European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 08, Issue 02, 2021

In-Silico Insights To Predict The Major Histocompatibility Complex Peptide Binders From Protein

Sonu Mishra[1], Virendra Gomase[2]

[1,2]*Department of Biotechnology, Mewar University, Chittorgarh, Rajasthan, 312901, India*

E-mail : [1]*sonumishra1014@gmail.com,* [2]*gomase.viren@gmail.com*

ABSTRACT

The in-silico method is extensively utilized in the study of proteomics and genomics studies. The T-cell epitopes prediction is essential step in the development of peptide -based vaccines and diagnostic. The epitopes emanate as an emanation of intricate proteolytic mechanism within cell. Proceed to being perceived by T cells, an epitope is presented on the cell surface as a complex with a major histocompatibility complex protein. Henceforth, T-Cell identified epitopes are excellent binder of MHC. Therefore detection and identification of the MHC binders essential for target based study of drug. In recent study, we analyzed D. medinensis antigenic protein peptide binders to MHC-I and MHC-II molecules. The binding with MHC-I molecules are obtained with are 11mer_H2_Db, 10mer_H2_Db, 9mer_H2_Db, 8mer_H2_Db and for MHC-II are as I_Ab.p, I_Ad.p, I_Ag7.p, I_Ak.p .

Keywords- Artificial neural network; machine learning techniques; MHC binders; TAP; PSSM .

1. INTRODUCTION

Humans are exposed to several kinds of infectious agent ranges from viral, fungi, or bacteria. However, not all the individual who exposed to the infection get the disease, only few of them get infected and subsequently develop resistant to that illness, whereas some resist the illness and some acquired chronic infection. The different manifestation because of the host factors which are involved in different responses of immune to the infectious agents. The essential and important one is MHC which differentiates among self and non self agents and triggers a cascade of events to eliminate the agent from host. A set of genes codes for MHC molecules and form MHC complex. A fragments of protein like self protein, pathogenic protein or antigenic protein are taken and presented by MHC molecule and shown it on its surface which is further recognized by the immune system via immune cells and destroyed. Henceforth, identification of the MHC binders is important. To identify the MHC binders we have taken 527 amino acid sequences from the guinea worm disease agent that from *D. medinensis*. This human parasite causes dracunculiasis disease in the human. It completes its incubation within host in very non symptomatic manner for about one a half years or more, which is sometimes cause of the severe motility and morbidity in infected individuals. The intermediate host for this infection is Cyclops that is present in the contaminated water and eats the larvae of the guinea worm and become infectious. Whenever, any individual

consume this infected water gets catch hold of this infection and without any awareness of the host this larvae develop and mature into the adults. A female worm releases eggs into the environment once the blister burst out and causes the initiation of the second cycle of the infection [1-7]. However, Identification antigenic peptide that binds with MHC molecules will ameliorates the apprehension of specificity of immune responses against the pathogen, which could be helpful in vaccine development in forth coming future.

I (a). MHC Class I Antigen

On the surface of the nucleated cells, the presence of MHC class I molecules are seen which display epitopes peptide array for surveillance through the CD^{8+} T cell repertoire. CD^{8+} T cell responses, which are essential for the disease or infection control through efficiently discriminate among the healthy and the infected cell via peptides recognition that are linked with MHC-I (pMHC-I) molecules present on the cell surface. The eight to eleven amino acids length peptide derived in cytosol from protein antigen that arises through commonly cryptic translational reading frames [8]. The cytosolic synthesized protein subjected to proteasomal degradation, the resultant peptide translocated to ER and loaded onto MHC –I molecules [9]. The loading process of peptide onto MHC molecules class I stabilizes and gets enter to the location where the circulating CD8+ T cells are present via cell surface and this complex is referred as 'immune surveillance' [10-19]. TAP binding peptides prediction is for detection and recognition of the MHC class-1 restricted T cell epitopes.

