A Computational Assay to Design an Epitope-Based Peptide Vaccine Against Saint Louis Encephalitis Virus

Md. Anayet Hasan, Mehjabeen Hossain and Md. Jibran Alam

Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong, Bangladesh.

ABSTRACT: Saint Louis encephalitis virus, a member of the flaviviridae subgroup, is a culex mosquito-borne pathogen. Despite severe epidemic outbreaks on several occasions, not much progress has been made with regard to an epitope-based vaccine designed for Saint Louis encephalitis virus. The envelope proteins were collected from a protein database and analyzed with an *in silico* tool to identify the most immunogenic protein. The protein was then verified through several parameters to predict the T-cell and B-cell epitopes. Both T-cell and B-cell immunity were assessed to determine that the protein can induce humoral as well as cell-mediated immunity. The peptide sequence from 330–336 amino acids and the sequence REYCYEATL from the position 57 were found as the most potential B-cell and T-cell epitopes, respectively. Furthermore, as an RNA virus, one important thing was to establish the epitope as a conserved one; this was also done by *in silico* tools, showing 63.51% conservancy. The epitope was further tested for binding against the HLA molecule by computational docking techniques to verify the binding cleft epitope interaction. However, this is a preliminary study of designing an epitope-based peptide vaccine against Saint Louis encephalitis virus; the results awaits validation by *in vitro* and *in vivo* experiments.

KEYWORDS: epitope, computational tools, humoral, cell-mediated immunity, conservancy

CORRESPONDENCE: anayet_johny@yahoo.com

Introduction

Saint Louis encephalitis virus (SLEV) is related to Japanese encephalitis virus and belongs to the family Flaviviridae (Group B arborviruses), genus *Flavivirus*. It mainly affects the United States and, in occasional cases, Mexico and Canada.[1,2] The principle reservoirs of SLEV include domestic fowl and wild birds. The virus is transmitted to humans from wild birds via mosquito vectors (*Culex tarsalis*, *C pipiens*, and *C quinquefasciatus*).[3] SLEV is a single stranded, enveloped, positive-sense ribonucleic acid (RNA) virus and has a relatively conserved nucleotide sequence.[4] In all, 4437 cases of SLEV infection were reported across the United States since 1964 with an average of 128 cases per year. The fatality rate caused by the disease ranges from 5% to 15%. SLEV infection fatality rates are generally lower in children than adults, but the cases in which children were infected, the encephalitis

rate was higher. The severity of the disease increases with age, and the disease is most prevalent in people over the age of 60. Hypertension and vascular diseases may act as risk factors for infection. Encephalitis is recognized as the emerging disease in the United States.[5] The risk of exposure to SLEV infection in urban areas is increasing, as the deterioration of inner cities can make new habitats for mosquitoes. There is no specific treatment for SLEV infection; care is based on only a few symptoms. No vaccine against SLEV or specific antiviral treatment for clinical SLEV infections is available.[6] For all these reasons, it is important to develop a new vaccine for Saint Louis encephalitis virus, and despite no technical difficulties having been reported, new vaccine development has not yet been achieved. These viruses contain 3 structural proteins: nucleocapsid, premembrane (prM), and envelope (E). The envelope protein is particularly important for vaccine

development as it mediates viral entry by interacting with host cell surface receptors and is also the primary target of adaptive immune response.[7] It would be of great value if the protective epitopes associated with this envelope protein were well known, and, if they could be produced synthetically, it would be possible to offer consistent cost and quality advantages over the current treatment.[8]

Synthetic peptides are being tested as potential vaccines in a wide range of infectious diseases. The use of synthetic peptides or epitopes could offer an important contribution to protection against SLEV.[9] Current immunoinformatics tools are able to predict human B-cell and T-cell linear epitopes with high accuracy. These tools are playing a vital role in understanding the molecular basis of immunity and, notably in the development of epitope based-peptide vaccines, immunotherapy against cancer and autoimmune diseases.[10–13] This approach could save substantial time and cost, especially for researchers in countries with limited resources. The concept of computer-aided vaccine design has proven to be essential for combating diseases such as multiple sclerosis,[14] malaria,[15] and tumors.[16] For epitope-based vaccine design, identification of HLA ligands and T-cell epitopes are the most crucial steps.[17] T-cell epitope prediction tools assist in identifying allele-specific binding peptides, thus reducing the number of potential peptides to be considered as vaccine candidates. These tools help researchers to identify regions with high concentrations of T-cell epitopes or immunological "hot spots" and thus open a new window for rationality-based vaccine design. A number of methods have been developed to facilitate the identification of proteasomal peptide cleavage sites, major histocompatibility complex (MHC) binding peptides, and transporters associated with antigen presentation (TAP) molecules.[18–21] These methods use a variety of statistical and machine learning approaches to assist computational prescreening of immunogenic epitopes for vaccine design. Similar immunoinformatics approaches are being developed for other organisms and will constitute the base for the design of a new generation epitope vaccines. Our present study concerns the computational prediction of immunogenic epitopes of the Saint Louis encephalitis virus and design of a synthetic peptide vaccine candidate. We hypothesize that this envelope protein comprises a linear epitope with an efficacy close to 100%.

