Original Article

In silico prediction of B- and T- cell epitope on Lassa virus proteins for peptide based subunit vaccine design

Sitansu Kumar Verma, Soni Yadav, Ajay Kumar[1]

Department of Biotechnology, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, [1]Department of Biotechnology, Faculty of Engineering and Technology, Rama University, Kanpur, Uttar Pradesh, India

Abstract

Background: Lassa fever is a severe, often-fatal and one of the most virulent disease in primates. However, the mechanism of escape of virus from the T-cell mediated immune response of the host cell is not explained in any studies yet. In our studies we had aimed to predict B- and T- cell epitope of Lassa virus protein, for impaling the futuristic approach of developing preventive measures against this disease, further we can also study its presumed viral- host mechanism.

Materials and Methods: Peptide based subunit vaccine was developed from all four protein against Lassa virus. We adopted sequence, 3D structure and fold level *in silico* analysis to predict B-cell and T-cell epitopes. The 3-D structure was determined for all protein by homology modeling and the modeled structure validated.

Results: One T-cell epitope from Glycoprotein (WDCIMTSYQ) and one from Nucleoprotein (WPYIASRTS) binds to maximum no of MHC class I and MHC class II alleles. They also specially bind to HLA alleles namely, A*0201, A*2705, DRB*0101 and DRB*0401.

Conclusions: Taken together, the results indicate the Glycoprotein and nucleoprotein are most suitable vaccine candidates against Lassa virus.

Key Words: B-cell, homology modeling, Lassa virus, T-cell

Address for correspondence:
Dr. Sitansu Kumar Verma, Department of Biotechnology, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh - 474 005, India.
E-mail: sitansumtech@gmail.com

Received: 28.02.2014, **Accepted:** 05.01.2015

INTRODUCTION

Lassa fever is an acute viral zoonotic illness caused by Lassa virus, a member of the *Arenaviridae* family and responsible for a severe hemorrhagic fever characterized by fever, sour throat, muscle pain, nausea. Lassa fever was first described in Sierra Leone in 1950s but the virus responsible for the disease was not identified until 1969 when two missionary nurses died in Nigeria, West Africa, and the cause of their illness was found to be Lassa virus, named after the town in Nigeria (Lassa in the Yedseram River valley) where the first cases were isolated.[1,2] There are an estimated 300,000 to 500,000 cases of Lassa fever each year[3,4] with a mortality rate if 15–20% for hospitalized patient. Mortality rate has become high as 50% during epidemic and 90% in third month pregnancy for the expectant mother and the fetus both.[3,2] Since then, a number of outbreaks of Lassa virus infection were reported in various parts of Nigeria including Jos, Zonkwua, Onitsha, Owerri, Abo Mbaise, Lafiya and Epkoma.[4-8] Epidemics of Lassa fever were also documented in other West African countries including

Access this article online	
Quick Response Code:	Website: www.advbiores.net
	DOI: 10.4103/2277-9175.166137

How to cite this article: Verma SK, Yadav S, Kumar A. *In silico* prediction of B- and T- cell epitope on Lassa virus proteins for peptide based subunit vaccine design. Adv Biomed Res 2015;4:201.

Liberia, Sierra Leone, Guinea, Mali and Senegal.[1,7,9] A few cases of the importation of Lassa virus into other parts of the world for example by travelers were documented.

Morphologically, Lassa virus consists of enveloped particles that vary in diameter from approximately 60 to more than 300 nm, with mean particle size of 92 nm. The virus is approximately spherical, enveloped particles that rang in diameter from 50 to 300 nm. Lassa virus having ambisense genomic organization (two viral genes separated by an intergenic region), negative sense, bisegmented, ssRNA genome S (small, ~3.4 kb) and L (large, ~7.2 kb) segments.[10,11] The small segment encodes the 75 KDa glycoprotein precursors (GPC) and the 63 KDa nucleoprotein (NP). After part translation modification GPC is cleaved into GP1 and GP2, the large segment of RNA encodes the 11KDa Z protein, which binds zinc and as Matrix protein and the 200KDa L protein, which acts as matrix protein. Replication and transcription of the genome occur in the cytoplasm of an infected cell and both take place within Rib nucleoprotein complex.[11,12]

Lassa virus is transmitted to human being from the rodent reservoir *Mastomys natalensis*, by direct content with infected tissue, food contaminated with excreta. Mostomys natalensis, is a common rodent in village houses, is therefore primary human infection and common (r) Lassa fever may also spread through person to person contact. Lassa virus trasmition occurs when a healthy person comes in contact with virus in the blood, secretion, tissue or excretion of any infected individual. The virus cannot be transmitting through skin to skin contact without exchange of body fluid. Lassa virus enters the cell via the alpha-dystroglucan receptor.[13] Furthermore, ribavirin should be made available in hospitals and health centers in the endemic areas particularly in rural communities. This would help to control the disease. The aim of this study is to design the peptide based vaccines for Lassa virus.

