

# Prediction of Major Histocompatibility Complex Binding Peptides and Epitopes from Fatty-Acid-Binding Protein of the Human Blood Fluke *Schistosoma Japonicum*

Somnath Waghmare<sup>1\*</sup> and Ramrao Chavan<sup>2</sup>

<sup>1</sup>Department of Zoology, Nowrosjee Wadia College of Arts and Science, Pune, India

<sup>2</sup>Department of Zoology, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad, India

## Abstract

*Schistosoma japonicum* are blood flukes of humans that cause chronic, highly debilitating diseases involving extensive liver damage. In the present study, fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum* is being used to find out highly suitable MHC binding peptides and epitopes. MHC molecules are cell surface proteins, which take active part in host immune reactions and involvement of MHC class I in response to almost all antigens and it give effects on specific sites. Predicted MHC binding regions acts like red flags for antigen specific and generate immune response against the parent antigen. So a small fragment of antigen can induce immune response against whole antigen. This theme is implemented in designing subunit and synthetic peptide vaccines. Fragments identified through this approach tend to be high efficiency binders, in which larger percentage of their atoms are directly involved in binding as compared to larger molecules. Binding ability prediction of peptides to major histocompatibility complex (MHC) class I & II molecules is important in vaccine development from fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum*.

**Keywords:** Epitopes; Fatty-acid-binding protein; *Schistosoma japonicum*; MHC binders; Support vector machine

**Abbreviations:** MHC: Major Histocompatibility Complex; PZQ: Praziquantel; TAP: Transporter Associated with Antigen Processing; SVM: Support Vector Machine; ANN: Artificial Neural Networks

## Introduction

Schistosomiasis, caused by trematode blood flukes of the genus *Schistosoma*, is recognized as the most important human helminth infection in terms of morbidity and mortality. Infection follows direct contact with freshwater harbouring free-swimming larval (cercaria) forms of the parasite. Despite the existence of the highly effective antischistosome drug praziquantel (PZQ), schistosomiasis is spreading into new areas, and although it is the cornerstone of current control programs, PZQ chemotherapy does have limitations. In particular, mass treatment does not prevent reinfection. Furthermore, there is increasing concern about the development of parasite resistance to PZQ. Consequently, vaccine strategies represent an essential component for the future control of schistosomiasis as an adjunct to chemotherapy. An improved understanding of the immune response to schistosome infection, both in animal models and in humans, suggests that development of a vaccine may be possible [1]. Approximately 200 million people in 74 countries are infected with schistosomes; 120 million are symptomatic, and 20 million suffer severe illness [2,3].

Schistosomiasis is the most important human helminth infection in terms of morbidity and mortality; a recent meta-analysis assigned 2 to 15% disability weight to the disease [4]. There is also emerging evidence that schistosome infections may impact the etiology and transmission of human immunodeficiency virus/AIDS (HIV/AIDS) [5-7]. In particular, the possible interaction between schistosomiasis and HIV/AIDS is receiving increasing attention, given the role of immune responses in both diseases and the geographic overlap in distribution; low CD4<sup>+</sup> T-cell counts resulting from HIV infection may increase susceptibility to schistosome infection and influence egg excretion [8-10]. Thus, schistosomiasis imposes a high socioeconomic burden on many affected developing countries.

MHC molecules have been well characterized in terms of their role in immune reactions. They bind to some of the peptide fragments generated after proteolytic cleavage of antigen [11]. These binding sites are antigen specific and generate immune response against the parent antigen. Prediction methods to find out the small peptides fragments from a protein which may represent the whole protein and excite the immune response are available [12]. The present paper deals with the possibilities of exploiting fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum* to find out the highly suitable MHC binding peptide and have high affinity to TAP binding peptides that can be used for inducing cross protection and as immunogen to produce antiserum for the development of sero-diagnostics for *Schistosoma japonicum*.

## Materials and Methods

### Protein sequence used

For recognition of immunologically relevant regions, hydrophilicity, antigenicity, solvent accessible regions and MHC class peptide binding of the fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum* (GenBank accession no. AAA64426.1) had been considered.