Materials and Methods

Protein sequence retrieval. The outer membrane protein (envelope proteins) sequences of Saint Louis encephalitis were retrieved from UniProtKB (www.uniprot.org/) in FASTA format. The sequence was analyzed with a view to recognizing the immunologically relevant regions, which is done by studying antigenicity, solvent accessible regions, and MHC class I and II binding sites.

Identification of antigenic proteins. The envelope proteins were then submitted in the VaxiJen[22] with default parameters to isolate the antigenic proteins. All the antigenic

proteins with their respective score were then filtered in Excel. A single antigenic protein with highest antigenicity scores was selected for further evaluation.

Secondary structure prediction. ExPASy's secondary structure prediction server (http://web.expasy.org/protparam/) and selfoptimized prediction method with alignment (SOPMA)[23] were used to get an idea about the secondary structure of the chosen protein. Several parameters given by the ProtParam tool, for example, 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 executed for this purpose. Secondary structure prediction of a protein aims to predict solvent accessibility, transmembrane helices, globular regions, and coiled-coil regions. The objectives of this step were to determine the protein's stability and function.

Identification of T-cell epitope. Linear peptides as T-cell epitopes were selected by NetCTL server.[24] The whole protein sequence was submitted in the server, which was based on neural network architecture, and all the super types were chosen. The parameter was set at 50 to have sensitivity and specificity of 0.89 and 0.94, respectively. This allowed the opportunity to focus on more epitopes critically. A combined algorithm of MHC-1 binding, transporter of antigenic peptide (TAP) transport efficiency and proteosomal cleavage efficiency were involved to predict overall scores. Based on this overall score, 5 best epitopes were selected for further experimentation.

MHC-1 binding prediction,[25] a different prediction method, was available in the Immune Epitope Database (IEDB). The Stabilized Matrix Base Method (SMM) was used to calculate IC50 values of peptide binding to MHC-1 molecules. For all the alleles, peptide length was set to 9 amino acids prior to the prediction. The alleles having binding affinity IC50 less than 200 nm were chosen for further consideration.

For the selected epitopes, a web-based tool, proteasomal cleavage/TAP transport/MHC class 1 combined predictor,[26] was also implemented to predict proteasomal cleavage score, TAP score, processing score, and MHC-1 binding score using SMM for each peptide.

Prediction of epitope conservancy. Epitope conservancy for individual peptides was predicted using the epitope tool from the IEDB analysis resource.[27]

Docking simulation study. The predicted epitope was further tested by *in silico* docking simulation to find out whether or not this peptide will bind to the HLA molecules when it will be applied *in vivo*. To accomplish the simulation, 3 LKN, a crystal structure of the HLA-B*3501 molecule, was retrieved from the Protein Data Bank (PDB).[28] Prior to docking simulation, the influenza NP418 epitope, which was complexed in the binding groove of HLA-B*3501, was removed by using AutoDockTools. Docking simulation for the predicted epitope with prepared HLA-B*3501 was done by using Autodock Vina.[29] Then again influenza NP418 epitope was

docked with HLA-B*3501 mentioned earlier, to compare the binding energy of the epitope.

Identification of B-cell epitope. The B-cell epitope is the portion of the antigen which interacts with B-lymphocytes. As a result, the B-lymphocyte is differentiated into antibody-secreting plasma cell and memory cells. In other terms, the objective of the B-cell epitope is to ultimately stimulate the B cell to synthesize the antibody specific for it (primary humoral response) or to convert the naïve B cell into a memory B cell and make it ready to produce specific antibody in later encounters.[30]

A B-cell epitope is characterized by being hydrophilic, accessible, and in a flexible region of an immunogen.[31] Thus, Kolaskar and Tongaonkar antigenicty scale,[32] Emini surface accessibility prediction,[33] Parker hydrophilicity prediction,[34] Karplus and Schulz flexibilty prediction[35] and Bepipred linear epitope prediction analysis were done computationally from the IEDB analysis resource.[36] In several experimental studies, it was found that the antigenic parts of a protein belong to the beta turn regions.[37] Therefore, the Chou and Fashman beta turn prediction tool[38] was used. The results from all these sites were cross-referenced, and apparently common findings were taken as the B-cell epitopes.

Results

Protein retrieval and antigenic protein prediction. A total of 75 envelope proteins were retrieved from UniprotKB in FASTA format and then subjected to VaxiJen to predict the most immunogenic protein. UniprotKB id: B6C6 W4 was predicted as the most immunogenic protein with a total score of 0.6591, which is a glycoprotein among all envelope proteins. Finding of this prediction coincides with a previous finding where envelope glycoprotein was presented as an immunogenic protein.[39]

Secondary structure analysis. The most immunogenic protein was analyzed for its secondary structural characteristics, and the properties shown in Table 1 and Figure 1 were found. The protein was characterized as a neutral charged protein, as the number of positively charged residues (Arg+Lys) and negatively charged residues (Asp+Glu) were the same (ie, 46). The total amino acid distribution showed that alanine,

Table 1. Secondary structural analysis of one envelope protein (UniprotKB id: B6C6 W4) of Saint Louis encephalitis virus.

