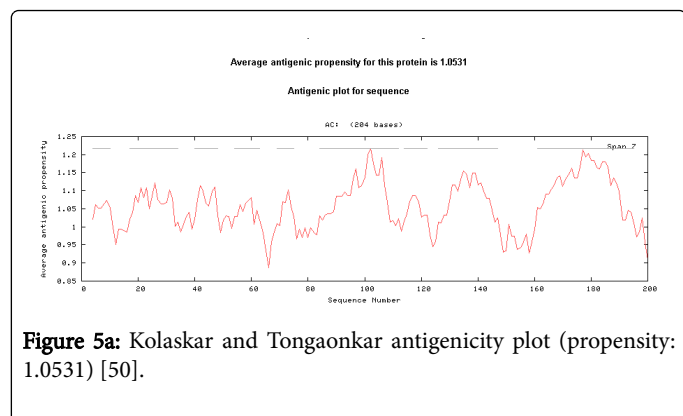


Figure 5a: Kolaskar and Tongaonkar antigenicity plot (propensity: 1.0531) [50].

Roseman [69] result found high in Position: 180 Score: 1.491 (max) 177-TVFLLI-183 (Figure 10).

Wilson et al. [70] in Position: 180, Score: 5.411 (max) 177-TVFLLI-183 (Figure 11).

n	Start Position	Sequence	End Position
1	4	GSMYLIIF	10
2	17	VGAGLSILIRAEELCKPGF	34
3	40	QLYNAVITS	48
4	54	IFFVMVPSLI	63
5	69	WMVPLML	75
6	84	RLNNVSYWLMFVSLMLILSACLVDSSCGT	112
7	114	WTIYPPLST	122
8	126	PGNSVDLAIFSLHCSGVSSILG	147
9	161	ASISLEHLSLFVWTVFVTVLLILTLPLVLAGAITM	195

Figure 5b: The 9 antigenic determinants.

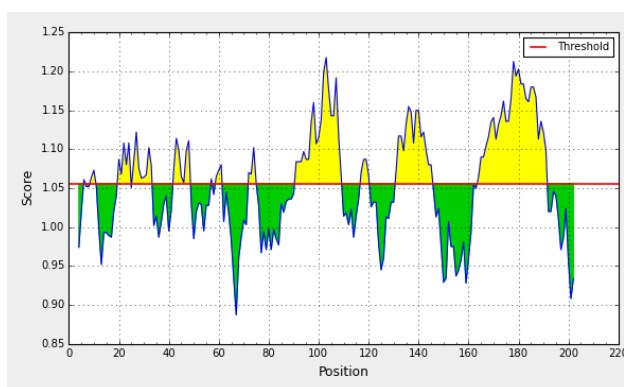


Figure 6: Kolaskar and Tongaonkar antigenicity plot.

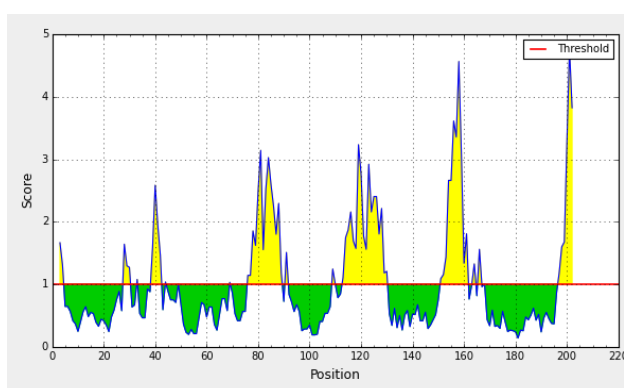


Figure 7: Emini surface accessibility.

### Prediction of MHC binding peptide

The binding of peptides to a number of different alleles using PSSM were identified from cytochrome c oxidase subunit I (mitochondrion) protein of *Dracunculus medinensis* having 247 nonamers. We have predicted MHC-I peptide binders of cytochrome c oxidase subunit I (mitochondrion) from *Dracunculus medinensis*.

