

Full Length Research Paper

## Selection of T cell epitopes from *S. mansoni* Sm23 protein as a vaccine construct, using Immunoinformatics approach

Onile Olugbenga Samson<sup>1,3\*</sup>, Fadare Shadrach Olaleka<sup>2</sup> and Omitogun G. Ofelia<sup>1,2</sup>

<sup>1</sup>Biotechnology Unit, Department of Biological Sciences, Elizade University, Ilara-Mokin, Ondo State, Nigeria.

<sup>2</sup>Microbiology Unit, Department of Biological Sciences, Elizade University, Ilara-Mokin, Ondo State, Nigeria.

<sup>3</sup>Centre for Biotechnology Research and Development, Department of Biological Sciences, Elizade University, Ilara-Mokin, Ondo State, Nigeria.

Received 20 July, 2016; Accepted 19 January, 2017

Schistosomiasis, a neglected and most prevalent tropical diseases after malaria, have been a threat to people living in endemic areas. With regards to possible resistance to the popular drug (praziquantel) use for treatment of schistosomiasis, the need for a permanent vaccinating approach has been justified. This study uses an *in silico* approach to identify potential target vaccine candidate or T cell epitopes (T cell response activating epitope) for the treatment of schistosomiasis. This research therefore identified some candidate T cell epitopes from Sm23 protein of *Schistosoma mansoni* using immunoinformatics tools. Nonameric epitopes like <sup>85</sup>YMYAFFLVV<sup>93</sup>, <sup>83</sup>MLYMYAFFL<sup>91</sup>, <sup>8</sup>MRCLKSCVF<sup>16</sup>, <sup>41</sup>SQYGDNLHK<sup>49</sup> and <sup>104</sup>VAVVYKDRI<sup>112</sup> was found to exhibit strong binding affinity with some human leukocyte antigen (HLA). The predicted epitope was found to have no similarity with human proteome, a good attribute that is conferred on any good vaccine candidate. The predicted epitopes provide promising drug candidates and could be tested by wet laboratory as targeted vaccine against *S. mansoni* infection.

**Key words:** Schistosomiasis, T-cell epitopes, human leukocyte antigen (HLA), vaccines.

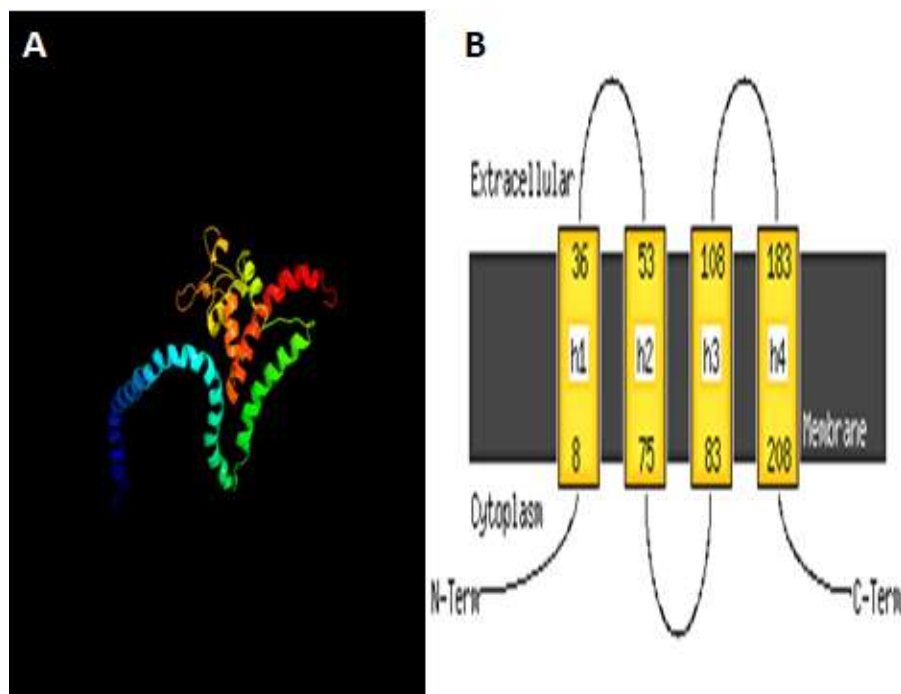
### INTRODUCTION

Schistosomiasis is well-known as a significant public health problem in tropical and sub-Saharan Africa countries, despite the existence of effective drugs against the parasite. An estimated 732 million persons are vulnerable to schistosoma infection worldwide in

renowned transmission areas (Onile et al., 2014; Adenowo et al., 2015). In 2008, 17.5 million people were treated globally for schistosomiasis, 11.7 million of these are from sub-Saharan Africa only. Approximately, 120 million individuals in sub-Saharan Africa have

\*Corresponding author. E-mail: onileg@yahoo.com or olugbenga.onile@elizadeuniversity.edu.ng. Tel: +2348133109389.