CRITERIA	ASSESSMENT
Number of Amino Acids	501
Molecular Weight	54056.4 Da
Isoelectric pH	7.24
No. of Negatively charged residues (Asp+Glu)	46
No. of Positively charged residues (Arg+Lys)	46
Formula	$C_{2401}H_{3753}N_{653}O_{726}S_{21}$
Extinction Coefficient	77640
Instability Index	25.04
Aliphatic Index	9.02

glycine, leucine, lysine, serine, threonine, and valine constitute 57% of the protein. Though all these amino acids except lysine have acidic pI (from 6.0 for Alanine to 5.6 for valine), the net pI of the protein tended to be slightly basic (pI 7.24), maybe due to the relatively higher abundance of lysine (5.2%) with a very high basic pI (9.74). The basic character of the protein ascribes the protein to be a nonallergen.[40]

The secondary character shows an important feature of the protein as an antigenic one. In several experiments, it was shown that the antigenic part of the protein is more likely to belong to the beta sheet region.[41] Regarding this fact, it was found that this protein contained the highest percentage (11%) of the threonine residue, which prefers to lie in the beta sheet, indirectly indicating the probability of the protein's antigenicity. Moreover, the grand average of the hydrophobicity rule (GRAVY) of the protein linear sequence was predicted to be negative (−0.079) showing the protein as hydrophilic in nature and predicting most of the residues to be present in the surface. In addition, the aliphatic and the instability index of the protein described it as an aliphatic and stable protein, respectively.

T-cell epitope prediction. The NetCTL prediction tool predicted epitopes from the given protein sequence according to all supertypes. But only 5 most potential epitopes were chosen on the basis of high combinatorial scores.
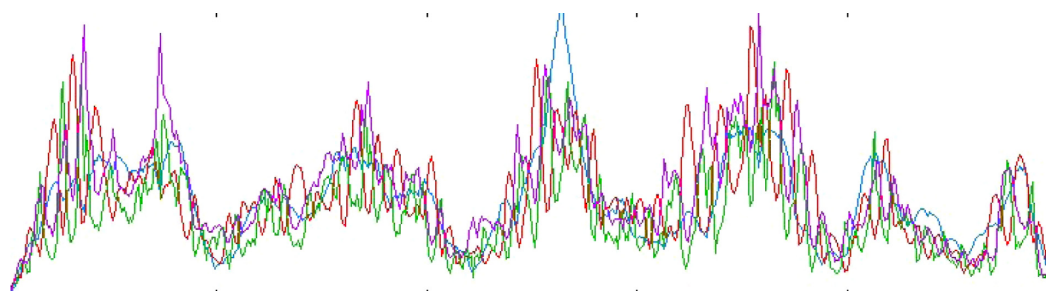
Figure 1. Secondary structure plot of envelope protein (UniprotKB id: B6C6 W4) of Saint Louis encephalitis virus. Here, helix is indicated by blue, while extended strands and beta turns are indicated by red and green, respectively.

The SMM-based IEDB MHC-1 binding prediction tool retrieved 406 possible MHC-1 allele interactions with the 5 T-cell epitopes that were selected before. The MHC-1 alleles for which the epitopes showed higher affinity (IC50 < 200) were selected for further analysis (Table 2). Binding affinity of the epitopes with the MHC-1 alleles had an inverse relationship with the IC50 values.

The proteasome complex includes enzymes that cleave the peptide bonds, converting the proteins into peptide. The peptide molecules from proteasome cleavage associate with class-1 MHC molecules and the peptide-MHC molecule then transport to the cell membrane where they are presented to T helper cell. Before that, TAP (transporter of antigenic peptides) transfers it to the endoplasmic reticulum. In this case, the total score was predicted, and the higher the score meant the higher the efficiency of all these processes. The predicted total score of all these processes are summarized in Table 2.

Among the 5 T-cell epitopes, a 9-mer epitope REY-CYEATL was found to interact with most MHC-1 alleles including HLA-B*40:01, HLAC*03:03, HLA-A*32:07, HLA-A*68:23, HLA-B*27:20, HLA-A*02:17, HLA-B*40:13, HLA-C*12:03, HLA-A*02:50, HLA-B*15:02, HLA-B*40:02, HLA-B*15:03.

The epitope conservancy analysis of the 5 T-cell epitopes was done. In this case, REYCYEATL showed the highest conservancy of 63.51%, while the other 4 showed 40.54%.
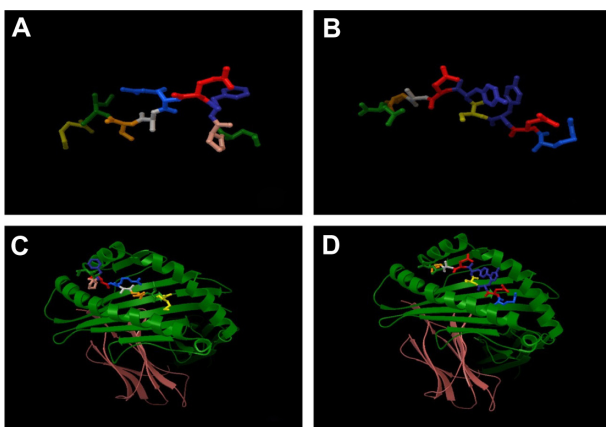
Docking simulation results. The center box coordinates for HLA-B*3501 binding groove were X:3.659, Y:-15.259, and Z:-36.372, and the grid box sizes were X:36, Y:22, and Z:18. All the measurements were spaced at the scale of 1 angstrom. These coordinates were chosen and calculated to aid the epitope sequence to bind to the binding groove of the receptor. The binding energy of predicted epitope with HLA-B*3501 receptor was found to be -7.6 kcal/mol (Fig. 2, C and D). This binding energy was compared with the binding energy of influenza

NP418 epitope to HLA B*3501 and found to be the same as the predicted epitope (-7.6 kcal/mol, Figure 2, A and B). Similar binding energy of both the simulations indicates the satisfactory accuracy of the predicted epitope.

