**In silico** screening and epitope mapping of leptospiral outer membrane protein—Lsa46

Junaida M. Ibrahim, Shanitha A., Achuthsankar S. Nair, Oommen V. Oommen and Perumana R. Sudhakaran

Department of Computational Biology & Bioinformatics, University of Kerala, Thiruvananthapuram, Kerala, India

Communicated by Ramaswamy H. Sarma.

**ABSTRACT**

Leptospirosis is one of the neglected diseases caused by the spirochete, *Leptospira interrogans*. Leptospiral surface adhesion (Lsa) proteins are surface exposed outer membrane proteins present in the pathogen. It acts as laminin and plasminogen binding proteins which enable them to infect host cells. The major target for the development of vaccine in the current era focuses on surface exposed outer membrane proteins, as they can induce strong and fast immune response in hosts. Therefore, the present study mapped the potential epitopes of the Leptospiral outer membrane proteins, mainly the surface adhesion proteins. Protein sequence analysis of Lsa proteins was done by in silico methods. The primary protein sequence analysis revealed Lsa46 as a suitable target which can be a potent Leptospiral vaccine candidate. Its structure was modelled by threading based method in I-TASSER server and validated by Ramachandran plot. The predicted epitope’s interactions with human IgG, IgM(Fab) and T-cell receptor TCR(β) were performed by molecular docking studies using Biovia Discovery studio 2018. One of the predicted B-cell epitopes and the IgG showed desirable binding interactions, while four of the predicted B-cell epitopes and T-cell epitopes showed desirable binding interactions with IgM and TCR respectively. The molecular dynamic simulation studies carried out with the molecular docked complexes gave minimized energies indicating stable interactions. The structural analysis of the entire simulated complex showed a stable nature except for one of the Epitope-IgM complex. Further the binding free energy calculation of eight receptor-ligand complex predicted them energetically stable. The results of the study help in elucidating the structural and functional characterization of Lsa46 for epitope-based vaccine design.

**Introduction**

Leptospirosis is one of the wide spread and neglected zoonotic diseases world-wide. The causative organism of Leptospirosis is mainly *Leptospira interrogans*. The genus Leptospira consists of genetically diverse group of non-pathogenic or saprophytic, pathogenic as well as intermediate species (Mohammed et al., 2011; Nascimento et al., 2004). They are classified in to more than 350 serovars including pathogenic and saprophytic strains (Karpagam & Ganesh, 2020). The mode of infection is through direct skin contact with the soil, water and food contaminated with the urine of infected animals or through mucous membrane of conjunctiva or oral cavity. Infections caused by the leptospirosis include wide spectrum of clinical manifestations from less severe symptoms like fever, chills, head ache and myalgias to more severe symptoms like jaundice, renal failure and haemorrhage (Bharti et al., 2003; Fraga et al., 2011). The severe form of leptospirosis is called Weils disease. This zoonotic infestation is mainly reported in Caribbean, Central and South America, South East Asia and Oceania (Soo et al., 2020). According to the Ministry of Health and Social Services report, Seychelles has reported the highest annual incidence of leptospirosis (Pappas et al., 2008). Leptospirosis is endemic to most of the states in India, especially in Andaman & Nicobar Islands, Kerala, Maharashtra, Gujarat and Tamil Nadu (Sehgal et al., 1995; Vijayachari et al., 2008).

The common treatment available for the disease is the usage of antibiotics such as β-lactams and macrolides (Goorant, 2016) and also as a mode of management of the disease, several peptides as vaccines have been developed. There have been numerous descriptions of problems associated with bacterin-type vaccines, including severe side effects (Martinez et al., 2004; Yan et al., 2003). There are more than 20 subunit vaccines identified, but failed to provide protection in clinical trials. (Prasad et al., 2020). The currently available vaccines are not serovar-specific and required regular annual booster immunizations (Dellagostin et al., 2011; Fraga et al., 2011), so there is a need for a well conserved vaccine candidate which can show cross protection against number of serovars of pathogenic leptospires. Full length genome sequence of many of the leptospiral species and different serovars are available currently; this information and datasets can be used for the identification of successful
leptospiral vaccine candidates (Grassmann et al., 2017). One of the pathogenic serovars of *Leptospira interrogans* i.e. *L. interrogans* serovar Copenhageni strain L1-130 genome revealed the occurrence of many proteins, both surface exposed and outer membrane proteins having trans-membrane helices. These show a high degree of conservation among the other pathogenic species of leptospires.

Most of the membrane proteins can mediate host-pathogen interactions, in which many of them are surface exposed proteins (Adler, 2015a; Haake & Matsunaga, 2010). The surface exposed proteins are considered as ideal vaccine candidate due to the high susceptibility to antibody recognition and protective immune response elicitation in the host organisms (Koizumi & Watanabe, 2005; Zeng et al., 2017). The surface exposed proteins present in the pathogens mainly mediate the adherence to host tissues during infection. Leptospiral outer membrane proteins Lsa46 and Lsa77 showed their capacity to induce protection and sterilizing immunity against hamster model which were infected with virulent leptospires. The Lsa46 and Lsa77 were earlier characterized as plasminogen and laminin binding proteins (Teixeira et al., 2018). Recombinant Lsa46 and Lsa77 proteins were reactive to laminin, plasminogen and plasma fibronectin; but showed a less binding with other ECM components like collagen, elastin and fibronectin and other plasma components like fibrinogen, vitronectin, C4BP etc. when tested by ELISA. The immunofluorescence assay with Lsa46 and Lsa77 revealed their presence in bacterial cell surface (Teixeira et al., 2015) also. Lsa25 and Lsa33 are two other leptospiral surface adhesion proteins that mediate colonization of host by binding to ECM component laminin (Domingos et al., 2012). Lsa24, one of the membrane proteins found in the *L. interrogans* plays a role in mediating its adhesion to the host cell. The penetration of the pathogen in to the host is one of the crucial steps causing the colonization of pathogen in the host, mainly by adhesion-mediated binding. Lsa24 is the first leptospiral adhesion protein with laminin binding properties reported (Barbosa et al., 2006).

