

# Knowledge based threading approach to model the MHC-binders to MHC class I alleles in Hepatitis E Virus: a clue for epitope-based vaccine design

B Rathi, A Sarangi, S Naik

## Citation

B Rathi, A Sarangi, S Naik. *Knowledge based threading approach to model the MHC-binders to MHC class I alleles in Hepatitis E Virus: a clue for epitope-based vaccine design*. The Internet Journal of Medical Informatics. 2009 Volume 5 Number 2.

## Abstract

Identification of cytotoxic T lymphocytes (CTL) is a crucial step for designing a peptide based vaccine. T-cell recognition of the peptide-MHC complex is a pre-requisite for initiating the immunological events which leads to the immune response. In the present study CD8+ T-cell epitope for capsid protein of hepatitis E virus (HEV) were identified and knowledge based threading approach has been implemented for MHC binding peptides. The study demonstrated structural analysis of interactions in protein-peptide complexes can lead to insight into the mode of substrate recognition. The compatibility of the peptide to bind to MHC complex is evaluated by pairwise amino acid contact potential. Threading approach has been applied for peptide binding to two MHC class I alleles for the capsid protein of HEV whose crystal structure is available in Protein Data Bank. The threading approach could be extended to screen a large number of MHC alleles for the prediction of T-cell epitopes using data generated for molecular modeling of peptide MHC complex.

## INTRODUCTION

Hepatitis E virus (HEV) infection is a major cause of acute viral hepatitis in several developing countries, in particular those in South and Southeast Asia, North Africa, the Middle East, etc [1,2]. Infection is usually self-limiting and chronicity is not known. The genome of HEV is of 7.5 Kb having three overlapping reading frames. ORF1 (5 Kb) is located at the 5' end of the genome and encodes a polyprotein of 1690 amino acids. ORF2 doesn't overlap ORF1 and is located at the 3' end of the genome and encodes for the principal and probably only structural protein, the capsid protein of 660 amino acids. ORF3 extensively overlaps ORF2 and the shortest ORF a small immunogenic encoding 123 amino acids phosphoprotein.

T-cell immunemhc-tbl response may be helpful in providing protection against viruses. Thus, viral proteins that contain T-cell epitopes may prove useful as potential vaccines, by stimulating T-cell immunity and by providing T-cell help for antibody production. Further, activation of T-cell immune response may down regulate viral replication through a cytokine-mediated, noncytolytic pathway, as has been shown to occur in hepatitis B and C virus infections [3].

Studies have demonstrated that structural analysis of interactions in protein-peptide complexes can lead to novel insight into the mode of substrate recognition. Molecular modeling of peptide-MHC and peptide-kinase interactions have been carried out by several groups using ab initio docking [4] and MD simulation approach [5]. However, the compute intensive nature of these calculations has limited such studies to few protein-peptide complexes. Since knowledge-based methods are less compute intensive, development of suitable knowledge-based tools for modeling protein-peptide complexes would permit quick structural analysis of MHCs with their substrate peptides.

In the current study, CD8+ T-cell epitopes were identified by neural network approach implemented in NetMHCpan [6], knowledge based threading approach is applied by using MODPROPEP web server [7] and the potential interaction energies [8] of the epitopic sequence with specific MHC molecule were evaluated using statistical pairwise contact potentials, MJ [9]

## MATERIAL & METHODS

### RETRIEVAL OF THE PROTEIN SEQUENCES OF

## **HEV**

The HEV capsid protein is downloaded from National centre for biotechnology ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). It is the antigenic peptide evaluated for the binders and non-binders of MHC.

## **CD8+ T-CELL EPITOPE PREDICTION**

NetMHCpan server [6] is one of the most accurate prediction servers currently available and has been used for the prediction of CD8+ T-cell epitopes. The server predicts binding for more than 80 MHC molecules using artificial neural networks (ANNs). Artificial neural networks (ANNs) take into account the identity of each amino acid residue as well as the interactions between adjacent amino acids in a potential epitope. ANN for a particular MHC molecule is trained to recognize a peptide sequence as input and the binding affinity for the sequence with the MHC molecule as output. Once an ANN is trained for a particular molecule, it can predict the binding affinity of novel peptide sequences.

## **KNOWLEDGE BASED THREADING FOR MHC-BINDERS**

The CD8+ T-cell epitopes were predicted through NetMHCpan. The MHC-binders obtained through the neural network approach were used for the knowledge based threading. The CD8+ T-cell epitopic peptides of HEV capsid protein were threaded through the backbone coordinates of known peptide fold in the MHC groove, The amino-acid sequence of the peptide is threaded onto the coordinates of the peptide in the template using MODPROPEP web server [7] and their interaction energies are evaluated using statistical pairwise contact potentials. Threading of the peptides in MHC groove was done for two MHC class I HLA-A alleles i.e. HLA A\*0201, HLA A\*6801 whose crystal structure for protein-peptide interaction information available in Protein Data Bank.