**\*Corresponding author:** Dr. Somnath Waghmare, Assistant Professor, Department of Zoology, Nowrosjee wadia College of Arts and Science, Pune-1, India, Tel: +91 9881926518; E-mail: [drsomnathwaghmare@gmail.com](mailto:drsomnathwaghmare@gmail.com)

**Received** May 07, 2012; **Accepted** June 21, 2012; **Published** June 23, 2012

**Citation:** Waghmare S, Chavan R (2012) Prediction of Major Histocompatibility Complex Binding Peptides and Epitopes from Fatty-Acid-Binding Protein of the Human Blood Fluke *Schistosoma Japonicum*. Metabolomics 2:113. doi:10.4172/2153-0769.1000113

**Copyright:** © 2012 Waghmare 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.

## Prediction of secondary structure of protein and its antigenicity

The secondary structure diagram based on Garnier algorithm provides additional information about possible sequence accessibility [13]. The aim of secondary structure prediction is to provide the location of alpha helices, and beta strands within a protein or protein family. Residue conformational propensities, sequence edge effects, moments of hydrophobicity, position of insertions and deletions in aligned homologous sequence, moments of conservation, auto-correlation, residue ratios, secondary structure feedback effects, and filtering [12,14] are the important concepts involved in secondary structure prediction.

Antigenicity prediction tools adopted in this study predict those segments from fatty-acid-binding protein that are likely to be antigenic by eliciting an antibody response using Hopp and Woods [15], Welling [16], Parker [17], B-EpiPred Server [18] and Kolaskar and Tongaonkar [19].

### Targeting the location in solvent accessible regions

Protein antigenicity is a surface property. Antigenic epitopes can be located as those segments of primary structure that are markedly hydrophilic [15]. Hydrophilicity plots provide a measure of distribution of polar and apolar amino acid residues within the protein sequence. The Kyte-Doolittle scale [20] provides a measure of hydrophobicity with each amino acid. Similarly Hopp-Woods scale was used to predict potential antigenic sites. This may be useful in predicting membrane-spanning domains, potential antigenic sites and regions that are likely exposed on the protein surface [21-24].

### Prediction of MHC binding peptide

Prediction methods for identifying binding peptides could minimize the number of peptides required to be synthesized and assayed, and thereby facilitate the identification of potential epitopes [12]. Several methods have been used to predict MHC binding peptides, including those based on binding motifs [25,26], quantitative matrices [27], artificial neural networks (ANNs) [28-30] and support vector machine (SVM) [27,31,32]. Binding motifs specify which residues at given positions within the peptide are necessary or favourable for binding to a specific MHC molecule [33]. In this study, prediction of MHC peptide binding is performed using neural networks trained on C terminals of known epitopes. An elegant machine learning technique i.e. SVM based method is used for prediction of promiscuous MHC class II binding peptides. In SVM based method, the average accuracy is reported to be high as compared to other methods since SVM can handle noise or non linearity in data very well [27,32,34]. The predicted peptides from fatty-acid-binding protein under study and their affinity to TAP binding peptides are determined by the scoring based on the average score / affinity of an amino acid at particular position and calculated as follows:  $A_{i,r}$  = Average affinity of peptides having residues  $r$  in position  $i$ , where  $A_{i,r}$  is the matrix entry of residue  $r$  in position  $i$ ,  $r$  may be any natural amino acid and  $i$  varies from 1 to 9 [27].

## Results

Study refers to the Sj-FABPc fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum* having 132 aa protein described under material and methods.

### Determination of antigenic peptides

Parameters such as hydrophilicity, flexibility, accessibility, turns,

exposed surface, polarity and antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes. Hydrophobicity (or hydrophilicity) plots are designed to display the distribution of polar and non-polar residues along a protein sequence. In our study, antigenic determinants have been targeted by locating the positive peaks in hydrophilicity plots, thus identifying the regions of maximum potential antigenicity. Hopp-Woods scale [15] was used for predicting potential antigenic sites of protein which is essentially a hydrophilic index, with non-polar residues assigned negative values (Figure 1). Welling antigenicity plot [16] gives antigenicity value as the log of the quotient between percentage in a sample of known antigenic regions and percentage in average proteins (Figure 2). Parker [17], Kolaskar and Tongaonkar antigenicity methods [19] and B-EpiPred Server [18] were also studied (Figures 3-5).