MATERIALS AND METHODS

Proteome data retrieval and analysis

Z protein (NP_694871), Nucleoprotein (NP_694869), L protein (NP_694872), and Glycoprotein (NP_694870) are the protein of Lassa Virus available in the NCBI Protein data base and hence it is used for this analysis. Bioinformatics tools were used for the analysis of proteome of Lassa virus. The expected molecular weight, isoelectric point (pI) highly repeated amino acids (%) of repetition values were calculated using ExPaSy (http://www.expasy.org/).[14,15] All four proteins were analyzed for antigenicity using VaxiJen web server [Figure 1].

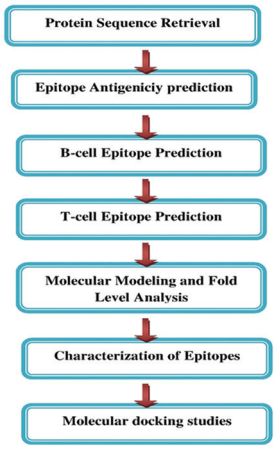

Figure 1: *Insilco* methodology for B and T-cell epitope prediction

B-cell and T-cell epitope prediction

All of these targeted proteins of the Lassa virus strain found namely Z protein Nucleoprotein L protein and Glycoprotein were analyzed for the B-cell epitopes using BCPred.[16] Identify common B-cell epitope by using both BCPred and AAP prediction method of BCPred.[17] B-cell epitope with >0.8 BCPred and >0.4 Vaxijen score were re-elected for the identification of T cell epitope. These selected epitope further subjected to ProPred1 and ProPred analysis.[18-21] Both (Propred1 and Propred) are matrix-based method that allows prediction of MHC binders for various alleles based on the multiplication and additional matrices, proteosome cleavage site, simultaneously. ProPred1 allow predicting MHC class I binding peptide (CTL epitope) for 47 Allele and ProPred to predict MHC class II binding peptide (HTL peptide) for 51 alleles. Common epitope that bind both the MHC classes were selected for further analysis.

These epitopes were analyzed with VaxiJen v2.0 web server. The IC_{50} (inhibitory concentration 50) value of corresponding epitope was deduced from MHCpred server.[22] Epitope having less than 1000 nM IC_{50} values for DRB1*0101 of MHC class II were selected. T epitope designer and MHCpred are the second screening method. T-epitope Designer is a structure based QSAR simulation method to predict HLA-peptide binding based on virtual binding pocked using X-ray crystal structure of HLA-peptide complex.[23] In the second screening, the following selection criteria were used: i) Binding with large number of HLA alleles (>1000), (ii) must bind to A*0201, A*0204 and A*2705. Predicted epitope well further analyzed for fold level topology.

Molecular modeling and fold level analysis

The 3D structure of Glycoprotein was not available in database PDB. Query glycoprotein and template

3MKO after alignment were used as input in Modeller program, giving five models for target. Modeller objective function and dope score helped in the validation of the model of glycoprotein. The validated glycoprotein models were chosen for further studies and refinement. The most acceptable model of NDM-1 was finalized by Ramachandram Plot, providing the position of residue in particular segment based on phi (Ø) and psi (ψ) angles between N-Cα and Cα-C atoms of residue. The theoretical model generated was validated by using the programs PROCHECK[24,25] and ProSA.[26] PROCHECK is a suite of programs to check the stereo chemical quality of protein structure. It includes parameters such as bond length, bond angle, main chain and side chain properties and residue-by residue properties to assess whether the geometry of the residues in a given protein structure is normal or unusual.[25] ProSA (Protein Structure Analysis) program exploits interactive web- based applications to check the three-dimensional models of protein structures for potential errors by displaying scores and energy plots.

Pepitope servers were used to analyze the folding and cluster of selected epitopes in folded protein PepSurf and Mapitope algorithms used by this server to predict fold level topology. PepSurf algorithm helps to map the epitope onto the surface of the antigen.[27] By Mapitope algorithms epitopes shared by the entire set of peptides are detected by the following steps user carried out to predict fold level topology. i) Prediction algorithm was executed and ii) 3D structure of predicted epitope are visualized using 3D structure viewer.