In this method, binding affinity of a peptide is predicted by

the total energy of interaction with contact residues. The nearest atom  $< 4 \text{ \AA}$  criterion was used to determine the contacts of the peptide in the available template co-crystal structure [10], because the nearest atom  $< 4 \text{ \AA}$  distance criterion to determine the contacting residues gives a better prediction compared to C-beta  $< 7.0 \text{ \AA}$  distances [11]. Energy values for amino acid-to-amino acid interactions were taken from the table of statistical pairwise contact potentials derived by MJ.

In case of non availability of crystal structures for a given MHC protein, the program can model its structure in complex with peptides of desired sequence using the crystal structure of the closest homologous protein-peptide complexes.

In order to analyze the interactions between the peptide and the protein, the residues of the MHC, which are in contact with different side chains of the modeled peptide, were identified by using a distance-based cutoff. Based on these contact residues, putative binding pockets were defined for each of the residues in the peptide. Jmol java applet interface was implemented for the visualization of binding pockets in modeled complex.

## **RESULTS & DISCUSSION**

The CD8+ T-cell epitopic sequence of the capsid protein of HEV were threaded into the crystal structure of MHC class I peptide complexes. Two MHC class I alleles i.e. HLA A\*0201, HLA A\*6801 were screened for identification of the MHC binders using NetMHCpan web server. The statistical potential matrices MJ was used to obtain an estimate of the binding affinity of the threaded T-cell epitopes of HEV to the MHC class I alleles. Table 1 gives the list of the T-cell epitopes identified as weak or strong binders of MHC molecule by NetMHCpan and the binding affinities predicted by MJ algorithm for distance criteria i.e. Nearest atom  $< 4.0 \text{ \AA}$  to define the contacting residues.

**Knowledge based threading approach to model the MHC-binders to MHC class I alleles in Hepatitis E Virus: a clue for epitope-based vaccine design**

**Figure 1**

Table 1. List of binders identified by NetMHCpan and prediction of binding energies by threading approach using scoring matrix MJ with distance criterion nearest atom < 4 Å.

MHC Class I allele/peptide length	Position of amino acids	Binders predicted by NetMHCpan	CD8+ T-Cell Epitopes	Miyazawa–Jernigan (MJ) Nearest atom < 4.0 Å	
HLA-A*0201/9	6-14	Weak Binder	ILLLLMFL	-220.54	
	9-17	Weak Binder	LLLMFLPML	-198.26	
	11-19	Weak Binder	LMFLPMLPA	-165.87	
	45-53	Weak Binder	RVDSQPF AI	-132.40	
	48-56	Weak Binder	SQPF AIPYI	-124.29	
	51-59	Weak Binder	FAIPYIHPT	-133.30	
	55-63	Weak Binder	YIHPTNPFA	-129.31	
	62-70	Weak Binder	FAPDV TAAA	-93.89	
	146-154	Weak Binder	SVATGTNLV	-121.55	
	152-160	Weak Binder	NLVLYAAPL	-145.12	
	156-164	Weak Binder	YAAPLSPLL	-135.62	
	276-284	Weak Binder	VMLCIHGSL	-171.2	
	277-285	Weak Binder	MLCIHGSLV	-165.86	
	296-304	Weak Binder	ALGLLDFAL	-165.87	
	298-306	Strong Binder	GLLDFALEL	-168.41	
	349-357	Weak Binder	FMKDLYF TS	-159.42	
	367-375	Weak Binder	GIALTLFNL	-168.34	
	371-379	Weak Binder	TLFNLADTL	-152.28	
	379-387	Weak Binder	LLGGLPTEL	-154.51	
	393-401	Weak Binder	GQLFY SRPV	-119.40	
	394-402	Weak Binder	QLFY SRPVV	-156.08	
	411-419	Strong Binder	KLYTSVENA	-117.89	
	469-477	Weak Binder	VLWLSLTAA	-169.04	
	492-500	Strong Binder	YVSDSVTLV	-142.95	
	498-506	Weak Binder	TLVNVATGA	-106.22	
	512-520	Weak Binder	SLDWTKVTL	-147.50	
	529-537	Strong Binder	QQYSKTFV	-150.50	
543-551	Weak Binder	KLSFWEAGT	-110.70		
579-587	Weak Binder	AISTYTTSL	-132.40		
586-594	Weak Binder	SLGAGP VSI	-130.68		
599-607	Strong Binder	VLAPHSALA	-128.86		
	<b>IAKJ*</b>		ILKEPVHGV	-151.89	
HLA-A*6801/10	66-75	Strong Binder	VTAAAGAGPR	-106.1	
	90-99	Weak Binder	QAQRPAVASR	-90.26	
	124-133	Weak Binder	DVDSRGAILR	-126.89	
	173-182	Weak Binder	HIMATEASNY	-152.86	
	180-189	Weak Binder	SNYAQYRVAR	-124.79	
	186-195	Weak Binder	RVARATIRYR	-130.53	
	224-233	Weak Binder	MNSITSTDVR	-109.52	
	241-250	Weak Binder	ASELVIPSER	-112.41	
	245-254	Weak Binder	VIPSERLHYR	-155.58	
	310-319	Weak Binder	NLTPGNTNTR	-116.71	
	319-328	Weak Binder	RVSRYSTAR	-128.88	
	323-332	Strong Binder	YSSTARHRLR	-129.59	
	324-333	Weak Binder	SSTARHRLRR	-112.47	
	342-351	Strong Binder	TTAATRFMK	-120.01	
	357-366	Strong Binder	STNGVGEIGR	-107.34	
	390-399	Weak Binder	SAGGQLFYSR	-117.61	
	503-512	Weak Binder	VATGAQAVAR	-111.27	
	509-518	Weak Binder	AVARSLDWTK	-141.87	
	535-544	Weak Binder	IFFVLPLRGK	-155.81	
	545-553	Weak Binder	LSFWEAGTTK	-131.51	
	640-649	Weak Binder	FQSTVAELQR	-138.49	
	642-651	Weak Binder	STVAELQRLK	-132.48	
		<b>ITMC*</b>		EVAPPEYHRK	-121.57