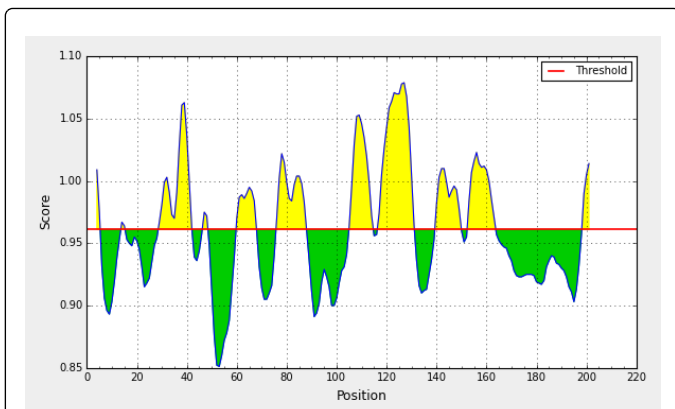


Figure 8: Karplus and Schulz flexibility prediction.

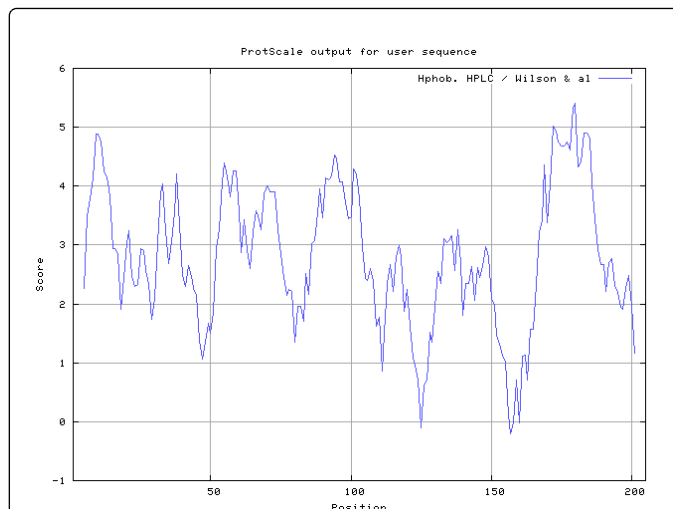


Figure 11: Hydrophobicity / HPLC plot of Wilson et al. [70].

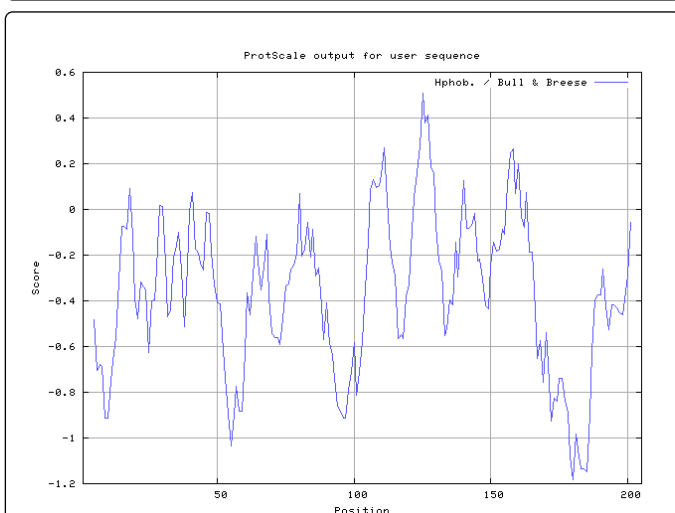


Figure 9: Bull and Breese use surface tension to measure hydrophobicity.

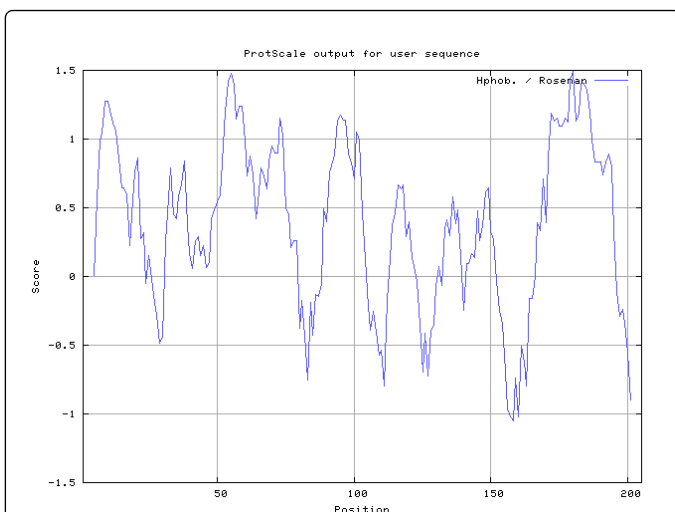


Figure 10: Hydrophobicity plot of Roseman [69].