### Secondary alignment

For the protein under study, secondary structure was predicted using Garnier-Osguthorpe-Robson (GOR) method [13]. It assumes that the amino acids flanking the central amino acid also influence the secondary structure. Values for alpha helix, beta sheet, turns and

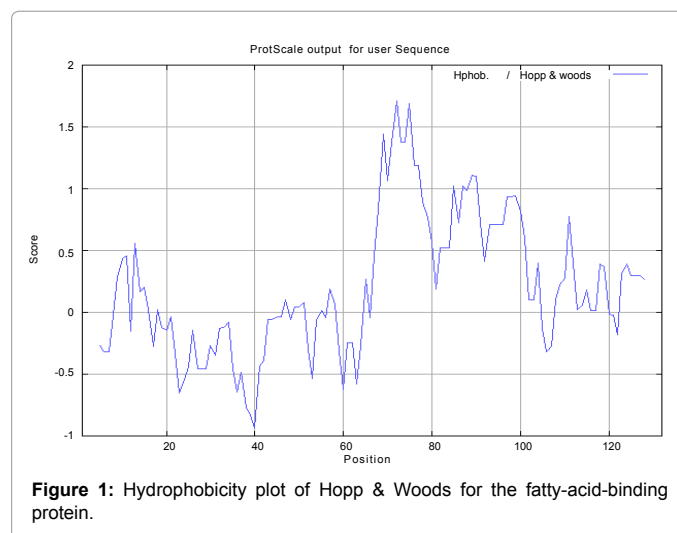


Figure 1: Hydrophobicity plot of Hopp & Woods for the fatty-acid-binding protein.

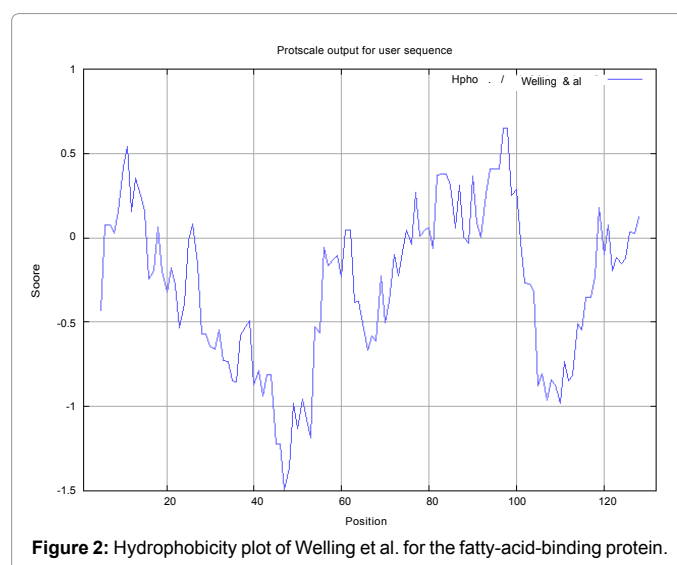
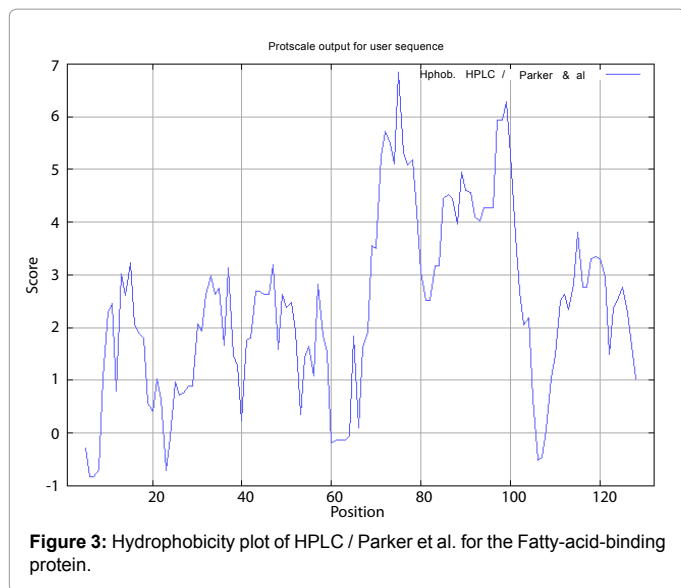


Figure 2: Hydrophobicity plot of Welling et al. for the fatty-acid-binding protein.



of the four possible conformations alpha, beta, reverse turn or coils can be calculated and the conformation with the largest likelihood may be assigned to the residue.