The present study focuses on *in silico* predicting of leptospiral surface adhesion proteins to find an alternative vaccine candidate with cross protection for the management of the disease. Several online servers and tools for the sequential and structural analyses of different Lsa proteins were employed in the study. On the basis of the computational analysis, one of the Lsa proteins, Lsa46 was chosen as a better candidate and is used for molecular docking and dynamic simulation. The stability of receptor–ligand binding was also analysed using MMPBSA analysis. The antigenic epitopes of Lsa46 exhibit stable molecular interactions with IgG, IgM and TCR; and these peptides may have significant role in eliciting immunogenic response in host cells.

**Methodology**

**Dataset and tools**

The nucleotide and protein sequences of fifteen Leptospiral surface adhesion proteins were retrieved from NCBI (NCBI Resource Coordinators, 2018) and UniProt (UniProt Consortium, 2018) databases respectively. All the fifteen Lsa proteins were selected from a single pathogenic species, *Leptospira interrogans* serogroup Icterohaemorrhagiae serovar copenhageni strain Fiocruz L1-130. These Leptospiral surface adhesion proteins were Lsa20, Lsa21, Lsa23, Lsa24, Lsa25, Lsa26, Lsa27, Lsa30, Lsa33, Lsa36, Lsa37, Lsa46, Lsa63, Lsa66 and Lsa77. The retrieved sequences were analysed with multiple sequence alignment tool, Clustal W to find the evolutionary relationship of the sequences which was implemented in MEGA10 software. The molecular phylogenetic tree was illustrated using Molecular Evolutionary Genetic Analysis (MEGA10) software (Kumar et al., 2018). The Phylogenetic tree gives a 2D graph which represents the evolutionary relationship among genes, showing which of the genes are closely related.

The similarity and conserved nature of the protein sequences were studied using the BlastP (Altschul et al., 1997) analysis. The physicochemical parameters like GRAVY (Grand Average of hydropathicity), molecular weight, amino acid composition, instability index, aliphatic index, theoretical pl and atomic composition of all selected proteins were analysed using ProtParam tool (Wilkins et al., 1999). VirulentPred tool (Garg & Gupta, 2008) was used to determine whether the proteins selected were virulent or not. The prediction of sub-cellular localisation of all the proteins was done using CELLO v.2.5 (Yu et al., 2006) and PSORTb v.3.0 (Yu et al., 2010) tools. The antigenicity prediction was carried out by using the tool VaxiJen 2.0 (Doytchinova & Flower, 2007), with a threshold of 0.4. The linear B-cell epitopes of the fifteen proteins were mapped from the available protein sequences using the tools BCPred 1.0 (El-Manzalawy et al., 2008) and ABCPred (Saha & Raghava, 2006) by keeping the threshold value as default. The MHC-I and MHC-II epitopes were predicted using IEDB (Jurtz et al., 2017; Reynisson et al., 2020) and IFN-gamma--inducing MHC-II epitopes prediction were done with IFNepitope prediction server (Dhanda et al., 2013). The immunogenicity of selected class I MHC binding epitopes was predicted using IEDB Class I immunogenicity prediction tool (Calis et al., 2013). The secondary structures of the proteins were studied using the online tool PSIPRED (Buchan & Jones, 2019). A free online tool SOSUI (Hirokawa et al., 1998) was used to determine whether the protein is soluble or transmembrane in nature. The presence of signal peptide regions on the protein sequences was predicted using Phobius signal predictor and PrediSi (PREDiction of Signal peptides) signal peptide predictor (Nielsen et al., 1997). Conserved protein domains and their families to which they belong to were analysed from Interproscan (Jones et al., 2014) and CDD of NCBI (Marchler-Bauer et al., 2015).

**Secondary structure analysis**

The 3D structure of the protein was modelled using threading-based modelling method. For this I-TASSER server (Zhang, 2008) was employed. The derived 3D structure was visualized using protein structure visualizer PyMol (DeLano, 2002) and also validated using the Whatif (Vriend, 1990) and procheck tool (Morris et al., 1992). From the derived 3D
structure, the B-Cell epitopes were mapped using the Ellipro tool (Ponomarenko et al., 2008).

**Molecular docking**

Molecular docking studies using Biovia Discovery studio 2018 software were used for finding the binding conformation of ligand with respect to the target protein Lsa46. Protein-Protein docking was performed with modelled Lsa46 target against two human immunoglobulins—IgG and IgM and also against human T-Cell Receptor (TCR).

**Molecular dynamic simulation**

Dynamic simulation of the docked complex was done using the software GROMACS 2018.1 (Lemkul, 2018) to examine the stability of the docked complexes. CHARMM36 (Huang & MacKerell, 2013) was used as the Force Field for all the simulation studies. The system was prepared for running MD simulation by creation of a periodic boundary/simulation box, solvation and adding ions. Energy Minimization, Equilibration and Production dynamics were performed. The results were saved every 2 fs. The production MD simulations for all the selected docked poses were performed for 100 ns and the results are stored in the simulation trajectory, from which structural and energetic properties were calculated and analysed.