Characterization of epitopes

DISTELL server was used to design the 3D structures of the predicted binding peptides based on the similarity with PDB temples.[28] After designing the structures epitopes were then validated with PROCHECK and ProSA. For prediction of domain, motif and functionality of epitope ProFunc, Motif Scan and InterProScan are used. ProteinDigest web server was used to predict pI, molecular weight and enzyme degradation site.[29]

Molecular docking studies

After designing the epitopes structures molecular docking of selected alleles and epitopes was performed with the help of Autodock.[30] This program used a simulated annealing approach to explore the conformation space between the ligand and target protein. The energy evaluation process is done by using grid-based molecular affinity potential, the docking was performed on the basis of Lamarckian Genetic Algorithm. The PMV (Python Molecular Viewer) was used for the visualization of Binding, position, H-bonding between the selected peptides and alleles.

RESULTS

Physicochemical analysis and antigenicity prediction

The L protein has the highest molecular weight of about 253431.9 KDa which consists of Leucine (L) a neutral nonpolar amino acid residue has the highest percentage of repetition (11.9%). The physicochemical properties of putative proteins are given in Table 1. The pI value of any protein indicates the stability of protein in that particular isoelectric point. Isoelectric points of these proteins were ranged 6.17 to 8.95.

Prediction of B and T-cell epitope

Epitopes which are capable to induce both type immunity (B-cell and T-cell) are known to be good vaccine candidate. For identification of these epitopes amino acid sequence of all four proteins were subjected to BCPred for B- cell epitope prediction. B-cell epitope prediction is the initial step for vaccine construct. BCPred is web based method that uses a novel method of subsequence kernel which was used to predict linear B-cell epitope from each protein. BCPed and AAP (amino acid pair) methods are used to predict fixed length epitope [Table 2]. Finally, two out of three B-cell epitope from Z protein, seven out of ten epitope from Glycoprotein, seven out of fourteen epitope from Nucleoprotein and sixteen out of twenty five epitope from L protein were selected for further analysis. These selected epitopes were analyzed for T-cell epitope identification. On the first level, sequence based 2D screening propred 1, propred and MHCPred were used to identify the potential T-cell epitope. Default ProPred1 and Propred parameters were used to determine the T-cell epitope [Table 3].

Common epitopes that bind with both MHC class I and II, and interact with DRB1*0101 are shown in Table 4. Epitopes with best Vaxijen and MHCPred score were selected for further analysis. T-epitope

Table 1: It comprises the data of Lassa virus proteins, molecular weight, pI and percentage of highly repeated amino acid residues in individual protein with antigenisity score

Proteins name	Gene ID	PDB ID	Length	M. wt	pI value	Highly repeated amino acid	Antigenisity score
Z protein	23343513	2M1S	99	10675.4	8.95	Pro (15.2%)	0.5953
Glycoprotein	23343511	3Q7B	491	55813.4	7.54	Leu (11.4%)	0.6309
Nucleoprotein	23343510	-	569	62998.2	8.64	Leu (12.0%)	0.4190
L protein	23343514	4MIW	2218	253431.9	6.17	Leu (11.9%)	0.4417

Table 2: B-cell epitope prediction using BCPred (BCPred+AAP) and anigenicity of peptide using Vaxijen