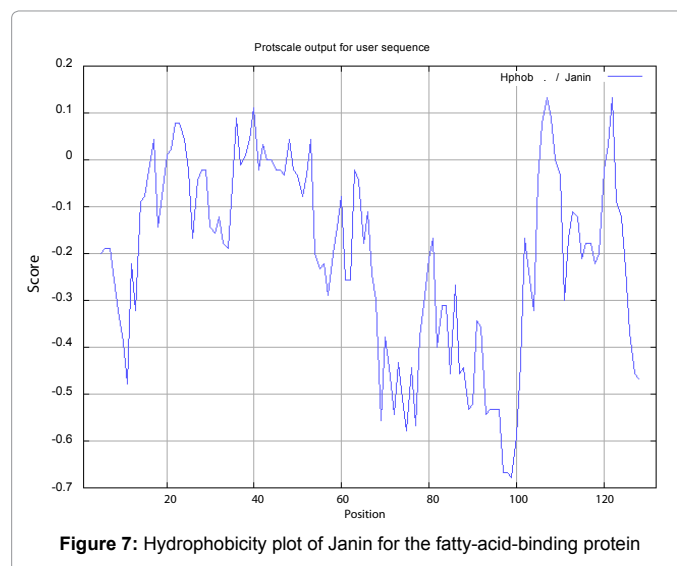
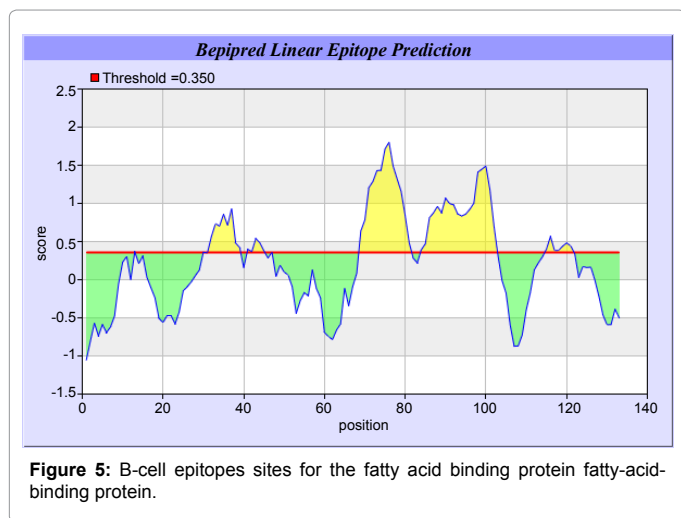
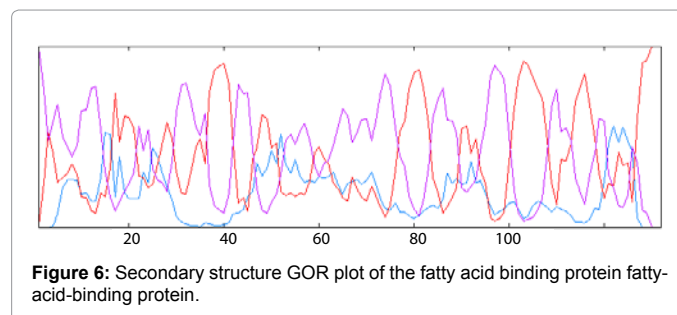
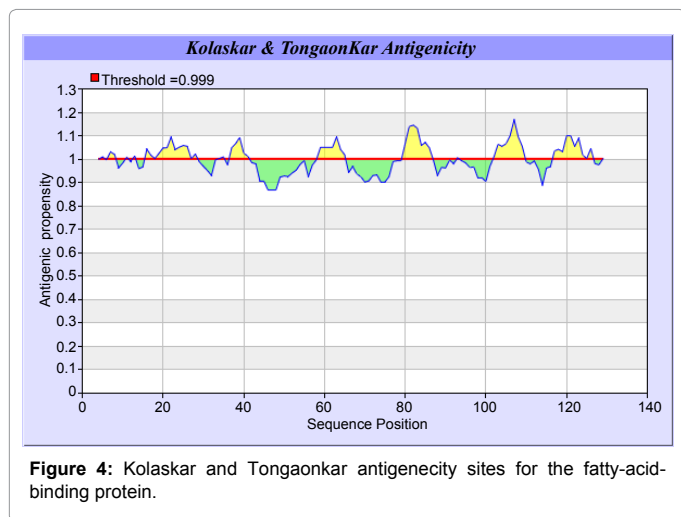
### Solvent accessible regions