**Trajectory analysis**

Trajectory analysis of the simulated protein complexes were done to explore the time series analysis of structural and energetic properties. The trajectory was analysed to generate RMSD and RMSF graphs. The structural stability of the simulated complexes were analysed through residual backbone RMSD and RMSF analysis for the entire molecule (Reva et al., 1998).

**Binding energy calculations**

The binding free energies of the receptor and ligand complexes were calculated using Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) (Wang et al., 2019) using g_mmpbsa software (Kumari et al., 2014). From the trajectories of each system, the frames which had shown stable conformations (after attaining equilibrium) at different time intervals were processed. The energy calculation provides Van der Waals energy, electrostatic energy, polar solvation energy and SASA (Solvent-Accessible Surface Area) energy other than binding free energy of the complexes.

**Statistical analysis**

The one-way ANOVA was performed using Graph Pad Prism 9.2.0 software. A p-value of <0.05 was considered as statistically significant.

**Results**

**Phylogenetic analysis**

The phylogenetic analysis conducted with fifteen Leptospiral surface adhesion proteins sequences, revealed the patterns of relatedness. All the amino acid sequences were retrieved from the central access point of extensive curated protein information database UniProt KB (UniProt Knowledgebase). The phylogenetic tree constructed using distance based method, Neighbour Joining (NJ) method (Figure 1) showed that Lsa46 and Lsa77 are clustered into the same group with a bootstrap value of sixty one. The BlastP was used to find the homologous and conserved sequences. All the fifteen selected Leptospiral surface adhesion proteins which were subjected to BlastP analysis showed that out of 15 Lsa proteins, 11 of them were well conserved among other strains and serovars. The Leptospiral surface adhesion proteins, Lsa20, Lsa33, Lsa66, Lsa63, Lsa36, Lsa26, Lsa23, Lsa77 and Lsa46 showed greater sequence conservation, when compared to other selected proteins.

**Physicochemical characterisation**

The online tool ProtParam analysis gave an account of the physicochemical parameters of the Leptospiral surface adhesion proteins (Table 1). The molecular weights of proteins were in the range of 22 and 80 kDa, in which Lsa20 has the least and Lsa77 the highest. The maximum number of amino acid residues in the sequences was identified and seven out of 15 proteins showed polar amino acid residues as the maximum frequency. The proteins Lsa21, Lsa24, Lsa25, Lsa26, Lsa37, Lsa66 and Lsa46 have the maximum number of polar amino acids such as Serine, Threonine and Asparagine. Lsa20, Lsa21, Lsa63 and Lsa77 were predicted as positively charged protein with a maximum number of Arginine and Lysine residues than the negatively charged Aspartic acid and Glutamic acid residues. While Lsa23 was predicted to be neutral with equal number of positively and negatively charged residues, all the other twelve proteins were predicted to be negatively charged. The predicted theoretical pl of Lsa63 was high with a value of 8.42 making it a basic protein, while Lsa24 has lower pl value of 4.93 making it a more acidic protein. All the proteins predicted were rich in amino acids with aliphatic side chains.

The instability index was calculated using ProtParam, the tool that classifies the proteins with instability index lower than 40 as stable. Instability index of proteins Lsa66, Lsa46, Lsa20, Lsa21, Lsa24, Lsa25, Lsa26, Lsa27, Lsa36 and Lsa37 was predicted to be stable. Instability index predicted for Lsa23, Lsa30, Lsa33, Lsa63 and Lsa77 were greater than 40, hence calculated as unstable. The GRAVY was calculated to be greater negative value on Lsa63 (−0.955) followed by Lsa33 (−0.75), Lsa77 (−0.589), Lsa27 (−0.565), Lsa66 (−0.552) and Lsa46 (−0.545), all others have comparatively low values. The higher negative values of GRAVY indicate that these proteins can have better interaction with water molecules.
Protein characterisation by in silico analysis

The other distinguishing features of the Lsa proteins were determined from predicted subcellular localisation, antigenicity, virulent nature, B-cell and T-cell epitopes and domain family.

All Lsa proteins predicted were virulent except Lsa20 and Lsa27. Lsa21, Lsa24, Lsa25, Lsa26, Lsa30, Lsa37, Lsa46, Lsa63 and Lsa66 have higher virulence scores with the default threshold value. In case of bacterial pathogens, the localization of proteins is crucial that the proteins which are presented on the cell surface are regarded as primary target of drugs or as potential vaccine candidates. This is taken as the second criterion for the identification of the vaccine candidate. The subcellular localization of the Lsa proteins was predicted using two online tools CELLO v.2.5 and PSORTb 3.0.

The CELLO v.2.5 results revealed that Lsa20, Lsa23 Lsa27 and Lsa33 are cytoplasmic; Lsa21 as extracellular outer membrane; Lsa24 and Lsa37 as extracellular and Lsa25, Lsa26, Lsa30, Lsa46, Lsa63, Lsa66, Lsa77 as outer membrane in localization. The Lsa46 showed the highest CELLO prediction score of 3.79, compared to other 14 Lsa proteins. While the PSORTb 3.0 showed that only Lsa46 and Lsa77 as outer membrane proteins; Lsa20, Lsa23 and Lsa33 as cytoplasmic; whereas all others are predicted as unknown, indicating that they may have multiple localisation sites. A set of positive known outer membrane and non-outer membrane proteins was also analysed (Table 2) to validate the protein localization prediction.