Proteins	Amino acid position	BCPred epitope sequence	Gravy score	BCPred score	Vaxijen score
Z protein	21	PDATHLGPQFCKSCWFENKG	−0.760	0.86	0.9492
	72	PTKLRPSAAPTAPPTGAADS	−0.560	1	0.6139
	6	AKAPESKDSPRASLIPDATH	−0.920	1	0.4161
Glycoprotei	471	MGICSCGLYKQPGVPVKWKR	−0.175	0.994	0.4940
	95	IMVGNETGLELTLTNTSIIN	0.455	0.988	0.9224
	260	GTFTWTLSDSEGKDTPGGYC	−0.755	0.973	0.7164
	74	NMETLNMTMPLSCTKNNSHH	−0.865	0.894	0.7812
	148	NQYEAMSCDFNGGKISVQYN	−0.785	0.879	0.7293
	203	LDSGRGNWDCIMTSYQYLII	0.020	0.807	1.2077
	293	FGNTAVAKCNEKHDEEFCDM	−0.815	1	0.6083
	258	LLGTFTWTLSDSEGKDTPGG	−0.435	1	0.5385
	96	MVGNETGLELTLTNTSIINH	0.070	0.96	0.8940
	208	GNWDCIMTSYQYLIIQNTTW	−0.175	0.938	1.1256
Nucleoprotein	339	ESDGKPQKADSNNSSKSLQS	−1.870	0.996	0.4149
	381	DPNAKTWMDIEGRPEDPVEI	−1.140	0.995	0.7433
	512	VVVEKKKRGGKEEITPHCAL	−0.585	0.995	0.8202
	760	MGICSCGLYKQPGVPVKWKR	−0.175	0.94	0.4940
	140	LLNMIGMSGGNQGARAGRDG	−0.335	0.977	1.0227
	43	DFSEVSNVQRLMRKERRDDN	−1.730	0.951	0.7703
	311	CLSGDGWPYIASRTSITGRA	−0.120	0.756	0.4334
	321	ASRTSITGRAWENTVVDLES	−0.425	1	0.7243
	109	LKSKVIRTERPLSAGVYMGN	−0.280	1	0.4308
	57	ERRDDNDLKRLRDLNQAVNN	−1.975	1	0.5411
	760	MGICSCGLYKQPGVPVKWKR	−0.175	1	0.4940
	142	NMIGMSGGNQGARAGRDGVV	−0.295	1	0.9323
	582	FGNTAVAKCNEKHDEEFCDM	−0.815	1	0.6083
	509	HTGVVVEKKKRGGKEEITPH	−1.205	0.996	1.0983
L Protein	739	STNPADTATEEELDDMVYNA	−0.925	0.988	0.4411
	304	NVDEGNEKRGNQTIGECQVQ	−1.495	0.978	0.5006
	468	PTISYQRTEEETFPYVMGDV	−0.675	0.972	0.6184
	1659	NGIQYWQVPLELRNGSGGES	−0.810	0.951	0.8711
	2021	SGPTFKGKSAWYVDTEIINE	−0.580	0.951	0.7039
	1857	VATTRTLGSFTWFPHKTSVP	−0.080	0.917	0.5513
	702	NVSYMCHFITKETPDRLTDQ	−0.760	0.915	0.4975
	2064	QFTEYDFVLVGPCTEPTPLV	0.285	0.914	1.0446
	985	CNAVYENSRLKQKYFYCGHM	−0.755	0.9	0.5966
	1181	SLAHVSYSMDHSKWGPMMCP	−0.240	0.873	1.0187
	138	IPLIDGRTSFYNEQIPDWVN	−0.425	0.87	0.8509
	551	YQKTGECSKCYAINDNKVGE	−1.105	0.863	0.6133
	1796	SLVSHIVKWKREEHYIVLSS	−0.095	0.849	0.4580
	1619	GLCRTLGSKCVRGPNKENLY	−0.615	0.803	0.4306
	2052	TPVEIVVDMERLQFTEYDFV	0.140	1	1.1971
	460	DFDVSGVVPTISYQRTEEET	−0.630	1	0.6240
	735	FGHVSTNPADTATEEELDDM	−0.815	1	0.6620
	1823	RTHEPMVEERVVSTSDAVDN	−0.865	1	0.4316
	506	VNSMKTSSTVKLRQNEFGPA1	−0.670	1	0.9756
	1657	NSNGIQYWQVPLELRNGSGG	−0.810	1	0.6851
	2141	WSGVDIVSTLRAAAPSCEGI	0.560	1	0.5917
	1285	HISSILDMGQGILHNTSDFY	0.010	0.999	0.5058
	2020	LSGPTFKGKSAWYVDTEIIN	−0.215	0.999	0.7419
	2189	MVSSGGKLRLKGRTCEELT	−0.575	0.994	0.6502
	1099	YSQESPQSYDSVGPDTGRLK	−1.420	1	0.5213

Designer was used to predict identified epitope binding abilities to >1000 MHC allele. T-epitope Designer is implemented based on a MHC-peptide prediction model described recently.

Table 3: Common epitope (B and T-cell) along with their various parameters (selected epitope are highlighted in bold)

Protein	Epitope	Amino acid position	Vaxijen score	IC50 value of epitope for DRB1* 0101 (MHC pred)	Total no of MHC binding allele (propred1 and propred)
Z protein	PDATHLGPQ	21	1.7474	105.44	98
	PTKLRPSAA	**72**	**1.0205**	**10.26**	**98**
Glycoprotein	FTWTLSDSE	62	0.5639	5.01	98
	NMETLNMTM	**74**	**1.0555**	**103.75**	**91**
	YEAMSCDFN	150	0.5134	150.66	98
	LDSGRGNWD	203	1.2796	150.66	50
	LLGTFTWTL	258	0.5967	267.30	89
	METLNMTMP	76	0.5457	27.23	53
	WDCIMTSYQ	**210**	**1.1117**	**70.96**	**64**
Nucleoprotein	ESDGKPQKA	339	0.4261	17.10	94
	WMDIEGRPE	387	1.5996	78.16	35
	YKQPGVPVK	479	1.3483	123.03	36
	LLNMIGMSG	140	0.7761	564.94	61
	WPYIASRTS	**317**	**1.3451**	**293.76**	**64**
	ASRTSITGR	321	1.4497	165.20	62
	MSGGNQGAR	146	0.5655	285.10	59
L Protein	STNPADTAT	739	0.5346	121.06	67
	YSQESPQSY	1099	0.4556	10.84	76
	NVDEGNEKR	304	0.7064	143.55	60
	FKGKSAWYV	2025	1.3307	99.54	61
	LVGPCTEPT	2072	0.6472	240.99	59
	YENSRLKQK	989	0.8622	25.00	57
	WKREEHYIV	1804	1.2097	363.08	45
	VVDMERLQF	**2057**	**1.1078**	**38.19**	**62**
	IPLIDGRTS	**138**	**1.6689**	**765.60**	**66**
	FGHVSTNPA	**735**	**0.7505**	**3.52**	**75**
	MVEERVVST	1828	0.4598	89.74	60
	MKTSSTVKL	509	0.5398	88.72	46
	VPLELRNGS	1666	1.9587	161.81	47
	ILDMGQGIL	1289	1.0798	45.71	75
	LSGPTFKGK	2020	1.5080	349.95	60
	VSYMCHFIT	703	0.5034	1210.60	64