I (b). Proteasomal Degradation

In the process of the antigen-presentation, the proteasomal depravity phase is essential to coordinate the balance among intracellular proteins [20]. Through the activity of the proteinase within the proteosome the protein gets cleaved into oligopeptides [21]. This oligopeptides gets bounded to the TAP, which afterward translocate these peptides into the ER.

I (c).TAP mediated peptide transport into ER (Endoplasmic Reticulum)

The heterodimeric transmembrane protein TAP belongs to ABC transporter protein family which transports antigenic peptide protein into ER [22] on account of majority of the MHC binding peptides are ineffectual to disseminate across membrane, but TAP protein is adequate to transporting these peptide within the ER where it binds to MHC class I molecules and form MHC-peptide complexes. This complex sooner will get translocated to antigen presenting cells surface [23] which are identified via by T-cell receptors that trigger an immune response.

I (d). MHC Class II Antigen

The difficulty encounter while predicting the peptides binding to a MHC class II molecule because of its different side chains and larger length occurrence in the extracellular antigen presentation [24-26]. In the MHC class II antigen presentation process, antigenic protein are firstly took in though antigen-presenting cells via endocytosis or phagocytosis process.Once it is ingested, cathepsins (a class of protease into oligopeptides) acts on it and cleaved in the endosomes , which is than fused with lysosomes that bears MHC class II molecules [27] and present them on the cell surface for realization by T cells [28-36]. The inflammation and swelling due to phagocytes, T cell induces a response of immune via B-Cell activation. Since MHCs have essential part to play in immune system through triggering cellular and humoral immunity against protein from *D. medinensis* and can be applied for insuring particular immunological procedures by producing peptides to get attached to specific MHC alleles and

this linking affinity to can further utilized for designing synthetic peptide vaccines in near future [37-40].

2. MATERIALS AND METHODS

Predictions of MHC class I binding peptide

The neural network trained on C terminal methodology is applied to find the MHC binding peptide of known epitope. RANKPEP prediction tool is used identify peptide binders to MHCI molecules from protein sequence.

Prediction of Antigenic Peptides by Cascade SVM based TAPPred method

TAP binders were predicted through utilizing the specific tool TAPPred, which predict the binders from protein on the basis of the properties of the amino acids and the sequence specifications. The obtained MHC-I binding regions [Table-3] with the binding affinity higher of targeted protein from *D. medinensis* having 527 amino acids, which shows 519 nonamers.

Predictions of MHC class II binding peptide

MHC peptide binding of targeted protein from *D. medinensis* predicted using neural networks trained on C terminals of known epitopes through RANKPEP. We predict peptide binders to MHCII molecules from protein sequences or sequence alignments using PSSMs. MHC molecule binds to some of the peptide fragments generated after proteolytic cleavage of antigen.

3. RESULTS AND INTERPRETATION

In this present work, we predict the peptide binders of targeted protein from D. medinensis sequence to MHC-I molecules are as 8mer_HLA_A0201, 9mer_HLA_A01, 10mer_HLA_A0201, 11mer_HLA_A0201 [Table-1]. The peptide fragment arises from the proteolytic cleavage gets binds to MHC molecule with high-efficiency. The tranlocation take place through TAP transporter from cytosol to ER. TAP binds and translocates selective peptides for binding to specific MHC molecules. Therefore, predicting peptide binding affinity toward the TAP transporter is essential to identify the MHC class-1 restricted T cell epitopes. Cascade based support vector machine shows 160 High affinity TAP binder residues at N and C termini using sequence and properties of the amino acids of protein from *D. medinensis* [Table-3]. This method integrates prediction of peptide MHC class I binding; proteasomal C terminal cleavage and TAP transport efficiency by using sequence and properties of the amino acids. We also found the binding of peptides to different alleles by using Position Specific Scoring Matrix for target protein from *D. medinensis* 527 residues long with 519 nonamers having antigenic MHC binding peptides. PSSM based server will predict the peptide binders of protein sequence from *D. medinensis* to MHCII molecules are as I_Ab.p, I_Ad.p,I_Ag7 which are found antigenic epitopes region in protein [Table-2].