The antigenicity of the Lsa proteins predicted using the tool VaxiJen 2.0, showed that all the selected proteins were antigenic with the default threshold value. The probable antigenic scores of all proteins are given in Table 3. The

---

**Table 1.** The physicochemical characteristics of Lsa proteins.

| Name of protein | No. of amino acids | Molecular weight (kDa) | Theoretical pl | Max frequency of amino acid residues (%) | Least frequency of amino acid residues (%) | Instability index | Aliphatic index | GRAVY |
|-----------------|--------------------|------------------------|---------------|----------------------------------------|------------------------------------------|------------------|----------------|-------|
| Lsa20           | 197                | 22.37                  | 7.66          | Leu(11.7)                               | Trp, His, Cys (1.5)                       | 35.71            | 98.98          | -0.12 |
| Lsa21           | 182                | 20.15                  | 7.63          | Thr(11)                                 | His (0.5)                                | 23.79            | 95.27          | 0.02  |
| Lsa23           | 212                | 24.80                  | 7.01          | Leu(10.8)                               | Trp (0.9)                                | 42.59            | 81.84          | -0.48 |
| Lsa24           | 240                | 25.63                  | 4.93          | Asn(11.2)                               | Met, His(0.4)                            | 22.79            | 84.54          | -0.17 |
| Lsa25           | 227                | 25.44                  | 4.95          | Thr(10.6)                               | Trp, Met, His (1.8)                      | 17.4             | 73.04          | -0.22 |
| Lsa26           | 249                | 28.36                  | 5.61          | Ser(9.6)                                | His(0.8)                                 | 36.23            | 86.91          | -0.13 |
| Lsa27           | 224                | 25.56                  | 6.46          | Lys(11.2)                               | Trp(0.9)                                 | 33.37            | 78.39          | -0.565|
| Lsa30           | 292                | 32.69                  | 5.39          | Leu(10.6)                               | Trp, Met (0.7)                           | 40.44            | 94.42          | -0.10 |
| Lsa33           | 309                | 34.68                  | 5.20          | Lys(13.9)                               | Cys (0.6)                                | 40.54            | 83.27          | -0.75 |
| Lsa36           | 360                | 38.83                  | 6.08          | Gly(11.7)                               | Cys, Trp(0.6)                            | 33.44            | 84.86          | -0.083|
| Lsa37           | 357                | 36.93                  | 5.27          | Ser(13.2)                               | Arg(0.6)                                 | 32.64            | 87.11          | 0.030 |
| Lsa46           | 428                | 49.22                  | 5.48          | Ser(10.3)                               | Cys (0.2)                                | 39.80            | 80.84          | -0.55 |
| Lsa63           | 557                | 63.67                  | 8.42          | Lys(16.2)                               | Met, Ser(0.4)                            | 44.29            | 74.51          | -0.96 |
| Lsa66           | 589                | 67.89                  | 5.41          | Ser(11.5)                               | Cys (0.2)                                | 34.72            | 79.22          | -0.55 |
| Lsa77           | 687                | 77.95                  | 7.58          | Ile and Leu(8.6)                        | Cys (0.1)                                | 41.76            | 79.77          | -0.59 |

The physicochemical characteristics of Lsa proteins were identified by ProtParam tool. The summarized ProtParam results of all selected Leptospiral surface adhesion proteins are tabulated. The physicochemical parameters like molecular weight (kDa), acidic and basic nature of protein, theoretical pl, instability index, GRAVY score and aliphatic index were predicted from amino acid sequences of the proteins as described in the text.

---

**Protein characterisation by in silico analysis**

The other distinguishing features of the Lsa proteins were determined from predicted subcellular localisation, antigenicity, virulent nature, B-cell and T-cell epitopes and domain family.

All Lsa proteins predicted were virulent except Lsa20 and Lsa27. Lsa21, Lsa24, Lsa25, Lsa26, Lsa30, Lsa37, Lsa46, Lsa63 and Lsa66 have higher virulence scores with the default threshold value. In case of bacterial pathogens, the localization of proteins is crucial that the proteins which are presented on the cell surface are regarded as primary target of drugs or as potential vaccine candidates. This is taken as the second criterion for the identification of the vaccine candidate. The subcellular localization of the Lsa proteins was predicted using two online tools CELLO v.2.5 and PSORTb 3.0.

The CELLO v.2.5 results revealed that Lsa20, Lsa23 Lsa27 and Lsa33 are cytoplasmic; Lsa21 as extracellular outer membrane; Lsa24 and Lsa37 as extracellular and Lsa25, Lsa26, Lsa30, Lsa46, Lsa63, Lsa66, Lsa77 as outer membrane in localization. The Lsa46 showed the highest CELLO prediction score of 3.79, compared to other 14 Lsa proteins. While the PSORTb 3.0 showed that only Lsa46 and Lsa77 as outer membrane proteins; Lsa20, Lsa23 and Lsa33 as cytoplasmic; whereas all others are predicted as unknown, indicating that they may have multiple localisation sites. A set of positive known outer membrane and non-out membrane proteins was also analysed (Table 2) to validate the protein localization prediction.