B-cell epitope identification. Characteristic features of the B-cell epitope include hydrophilicity, surface accessibility,

Figure 2. Docking simulation results generated by Autodock Vina. (**A**) Structure of influenza NP418 epitope "LPFERATIM", (**B**) Structure of predicted epitope "REYCYEATL", (**C**) Docking of influenza NP418 epitope with HLA B*3501, (**D**) Visualization of docking results of "REYCYEATL" with HLA B*3501.

Table 2. Most potential 5 T-cell epitopes with Interacting MHC-1 alleles, total processing score and epitope conservancy result.

EPITOPE	INTERACTING MHC-1 ALLELE WITH AN AFFINITY <200 (TOTAL SCORE OF PROTEASOME SCORE, TAP SCORE, MHC SCORE, PROCESSING SCORE AND MHC-1 BINDING)	EPITOPE CONSERVANCY ANALYSIS RESULT
RSGINTEDY	HLA-B*27:20(1.46), HLA-A*32:07(1.27), HLA-C*12:03(1.10), HLA-A*68:23(1.06), HLA-B*40:139(0.76), HLA-A*32:15(0.66), HLA-B*15:17(0.52), HLA-C*03:03(0.22)	40.54%
IGKAVHQVF	HLA-C*12:03(1.32), HLA-B*15:03(1.17), HLA-B*27:20(1.10), HLA-A*02:50(0.25), HLA-A*32:07(0.66), HLA-A*68:23(0.53), HLA-B*40:13(0.51), HLA-C*03:03(0.38), HLA-C*07:02(0.36), HLA-B*15:01(0.16)	40.54%
REYCYEATL	HLA-B*40:01(1.05), HLA-C*03:03(1.01), HLA-A*32:07(0.79), HLA-A*68:23(0.63), HLA-B*27:20(0.63), HLA-A*02:17(0.39), HLA-B*40:13(0.39), HLA-C*12:03(0.38), HLA-A*02:50(0.34), HLA-B*15:02(-0.13), HLA-B*40:02(-0.15), HLA-B*15:03(-0.21)	63.51%
WTSPATTDW	HLA-A*68:23(0.81), HLA-B*15:17(0.66), HLA-C*03:03(0.45), HLA-B*58:01(.18), HLA-A*32:0(0.10), HLA-C*12:03(0.04), HLA-B*27:20(0.02), HLA-A*32:15(-0.19), HLA-B*40:13(-0.19), HLA-B*57:01(-0.28)	40.54%
QEGALHTAL	HLA-B*27:20(0.47), HLA-C*03:03(0.23), HLA-A*68:23(0.08), HLA-B*40:01(0.06), HLA-A*02:50(-0.22), HLA-C*12:03(-0.31), HLA-B*15:02(-0.16), HLA-B*40:13(-0.13), HLA-A*32:15(-0.31), HLA-A*32:07 (-0.36)	40.54%

and beta-turn prediction. Thus, several prediction methods were used to determine the epitope considering all these criteria.

Kolaskar and Tongaonkar antigenicity prediction method functions on the basis of physiochemical properties of amino acids and abundances in experimentally known epitopes.

The antigenic propensity for the protein was 1.00; all the values greater than 1.00 are potential antigenic determinants. Thirteen epitopes were found to satisfy the threshold value set prior to the analysis, and they have the potential to evoke the B-cell response. The peptide from PCRVPIS-VTAN from 335 to 345 was found to have the highest antigenic propensity score. The results are summarized in Table 3A and Figure 3.

Hydrophilic regions are likely to be exposed to the outer surface and most likely to evoke B-cell antigenecity. For this purpose, the protein was subjected to Emini surface accesibilty prediction (Table 3B) and Parker hydrophilicity prediction (Fig. 4). From these 2 prediction methods, the regions around 70 to 80 (more specifically 77–87 in the Emini surface accessibility prediction method), 140 to 156 (more specifically 146–156, surface accessibility prediction method), 240 to 250 (more specifically 242–250 in the Emini surface accessibility prediction method), and 320 to 330 (more specifically 327–332 in the Emini surface accessibility prediction method) were found as hydrophilic. And thus, these regions can be said to be antigenic.

It was found experimentally that the segment of protein that interacted with the antibody was flexible. For this reason, the flexibility of the segment was also counted in the search of antigenicty. In the Karplus and Schulz flexibility prediction method, the region of 330 to 340 was found to be the most flexible (Fig. 5).