Table -1 Peptide binders of 527 amino acid long protein sequence from D. medinensis to MHC-I molecules, having C-terminal ends are proteosomal cleavage sites. Matrix: 8mer_HLA_A0201.p.mtx; Consensus: ILGIHCWV; Optimal Score: 42.897;Binding Threshold: 20.81, Matrix: 9mer_HLA_A01.p.mtx;Consensus: YTDPGFWYY;Optimal Score: 50.449;Binding Threshold: 25.66; Matrix: 10mer_HLA_A0201.p.mtx; Consensus: WLWWFWWWCV; Optimal Score: 41.367;Binding Threshold: 9.69; Matrix: 11mer_HLA_A0201.p.mtx; Consensus: YLWCWLWWYCV; Optimal Score: 49.99; Binding

Threshold: 20.21 (All rows highlighted in red represent predicted binders &A peptide highlighted in violet has a C-teminus predicted by the cleavage model used).

MHC-I Allele	RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
8mer_H2_Db	1	443	YLF	ILGLVLCV	VFF	811.09	29.396	68.53 %
8mer_H2_Db	2	440	SLF	YLFILGLV	LCV	919.18	17.404	40.57 %
8mer_H2_Db	3	191	LPK	AMSAPTPV	SAL	754.9	16.371	38.16 %
8mer_H2_Db	4	296	LFM	QMGYIIHL	CGG	956.17	14.029	32.70 %
8mer_H2_Db	5	330	TFV	SLMCLCGL	FFL	821.08	13.399	31.24 %
9mer_H2_Db	1	432	SLL	YSDYYSLFY	LFI	1202.31	23.709	47.00 %
9mer_H2_Db	2	70	GFY	MMEDYNFNY	FCV	1208.33	20.472	40.58 %
9mer_H2_Db	3	457	KFG	SFDVKYKFY	GDL	1178.36	18.519	36.71 %
9mer_H2_Db	4	239	LFT	MFSSGLMAY	FEF	988.19	17.688	35.06 %
9mer_H2_Db	5	174	FIL	STFTKSAQY	PFS	1014.1	15.793	31.30 %
10mer_H2_Db	1	11	FGV	ILLCFLLLFV	FFY	1175.59	25.117	60.72 %
10mer_H2_Db	2	442	FYL	FILGLVLCVV	FFK	1057.4	21.532	52.05 %
10mer_H2_Db	3	407	VSL	VLVVFSIVFL	WWL	1117.44	18.206	44.01 %
10mer_H2_Db	4	361	WSL	FLVFLFFFSI	LLT	1261.59	18.042	43.61 %
10mer_H2_Db	5	102	CIS	MLIFWDLLGV	SSY	1165.49	17.84	43.13 %
11mer_H2_Db	1	107	IFW	DLLGVSSYFLV	LYY	1194.4	24.106	48.22 %
11mer_H2_Db	2	54	FFC	LLVMVVGSVVV	YSG	1096.42	21.268	42.54 %
11mer_H2_Db	3	370	FFS	ILLTYLYCYRL	MKG	1415.77	19.589	39.19 %
11mer_H2_Db	4	226	FLD	FVLDFMFFVGL	FTM	1316.63	18.014	36.04 %
11mer_H2_Db	5	69	SGF	YMMEDYNFNYF	CVV	1518.69	17.301	34.61 %

* The RANKPEP consists of a list of selected peptides binding potential (score) to the MHC molecule from the query given at a selected threshold. Peptides shown here contain a C-terminal residue that is predicted to be the result of proteasomal cleavage and also focus on the prediction of conserved epitopes that help to avoid immune evasion resulting from mutation. Proteasomal cleavage options are only applied to the prediction of MHCI-restricted peptides.