The antigenicity of the Lsa proteins predicted using the tool VaxiJen 2.0, showed that all the selected proteins were antigenic with the default threshold value. The probable antigenic scores of all proteins are given in Table 3. The
suitability of the tool was confirmed by analysing the prediction values of positive controls (known antigens) and negative controls (non-antigens) with the Lsa proteins. The Linear B-cell epitopes present in the protein sequences were mapped using the tools BCPred 1.0 and ABCPred. Epitopes having score 0.85 and greater were selected. Antigenicity of the selected epitopes was assessed using VaxiJen 2.0 tool. T-cell epitopes, MHC class-I binding and MHC class-II binding epitopes were predicted using IEDB epitope prediction tools. All the IEDB frequently occurring reference allele were taken for the prediction. The list of selected alleles used for MHC-I and MHC-II epitopes are given in Supplementary Data 1. The lowest IC50 prediction score and low adjusted rank are regarded as good MHC-I and MHC-II binders. The predicted epitopes having lowest IC50 prediction score below 10 and 20 were selected for MHC-I and MHC-II respectively.

Table 2. The subcellular localisation prediction of Lsa proteins.

| Name of proteins | Name of organism | CELLO prediction                          | PSORTb prediction       |
|------------------|------------------|------------------------------------------|-------------------------|
| Lsa20            | L. interrogans   | cytoplasmic (3.72)                       | Cytoplasmic (8.96)      |
| Lsa21            | L. interrogans   | Extracellular (2.12) outer membrane (1.54)| Unknown                 |
| Lsa23            | L. interrogans   | Cytoplasmic (3.01)                       | Cytoplasmic (8.96)      |
| Lsa24            | L. interrogans   | Extracellular (4.12)                     | Unknown                 |
| Lsa25            | L. interrogans   | outer membrane (2.61)                    | Unknown                 |
| Lsa26            | L. interrogans   | outer membrane (2.93)                    | Unknown                 |
| Lsa27            | L. interrogans   | Cytoplasmic (3.41)                       | Cytoplasmic (8.97)      |
| Lsa30            | L. interrogans   | outer membrane (2.79)                    | Unknown                 |
| Lsa33            | L. interrogans   | Cytoplasmic (1.44)                       | Cytoplasmic (8.96)      |
| Lsa36            | L. interrogans   | Extracellular (2.23)                     | Unknown                 |
| Lsa37            | L. interrogans   | Extracellular (4.1)                      | Unknown                 |
| Lsa46            | L. interrogans   | outer membrane (3.794)                   | Outermembrane (10)      |
| Lsa63            | L. interrogans   | Outermembrane (2.27)                     | Unknown                 |
| Lsa66            | L. interrogans   | Outermembrane (3.68)                     | Unknown                 |
| Lsa77            | L. interrogans   | Outermembrane (3.04)                     | Outermembrane (10)      |

The subcellular localisation prediction score of the proteins, known outer membrane proteins and non-outer membrane proteins sequences were taken from Uniprot database. The subcellular localisation prediction of the proteins was carried out with two online tools—CELLO and PSORTb. Only Lsa46 and Lsa77 gave prediction as outer membrane in localisation. Lsa20, Lsa23, Lsa27 and Lsa33 were predicted as cytoplasmic by both the tools.

Table 3. Antigenicity prediction of Lsa proteins.

| Name of proteins | Name of organism   | VaxiJen scores | Prediction |
|------------------|--------------------|----------------|------------|
| Lsa20            | L. interrogans     | 0.67           | Antigen    |
| Lsa21            | L. interrogans     | 0.72           | Antigen    |
| Lsa23            | L. interrogans     | 0.53           | Antigen    |
| Lsa24            | L. interrogans     | 0.61           | Antigen    |
| Lsa25            | L. interrogans     | 0.59           | Antigen    |
| Lsa26            | L. interrogans     | 0.41           | Antigen    |
| Lsa27            | L. interrogans     | 0.47           | Antigen    |
| Lsa30            | L. interrogans     | 0.55           | Antigen    |
| Lsa33            | L. interrogans     | 0.64           | Antigen    |
| Lsa36            | L. interrogans     | 0.77           | Antigen    |
| Lsa37            | L. interrogans     | 0.83           | Antigen    |
| Lsa46            | L. interrogans     | 0.51           | Antigen    |
| Lsa63            | L. interrogans     | 0.81           | Antigen    |
| Lsa66            | L. interrogans     | 0.62           | Antigen    |
| Lsa77            | L. interrogans     | 0.50           | Antigen    |
| Collagen adhesin | Staphylococcus aureus | 0.72       | Antigen    |
| A-type flagellin | Pseudomonas aeruginosa | 0.76       | Antigen    |
| Heparin-binding hemagglutinin | Mycobacterium tuberculosis | 0.68 | Antigen |

Positive controls (known antigens)

Negative controls (non-antigen)

L-lactate dehydrogenase | Lactobacillus casei | 0.34 | Non-antigen |
Phosphomethylpyrimidine synthase | Caulobacter vibrioides | 0.38 | Non-antigen |
Beta-galactosidase | Thermus thermophilus | 0.34 | Non-antigen |

The antigenicity prediction of all the proteins was done using VaxiJen tool. Known antigens were selected from antigenic database—Antigen DB. The threshold value for the prediction was set as 0.4 for bacteria which is the default value. The predicted antigenic value above 0.4 was regarded as probable antigen and less than that was regarded as non-antigen.
Antigenicity of the selected epitopes was also done. The IFNγ inducers were also predicted for Class II MHC epitopes.

IFNγ-inducing epitopes were selected from all predicted MHC-II epitopes using IFNepitope predict server. The MHC-II epitopes with acceptable antigenicity score and positive IFNγ prediction score were taken for docking studies. The antigenic MHC-I epitopes having positive immunogenicity scores were selected for further molecular docking analysis.