Admitting the effect of the beta sheet region in antigenicity, Chou and Fashman beta-turn prediction was done. In this result, approximately100 to 105 and approximately 330 to 340 regions were shown to have the highest beta-turn regions (Fig. 6).

The single-scale amino acid propensity profile cannot always reliably predict the antigenic epitopes, and even the best antigenic prediction method could yield a slightly better score than the receiver operating characteristics (ROC) plot.[42] Because of this, a machine learning process, Bepipred, was done to predict the epitopes with increased success (Table 3C).

By cross-referencing all the data, we hypothesized that the peptide sequence from 330–336 amino acids are capable of inducing the desired immune response as B-cell epitopes.

Discussion

With the advancement of sequencing technologies and the huge disclosure of protein data, development of an epitope-based peptide vaccine has become a key motif in case of viral vaccine

Table 3. Potential B-cell peptide epitope generated by Kolaskar and Tongaonkar antigenicity, Emini surface accessibility prediction and Bepipred linear epitope prediction. **Table 3A**. Kolaskar and Tongaonkar antigenicity.

NO.	START POSITION	END POSITION	PEPTIDE	LENGTH
1	20	36	WIDLVLEGGSCVTVMAP	17
2	53	75	LATVREYCYEATLDTLSTVARCP	23
3	88	99	PTFVCKRDVVDR	12
4	113	124	IDTCAKFTCKSK	12
5	136	146	KYEVAIFVHGS	11
6	166	175	RFTISPQAPS	10
7	185	191	TVTIDCE	7
8	200	207	DYYVFTVK	8
9	249	256	KQTVVALG	8
10	260	298	GALHTALAGAIPATVSSSTLTLQSGHLKCRAKLDKVKIK	39
11	304	310	MCDSAFT	7
12	321	329	GTVIVELQY	9
13	335	345	PCRVPISVTAN	11
14	349	361	LTPVGRLVTVNPF	13
15	372	378	MVEVEPP	7
16	380	388	GDSYIVVGR	9
17	405	410	GKALAT	6
18	416	423	QRLAVLGD	8
19	431	437	IGGVFNS	7
20	439	448	GKAVHQVFGG	10
21	461	473	TQGLLGALLLWMG	13

Table 3B. Emini surface accessibility prediction.

NO.	START POSITION	END POSITION	PEPTIDE	LENGTH
1	35	41	APEKPTL	7
2	77	87	TGEAHNTKRSD	11
3	146	156	STDSTTHGNYF	11
4	159	165	IGKNQAA	7
5	195	202	GINTEDYY	8
6	224	238	PWTSPATTDWRNRET	15
7	242	250	FEEPHATKQ	9
8	311	317	FSKNPAD	7
9	327	332	LQYTGS	6
10	389	394	GTTQIN	6
11	396	401	HWHKEG	6
12	410	415	TTWKGA	6

Table 3C. Bepipred linear epitope prediction.

NO.	START POSITION	END POSITION	PEPTIDE	LENGTH
1	9	18	RDFVEGASGA	10
2	34	41	MAPEKPTL	8
3	73	89	RCPTTGEAHNTKRSDPT	17
4	98	103	DRGWGN	6
5	112	114	SID	3
6	123	126	SKAT	4
7	146	154	STDSTTHGN	9
8	159	162	IGKN	4
9	168	183	TISPQAPSFTANMGEY	16
10	193	198	RSGINT	6
11	224	235	PWTSPATTDWRN	12
12	244	250	EPHATKQ	7
13	257	261	SQEGA	5
14	271	276	PATVSS	6
15	311	321	FSKNPADTGHG	11
16	329	337	YTGSNGPCR	9
17	362	369	ISTGGANN	8
18	376	381	EPPFGD	6
19	390	392	TTQ	3
20	398	405	HKEGSSIG	8
21	412	413	WK	2
22	426	428	WDF	3

preparation. Although Saint Louis encephalitis,since its first recognition in 1933, has been responsible for 10 000 cases and 1000 deaths, the production of the vaccine has not received attention.[2] One study in 1995 was done with the virus in this prospect and now with the help of bioinformatics tools, this admittedly preliminary study was done from a different perspective.[43]

The objective of the study was to investigate the most immunogenic proteins of the Saint Louis encephalitis virus and to find out the T-cell and B-cell epitopes with regard to a burning question: the viral protein mutation. Another positive outcome of the study would be the cost. Currently, where a significant amount of resources are spent for the vaccine development for viruses, this dry lab approach can be proved as a worthy preliminary step not only towards an effective proposition, but also a less costly one. Thus, we believe that this perspective is well worth attempting.