Table 4: 3D QSAR based T-cell epitope prediction using T epitope designer

Protein	Epitope	% of binders	Lowest score	Highest score
Z protein	PTKLRPSAA	68.56	10.12 (B*1512)	2078.68 (C*0702)
Glycoprotein	NMETLNMTM	54.10	13.49 (A*0233)	1118.51 (A*0217)
	WDCIMTSYQ	96.57	2.22 (C*0805)	1519.41 (B*0817)
Nucleoprotein	WPYIASRTS	99.82	1015.00 (A*0104)	2746.34 (B*0709)
L Protein	VVDMERLQF	98.95	4.05 (A*3206)	2005.83 (C*0107)
	IPLIDGRTS	93.87	3.61 (A*0108)	1586.95 (B*0704)
	FGHVSTNPA	96.65	46.55 (A*0240)	2354.75 (A*0240)

Modeling and fold level analysis

Modeller 9.10 generated the 3-D model of query glycoprotein protein at resolution of 1.80 Á based on template protein 3MKO with sequence query coverage 26%. PROCHECK analysis of the model structure from Modeller 9.10 through Ramachandran plot showed that 96.1% of the residues were in favored and additional allowed regions [Figure 2]. Qualitative assessment of the model through ProSA analysis revealed that the model matched with the NMR region of the plot with Z score of -8.02. GROMOS96 which was used for energy minimization optimized the model structure from initial energy -3568.30 KJ/mol to final energy of -9434.68 KJ/mol.

Positional fold level topologies of predicted epitope on theoretical models of proteins were predicted using Pepitope web server. Figure shows that the epitopes were present within the cluster on the surface of the

protein. The glycoprotein is antigenic and one epitope WDCIMTSYQ from cluster I (score: 8.1608, Residue No. 8) was found to be antigenic (Vaxijen score: 1.1117) and can bind 64 MHC molecule of both MHC I and II molecule. The IC_{50} values of this epitope for DRB1*0101 and DRB1*0401 were 70.96 and 186.64, respectively, which indicates a good inhibition this epitope has also been found to bind all selected MHC molecules (A*0201, A*0204, B*2705) and 96.57% HLA molecule of T epitope designer shown in Table 5. The Nucleoprotein is also antigenic and one epitope WPYIASRTS from cluster I (Score: 10.301 Residue No. 8) was found to be antigenic (VaxiJen score: 1.3451) and can bind 64 MHC molecule of both MHC class I and II. The IC_{50} values of this epitope for DRB1*0101 and DRB1*0401 were 293.76 and 146.89, respectively, which indicates a good inhibition this epitope has also been found to bind all selected MHC molecules (A*0201, A*0204, B*2705) and 96.57% HLA molecule of T epitope designer [Figure 3].

Epitope characterization

After designing the 3D structures both the tools (PROCHECK and ProSA-web) are used for model validation. Furthermore no motif or domain could be assigned using ProFunc, Motif Scan and InterProScan for both proteins. Calculated M. wt. and pI of the WDCIMTSYQ from glycoprotein were 1146.30 Da and 3.08 respectively. Protein was found to be un digested with Tripsin, Clostripain, Proline Endopept, Staph Protease, Trypsin K, Trypsin R, as analyzed by ProteinDigest server for epitope WPYIASRTS from nucleoprotein, M. wt. and pI were 1080.21 and 8.75 respectively, and found to be undigested by Trypsin, Cyanogen_Bromide, Staph_Protease, Trypsin_K, AspN.