Table -2 Cascade SVM based High affinity TAP Binders of 527 amino acid long protein sequence from *D. medinensis*

* TAPPred showing Cascade SVM based High affinity TAP Binders sites, their sequence, rank, position and scores are displayed in the tabular output are to be found 160 High affinity TAP Transporter peptide regions which represents predicted TAP binders residues which occur at N and C termini 527 amino acid long protein sequence from D. medinensis.

Peptide Rank	Start Position	Sequence	Score	Predicted Affinity
1	489	YSIMKFGDF	8.651	High
2	483	SDCMVDYSI	8.648	High
3	161	ELVVSVALF	8.643	High
4	96	SNNCISMLI	8.64	High
5	337	LFFLGGSVS	8.639	High
6	305	GGQQDSRGY	8.636	High
7	262	SQIGFCFFG	8.626	High
8	502	SKIFVMGFS	8.62	High
9	417	WWLNYNSFV	8.616	High
10	217	CFSEVMFLD	8.613	High

Table - 3 Peptide binders of 527 amino acid long protein sequence from *D. medinensis* to MHC-II molecules.
Matrix: I_Ab.p.mtx, Consensus: YYAPWCNNA, Optimal Score: 35.632,Binding Threshold: 9.52; Matrix: I_Ad.p.mtx, Consensus: QMVHAAHAE, Optimal Score: 53.145, Binding Threshold: 7.10 ; Matrix: I_Ag7.p.mtx, Consensus: WYAHAFKYV, Optimal Score: 40.873, Binding Threshold: 7.54. (All rows highlighted in red represent predicted binders & a peptide highlighted in violet has a C-teminus predicted by the cleavage model used).

MHC-II Allele	RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
I_Ab	1	385	KGF	YYYCSSSLF	YSG	1114.3	12.219	34.29%
I_Ab	2	484	KWS	DCMVDYSIM	KFG	1058.3	12.095	33.94%
I_Ab	3	384	MKG	FYYYCSSSL	FYS	1114.3	11.744	32.96%
I_Ab	4	324	SVV	YIQTFVSLM	CLC	1083.3	11.388	31.96%
I_Ab	5	192	PKA	MSAPTPVSA	LVH	841.98	11.24	31.54%
I_Ad	1	89	FVF	SMVGVVFSN	NCI	921.07	16.144	30.38%
I_Ad	2	282	SFI	HMLSHAVFK	SCL	1051.3	11.598	21.82%

I_Ad	3	77	YNF	NYFCVVLSI	FVF	1039.3	7.574	14.25%
I_Ad	4	387	FYY	YCSSSLFYS	GGS	1038.2	7.329	13.79%
I_Ad	5	307	CGG	QQDSRGYVG	VGG	991.03	6.348	11.94%
I_Ag7	1	332	VSL	MCLCGLFFL	GGS	1028.4	10.512	25.72%
I_Ag7	2	33	LDF	SSLELLQFQ	FRL	1046.2	9.939	24.32%
I_Ag7	3	373	ILL	TYLYCYRLM	KGF	1207.5	9.4	23.00%
I_Ag7	4	457	KFG	SFDVKYKFY	GDL	1178.4	9.342	22.86%
I_Ag7	5	71	FYM	MEDYNFNYF	CVV	1224.3	8.271	20.24%

4. CONCLUSION

The protein sequence of 527 amino acid long antigenic proteins from *D. medinensis* were taken for the study and it was observed and noted that a minute fragment of the peptide can trigger immune response compare to the whole protein. We also studied that through increasing in affinity of MHC binding peptides might lead immunogenicity advancement of *D. medinensis* antigenic protein. Henceforth, this study can contribute in planning and formulating new synthetic vaccines with high quality and efficiency to cure the illness. Finally, precise prediction remains essential for the future designing of peptide synthetic vaccine. The overall conducted study is quite encouraging and predictions of MHC binding peptides remain essential step in vaccine formulation, and could be helpful in future investigation in cellular immunology, immune diagnostic methodology therapies and susceptibility of autoimmunity area.