T-cell and B-cell epitopes are the prime molecules of cell mediated and humoral immunity. Thus, for designing a vaccine for protective immunity, the first step should be

to determine these epitopes. All the vaccines at present are mostly based on B-cell immunity, whereas, in this case, a vaccine based on T-cell epitope was inspired by the fact that the host can mediate a strong immune response by CD8+ T cell to infected cell.[44] In this approach, an attempt was made to illuminate a potential but rather overlooked aspect of vaccine development.[45] T-cell based vaccine development is very promising in the sense that due to antigenic drift, the foreign particles can easily escape the antibody memory response. On the contrary, a T-cell immune response tends to be a lasting
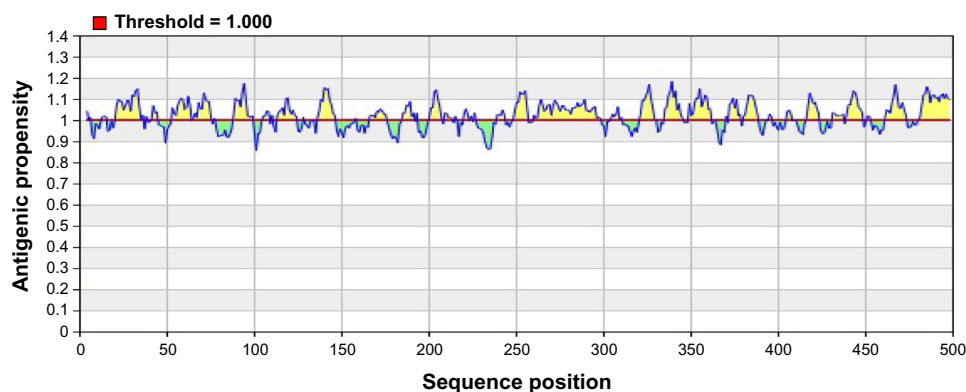
Figure 3. Kolashkar and Tongaonkar antigenicity prediction. Here the x-axis and y-axis represents sequence position and antigenic propensity. The threshold value is 1.000. The regions above the threshold are antigenic, shown in yellow.
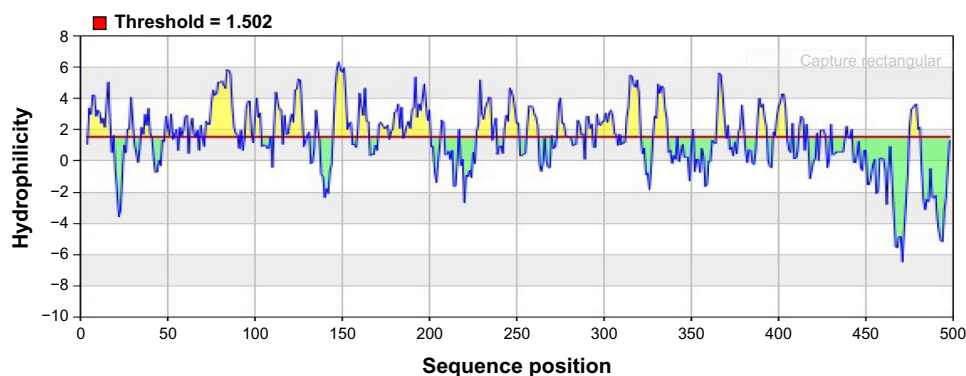
Figure 4. Parker hydrophilicity scale. Here the x-axis and y-axis represent sequence position and hydrophilicity scale, respectively. The threshold is 1.502. The regions above the threshold are hydrophilic, shown in yellow.

one. Along with the T-cell epitope, attention was given also to the B-cell epitope, which can induce primary and secondary humoral immunity.[46]

This study evolved around the cornerstone of 4 criteria of a good peptide epitope: good conservancy, T-cell and B-cell processivity, interaction with MHC-1 allele with good affinity, and population coverage.

The predicted epitope fulfilled all the criteria, while more explanation of some results is given here including the fact that as a RNA virus, Saint Louis encephalaitis has a higher tendency towards mutation resulting from the lack of proofreading activity of RNA polymerase.[47] So, while considering the segment to induce immunity in the human body against this virus, it is of utmost importance to choose a portion with relative conservancy. The most preferable way to do that in a dry lab will be to consider a portion present in all the strains of the virus. By securing a presence in a specific protein in all strains, it actually secures a place in evolution. But unfortunately, due to resource scarcity, among the 17 strains till identified,[1] the database includes only the protein sequence of strain MS1–7. Thus in this approach this problem was tried to solve from a different angle. The 5 most potential T-cell epitopes chosen by the server NetCTL were again subjected to the epitope conservancy tool.

One epitope, REYCYEATL, showed a conservancy of 63.51%, while the other 4 epitopes showed conservancy of only 40.54%. It was assumed that the REYCYEATL epitope could be used as a universal epitope rather than other 4, as it showed higher conservancy than all other potential epitopes.

In the fourth criteria, to make sure the interacting MHC alleles has good population coverage overall, 12 allele supertypes were set prior to the epitope prediction, where supertypes A*0101, A*0201, A*0301, A*2402, B*0702, and B*4403 showed a population coverage of 99%, including North American, where the disease is most prevalent.[48]

Most importantly, REYCYEATL also showed interaction with HLA-A2, which is more common in North America than the other parts of the world.[49] Other four Epitopes did not show interaction with the HLA-A2.