Molecular docking analysis

Some of the conserved peptides (WDCIMTSYQ, WPYIASRTS) with their interaction energies in kcal mol^{-1} are given in Table 6 and shown in Figures 4 and 5. These peptides are promiscuous HLA binders. It will be useful to include these peptides in a chimeric constructs containing both cytotoxic and

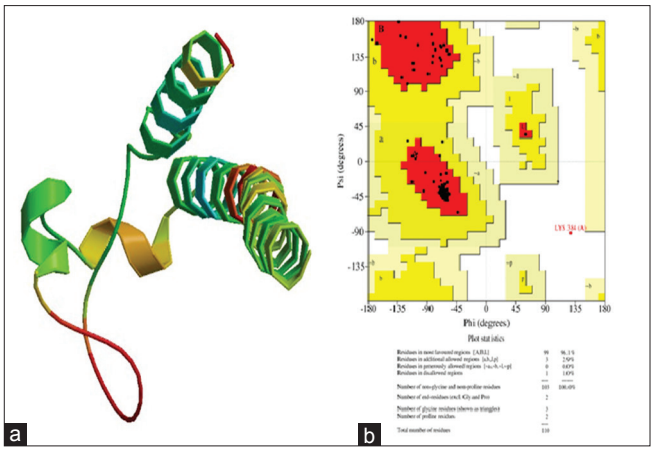

Figure 2: (a) 3D model of Glycoprotein, (b) Ramachandram plot for Glycoprotein model

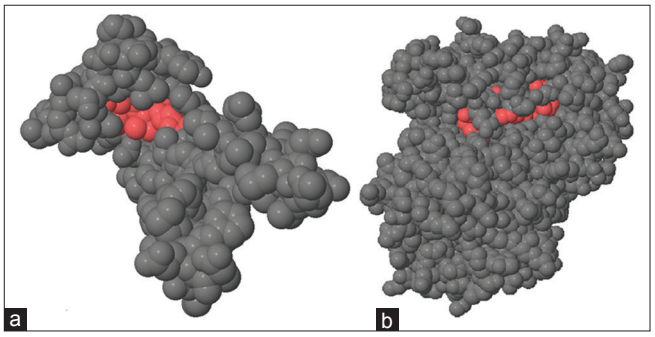

Figure 3: Fold level topology of epitope analyzed by Pepitope server (a) Glycopeotein epitope (WDCIMTSYQ) shown in red colour, (b) Nucleoprotein epitope (WPYIASRTS) shown in red color

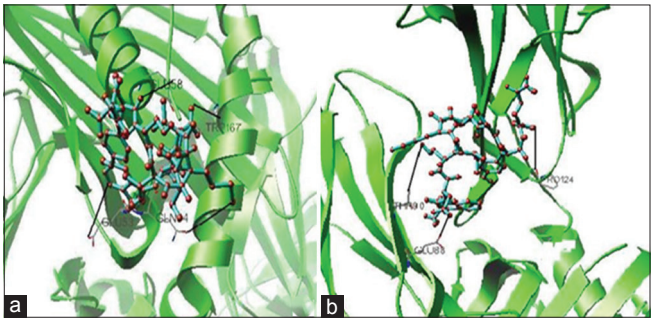

Figure 4: Two best docking conformation ((a). WDCIMTSYQ with, (b). WDCIMTSYQ with 1KLG) analyzed by Python Molecular Viewer (docked ligand shown by balls and sticks while hydrogen bonds shown by black sticks)

Table 5: The selected epitopes showing MHC binding and inhibition values predicted from 3D QSAR based T-epitope designer and MHCPred server

Protein	Epitope	T-epitope designer			MHCPred (IC_{50}) value	
		A*0201	A*0204	B*2705	DRB1*0101	DRB1*0401
Z protein	PTKLRPSAA	−398.56	−1196.00	−208.82	10.26	1006.93
Glycoprotein	NMETLNMTM	141.54	611.40	−1274.61	103.75	1309.18
	WDCIMTSYQ	354.54	473.42	1192.34	70.96	186.64
Nucleoprotein	WPYIASRTS	1509.02	1001.79	2262.06	293.76	146.89
	VVDMERLQF	472.32	277.89	−562.73	38.19	558.47
L Protein	IPLIDGRTS	199.63	−343.39	1345.75	765.60	1644.37
	FGHVSTNPA	46.55	−286.72	981.52	3.52	616.60

Table 6: The Docking results of interactions obtained with docking energy, build hydrogen bond and active site residues

Epitope	HLA	PDB ID	Interaction energies	Build hydrogen bonds	Active site residue
WDCIMTSYQ	A*0201	1AOS	−19.03	2	ARG27, TRP167
	B*2705	1HSA	−19.44	4	GLN54, GLU53, TRP167, GLU58
	DRB1*0101	1KLG	−19.81	3	GLU88,THR90,PRO124
	DRB1*0401	1J8H	−18.05	3	SER51, PRO199, GLY40
WPYIASRTS	A*0201	1AOS	−20.67	3	ASP61, TYR48,LYS103
	B*2705	1HSA	−17.94	3	PHE22, SER38, PRO43
	DRB1*0101	1KLG	−18.83	2	PRO114, ILE82
	DRB1*0401	1J8H	−19.56	3	HIS9, TYR191, PRO177