Abbreviations –
MHC - Major Histocompatibility Complex
TAP -Transporter associated with antigen processing
PSSM- Position Specific Scoring Matrices
SVM - Support Vector Machine

Conflict on Interest - None

REFERENCES

[1] Miillner A, Helfer A ,Kotlyar D. Oswald J, Efferth T (2011). Chemistry and pharmacology of neglected helminthic disease.Curr Med Chem., 18(5), 767-789.
[2] Ruiz-Tiben E, Hopkins D R. (2006). Dracunculiasis (Guinea worm disease) eradication . Adv Parasitol., 61, 275-309.
[3] Iriemenam N C, Oyibo WA, Fagbenro-Beyioku A F. (2008). Dracunculiasis – The saddle is virtually ended. Parasitol Res., 102(3), 343-347.
[4] Silkjaer T, Nyvold C G, Juhl-Christensen C, Hokland P, Nørgaard J M. (2013) .Mitochondrial cytochrome c oxidase subunit II variations predict adverse prognosis in cytogenetically normal acute myeloid leukaemia. Eur J Haematol., 91(4), 295-303.
[5] Hamblet N S, Ragland B, Ali M, Conyers B, Castora F J. (2006). Mutations in mitochondrial-encoded cytochrome c oxidase subunits I, II, and III genes detected in Alzheimer's disease using single-strand conformation polymorphism. Electrophoresis. 27(2), 398-408.

[6] Nwoke B E. (1992). Behavioral aspects and their possible uses in the control of dracunliasis (guinea-worm) in Igwun river basin area of Imo State, Nigeria. Angew. Parasitol., 33, 205–210.
[7] Muller R.1985. Life cycle of DracunculusMedinesis .In workshop on opportunities for control of dracunculiasis : contaminated papers, washinton, DC:National Academy Press.
[8] Shastri N, Schwab S, Serwold T. (2002). Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules. Annu Rev Immunol., 20: 463–493.
[9] Blum J, Wearsch P, Cresswell P. (2013) Pathways of antigen processing. Annu Rev Immunol., 31: 443–473.
[10] Rock K, Gramm C, Rothstein L, Clark, Stein R, Dick L, Hwang D, Goldberg AL. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell. 78:761–771.
[11] Bhasin M, Raghava G P.(2005). Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences Nucleic Acids Res., 33: W202-207.
[12] Vyas J M, Van der Veen AG &Ploegh H L. (2008).The known unknowns of antigen processing and presentation. Nat Rev Immunol., 8(8): 607-618.
[13] Kelly A, Powis S H, Kerr L A, Mockridge I, Elliott T, Bastin J, Uchanska-Ziegler B, Ziegler A, Trowsdale J, Townsend A. (1992). Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex. Nature. 13; 355(6361):641-4.
[14] Lautscham G, Rickinson A, Blake N. (2003). TAP-independent antigen presentation on MHC class I molecules: Lessons from Epstein--Barr virus. Microbes and Infection., 5:291–299.
[15] Nussbaum A K, Kuttler C, Tenzer S & Schild H. (2003). Using the World Wide Web for predicting CTL epitopes. Curr. Opin. Immunol., 15: 69–74.
[16] Lankat-Buttgereit B & Tampe R. (2002). The transporter associated with antigen processing: Function and implications in human diseases. Physiol. Rev., 2002; 82: 187–204.
[16] Rock KL, York IA & Goldberg AL.(2004). Post-proteasomal antigen processing for major histocompatibility complex class I presentation. Nature immunology. 5(7): 670-677.
[17] Hammerling G J, Vogt A B & Kropshofer H. (1999). Antigen processing and presentation: Towards the millennium. Immunol. Rev., 172: 5–11.
[19] Bhasin M, Ragahava G P S. (2007). A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. J Biosciences., 32(1):31–42.
[20] Brussic V, Bajic V B, Petrovsky N. (2004). Computational methods for prediction of T-cell epitopes--a framework for modelling, testing, and applications. Methods. 34:436–443.
[21] Reidesel H, Kolbeck B, Schmetzer O, Knapp E W. (2004). Peptide binding at class I major histocompatibility complex scored with linear functions and support vector machines. Genome Informatics. 15(1):198–212.
[22] Princiotta M F, Finzi D, Qian S B, Gibbs J, Schuchmann S, Buttgereit F, *et al*., (2003). Quantitating protein synthesis,degradation, and endogenous antigen processing. Immunity. 18(3): 343-354.
[23] Rock K L, York I A & Goldberg A L. (2004). Post-proteasomal antigen processing for major histocompatibility complex class I presentation. Nat Immunol., 5(7):670-677.