Some investigations were done using multiple computational tools, such as in the case of B-cell epitope identification, which was done using the parameters of hydrophilicity, flexibility, and beta-turn. This emphasizes the fact that although computational algorithms are available, it is still important to check all the criteria *in vitro* to assess the immunogenicity and identify the epitopes of a protein.
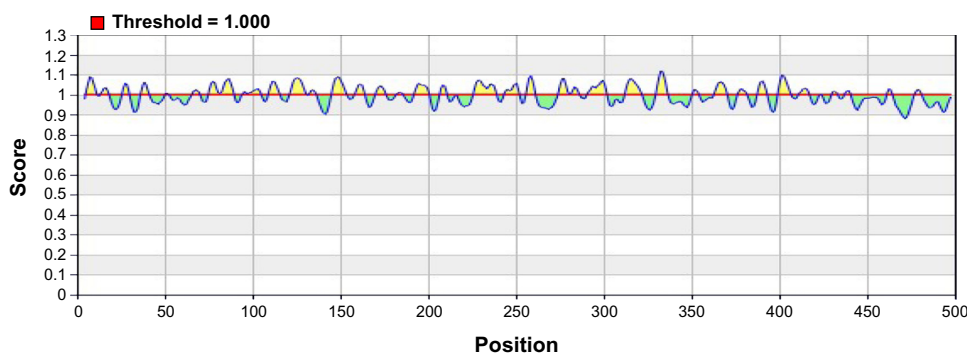
Figure 5. Karplus and Schulz flexibility prediction. Here x-axis and y-axis represent position and score, respectively. The threshold is 1.00. The flexible regions are shown in yellow.
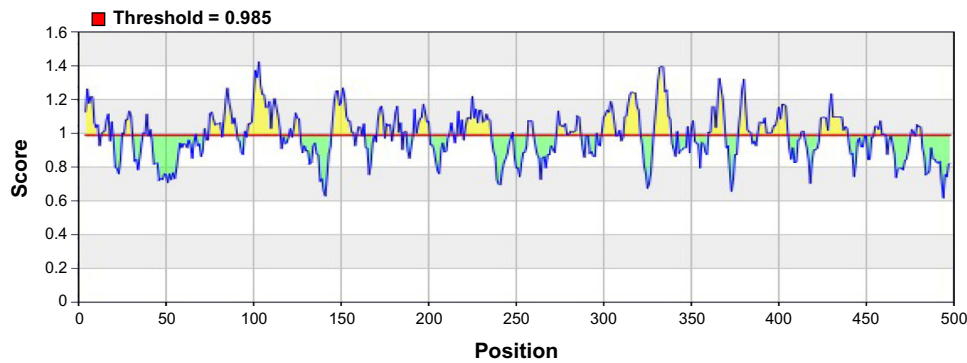
Figure 6. Chou and Fasman beta-turn prediction. The x-axis and y-axis represent position and score, respectively. The threshold is 0.985. The regions having beta-turns are shown in yellow.

The results are based on a careful sequence analysis and deposited data on various immune databases. The results suggest that the epitopes found here are good candidates for establishing a peptide vaccine that can trigger an efficacious immune response *in vivo*.

Conclusion and Future Perspective

To prove the effectiveness of mounting an immune response, both *in vivo* and *in vitro* studies are required along with this *in silico* study. To determine the binding affinity of the whole peptide, the binding chip assay for the HLA and peptide would also be useful.

Author Contributions

Conceived and designed the experiments: MAH. Analyzed the data: MAH, MH. Wrote the first draft of the manuscript: MAH, MH. Contributed to the writing of the manuscript: MAH, MH, MJA. Agree with manuscript results and conclusions: MAH, MH, MJA. Jointly developed the structure and arguments for the paper: MH, MJA. Made critical revisions and approved in all version: MAH. All authors reviewed and approved of the final manuscript.