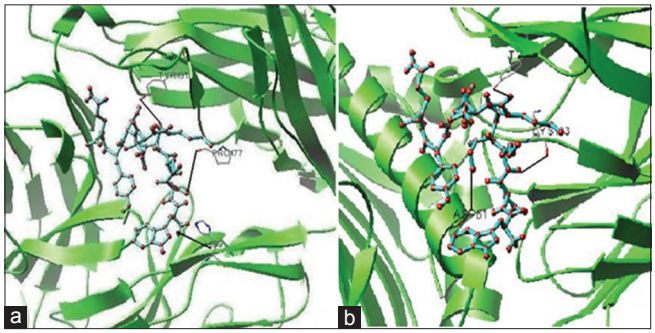

Figure 5: Two best docking conformation ((a). WPYIASRTS with 1J8H, (b). WPYIASRTS with 1AOS) analyzed by Python Molecular Viewer (docked ligand shown by balls and sticks while hydrogen bonds shown by black sticks)

helper epitopes. It is expected that though this T-cell vaccine would not prevent Lassa virus infection, it would aid in quick clearance of the virus and prevent the severe infection.

DISCUSSION

In the present study, three putative proteins of Lassa virus were used for the physicochemical analysis such as molecular weight, isoelectric point (pI value) and antigenic nature. Vaxijen was used to predict antigenisity, which is based on auto cross covariance (ACC) transformation of protein sequence into uniform vectors of principal aminoacid properties. The l00-CV (leave one – out cross validation) was used to identify antigenicity of protein with 91% sensitive, 82% accuracy and 72 specificity for viral sheces. The resultant protein was antigenic.

Epitopes which are capable to induce both type immunity (B-cell and T-cell) are known to be good vaccine candidate. Propred1 and ProPred are matrix-based methods that allow prediction of MHC binders for various alleles based on the multiplication and additional matrices, proteosome cleavage site, simultaneously. This is based on the observations made in previous studies which demonstrate that MHC binders having proteosome cleavage site at their C terminus have high potency to become T-cell epitopes.

MHCPred runs as a CGI server, and uses partial least square base approach for the identification of binding affinity to MHC molecule. MHCPred server generated IC_{50} nM values as output. A lover value of IC_{50} shows higher affinity with MHC molecule. The epitope bind with maximum number of alleles was selected for molecular docking analysis. The Pepitope server is a web based tool to predict discontinuous epitope based on set of peptide that have affinity against a monoclonal antibody or peptide. The server aligns a linear peptide sequence on to a 3D protein structure. The DISTILL model server was used to design the 3D structures of the predicted binding peptides. It is important to identify those peptides which are conserved across the various strains of Lassa virus and in this study that has been shown for the conserved peptides present in the constituent proteins of Lassa virus. The analysis reveals that there are number of suitable peptides from all four which may be included in the construction of poly epitopes T-cell vaccine.

CONCLUSION

The screening of putative epitopes using bioinformatics tools thus suggests that Glycoprotein and Nucleoprotein protein of Lassa virus could be used for preparation of immunological constructs. Molecular simulation and binding tests also suggest that the two nonameric epitopes WDCIMTSYQ and WPYIASRTS predicted and reported for the first time have considerable binding with MHC molecules and low energy minimization values providing stability to the peptide-MHC complex. These peptide construct will further undergo for wet lab studies, for the development of targeted vaccine against Lassa virus strains. Using a similar approach the short listing of candidate epitopes for vaccine design using other proteins can also be targeted that would reduce time and experimental expense.

ACKNOWLEDGEMENT

The authors are grateful to Director, Madhav Institute of Technology and Science, Gwalior, M.P., India for providing necessary facilities and encouragement. The authors

are also thankful to all faculty members of the Institute of Department of Biotechnology, Madhav Institute of Technology and Science, Gwalior, M.P., India for their generous support during the course of experimental work and manuscript preparation.