**Figure 1.** A. 3D structure of Sm23 protein; B. Transmembrane topology for Sm23 protein.

Sm23 protein as predicted by phyre 2 server is shown in Figure 1A and the secondary structure analyses of the queried protein have revealed that majority of the target protein consist mostly of  $\alpha$ -helices with few loops and no  $\beta$ -sheet (Figure 5). The signalP tool which incorporates a prediction of cleavage sites and signal/non-signal peptide predicted no presence of signal peptide in the protein. Further protein analysis also suggested its biological and molecular role in transmembrane transport activity and also its function in signal transduction activity (Tables 1 and 2).

### Prediction of specific HLA Class I T cell epitope

The predicted epitope with their HLA alleles have been summarized in Table 3. Identifying the affinity of interaction between the predicted T-cell epitope and HLA is essential, as the inclusion of such peptides in vaccine design will provide more population coverage and help reduce the number of peptides that need to be involved in the vaccine (Brusic et al., 2002; Shah et al., 2010). The results from predicted overlapping sequences of the consensus Sm23 protein to the human 33 Class I HLA alleles at various affinities identified 53 nonameric peptide sequences which bind both at low and high affinity to the different HLA alleles (Table 3). The observed binding affinities of the predicted epitopes with the HLA alleles shows that alleles like HLA B\_2705, HLA

B\_5102 and HLA B\_5201 bind to most of the peptides showing the tallest bar in the graph (Figure 4) and the rear peptides binder like HLA A1, HLA A24, HLA Cw\_0702, HLA B\_2702, B60 and B40 binds to very less peptides while some do not bind at all.

The epitope HLA\_A0201 bind to peptide sequences <sup>85</sup>YMYAFFLVV<sup>93</sup> and <sup>83</sup>MLYMYAFFL<sup>91</sup> at an affinity score of 2189 and 3707.145, respectively. Peptides <sup>8</sup>MRCLKSCVF<sup>16</sup> AND <sup>41</sup>SQYGDNLHK<sup>49</sup> bind to HLA B\_2705 both at affinity score of 1000. Sequence <sup>104</sup>VAVVYKDRI<sup>112</sup> binds to HLA B\_5102 at score 1200 and also bind to HLA B\_5101 (score 314.6) and B\_5103 (score 133.1). It is therefore necessary to include peptides that bind to the rare HLA alleles in a vaccine cocktail to achieve the desired population coverage (Parida et al., 2007). The choice of multiple computational tools in selecting HLA Class I T-cell epitope is to minimize the chances of failure in selection of a universal vaccine candidate while ensuring that only peptides with positive scores were selected. The prediction of presence of human-self peptides using HLA-PRED showed that the predicted epitopes are nonhuman but parasite specific (Figure 3).

### Anchor motifs and human peptide crosscheck

The sequence logo was predicted while considering the predicted peptides and using the anchor position of the

**Table 1.** The predicted biological function of Sm23.

GO term	Name	Prob.	SVM reliability
GO:0055085	Transmembrane transport	0.899	H
GO:0006810	Transport	0.886	H
GO:0016192	Vesicle-mediated transport	0.787	H
GO:0051649	Establishment of localization in cell	0.721	H
GO:0051641	Cellular localization	0.705	H
GO:0019222	Regulation of metabolic process	0.681	H
GO:0008104	Protein localization	0.679	H
GO:0050877	Neurological system process	0.662	H
GO:0045184	Establishment of protein localization	0.658	H
GO:0015031	Protein transport	0.625	H
GO:0002376	Immune system process	0.603	H
GO:0007166	Cell surface receptor signaling pathway	0.596	H
GO:0040011	Locomotion	0.587	H
GO:0098655	Cation transmembrane transport	0.584	H
GO:0043269	Regulation of ion transport	0.556	H
GO:0007268	Synaptic transmission	0.541	H
GO:0007155	Cell adhesion	0.539	H
GO:0046907	Intracellular transport	0.538	H
GO:0034613	Cellular protein localization	0.509	H