We found predicted MHC-I peptide binders of protein for 8mer\_H2\_Db alleles with the consensus sequence QNWNCCIT that yields the maximum score i.e. 52.494, 9mer\_H2\_Db with, the consensus sequence FCIHNCDDYM that yields the maximum score i.e. 50.365, 10mer\_H2\_Db with, the consensus sequence SGYYNFFWCL that yields the maximum score, i.e., 58.858, 11mer\_H2\_Db with, the consensus sequence CGVYNFYCCY that yields the maximum score, i.e., 79.495 (Table 2).

MHC-II peptide binders for I Ab with the consensus sequence YYAPWCNNA that yields the maximum score, i.e., 35.632, I\_Ad with the consensus sequence QMVHAAHAE that yields the maximum score, i.e., 53.145, MHC-II I\_Ag7 with the consensus sequence WYAHAFKYV that yields the maximum score, i.e., 40.873 for MHC II allele was tested. The prediction of high affinity binders are performed using a cascade SVM based TAPPred method, where more than 63 High affinity TAP Transporter peptide regions were obtained This high affinity binders represents predicted TAP binders residues which occur at N and C termini from *Dracunculus medinensis* antigen cytochrome c oxidase subunit I (mitochondrion). TAP is an important transporter that transports antigenic peptides from cytosol to ER. TAP binds and translocate selective antigenic peptides for binding to specific MHC molecules. The efficiency of TAP-mediated translocation of antigenic peptides is directly proportional to its TAP binding affinity. Thus, by understanding the nature of peptides, that bind to TAP with high affinity, is important steps in endogenous antigen processing. The correlation coefficient of 0.88 was obtained by using jackknife validation test. T cell immune responses are derived by antigenic epitopes; hence, their identification is important for design synthetic peptide vaccine. T cell epitopes are recognized by MHCI molecules producing a strong defensive immune response against *Dracunculus medinensis* antigen cytochrome c oxidase subunit I (mitochondrion). Therefore, the prediction of peptide binding to MHCI molecules by appropriate processing of antigen peptides occurs by their binding to the relevant MHC molecules. Because, the C-terminus of MHCI-restricted epitopes results from cleavage by the proteasome and thus, proteasome specificity is important for determining T-cell epitopes. Consequently, RANKPEP also focus on the prediction of conserved epitopes. C-terminus of MHCI-restricted peptides is generated by the proteasome, and thus moreover, these sequences are highlighted in

purple in the output results. Proteasomal cleavage predictions are carried out using three optional models obtained applying statistical language models to a set of known epitopes restricted by human

MHCI molecules as indicated as I\_Ab.p, I\_Ad.p, I\_Ag7.p, I\_Ak.p alleles, which is highlighted in red represent predicted binders (Table 3).