REFERENCES

1. Kramer LD, Presser SB, Hardy JL, Jackson AO. Genotypic and phenotypic variation of selected Saint Louis encephalitis viral strains isolated in California. *Am J Trop Med Hyg*. 1997;57(2):222–9.
2. Reisen WK. Epidemiology of St. Louis encephalitis virus. *Adv Virus Res*. 2003;61:139–83.
3. Shaman J, Day JF, Stieglitz M, et al. Seasonal forecast of St. Louis encephalitis virus transmission, Florida. *Emerg Infect Dis*. 2004;10(5):802–9.
4. Sejvar JJ, Bode AV, Curiel M, et al. Post-infectious encephalomyelitis associated with St. Louis encephalitis virus infection. *Neurology*. 2004;63(9):1719–21.
5. Reimann CA, Hayes EB, Guiseppi DC, et al. Epidemiology of neuroinvasive arboviral disease in the United States, 1999–2007. *Am J Trop Med Hyg*. 2008;79(6):974–9.
6. Rahal JJ, Anderson J, Rosenberg C, Reagan T, Thompson LL. Effect of interferon-alpha2b therapy on St. Louis viral meningoencephalitis: clinical and laboratory results of a pilot study. *J Infect Dis*. 2004;190(6):1084–7.
7. Trent DW, Qureshi AA. Structural and nonstructural proteins of Saint Louis encephalitis virus. *Journal of Virology*. 1971;7(3):379–88.
8. Testa JS, Philip R. Role of T-cell epitope-based vaccine in prophylactic and therapeutic applications. *Future virol*. 2012;7(11):1077–1088.
9. Seib KL, Dougan G, Rappuoli R. The key role of genomics in modern vaccine and drug design for emerging infectious diseases. *Plos Genetics*. 2009;5(10):e1000612.
10. Hammer J, Gallazzi F, Bono E, et al. Peptide binding specificity of HLA-DR4 molecules: correlation with rheumatoid arthritis association. *J Exp Med*. 1995;181(5):1847–55.
11. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*. 2006;65(1):40–8.
12. Segal NH, Parsons DW, Peggs KS, et al. Epitope landscape in breast and colorectal cancer. *Cancer Res*. 2008;68(3):889–92.
13. Stassar MJ, Raddrizzani L, Hammer J, et al. T-helper cell-response to MHC class II-binding peptides of the renal cell carcinoma-associated antigen RAGE-1. *Immunobiology*. 2001;203(5):743–55.
14. Bourdette DN, Edmonds E, Smith C,et al. A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Mult Scler*. 2005;11(5):552–61.
15. Lopez JA, Weilenman C, Audran R, et al. synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T lymphocyte immune response in humans. Implications for vaccination strategies. *Eur J Immunol*. 2001;31(7):1989–98.
16. Knutson KL, Schiffman K, Disis ML. Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J Clin Invest*. 2001;107(4):477–84.
17. Petrovsky N, Brusic V. Computational immunology: The coming of age. *Immunol Cell Biol*. 2002;80(3):248–54.
18. Brusic V, Bajic VB, Petrovsky N. Computational methods for prediction of T-cell epitopes—a framework for modelling, testing, and applications. *Methods*. 2004;34(4):436–43.
19. Peters B, Bulik S, Tampe R, et al. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol*. 2003;171(4):1741–9.
20. Bhasin M, Raghava GP. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci*. 2004;13(5):596–607.
21. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*. 2005;57(1–2):33–41.
22. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8(4).
23. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 1995;11(6):681–4.
24. Larsen MV, Lundegaard C, Lamberth K, et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*. 2007;8:424.
25. Buus S, Lauemøller SL, Worning P, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach. *Tissue Antigens*. 2003;62(5):378–84.

26. Lundegaard C, Nielsen M, Lund O. The validity of predicted T-cell epitopes. *Trends Biotechnol*. 2006;24(12):537–8.

27. Moutaftsi M, Peters B, Pasquetto V, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol*. 2006;24(7):817–9.

28. Gras S, Kedzierski L, Valkenburg SA, et al. Cross reactive CD8+ T-cell immunity between the pandemic H1N1–2009 and H1 N1–1918 influenza A viruses. *Proc Natl Acad Sci USA*. 2010;107(28):12599–604.

29. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem. 2010;31(2):455–61.

30. Deepak N, Singh K, Siddiqui Z, Nayak B, et al. Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. *J Immunol*. 2002;168(5):2371–82.

31. Fieser TM, John A, Tainer H, et al. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein a-helix. *Proc Natl Acad Sci*. 1987;84(23):8568–72.

32. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*. 1990;276(1–2):172–4.

33. Emini EA, Hughes JV, Perlow DS, et al. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*. 1985;55(3):836–9.

34. Parker JM, Guo D, Hodges RS. 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*. 1986;25(19):5425–32.

35. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins. *Naturwissenschaften*. 1985;72:212–3.

36. Andersen PH, Nielsen M and Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*. 2006;15(11):2558–67.

37. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* 1992;255(5047):959–65.

38. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol*. 1978;47:45–148.

39. Roehrig JT, Mathews JH, Trent DW. Identification of epitopes on the E glycoprotein of Saint Louis encephalitis virus using monoclonal antibodies. *Virology*. 1983;128(1):118–26.

40. Singh S, Taneja B, Salvi SS, Agrawal A. Physical properties of intact proteins may predict allergenecity or lack thereof. *PLoS One*. 2009;4(7):e6273.

41. Robson B, Pain RS. Analysis of the code relating sequence to conformation in globular proteins, the distribution of residue pairs in turns and kinks in the backbone chain. *Biochem J*, 1974;141(3):899–904.

42. Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Science*. 2009;14(1):246–8.

43. Venugopal K, Jiang WR, Gould EA. Immunity to St. Louis encephalitis virus by sequential immunization with recombinant vaccinia and baculovirus derived PrM/E proteins. *Vaccine*. 1995;13(11):1000–5.

44. Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. *J Virol*. 2004;78(15):8312–21.

45. Tan PT, Khan AM, August JT. Highly conserved influenza A sequences as T cell epitopes-based vaccine targets to address the viral variability. *Hum. Vaccine*. 2011;7(4),402–9.

46. Trainor NB, Crill WD, Roberson JA, et al. Mutation analysis of the fusion domain region of St. Louis encephalitis virus envelope protein. *Virology*. 2007;360(2):398–406.

47. Klavinskis LS, Whitton JL, Oldstone MB. Molecularly engineered vaccine which expresses an immunodominant T-cell epitope induces cytotoxic T lymphocytes that confer protection from lethal virus infection. *J Virol*. 1989,63(10):4311–6.

48. Sette S, Sidney J. Nine major HLA Class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*. 1999;50(3–4):210–2.

49. Middleton D, Menchaca L, Rood H, et al. New allele frequency database: http://www.allelefrequencies.net .*Tissue Antigens*. 2003;61(5):403–7.