**Table 2.** Predicted molecular function of Sm23 protein.

GO term	Name	Prob	SVM reliability
GO:0022857	Transmembrane transporter activity	0.972	H
GO:0022891	Substrate-specific transmembrane transporter activity	0.964	H
GO:0005215	Transporter activity	0.957	H
GO:0003824	Catalytic activity	0.893	H
GO:0022890	Inorganic cation transmembrane transporter activity	0.845	H
GO:0015075	Ion transmembrane transporter activity	0.842	H
GO:0004871	Signal transducer activity	0.792	H
GO:0046914	Transition metal ion binding	0.746	H
GO:0008324	Cation transmembrane transporter activity	0.742	H
GO:0016740	Transferase activity	0.730	H
GO:0050839	Cell adhesion molecule binding	0.717	H
GO:0016817	Hydrolase activity, acting on acid anhydrides	0.698	H
GO:0005198	Structural molecule activity	0.690	H
GO:0015267	Channel activity	0.640	H
GO:0016462	Pyrophosphatase activity	0.600	H
GO:0008270	Zinc ion binding	0.551	H
GO:0016757	Transferase activity, transferring glycosyl groups	0.531	H
GO:0004872	Receptor activity	0.529	H

**Table 2.** Cont.

GO:0030246	Carbohydrate binding	0.516	H
GO:0038023	Signaling receptor activity	0.511	H
GO:0005509	Calcium ion binding	0.502	H
GO:0005216	Ion channel activity	0.501	H

\*Tables 1 and 2 predictions terms represent terms predicted where SVM training includes assigned gene ontology (GO) terms across all evidence code types. SVM reliability is regarded as High (H) when MCC, sensitivity, specificity and precision are jointly above a given threshold; otherwise reliability is indicated as low (L).

**Table 3.** Peptides of consensus Sm23 protein present in surface exposed loop regions and their binding to different HLA alleles. The predicted.

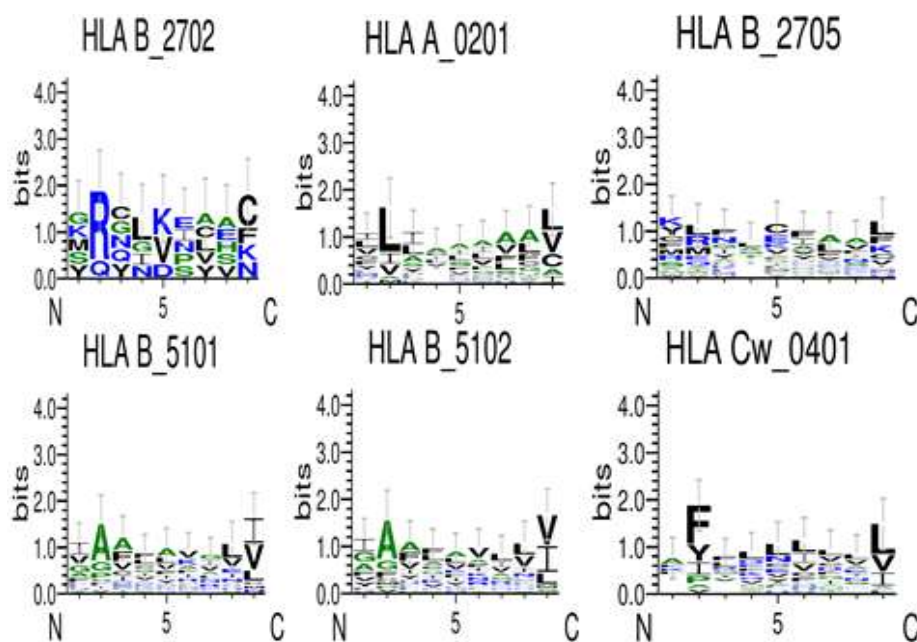
S/N	Position	Peptide	Score	HLA alleles	Looping
1	35	VEVKFSQY	45	HLA_A01	
2	83	MLYMYAFFL	3707.145	A_0201	I
			150	B_2705	
			51.408	A_0205	
3	16	FVLNIICLL	252	A_0205	
			73.172	A_0201	
4	118	ALMTGALDK	60	A3	III
5	86	MYAFFLVVL	200	A_2402	
			18	B_3902	I
			400	Cw_0404	
6	174	SVFGAFLKR	28	A_1101	III
			400	A_6801	I
			24	A_3101	
7	21	ICCLCSLV	24	B_1402	
8	210	RQIKEYENV	23	B13	
9	207	CLGRQIKEY	21	B_1510	I
			19.2	Cw_0702	
10	3	TGLTGMRCL	15	B_1510	
11	168	YTEGCVSVF	23	B_1516	I
			18	B_5701	
			48	B_5801	
12	129	KEITEFMNL	16	B18	I
			640	B_4001(B60)	I
			21	B_4901	
			26	B_4402	
			90	B_2705	
13	79	ENVCMLYMY	290	A26	
14	8	MRCLKSCVF	24	B_2705	
			18	B_2709	I
			200	B_2702	
			1000	B_2705	
15	126	KPTKEITEF	20	B_3501	
			60	B_3501	
			80	Cw_0401	
			20	B_5301	
16	116	IDALMTGAL	22	B37	I