MHC-I Allele	RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
8mer_H2_Db	1	147	SIL	GGINFMTT	VKN	821.93	13.596	25.90%
8mer_H2_Db	2	109	VDS	SCGTSWTI	YPP	812.92	10.984	20.92%
8mer_H2_Db	3	67	GGF	GNWMVPLM	LGA	906.15	10.568	20.13%
8mer_H2_Db	4	29	RAE	LCKPGFFF	GSG	940.18	10.043	19.13%
8mer_H2_Db	5	16	WSG	MVGAGLSI	LIR	728.9	8.059	15.35%
9mer_H2_Db	1	147	SIL	GGINFMTTV	KNM	921.06	16.377	32.52%
9mer_H2_Db	2	146	SSI	LGGINFMTT	VKN	935.09	14.44	28.67%
9mer_H2_Db	3	66	IGG	FGNWMVPLM	LGA	1053.33	14.031	27.86%
9mer_H2_Db	4	124	STS	GHPGNSVDL	AIF	876.92	12.987	25.79%
9mer_H2_Db	5	39	FGS	GQLYNAVIT	SHA	960.09	11.476	22.79%
10mer_H2_Db	1	76	LML	GAPDMSFPRL	NNV	1072.26	10.794	18.34%
10mer_H2_Db	2	146	SSI	LGGINFMTTV	KNM	1034.22	9.981	16.96%
10mer_H2_Db	3	153	NFM	TTVKNMRSAS	ISL	1076.22	9.867	16.76%
10mer_H2_Db	4	39	FGS	GQLYNAVITS	HAI	1047.17	7.555	12.84%
10mer_H2_Db	5	166	ISL	EHLSLFVWTV	FVT	1189.41	7.286	12.38%
11mer_H2_Db	1	146	SSI	LGGINFMTTVK	NMR	1162.39	20.946	26.35%
11mer_H2_Db	2	43	QLY	NAVITSHAIMM	IFF	1169.41	12.764	16.06%
11mer_H2_Db	3	139	SLH	CSGVSSILGGI	NFM	974.14	12.41	15.61%
11mer_H2_Db	4	160	NMR	SASISLEHLSL	FVW	1138.3	9.483	11.93%
11mer_H2_Db	5	31	ELC	KPGFFFSGGQL	YNA	1166.35	7.18	9.03%

**Table 2:** Promiscuous MHC ligands, having C-terminal ends are proteasomal cleavage sites of *Dracunculus medinensis*. (All rows highlighted in red represent predicted binders and A peptide highlighted in violet has a C-terminus predicted by the cleavage model used).

MHC -II Allele	RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
MHC -II I_Ab	1	42	GQL	YNAVITSHA	IMM	957.05	10.099	28.34%
MHC -II I_Ab	2	80	APD	MSFPRLNNV	SYW	1059.25	10.012	28.10%
MHC -II I_Ab	3	187	LTL	PVLAGAITM	LLM	854.07	9.53	26.75%
MHC -II I_Ab	4	90	NVS	YWLMPVSLM	LIL	1098.42	8.674	24.34%
MHC -II I_Ab	5	16	WSG	MVGAGLSIL	IRA	842.06	8.556	24.01%
MHC -II I_Ad	1	43	QLY	NAVITSHAI	MMI	907.03	18.11	34.08%
MHC -II I_Ad	2	166	ISL	EHLSLFVWTV	VFV	1090.28	8.935	16.81%
MHC -II I_Ad	3	157	TVK	NMRSASISL	EHL	960.12	7.307	13.75%
MHC -II I_Ad	4	140	LHC	SGVSSILGG	INF	757.84	5.106	9.61%
MHC -II I_Ad	5	135	LAI	FSLHCSGV	SIL	918.04	4.79	9.01%
MHC -II I_Ad	1	80	APD	MSFPRLNNV	SYW	1059.25	8.379	20.50%

L <sub>Ag</sub> <sub>7</sub>								
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	2	57	IFF	MVMP SLIGG	FGN	886.13	7.943	19.43 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	3	16	WSG	MVGAGLSIL	IRA	842.06	7.45	18.23 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	4	46	NAV	ITSHAIMMI	FFM	998.26	7.024	17.18 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	5	25	SIL	IRAELCKPG	FFF	968.19	6.141	15.02 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	1	107	CLV	DSSCGTSW T	IYP	901.93	17.05	42.73 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	2	65	LIG	GFGNWMVP L	MLG	979.19	16.596	41.59 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	3	11	LIF	GFWSGMVG A	GLS	870.02	11.251	28.20 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	4	37	FFF	GSGQLYNAV	ITS	889.96	9.513	23.84 %
MHC-II <sub>L<sub>Ag</sub><sub>7</sub></sub>	5	79	GAP	DMSFPRLNN	VSY	1075.21	8.672	21.73 %

**Table 3:** Prediction of MHC II ligands all rows highlighted in red represent predicted binders to the MHC-II Allele i.e. MHC-II<sub>L<sub>Ag</sub><sub>7</sub></sub>, MHC-II<sub>L<sub>Ad</sub></sub>, MHC-II<sub>L<sub>Ag</sub><sub>7</sub></sub>. (All rows highlighted in red represent predicted binders).