Denotes PDB ID of Template for MHC-peptide complex used in study.

The pair-wise potential helps to calculate the binding energies of the HEV peptide sequence upon the different structural templates of the MHC class I alleles. MJ pair-wise contact potential emphasis more on hydrophobic interaction for the MHC alleles that contain various pockets of hydrophobic characters. As most of the peptide are buried within the binding groove of the MHC molecule, threading approach is dependent on the template structure used, as the peptide would be more efficient binder if the binding pattern of the epitopic peptide is similar to the binding pattern of the template peptide [11].

The epitopic sequence QQYSKTFFV was predicted to be a strong binder with HLA-A\*0201 using NetMHCpan. Threading of the above epitopic sequence into the groove of the same HLA revealed the binding energy to be -150.50 which is nearly equivalent with that of the template peptide of MHC-peptide complex (1AKJ) i.e -151.89. The second epitopic sequence TTTAATRFMK was also predicted to be a strong binder for HLA-A\*6801 with binding energy -120.01 and the template peptide 1TMC has the binding energy of -121.57. The similar binding energy patterns reveal the fact that these epitope sequences efficiently fits in the volume defined by the MHC groove. The affinity of the peptide for the MHC complex depends on how efficiently the peptide fits in the space provided by the binding groove. This observation gives a clear picture that the fit of a given peptide to the conformations accessible in the bound form is an important factor to determine its binding affinity. The results of threading are very much dependent on the template structure used, as a peptide ranks high if its binding scheme is similar to the template peptide [11]. The peptides identified by threading approach could be used as potential vaccine candidates. It is important to validate these epitopic peptides for their immunogenicity in vitro.

## CONCLUSIONS

The knowledge based threading approach is advantageous for MHC alleles that lack binding data but have a solved structure in complex with a peptide, or alternatively, a structural model of the complex based on known structures. In such cases, the threading approach succeeds satisfactorily to narrow down the range of peptides that need to be tested for identifying the immunogenic ones. The threading approach can be extended to screen a large number of MHC alleles for the prediction of T-cell epitope generated by molecular-modeling of protein-peptide complex.

## ACKNOWLEDGEMENT

The Bioinformatics Facility at our institution, where the work was done, is supported by the Indian Council of Medical Research, New Delhi and the Department of Biotechnology, Government of India, New Delhi.

## References

1. Aggarwal R, Krawczynski K. Hepatitis E: an overview and recent advances in clinical and laboratory research. *J Gastroenterol Hepatol.*, 15: 9-20 (2000).
2. Krawczynski K, Aggarwal R. Hepatitis E. In: Schiff ER, Sorrell MF, Maddrey WC, eds. *Schiff's Diseases of Liver*, 9th edn. Philadelphia, PA: Lippincott-Raven., 877-889 (2002).
3. Ferrari c, Missale G, BoniC, Urbani S. Immunopathogenesis of hepatitis B. *J Hepato.*, 39: 67-73 (2003).
4. Tong, J.C., Tan, T.W. and Ranganathan, S. (2004) Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci.*, 13, 2523-2532.
5. Pohlmann, T., Bockmann, R.A., Grubmuller, H., Uchanska-Ziegler, B., Ziegler, A. and Alexiev, U. (2004) Differential peptide dynamics is linked to major histocompatibility complex polymorphism. *J. Biol. Chem.*, 279, 28197-28201.
6. netmhcpn
7. N. Kumar and D. Mohanty, *Nucleic Acids Res.*, 35: W549 (2007)
8. U. Sezerman, et al., *Protein Sci.*, 5: 1272 (1996)
9. S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.*, 256: 623 (1996)
10. D. R. Madden et al., *Cell*, 75: 693 (1993)
11. S.P. Singh and B.N. Mishra, *Bioinformation* 3(2): 72-82 (2008)

**Author Information**

**Bhawna Rathi**

Biomedical Informatics Centre, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow 226014, India

**Aditya N Sarangi**

Biomedical Informatics Centre, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow 226014, India

**Sita Naik**

Biomedical Informatics Centre, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow 226014, India