Protein domain prediction tools use protein sequence and biochemical properties such as hydrophobicity combined with algorithm to predict and identify domains. For the current analyses of the target protein, InterProScan and CDD of NCBI were used. The Lsa46 and Lsa66 belong to the OmpA_C superfamily, indicating it has a Peptidoglycan binding domain similar to the C-terminal domain of outer-membrane protein OmpA. Lsa77 belongs to OmpA superfamily. The NCBI CDD result of all the predicted family or superfamily of Lsa proteins are given in Table 4.

![Table 4. Super family prediction of Lsa proteins.](image)

| Name of proteins | Super family classification |
|------------------|-----------------------------|
| Lsa24            | DUF1554                     |
| Lsa25            | DUF1566                     |
| Lsa37            | DUF1565                     |
| Lsa33            | FeεR                        |
| Lsa46            | OmpA_C                      |
| Lsa63            | Borrelia-P83                |
| Lsa66            | OmpA_C                      |
| Lsa77            | OmpA                        |

All the 15 Lsa proteins were analysed with Interproscan and CDD of NCBI. Out of the 15 proteins analysed only 8 super families were matched. Lsa46, Lsa66 and Lsa77 were predicted to have match for outer membrane class of family. Lsa24, Lsa25 and Lsa37 were matched under protein with unknown function. Lsa33 matched for Ferrochrome family of proteins whereas Lsa63 classified as Borrelia P83/100 family protein.

![Figure 2. Secondary structure analysis of Lsa46 with PSIPRED prediction server showing the alpha helix, beta strand and the coils of amino acid sequence. The frequency of random coil is followed by extended strand and least frequent helix.](image)
Structure analysis

The secondary structure of protein was studied using the PSIPRED tool, which is an accurate and simple method for secondary structure prediction of proteins. This tool identifies and classifies each residue as alpha, beta or coil, as their functional properties depend upon their 3D structures. Figure 2 shows the secondary structure prediction of Lsa46. The results obtained from SOSUI prediction revealed that Lsa25, Lsa26, Lsa23, Lsa36, Lsa30, Lsa66 and Lsa77 are soluble proteins, whereas Lsa46, Lsa33, Lsa27, Lsa20 and Lsa63 are membrane proteins with only one transmembrane helix region and Lsa37, Lsa24 and Lsa21 are membrane proteins with two transmembrane helix regions. The presence of signal peptide region was predicted using Phobius signal predictor and PrediSi tool. Phobius signal predictor predicts the location of the transmembrane helices, intervening loop regions and also signal peptide region. PrediSi tool is employed to find whether the selected proteins are secreted or non-secreted protein. The results of phobius signal peptide predictor showed that Lsa24 and Lsa46 have no signal peptide region while other proteins are marked with the presence of signal peptide regions. The PrediSi results showed that the proteins Lsa20, Lsa21, Lsa23, Lsa25, Lsa27, Lsa33 and Lsa46 are non-secretory in nature.

Since there was no experimentally developed 3D structure available and also the sequence of Lsa46 was distantly related to the templates with low sequence coverage score, building the structure using homology modelling was not possible.

The 3D structure of Lsa46 protein was modelled from the available primary amino acid sequence. For this, threading based modelling with the I-TASSER server was done. This server modelled six 3D structures with different C-values. From this, Model 1 having the highest C-value was taken for further structural analysis. The selected model 1 has a C-value of 2.5. The derived structure was visualized in PyMoL (Figure 3(a)), and validated by Ramachandran plot analyses. This was done for analysing the stereo chemical quality and residue geometry of the selected modelled structure. The plot showed in the Whatif server indicated that majority of the residues were present in the allowed regions. The PROCHECK server was used to evaluate the protein backbone conformations (Figure 3(b)). The phi-psi torsion angle for 57.5% of residues of Lsa46 was in the most favourable region; 31%, 6.4% and 5% in additionally allowed, generously
allowed and disallowed regions respectively. This indicated that Lsa46 model is stereo chemically good and the model derived from I-TASSER was of good quality in terms of protein folding.

The modelled structure of Lsa46 was subjected to epitope prediction by both linear and discontinuous epitopes with Ellipro tool. Fourteen B-cell epitopes were mapped from the predicted 3D structure of Lsa46. This was done to find and map the 3D structure of epitope peptides. Out of the predicted 14 linear epitopes, 6 of them showed greater scores compared to others. These epitopes cover residues (EP1: 33–59), (EP2: 104–123), (EP3: 321–345), (EP4: 136–152), (EP5: 360–403) and (EP6: 251–270). From the Ellipro prediction, only one epitope ranging from sequence 33–59 was above the cut off score of 0.85.

Molecular interaction studies

Studies show that humoral immunity is dominant in protection from leptospiral infection (Koizumi & Watanabe, 2005) and both IgG and IgM are found in patients with different clinical conditions (Lessa-Aquino et al., 2017). Information on the role of cell-mediated immune responses in humans and animals against leptospirosis remains limited (Fraga et al., 2011; Klimpel et al., 2003). The molecular interaction of epitopes of the Lsa46 was studied by molecular docking. The amino acid sequence or crystallographic structure of whole IgM was not available in the databases thus the ligands selected against the target antigenic peptide were human Immunoglobulin G (IgG) (PDB ID:3AGV) (Nomura et al., 2010) and antigen binding fragment region (Fab) of human IgG.