RANKPEP report PSSM-specific binding threshold and is obtained by scoring all the antigenic peptide sequences included in the alignment from which a profile is derived, and is defined as the score value that includes 85% of the peptides within the set. Peptides whose score is above the binding threshold will appear highlighted in red and peptides produced by the cleavage prediction model are highlighted in violet. We also use a cascade SVM based TAPPred method which found 63 High affinity TAP Transporter peptide regions (Table 4).

Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	187	PVLGAITM	8.644	High
2	17	VGAGLSILI	8.644	High
3	9	IFGFWGMV	8.642	High
4	39	GQLYNAVIT	8.635	High
5	148	GINFMTTVK	8.631	High
6	16	MVGAGLSIL	8.627	High
7	6	MYLIFGFS	8.612	High

8	147	GGINFMTTV	8.61	High
9	103	ACLVDSSCG	8.599	High
10	156	KNMRSASIS	8.596	High

**Table 4:** Cascade SVM based high affinity TAP binders of COX1 from *Dracunculus medinensis*.

This represents predicted TAP binders residues which occur at N and C termini from *Dracunculus medinensis* (cytochrome c oxidase subunit I (mitochondrion)).

## Conclusion

MHC molecules are the cell surface proteins, which actively take part in the host immune responses against pathogens and reason of its involvement in the response to almost all variety of antigens and it gives effects on specific sites. By considering the above result we can conclude that the antigenic peptide that binds to MHC molecule are antigenic that means hydrophilic in nature. This means the increase in affinity of MHC binding peptides may result in enhancement of immunogenicity of *Dracunculus medinensis* antigen cytochrome c oxidase subunit I (mitochondrion) and could be helpful in the designing of synthetic peptide vaccine. This approach can help reduce the time and cost of experimentation for determining functional properties of *Dracunculus medinensis* antigen cytochrome c oxidase subunit I (mitochondrion). Overall, the results are encouraging, both the 'sites of action' and 'physiological functions' can be predicted with very high accuracies which ultimately facilitating the minimization of number of validation experiments.

## References

- Kosakyan A, Heger TJ, Leander BS, Todorov M, Mitchell EA, et al. (2012) COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the *Hyalospheniidae schultze*. Protist 163: 415-34.
- Rumbley J, Gennis RB, Horsman GJA, Barquera B, Ma J (1994) "The superfamily of heme-copper respiratory oxidases". J Bacteriol 176: 5587-5600.
- Hebert PD, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 270 Suppl 1: 96-99.
- Anderson S, Bankier AT, Barrell BG, deBruijn MHL, Coulson AR, et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.
- Wallace DC, Lott MT, Torroni A, Brown MD, Shoffner JM (1994) Report of the committee on human mitochondrial DNA. In: Cuticchia AJ, Pearson PL (eds): Human Gene Mapping, 1993: A Compendium. Baltimore: Johns Hopkins Univ Press (pub.), Pp: 813-845.
- Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Nat Acad Sci 77: 6715-6719.
- Case JT, Wallace DC (1981) Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. Somat Cell Genet 7: 103-108.
- Oliver NA, McCarthy J, Wallace DC (1984) Comparison of mitochondrially synthesized polypeptides of human, mouse, and monkey cell lines by a two-dimensional protease gel system. Somat Cell Molec Genet 10: 639-643.
- Oliver NA, Wallace DC (1982) Assignment of two mitochondrially synthesized polypeptides to human mitochondrial DNA and their use in the study of intracellular mitochondrial interaction. Molec Cell Biol 2: 30-41.