Table 3. Cont.

			21	B_4701	I
			40	B60	
17	108	YKDRIDSEI	20	B_3902	I
18	78	KENVCMLYM	40	B60	
19	97	AELAAAIVA	24	B_4101	
			24	B_4501	I
			20	B_5001	
			40	B61	
20	56	IAIIVVGI	314.6	B_5101	
			660	B_5102	
			110	B_5103	
21	12	KSCVFLNI	27	B_5802	
22	91	LVVLLIAEL	20	B7	
			47.6	A_0205	
23	153	YRGNVPASC	200	B_2705	
24	209	GRQIKEYEN	200	B_2705	
25	41	SQYGDNLHK	1000	B_2705	
26	181	KRNLVIVAC	600	B_2705	
27	112	IDSEIDALM	60	B_3701	
28	192	FGVCFQQL	120	Cw_0301	
			132	B_5102	
29	100	AAAIVAVVY	40.5	B_4403	
30	104	VAVVYKDRI	314.6	B_5101	
			133.1	B_5103	
			1200	B_5102	
31	204	IACCLGRQI	286	B_5101	
			220	B_5102	
			100	B_5103	
32	51	WQAAPIAII	120	B_5201	
33	15	VFVLLNIICL	400	Cw_0401	
34	39	KFSQVGDNL	200	Cw_0401	
35	87	YAFFLVVLL	500	B_5102	
			136	B_5101	
36	177	GAFLKRNLV	550	B_5102	
			200	B_5101	
			132	B_5103	
37	85	YMYAFFLVV	2188.816	A_0201	
			72	A_0205	
			75	B_2705	
38	28	VLIGAGAVV	650.311	A_0201	
39	22	CLLCSLVI	88.8	A_0201	
40	63	VIIIVSFL	63.167	A_0201	
			23.8	A_0205	
41	186	IVACVAFGV	60.006	A_0201	
42	198	QLLSIVAC	42.278	A_0201	
43	172	CVSVFGAFC	39.066	A_0201	
			28	A_0205	
44	58	IIVVGVIII	23.8	A_0205	
45	210	RQIKEYENV	180	B_2705	
46	51	WQAAPIAII	60	B_2705	

**Table 3. Cont.**

47	87	YAFFLVVLL	50	B_2705
48	149	GPDDYRGNV	220	B_5101
			100	B_5102
49	99	LAAAIVAVV	143	B_5101
			100	B_5102
			100	B_5103
50	96	IAELAAIV	143	B_5101
			55	B_5102
			121	B_5103
51	191	AFGVCFQQL	200	Cw_0401
52	42	KPTKEITEF	80	Cw_0401
53	88	AFELVLLI	50	Cw_0401