REFERENCES

1. Frame JD, Baldwin JM, Gocke DJ, Troup JM. Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. Am J Trop Med Hyg 1970;19:670-6.
2. McCormick JB, King IJ, Webb PA, Scribner CL, Craven RB, Johnson KM, *et al*. Lassa fever: Effective therapy with Ribavirin. New Eng J Med 1986;314:20-6.
3. McCormick JB, King IJ, Webb PA, Johnson KM, O'Sullivan R, Smith ES, *et al*. A case-control study of the clinical diagnosis and course of Lassa fever. J Infect Dis 1987;155:445-55.
4. Johnson KM, McCormick JB, Webb PA, Smith ES, Elliot LH, King IJ. Clinical virology of Lassa fever in hospitalized patients. J Infect Dis 1987;155:456-64.
5. Carey DE, Kemp GE, White HA, Pinneo L, Addy RF, Fom AL, *et al*. Lassa fever. Epidemiological aspects of the 1970 epidemic, Jos, Nigeria. Trans R Soc Trop Med Hyg 1972;66:402-8.
6. Bowen GS, Tomori O, Wulff H, Casals J, Noonan A, Downs WG. Lassa fever in Onitsha, East Central State, Nigeria in 1974. Bull World Health Organ 1975;52:599-604.
7. Monath TP. Lassa fever: A review of the epidemiology and epizootiology. Bull World Health Organ 1975;52:577-92.
8. Tim M. Public health leaflets on what your need to know about Lassa fever. Abuja, Nigeria: Federal Ministry of Health; 2000.
9. Monath TP, Maher M, Casals J, Kissling RE, Cacciapuoti A. Lassa fever in the Eastern Province of Sierra Leone, 1970–1972. II. Clinical observations and virological studies on selected hospital cases. Am J Trop Med Hyg 1974;23:1140-9.
10. Schmitz H, Kohler B, Laue T, Drosten C, Veldkamp PJ, Günther S, *et al*. Monitoring of clinical and laboratory data in two cases of imported Lassa fever. Microbes Infect 2002;4:43-50.
11. Gunther S, Emmerich P, Laue T, Kuhle O, Asper M, Jung A, *et al*. Imported Lassa fever in Germany: Molecular characterization of a new Lassa virus strain. Emerg Infect Dis 2000;6:466-76.
12. Lozano ME, Posik DM, Albarino CG, Schujman G, Ghiringhelli PD, Calderon G, *et al*. Characterisation of arenaviruses using a family-specific primer set for RT-PCR amplification and RFLP analysis. Its potential use for detection of uncharacterised arenaviruses. Virus Res 1997;49:79-89.
13. Keenlyside RA, McCormick JB, Webb PA, Smith E, Elliott L, Johnson KM. Case-control study of Mastomys natalensis and humans in Lassa virus-infected households in Sierra Leone. Am J Trop Med Hyg 1983;32:829-37.
14. Kyte J, Doolittle FR. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105-32.
15. Shehzadi A1, Ur Rehman S, Idrees M. Promiscuous prediction and conservancy analysis of CTL binding epitopes of HCV 3a viral proteome from Punjab Pakistan: An in silico approach. Virol J 2011;8:55.
16. El- Manzalawy Y, Dobbs D, Honavar V. Predicting liniar B-cell epitopes using string Kernels. J Mol Recognit 2008;21:243-55.
17. Chen J, Liu H, Yang J, Chou KC. Prediction of liniar B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 2007;33:423-8.
18. Singh H, Raghava GP. ProPred1: Prediction of promiscuous MHC Class-I binding sites. Bioinformatics 2003;19:1009-14.
19. Saraswat A, Shraddha, Jain A, Pathak A, Verma SK, Kumar A. Immuno-informatic speculation and computational modeling of novel MHC-II human leukocyte antigenic alleles to elicit vaccine for ebola virus. J Vaccin 2012;3:1-3.
20. Shekhar K, Dev K, Verma SK, Kumar A. In-silico screening and modeling of CTL binding epitopes of crimean congo hemorrhagic fever virus. Trends Bioinformatics 2011;514-24.
21. Verma SK, Yadav SP, Kumar A. In silico T cell antigenic determinants from proteome of H1N2 swine influenza A virus. Online J Bioinformatics 2011;12:371-8.
22. Gaun P, Doytchinova IA, Zygouri C Flower DR. MHCPred: A server for quantitative prediction of peptide MHC binding. Nuclic Acids Res 2003;31:362-4.
23. Kangueane P, Sakharkar MK. T-Epitope Designer: A HLA-peptide binding prediction server. Bioinformation 2005;1:21.
24. Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Protein Struct Funct Genet 1995;23:318-26.
25. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283-91.
26. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acid Res 2007;35:407-10.
27. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, Freund NT, *et al*. Pepitope: Epitope Mapping from affinity-selected peptides, Bioinformatics 2007;23:3244.
28. Baú D, Martin AJ, Mooney C, Vullo A, Walsh I, Pollastri G. Distill: A suite of web servers for the prediction of one-, two- and three- dimentional structural features of proteins. BMC Bioinformatics 2006;7:402.
29. Laskowski RA, Watsons JD, Thornton JM. ProFunc A server for predicting protein function from 3-D structure. Nucleic Acids Res 2005;33:89-93.
30. Goodsell DS, Morris GM, Halliday RS, Huey R, Belew RK, Olson AJ. Automated Docking using a Lamarckian Genetic Algorithm and Empirical Binding free energy function. J Comput Chem 1998;19:1639-62.

Source of Support: Nil, **Conflict of Interest:** None declared.