To predict potential antigenic sites of globular proteins, which are likely to be rich in charged and polar residues, solvent accessible scales are developed which delineate hydrophobic and hydrophilic characteristics of amino acids. The protein under study was exposed to Janin, Kyte & Doolittle, Abraham & Leo and Bull & Breese methods to predict its nature and prediction flexibility (Figures 7-10).

### Determination of MHC binding peptides

The binding between peptide epitopes and MHC protein(s) is an important event in the cellular immune response. SVMs are a class of learning based on non-linear modelling techniques with proven performance in a wide range of practical applications [35]. The prediction method used in our study is based on this elegant machine learning technique. The cascade support vector machine approach based on amino acid sequence and properties was used to predict MHCI and MHCII binding regions. In this assay, prediction of the binding affinity of fatty-acid-binding protein having 132 amino acids, showing 124 nonamers was performed. SVM was trained on the binary input of single amino acid sequence. The binding regions obtained are reported in Table 1 and Table 2.

Fifty Eight peptide regions were found to have high affinity to TAP binding peptides. The data presented in Table 1 showed top



coils are assigned for each residue (Figure 6). With the aid of these information parameters, likelihood of a given residue assuming each

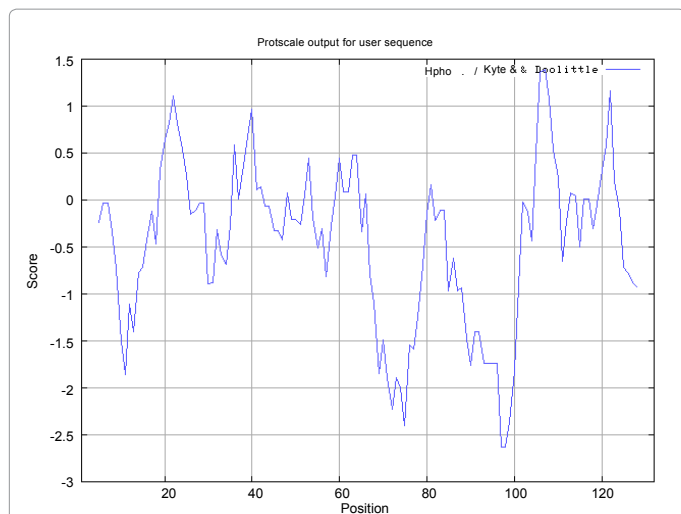


Figure 8: Hydrophobicity plot of Kyte & Doolittle for the fatty-acid-binding protein.

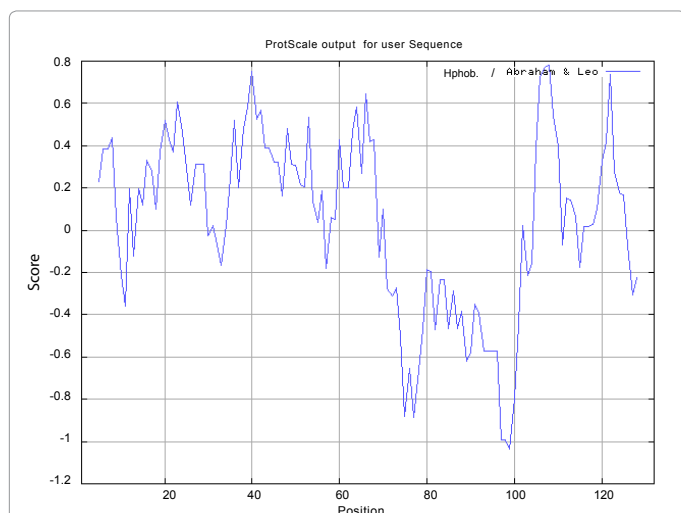


Figure 9: Hydrophobicity plot of Abraham & Leo for the fatty acid binding protein.