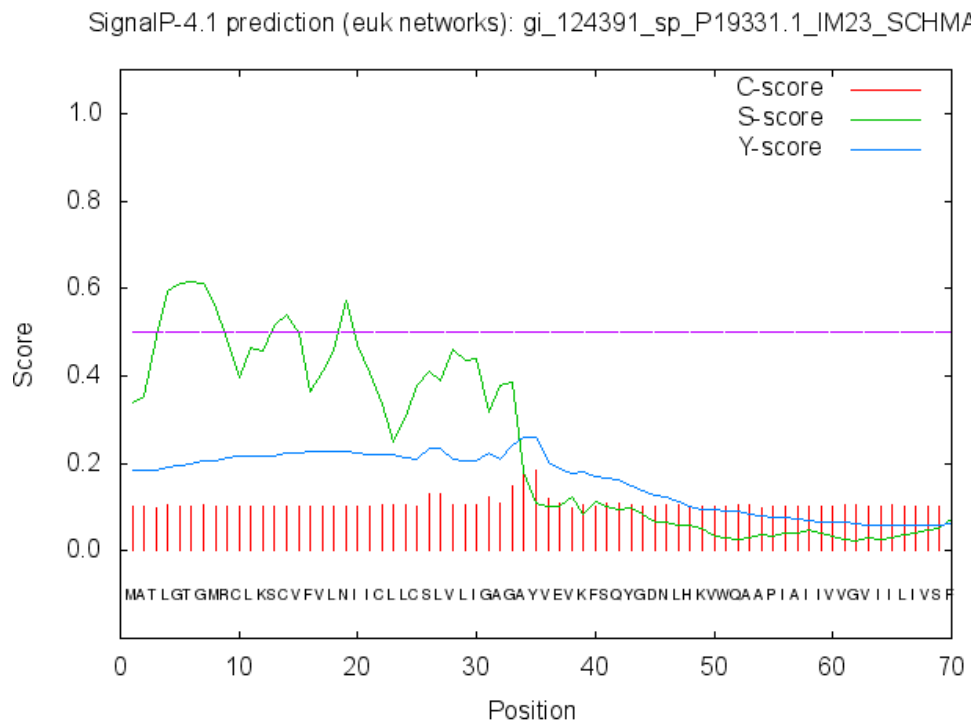
**Figure 2.** Sequence logo for binding of HLA alleles B\_2702, A\_0201, B\_2705, B\_5101, B\_5102, Cw\_0401 showing anchor motifs at position 2 and 9.

HLA alleles like HLA B\_5102, HLA B\_5101, HLA B\_2705, HLA B\_2702, HLA B\_0201 and HLA Cw\_0401 in order to find out the frequency of an amino acid at a particular position given in the SYFPEITHI database. HLAB\_2702 revealed the amino acids arginine and cysteine at position 2 and 9 as anchor motifs, while HLA B\_0201 showed two Leucine, HLA B\_2705 with Lysine and Leucine, HLA B\_5101- Alanine and Isoleucine, HLA B\_5102- Alanine and Valine and lastly HLA Cw\_0401- Phenylalanine and Leucine, all at the same position 2 and 9 as anchor motifs (Figure 2). The sequence conservation plot identified maximum peptides binding

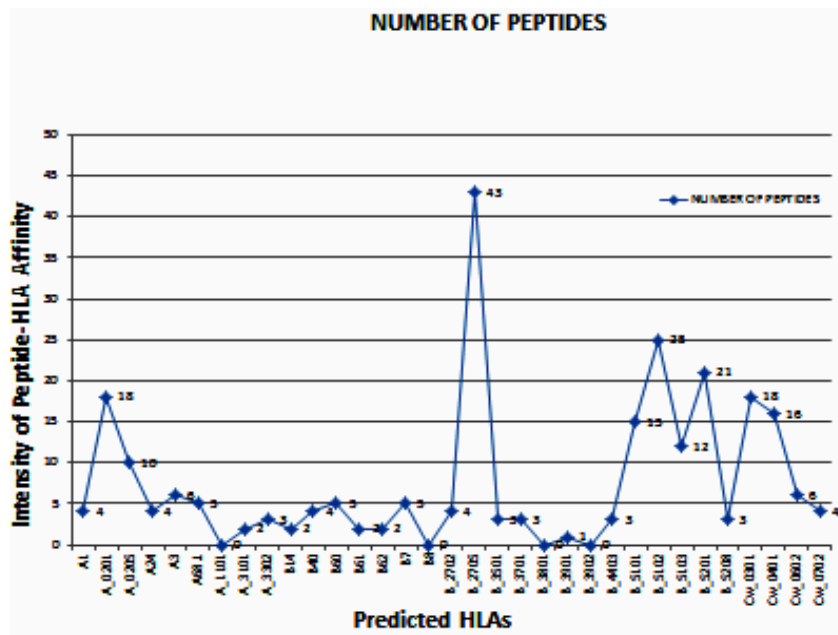
with HLA B\_2705 (Figure 4). The predicted peptides were identified as potential vaccine candidates when subjected to the HLApred server, as it was shown that all predicted peptide were completely dissimilar with human host. This is important in the prevention of possible autoimmune diseases when this vaccine is developed using the predicted peptide sequences.