Table 6. Docking parameters of Lsa46 epitope region 90–106 with IgG.

| Pose  | Z rank score | H-bond interacting residues | Bond distance (Å) |
|-------|-------------|-----------------------------|------------------|
| Pose6 | -104.32     | IgGLYS246—Lsa46:ASN106      | 1.94             |
|       |             | IgG:ASN389—Lsa46:GLU101     | 2.28             |
|       |             | IgG:GLN295—Lsa46:ASN104     | 2.91             |
|       |             | IgG:GLN295—Lsa46:SER105     | 2.75             |
|       |             | IgG:GLN295—Lsa46:ASN104     | 2.45             |
| Pose7 | -102.63     |                            |                  |
|       |             |                            |                  |
| Pose8 | -101.65     |                            |                  |
|       |             |                            |                  |
| Pose10| -98.94      |                            |                  |
|       |             |                            |                  |

List of desired poses and the corresponding z-rank scores, hydrogen bond interactions and the bond distance obtained after docking of Lsa46 EP: 90–106 with human IgG is given. From the list, Pose6 is the best and it has two desired hydrogen bond interactions with the epitope (EP: 90–106) of Lsa46 and IgG.

Figure 4. Visualisation of docked structures of Lsa46 EP 90–106 with IgG/IgM and EP 216–230 with TCR. The docked image of Lsa46 epitope region 90–106 with IgG/IgM and 216–230 with TCR depicted as Fig a, b and c, respectively. The human IgG/IgM and TCR are represented as green, blue and orange spheres respectively. Lsa46 is represented as white hard spheres. The epitope regions are the yellow hard spheres.
Immunoglobulin M (IgM) (PDB ID:1DEE) (Graille et al., 2000). Studies on cell-mediated immunity against leptospirosis reported that leptospires induce proliferation of both $\alpha\beta$ T cells and $\gamma\delta$ T lymphocytes in host. The most abundant type of TCR produced is $\alpha\beta$ T cells than $\gamma\delta$ T cells (Klimpel et al., 2003). Thus for finding the molecular interaction of Lsa46 epitopes with TCR, human TCR with only alpha beta subunits was used (PDB ID: 3MFF) (van Boxel et al., 2010). The 3D structure of the ligand proteins were retrieved from Protein Data Bank (PDB) (Berman et al., 2000) in PDB format.

Molecular interaction study of Lsa46 epitopes and IgG

Z-Dock (Protein-Protein) was employed for the docking studies. The ligand selected was IgG and target protein selected was Lsa46 for the receptor cavity docking. The receptor cavity z-docked results showed a total of 2000 poses and out of which the first ten poses are considered as the good and stable ones. All the selected poses showed stable z-rank scores.

Site defined docking studies were also done with Lsa46 protein. The epitope region for the docking analysis was taken based on the epitope score and antigenic scores. Selected B-cell epitopes having top four antigenic scores were taken as the target peptides for the docking analysis. Along with the antigenic epitopes, the predicted epitope having highest score of 0.999 but non-antigenic were also taken for docking studies. The list of selected antigenic B-cell epitopes of Lsa46 used for docking were retrieved from Protein Data Bank and their respective scores are listed in Table 5. Human IgG was taken as the ligand and protein for all site defined docking. The z-dock results of all epitopes with the ligand IgG generated a total of 2000 poses and the first ten poses were further analysed. Of all the docked epitopes, only one (EP: 90–106) showed the desired interaction with IgG molecule. The z-docked image of EP: 90–106 with IgG is given in Figure 4(a). Out of the ten poses analysed, pose6, pose7, pose8 and pose10 showed the Hydrogen bond interaction with epitope residues and IgG. The pose6 showed best z-rank score of $-104.32$, compared to pose7, 8 and 10 (Table 6). This z-rank score implies that
the docked complex has stable binding interactions. Two hydrogen bond interactions were seen with, Asn106 of epitope and Lys246 of IgG and the second interacting residues were Glu101 of epitope and Asn389 of IgG (Figure 5). The hydrogen bond distance of the interacting residues were 1.94 and 2.28 Å, respectively. All other docked epitope residues with IgG did not show the desired interactions. Thus for further analysis and studies, pose6 of the epitope (90–106)-IgG docked complex was used. Human protein albumin was blind docked with both Lsa46 and IgG separately as a control, in which both docked complex interactions are different as that of Lsa46 with IgG. This indicates that all the dockings were specific and based on the structure and chemical complementarity of the proteins selected.

**Table 7. Docking parameters of Lsa46 epitopes with IgM(Fab).**

| Pose     | Z rank score | H-bond interacting residues | Bond distance (Å) |
|----------|--------------|------------------------------|-------------------|
| Lsa46 epitope 90–106 with IgM Fab |             |                             |                   |
| Pose4    | −91.77       | Lsa46:ASN104—IgM Fab:GLN79  | 1.87              |
| Pose5    | −91.60       | Lsa46:ASN104—IgM Fab:LYS149 | 2.81              |
| Pose9    | −89.96       | Lsa46:GLU98—IgM Fab:ASN210  | 2.88              |
| Pose8    | −90.02       | Lsa46:SER105—IgM Fab:GLU195 | 2.57              |
| Pose10   | −89.83       | Lsa46:ASN104—IgM Fab:LYS149 | 3.13              |
| Lsa46 epitope 397–413 with IgM Fab |             |                             |                   |
| Pose8    | −90.02       | Lsa46:LYS410—IgM Fab:GLY2509| 2.68              |
| Pose10   | −89.83       | Lsa46:LYS410—IgM Fab:VAL2015| 1.91              |
| Lsa46 epitope 398–418 with IgM Fab |             |                             |                   |
| Pose8    | −90.02       | Lsa46:LYS410—IgM Fab:GLY2509| 2.68              |
| Pose10   | −89.83       | Lsa46:LYS410—IgM Fab:VAL2015| 1.91              |
| Lsa46 epitope 230–250 with IgM Fab |             |                             |                   |
| Pose6    | −91.58       | Lsa46:TYR236—IgM Fab:PRO2713| 2.10              |
| Pose8    | −90.02       | Lsa46:SER238—IgM Fab:ASN2716| 2.80              |
| Pose10   | −89.83       | Lsa46:SER238—IgM Fab:ASN2714| 3.08              |
| Pose8    | −90.02       | Lsa46:TYR236—IgM Fab:GLY2715| 2.76              |
| Pose10   | −89.83       | Lsa46:ASN240—IgM Fab:SER2525| 3.04              |
| Pose8    | −90.02       | Lsa46:ASN240—IgM Fab:GLN2503| 2.66              |
| Pose10   | −89.83       | Lsa46:ASN240—IgM Fab:SER2525| 2.95              |
| Pose8    | −90.02       | Lsa46:ARG241—IgM Fab:VAL2110| 1.68              |