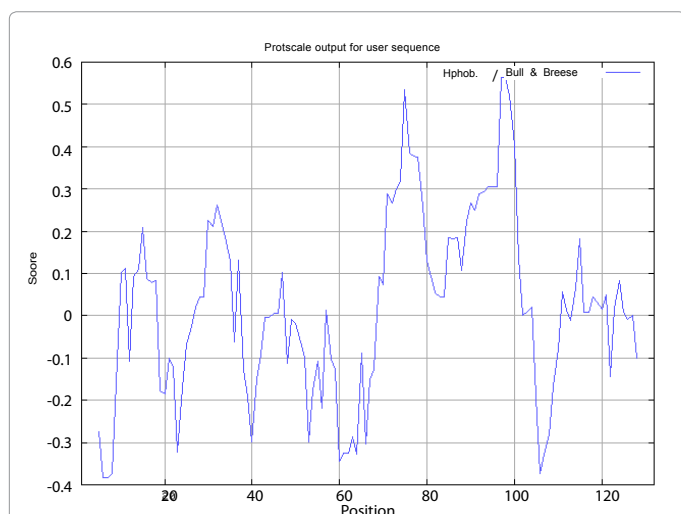


Figure 10: Hydrophobicity plot of Bull & Breese for the fatty-acid-binding protein.

fifteen peptide regions. Few of these were 98- DSKNTTVIV with score 8.634, 24- GVSWATRQI with score 8.630, 12- ESHNFDAVM with score 8.629, 85- TKDSESKIT with score 8.628 and 65- KFGEEFDEK with score 8.597, which are known as fatty acid binding protein TAP transporters.

The SVM based method for prediction of promiscuous MHC Class II binders are reported in Table 2. MHCII-IAb peptide regions, 111- GDTMKTTVT, 33- GNTVTPTVT, 46- GDTMTMLTE, 122- DVTAIRNYK; MHCII-IAd peptide regions, 104- VIVREIVGD, 56- TFKNLSVTF, 62- VTFKFGEEF, 88- SESKITHQ; MHCII-IAg7 peptide regions, 121- DDVTAIRNY, 24- GVSWATRQI, 15- NFDVMSKL, 61- SVTFKFGEE; and MHCII- RT1.B peptide regions, 41- TFTMDGDTM, 52- LTESTFKNL, 116- TTVTVDDVT, 92- ITHTQKDSK represent predicted binders from fatty acid binding protein under study. Table 3 shows the predicted antigenic epitopes from fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum*.

## Discussion

*Schistosoma japonicum* are blood flukes of humans that cause chronic, highly debilitating diseases involving extensive liver damage and requires attention. In the present study, B-EpiPred Server, Hopp

Peptide Rank	Start Position	Sequence	Score	
1	98	DSKNTTVIV	8.634	High
2	24	GVSWATRQI	8.630	High
3	12	ESHNFDAVM	8.629	High
4	85	TKDSESKIT	8.628	High
5	65	KFGEEFDEK	8.597	High
6	82	SVVTKDSSES	8.563	High
7	08	WKLSESHNF	8.524	High
8	53	TESTFKNLS	8.522	High
9	20	MSKLGVSWA	8.470	High
10	79	SVKSVTKD	8.441	High
11	31	QIGNTVTPT	8.423	High
12	113	TMKTTVTVD	8.392	High
13	116	TTVTVDVT	8.386	High
14	89	ESKITHQK	8.385	High
15	11	SESHNFDAV	8.373	High

Table 1: TAP Peptide binders of fatty-acid-binding protein.

MHC ALLELE	RANK	SEQUENCE	RESIDUE NO.	PEPTIDE
I-Ab	1	GDTMKTTVT	111	1.198
I-Ab	2	GNTVTPTVT	33	1.034
I-Ab	3	GDTMTMLTE	46	1.018
I-Ab	4	DVTAIRNYK	122	0.878
I-Ad	1	VIVREIVGD	104	0.455
I-Ad	2	TFKNLSVTF	56	0.439
I-Ad	3	VTFKFGEEF	62	0.391
I-Ad	4	SESKITHQ	88	0.325
I-Ag7	1	DDVTAIRNY	121	1.560
I-Ag7	2	GVSWATRQI	24	1.248
I-Ag7	3	NFDVMSKL	15	1.207
I-Ag7	4	SVTFKFGEE	61	1.040
RT1.B	1	TFTMDGDTM	41	1.002
RT1.B	2	LTESTFKNL	52	0.655
RT1.B	3	TTVTVDVT	116	0.642
RT1.B	4	ITHTQKDSK	92	0.616

Table 2: The SVM based method for prediction of Peptide binders to MHCII molecules of fatty-acid-binding protein.