## DISCUSSION

Investigation using bio-computational approach was



**Figure 3.** Signal P-NN result for the signal peptide of Sm23 protein using several artificial neural networks and Hidden Markov Models. No signal peptide was detected and it was predicted to be an integral membrane protein. Score C, S and Y represent cleavage site score, signal peptide score and combination score (derived from C and S scores), respectively.



**Figure 4.** The binding of the predicted peptides from Sm 23 protein to the HLA alleles HLA B\_2705, B\_5102 and B\_5201 bind to most of the peptides (tallest chart). The alleles like HLA; A1, A24, Cw\_0702, B\_2702, B60, B40 rarely bind to the predicted epitopes.





conducted on antigen presentation and identification of epitopes on different HLA molecules. Interest was on *S. mansoni*, as there is no efficacious vaccine reported against this parasite. The highly reported *S. mansoni* Sm23 protein was chosen in this study, in order to predict epitope that could be useful as possible vaccine candidate in eliciting immunological response against the parasite.

The blast search conducted proved that the homologous protein obtained was specific for *S. mansoni* pathogen. It is important to establish the subcellular localization of a pathogen protein to predict the most accessible pool of potential target using integrative approaches (Shah et al., 2010). Hence, the subcellular localization information obtained using *in silico* approaches (Pspred, proteinpredict and SignalP) confirmed the protein to function in cell surface signal pathways and substrate-specific transmembrane transporter activity. This information provides valuable clues regarding the protein molecular function in drug design as secreted or surface exposed protein and periplasmic protein were reported to be of primary interest due to their potentials as vaccine candidate and the ease with which they are accessible to drugs (Namrata et al., 2010). In all the 53 peptides epitopes predicted from the Sm23 protein with several computational tools like ProPred, SYFPEITHI, nHLAPred, only 5 promiscuous peptides recorded high immunogenicity. The need to use several analytic tools for peptide prediction became imperative to select only peptides that exhibit positive binding scores under extensive analysis by multiple methods because some peptide do bind when analysed by one algorithm and do not by another. All efforts to predict accurately peptide immunogenicity will help reduce several experimental efforts (Shah et al., 2010). It should also be noted that the prediction of peptide immunogenicity is influenced by many factors which include intrinsic physicochemical properties and extrinsic factors such as host immunological repertoire (Saffari et al., 2008, Namrata et al., 2010).

Namrata et al. (2010) suggested that predicted epitopes with high binding affinity are better suited for wet lab studies. It was observed that 4 out of the 5 predicted epitopes (<sup>85</sup>YMYAFFLVV<sup>93</sup>, <sup>83</sup>MLYMYAFFL<sup>91</sup>, <sup>8</sup>MRCLKSCVF<sup>16</sup> and <sup>104</sup>VAVVYKDRI<sup>112</sup>) with high affinities to the HLA alleles are rich with hydrophobic and charged amino acids residue thus making them a good choice for inclusion in an experimentally drug design study. It has been reported that MHC pockets favourably interact with hydrophobic or charged amino acid residues at carboxyl to enhance proper binding in pockets (Brusic et al., 2002). The interaction between peptide and a MHC molecule is mediated through anchors on the peptide side chains of amino acids at predetermined positions that protrude into complemen-

tary pockets of the class I groove (Parida et al., 2007; Petersen et al., 2004; Namrata et al., 2010). In peptide interaction with MHC class I, most of the bonding forces are provided by non-allele specific interactions such as the bonds between the peptides termini and the class I groove (Guo et al., 1993; Namrata et al., 2010). The MHC groove pockets determined the specificity of peptide interactions with fixed spacing from each other and the class I pockets are generally more specific in the amino acid residues that they bind (Brusic et al., 1995).