[24] Procko E & Gaudet R. (2009). Antigen processing and presentation: TAPping into ABC transporters. CurrOpinImmunol., 21(1):84-91.
[25] Yewdell J W, Reits E & Neefjes J. (2003) Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol. 3(12): 952-961.
[26] Stern LJ, Wiley DC. (1994). Antigenic peptide binding by class I and class II histocompatibility proteins. Behring Inst Mitt., (94):1-10.
[27] Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F. (1994) Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. J Exp Med., 180(6):2353-2358.
[28] Jardetzky T S, Brown J H, Gorga J C, Stern L J, Urban R G, Strominger J L, Wiley D C. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. Proc. Natl Acad. Sci., USA. 93: 734–738.
[29] Cresswell P. Assembly, transport, and function of MHC class II molecules. (1994). Annu Rev Immunol.,12: 259–293.
[30] Rudolph M G, Stanfield R L, Wilson I A. (2006). How TCRs bind MHCs, peptides, and coreceptors. Annu Rev Immunol., 24: 419–466.
[31] Nielsen M, Lund O, Buus S, Lundegaard C. (2010). MHC Class II epitope predictive algorithms. Immunology. 130:319–328.
[32] Nielsen M, Lundegaard C, Worning P, Hvid C S, Lamberth K, *et al*. (2004). Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. Bioinformatics. 20:1388–1397.
[33] Murugan N, Dai Y (2005). Prediction of MHC class II binding peptides based on an iterative learning model. Immunome Res., 1:6.
[34] Salomon J, Flower D R. (2006). Predicting Class II MHC-Peptide binding: a kernel based approach using similarity scores. BMC Bioinformatics. 7:501.
[35] Bordner A J, Mittelmann H D. (2010). Prediction of the binding affinities of peptides to class II MHC using a regularized thermodynamic model. BMC Bioinformatics. 11:41.
[36] Nielsen M, Lundegaard C, Lund O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics. 8:238.
[37] Gomase V S, Chitlange N R. (2012). Sensitive Quantitative Predictions of MHC Binding Peptides and Fragment Based Peptide Vaccines from Taeniacrassiceps. J Vaccines Vaccin., 3:131.
[38] Changbhale S S, Chitlange N R, Gomase V S, Kale K V. (2012). An Immunoinformatics Approach to Design Synthetic Peptide Vaccine from Dendroaspispolylepispolylepis Dendrotoxin-K(DTX-K). Journal of Environmental & Analytical Toxicology. 2(7), 157.
[39] Reche PA, Glutting JP, Zhang H, Reinherz EL.(2004). Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics. 56(6):405–419.
[40] Gomase V S, Chitlange N R, Changbhale S S, Kale K V. (2013). Prediction of Brugiamalayi antigenic peptides: candidates for synthetic vaccine design against lymphatic filariasis. Protein Pept Lett., 20(8):864-87.