No.	Start Position	End Position	Peptide	Peptide Length
1	13	13	S	1
2	30	39	RQIGNTVTPT	10
3	41	45	TFTMD	5
4	47	47	D	1
5	69	81	EFDEKTS DGRSVK	13
6	84	102	VTKDSESKITHTQKDSKNT	19
7	115	121	KTTVTVD	7

**Table 3:** Predicted antigenic epitopes from fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum*

and Woods, Welling, Parker, Kolaskar and Tongaonkar antigenicity scales were designed to predict the locations of antigenic determinants in fatty-acid-binding protein of the human blood fluke *Schistosoma japonicum*. High antigenicity of the fatty-acid-binding protein along with beta sheets regions, which have high antigenic response than helical region of this peptide are reported. The Janin hydrophobicity, Kyte & Doolittle hydrophobicity, Abraham & Leo and Bull & Breese hydrophobicity scales show hydrophilic index, with a polar residues assigned negative values. Peptide regions, 98- DSKNTTVIV with score 8.634, 24- GVSWATRQI with score 8.630, 12- ESHNFDAVM with score 8.629, 85- TKDSESKIT with score 8.628 and 65- KFGEEFDEK with score 8.597, which are known as fatty acid binding protein TAP transporters. It was observed that the highest ranked SVM based MHCII-IAb peptide region, 122- DVTAIRNYK; MHCII-IAD peptide region, 104- VIVREIVGD; MHCII-IAG7 peptide region, 121- DDVTAIRNY and MHCII- RT1.B peptide region, 116- TTVTVDVDT represented predicted binders from fatty-acid-binding protein.

Kolaskar and Tongaonkar antigenicity are the sites of molecules that are recognized by antibodies of the immune system for the fatty-acid-binding protein. The region of maximal hydrophilicity is likely to be an antigenic site, having hydrophobic characteristics, because C- terminal regions of fatty-acid-binding protein is solvent accessible and unstructured. Antibodies against those regions are also likely to recognize the native protein. Seven antigenic determinant sites in the fatty acid binding protein sequence were predicted. The highest pick is recorded between sequence of amino acid in the regions 102-VTKDSESKITHTQKDSKNT-19, 81-EFDEKTS DGRSVK-13 and 39-RQIGNTVTPT-10 (Table 3). The average propensity for the fatty-acid-binding protein is found to be 1.014. All residues having above 1.0 propensity are always potentially antigenic.

Fragment identified through this approach tend to be high efficiency binders, in which larger percentage of their atoms are directly involved in binding as compared to larger molecules. These fragments may, therefore, be used in cross protection and to develop human blood fluke *Schistosoma japonicum* specific antibodies that can be exploited in serodiagnostics.

## References

- McManus DP, Loukas A (2008) Current Status of Vaccines for Schistosomiasis. *Clin Microbiol Rev* 21: 225–242.
- Chitsulo L, Engels D, Montresor A, Savioli L (2000) The global status of schistosomiasis and its control. *Acta Trop* 77: 41-51.
- Ross AG, Bartley PB, Sleigh AC, Olds GR, Li Y, et al. (2002) Schistosomiasis. *N Engl J Med* 346: 1212-1220.
- King CH, Dickman K, Tisch DJ (2005) Reassessment of the cost of chronic helminth infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet* 365: 1561-1569.
- Brown M, Mawa PA, Joseph S, Bukusuba J, Watera C, et al. (2005) Treatment

of *Schistosoma mansoni* infection increases helminth-specific type 2 cytokine responses and HIV-1 loads in coinfected Ugandan adults. *J Infect Dis* 191: 1648-1657.