## Conclusion

The study therefore concluded that with predicted T-cell epitopes from Sm23, a transmembrane transporter protein using *in silico* approach can be targeted for vaccine constructs while considering its expression throughout all schistosome life stages. Also, all high binding affinity epitopes will be suited for wet laboratory experiment while understanding that vaccine formulations against schistosomiasis can be achievable only if the epitopes are able to elicit immune response under *in vitro* studies.

## CONFLICT OF INTEREST

The authors have not declared any conflict of interest.

## REFERENCES

- Adenowoa AF, Oyinloye BE, Ogunyinka BI, Kappo AP (2015). Impact of human schistosomiasis in sub-Saharan Africa. *Braz. J. Infect. Dis.* 19(2):196-205.
- Brusic V, Petrovsky N, Zhang G, Bajic VB (2002). Prediction of promiscuous peptides that bind HLA Class I molecules, *Immunol. Cell Biol.* 80:280-285.
- Brusic V, Rudy G, Harrison LC (1995). Prediction of MHC Binding Peptides Using Artificial Neural Networks. *Complex. Int.* 231:1-9.
- Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013). Scalable web services for the PSIPRED Protein Analysis Workbench *Nucleic Acids Res.* 41(W1): W340-W348.
- Carina SP, Vicente PM, Natan RGA, Bárbara CPF, Suellen BM, Vasco A, Sergio CO. (2011). Computational Vaccinology: An Important Strategy to Discover New Potential *S. mansoni* Vaccine Candidates. *J. Biomed. Biotechnol.* 9p.
- Fonseca CT, Carvalho GBF, Alves CC, de Melo TT (2012). Schistosoma Tegument Proteins in Vaccine and Diagnosis Development: An Update. *J. Parasitol. Res.* Article ID 541268, pp. 8.
- Guo HC, Madden DR, Silver ML, Jardetzky TS, Gorga JC, Strominger JL, Wiley DC (1993). Comparison of the P2, Specificity Pocket in Three Human Histocompatibility Antigen HLA- A\* 6801, HLA- A\* 0201 and HLA-B \*2705. *Proc. Natl. Acad. Sci. USA* 90:8053-8057.
- Jones DT (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195-202.
- Namrata M, Priyanka C, Ankita M, Kavita S (2010). Structural simulation of MHC-peptide interactions using T-cell Epitope in Iron-acquisition protein of *N. meningitidis* for vaccine design. *J. Protein Proteom.* pp. 53-63.
- Nugent T, Jones DT (2009). Transmembrane protein topology

- prediction using support vector machines. *BMC Bioinform.* 10(1):159.
- Onile OS, Otarigho B, Anumudu CI (2014). Understanding the Phylogenetics and Evolution of Genus *Schistosoma*- Africa and Asia Stand Point. *Ann. Res. Rev. Biol.* 4(10):1703-1712.
- Parida R, Shaila MS, Mukherjee S, Chandra NR, Nayak R (2007). Computational analysis of proteome of H5N1 avian influenza virus to define T cell epitopes with vaccine potential. *Vaccine* 25:7530-7539.
- Parker KC, Bednarek MA, Coligan JE (1994). Scheme for ranking potential HLA A2 binding peptides based on independent binding of individual peptides side-chains. *J. Immunol.* 152:163
- Petersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferri TE (2004). UCSF Chimera: A Visualization System. *J. Comput. Chem.* 25:1605-1612.
- Petersen TN, Søren B, Gunnar von H, Henrik N (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8:785-786.
- Rammensee HG, Bachmann J, Emmerich NN, Bachor OA and Stevanovic S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenet.* 50:213-219.
- Reynolds SR, Shoemaker CB, Harn DA (1992). T and B cell epitope mapping of SM23, an integral membrane protein of *Schistosoma mansoni*. *J. Immunol.* 149:3995-4001.
- Saffari B, Mohabatkar H, Mohsenzadeh S (2008). T and B cell epitopes prediction of Iranian Saffron (*Crocus sativus*) Profiling by Bioinformatics Tools. *Protein Peptide Lett.* 15:280-285.
- Shah K, Chaubey P, Mishra N (2010). Bioinformatics approach for screening and modelling putative T cell epitopes from Por B protein of *Neisseria meningitidis* as vaccine constructs. *Indian J. Biotechnol.* 9:351-359.