10. Wallace DC, Yang J, Ye J, Lott MT, Oliver NA, et al. (1986) Computer prediction of peptide maps: assignment of polypeptides to human and mouse mitochondrial DNA genes by analysis of two dimensional-proteolytic digest gels. Am J Hum Genet 38: 461-481.
11. Ching E, Attardi G (1982) High-resolution electrophoretic fractionation and partial characterization of the mitochondrial translation products from HeLa cells. Biochemistry 21: 3188-3195.
12. Hare JF, Ching E, Attardi G (1980) Isolation, subunit composition and site of synthesis of human cytochrome c oxidase. Biochemistry 19: 2023-2030.
13. Bernstein C, Facista A, Nguyen H, Zaitlin B, Hassounah N, et al. (2010) Cancer and age related colonic crypt deficiencies in cytochrome c oxidase I. World J Gastrointest Oncol 2: 429-42.
14. Legros F, Malka F, Frachon P, Lombès A, Rojo M (2004) Organization and dynamics of human mitochondrial DNA. J Cell Sci 117: 2653-2662.
15. Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136: 507-13.
16. Satoh M, Kuroiwa T (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp Cell Res 196: 137-40.
17. Menzies RA, Gold PH (1971) The turnover of mitochondria in a variety of tissues of young adult and aged rats. J Biol Chem 246: 2425-2429.
18. Miwa S, Lawless C, von Zglinicki T (2008) Mitochondrial turnover in liver is fast in vivo and is accelerated by dietary restriction: application of a simple dynamic model. Aging Cell 7: 920-923.
19. Ngui R, Mahdy MA, Chua KH, Traub R, Lim YA (2013) Genetic characterization of the partial mitochondrial cytochrome oxidase c subunit I (cox 1) gene of the zoonotic parasitic nematode, Ancylostomaceylanicum from humans, dogs and cats. Acta Trop 128: 154-157.
20. Callejón R, de Rojas M, Ariza C, Ubeda JM, Guevara DC, et al. (2009) Cytochrome oxidase subunit 1 and mitochondrial 16S rDNA sequences of Trichuris skrjabini (Tricocephalida: Trichuridae). Parasitol Res 104: 715-716.
21. Muller R (1971) Dracunculosis and dracunculiasis. Adv Parasitol 9: 73.
22. Muller R (1979) Guinea worm disease: epidemiology, control and treatment. Buel WHO 57: 683.
23. Jones HI, Mulder E (2007) *Dracunculus mulbusn.* sp.(nematode: Spirurida) from water Python Liasis fuscus (sperpentes: Boidae) in northern. Australia. Syst Parasitol 66: 195.
24. Moravec F, Santos CP (2009) *Dracunculus brasiliensis* sp. n. (Nematoda: Dracunculidae) from the anaconda. Eunectes murinus (Ophidia: Boidae). Parasitol Res 104: 589.
25. Bimi L (2005) Guinea worm infection of urinary bladder manifesting as obstructiveuropathy in rural Maharashtra. Trop DOct 35: 242.
26. Greenaway C (2004) Dracunculiasis (Guinea worm disease). CMAJ 170: 495-500.
27. Millner A, Helfer A, Kotlyar D, Oswald J, Efferth T (2011) Chemistry and pharmacology of neglected helminthic disease. Curr Med Chem 18: 767-789.
28. Tiben RE, Hopkins DR (2006) Dracunculiasis (Guinea worm disease) eradication. Adv Parasitol 61: 275-309.
29. Iriemenam NC, Oyibo WA, Beyioku AFF (2008) Dracunculiasis-The saddle is virtually ended. Parasitol Res 102: 343-347.
30. Flower DR (2008) "Vaccines: how they work," in Bioinformatics for Vaccinology. Wiley-Blackwell, Oxford, UK, pp: 73-112.
31. Michael B (1997) Peptide Binding by Class I and Class II MHC Molecules. Biopoly, pp: 281-302.
32. Marrack P, Scott-Browne JP, Dai S, Gapin L, Kappler JW (2008) Review Evolutionarily conserved amino acids that control TCR-MHC interaction. Annu Rev Immunol 26: 171-203.
33. Chapman HA (1998) Endosomal proteolysis and MHC class II function. Curr Opin Immunol 10: 93-102.
34. Watts C (2004) The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. Nat Immunol 5: 685-692.
35. Neeffes J, Jongasma ML, Paul P, Bakke O (2011) Review towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11: 823-36.
36. Kumar M, Gromiha MM, Raghava GP (2007) Identification of DNA-binding proteins using support vector machines and evolutionary profiles. BMC Bioinformatics 8: 463.
37. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res 33: 154-159.
38. Sayers EW (2012) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 40: 13-25.
39. Gill SC, Hippel VPH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem 182: 319-326.
40. Edelhoch H (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6: 1948-1954.
41. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 11: 2411-2423.
42. Ciechanover A, Schwartz AL (1989) How are substrates recognized by the ubiquitin-mediated proteolytic system? Trends Biochem Sci 14: 483-488.
43. Varshavsky A (1997) The N-end rule pathway of protein degradation. Genes Cells 2: 13-28.
44. Guruprasad K, Reddy BVB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4: 155-161.
45. Ikai AJ (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88: 1895-1898.
46. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-132.
47. Hoop TP, Woods KR (1978) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci 78: 3824.
48. Welling GW, Weijer WJ, van der Zee R, WellingWester S (1985) Prediction of sequential antigenic regions in proteins. FEBS Lett 188: 215-218.
49. Parker KC, Bednarek MA, Coligan JE (1994) cheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J. Immunology 152:163-75.
50. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276: 172-174.
51. Gomase VS, Chitlange NR (2012) Prediction of MHC Class Antigen Peptides from *Echinococcus multilocularis*. Application of Computer Intelligence. 1: 191.
52. Gomase VS, Kale KV, Chikhale NJ, Changbhale SS (2007) Prediction of MHC Binding Peptides and Epitopes from Alfalfa mosaic virus. Curr Drug Discov Technol 4: 117-215.
53. Sherkhane AS, Changbhale SS, Chitlange NR, Waghmare S, Gomase VS, et al. (2012) Prediction of Major Histocompatibility Complex Binding Peptides and Epitopes from Najanaja Cardiotoxin (CTX). Drug Invention Today 4: 435-438.
54. Sonu M, Gomase VS (2015) Prediction of antigenic epitope from *D. medinensis*: new paradigm of synthetic vaccine development. International Conference on "Recent Research Development in Environment, Social Sciences and Humanities"(ICRRDESH-15), pp: 103-107.
55. Mishra S, Gomase VS, Kale KV (2016). The Analytical study of antigenicity and Mhc Binder of the Cytochrome C Oxidase Subunit I. IJATES, pp: 585-600.
56. Sonu M, Gomase VS (2015) Analysis of hydrophobicity and antigenic epitope prediction from *D. medinensis*. International Conference on "Technologies for Sustainability-Engineering, information Technology,