- Brown M, Miro G, Nkurunziza P, Watera C, Quigley MA, et al. (2006) *Schistosoma mansoni*, nematode infections, and progression to active tuberculosis among HIV-1-infected Ugandans. *Am J Trop Med Hyg* 74: 819-825
- Kjetland EF, Ndhlovu PD, Gomo E, Mduluzi T, Midzi N, et al. (2006) Association between genital schistosomiasis and HIV in rural Zimbabwean women. *AIDS* 20: 593-600.
- Ganley-Leal LM, Mwinzi PN, Cetre-Sossah CB, Andove J, Hightower AW, et al. (2006) Correlation between eosinophils and protection against reinfection with *Schistosoma mansoni* and the effect of human immunodeficiency virus type 1 coinfection in humans. *Infect Immun* 74: 2169-2176.
- Gryseels B, Polman K, Clerinx J, Kestens L (2006) Human schistosomiasis. *Lancet* 368: 1106-1118.
- Karanja DM, Hightower AW, Colley DG, Mwinzi PN, Galil K, et al. (2002) Resistance to reinfection with *Schistosoma mansoni* in occupationally exposed adults and effect of HIV-1 co-infection on susceptibility to schistosomiasis: a longitudinal study. *Lancet* 360: 592-596.
- Kumar M, Gromiha MM, Raghava GP (2007) Identification of DNA-binding proteins using support vector machines and evolutionary profiles. *BMC Bioinformatics* 8: 463.
- Gomase VS, Kale KV, Shyamkumar K (2008) Prediction of MHC Binding Peptides and Epitopes from Groundnut Bud Necrosis Virus (GBNV). *J Proteomics Bioinform* 1: 188-205.
- Garnier J, Gibart JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* 266: 540-53.
- Robson B, Garnier J (1993) Protein structure prediction. *Nature* 361: 506. Hopp TP, Woods KR (1981) Prediction of Protein Antigenic Determinants from Amino Acid Sequences. *Proc Natl Acad Sci USA* 78: 3824-3828.
- Hopp TP, Woods KR (1981) Prediction of Protein Antigenic Determinants from Amino Acid Sequences. *Proc Natl Acad Sci USA* 78: 3824-3828.
- Welling GW, Weijer WJ, Van Der ZR, Welling WS (1985) Prediction of sequential antigenic regions in proteins. *FEBS Lett* 188: 215-18.
- Parker JMR, Guo D, Hodges RS (1986) New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites. *Biochemistry* 25: 5425-5431.
- Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res* 2: 2.
- Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276: 172-174.
- Kyte J, Doolittle RF (1982) A Simple Method for Displaying the Hydrophobic Character of a Protein. *J Mol Biol* 157: 105-132.
- Gomase VS (2006) Prediction of Antigenic Epitopes of Neurotoxin Bmbktx1 from *Mesobuthus martensii*. *Curr Drug Discov Technol* 3: 225-229.
- Janin J (1979) Surface and inside volumes in globular proteins. *Nature* 277: 491-492.
- Abraham DJ, Leo AJ (1987) Extension of the fragment method to calculate amino acid zwitterions and side chain partition coefficients. *Proteins* 2: 130-152.
- 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.
- Van Ender P M, Riganelli D, Greco G, Fleischhauer K, Sidney J, et al. (1995) The peptide-binding motif for the human transporter associated with antigen processing. *J Exp Med* 182: 1883-1895.
- Adams HP, Koziol JA (1995) Prediction of Binding to MHC class I molecules. *J Immunol Methods* 185: 181-190.
- Bhasin M, Raghava GPS (2003) Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci* 13: 596-607.

28. Brusic V, Rudy G, Harrison LC (1994) Prediction of MHC binding peptides using artificial neural networks Complex system: Mechanism of Adaptation IOS Press/Ohmsha 253-260.
29. Brusic V, Rudy G, Harrison LC (1995) Prediction of MHC binding peptides using artificial neural networks. *Complexity International* 2.
30. Brusic V, Rudy G, Honeyman G, Hammer J, Harrison L (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 14: 121-130.
31. Donnes P, Elofsson A (2002) Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics* 3: 25.
32. Ding CH, Dubchak I (2001) Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 17: 349-358.
33. Rotzschke O, Falk K, Stevanovic S, Jung G, Rammensee H (1992) Peptide motifs of closely related HLA class I molecules encompass substantial differences. *Eur J Immunol* 22: 2453-2456.
34. Brown M, Grundy W, Lin D, Cristianini N, Sugnet C, et al. (2000) Knowledge based Analysis of Microarray Gene Expression Data Using Support Vector Machines. *Proc Natl Acad Sci USA* 97: 262-267.
35. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge, UK.