- Management and the Environment”(SUSTECH-15), Journal: IJATES, pp: 44-51.
57. Sonu M, Gomase VS (2015) Analysis Of Hydrophobicity And Antigenicity Of Heat Shock Protein 70 From GWD. 2nd International Conference on “Recent Innovations in science, Engineering and Management” (ICRISEM-15), IJATES, pp: 25-33.
58. Mishra S, Gomase VS (2016) Study Hydrophobicity and Antigenicity of Cytochrome C Oxidase Subunit II from *D. medinensis*: New Prototype of Synthetic Vaccine Development. *Immunochem Immunopathol* 2: 113.
59. Mishra S, Gomase VS (2016) Cytochrome B Analysis For Hydrophobicity, Surface Accessibility And Antigenicity. *IJSTM*, pp: 25-35.
60. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, et al. (1993) Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell* 74: 929.
61. Reche PA, Glutting JP, Reinherz EL (2002) Prediction of MHC Class I Binding Peptides Using Profile Motifs. *Human Immunology* 63: 701-709.
62. Reche PA, Reinherz EL (2003) Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol* 331: 623-41.
63. Craiu A, Akopian T (1997) Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *ProcNatlAcadSci U S A* 94: 10850-10855.
64. Pieters J (2000) "MHC class II-restricted antigen processing and presentation." *AdvImmunol* 75: 159-208.
65. Bhasin M, Raghava GPS (2004) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 13: 596-607.
66. Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *JVirol* 55: 836-839.
67. Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins: a tool for the selection of peptide antigen. *Naturwissenschaften* 72: 212-213.
68. Bull HB, Breese K (1974) Surface tension of amino acid solutions: A hydrophobicity scale of the amino acid residues. *Arch Biochem Biophys* 161: 665-670.
69. Roseman MA (1988) Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol* 200: 513-522.
70. Wilson J, Honegger A, Stotzel RP, Hughes GJ (1981) The behaviour of peptides on reverse-phase K. supports during high-pressure liquid chromatography. *Biochem J* 199: 31-41.