List of desired poses and their corresponding z-rank scores, hydrogen bond interactions and their bond distance of Lsa46 Epitopes with human IgM(Fab). The pose 4 of EP: 90–106, pose 6 of EP: 230–250 and pose 8 of EP: 397–413 and EP: 398–418 with ligand have desired interactions and highest z-rank scores.

**Molecular interaction study of Lsa46 epitopes and IgM(Fab)**

The patients affected with leptospirosis show different antibody profiles with respect to different clinical outcomes. Considerable amount of IgM and IgG were found during severe and mild conditions of the disease (Lessa-Aquino et al., 2017). Blind dock (receptor cavity dock) as well as site specific docking was done to examine the molecular interactions between the Lsa46 antigenic epitopes and IgM(Fab). All the docking analyses were done with Lsa46 as target and IgM(Fab) as ligand protein. A total of 2000 poses were generated in receptor cavity dock, all the selected first ten poses showed stable z-rank scores. *Staphylococcus aureus* protein-A was blind docked with the target protein Lsa46 as a control and was different from that of Lsa46 and IgM(Fab) in interacting residues. The IgM(Fab) structure retrieved from PDB had bound Staph-A protein, this was used as a target instead of human protein albumin for control. IgM(Fab) without Staph-A protein was used for other site specific and receptor cavity dockings. The z-dock image of one of the site specific docking of IgM(Fab) and Lsa46 was shown in Figure 4(b).

The epitopes selected were same as those for IgG for site specific docking. The first ten poses of all z-dock results of epitopes with IgM(Fab) were analysed for selecting desirable interactions. Of the five epitopes docked, only four showed the desired interaction with IgM(Fab) molecule. The epitopes which showed the preferable interactions are (EP: 90–106), (EP: 397–413), (EP: 398–418) and (EP: 230–250). The hydrogen bond interacting residues of IgM(Fab) with Lsa46 epitopes, different poses and corresponding z-rank score and bond distance are tabulated in Table 7. The poses in each docked epitope with the ligand having the best z-rank score were taken for further analysis. The z-docked images of four epitopes H-bond interactions with IgM(Fab) are shown in Figure 6.

**Molecular interaction study of Lsa46 epitopes and TCR(αβ)**

The molecular docking studies were performed with target Lsa46 and ligand human TCR(αβ). The receptor cavity and site specific docking was done. Predicted class-I MHC epitopes and class-II MHC epitopes of target were docked against the ligand for the site specific docking. The list of T-cell epitopes (both MHC-I and MHC-II) and their corresponding IC50, adjusted rank and antigenicity scores are mentioned in Table 8. The resultant z-docks generated all the possible 2000 poses for each epitope of Lsa46 and TCR(αβ). Human albumin protein was docked with TCR(αβ) as a control. The z-
Figure 6. H-bond interaction of docked Lsa46 epitopes with IgM(Fab). Docking result of four Lsa46 epitopes with IgM(Fab) showing the hydrogen bond interactions of best poses. (a) Pose 4 interactions of IgM(Fab) and EP 90–106. (b) Pose 8 interactions of IgM(Fab) and EP 398–418. (c and d) Two H-bond interactions of pose 8 of IgM(Fab) and EP 397–413. (e) Pose 6 interactions of IgM(Fab) and EP 230–250. H-Bond is represented as bright green cylinders. The Lsa46 is represented as red lines and IgM(Fab) is represented as blue lines.
dock image of one of the site specific docking is shown in Figure 4(c).

The first ten docked poses were selected and analysis was done based on their highest z-rank score and corresponding intermolecular H-bond interactions. Four of the T-cell epitopes showed H-bond interactions with the ligand, they are (EP: 341–358), (EP: 216–230), (EP: 181–189) and (EP: 133–144). The poses and corresponding z-rank score, interacting residues and H bond distance are given in Table 9. The poses with best z-rank score were selected for dynamic simulation analysis. The z-docked images of four epitopes, H-bond interactions with TCR(αβ) are as in Figure 7.

Molecular dynamic simulation

The docked complexes were subjected to molecular dynamic simulation and energy minimization studies to obtain a stable conformation. The standard dynamic cascade of the different epitopes docked with IgG/IgM(Fab) and TCR(αβ) showed stable minimized energy conformation. This implied that the docked complexes were of stable minimized structure and orientation.

Trajectory analysis

The RMSD and RMSF plots show whether the simulated complex is stable by analysing stability and predicting conformational changes of the protein. The RMSD and RMSF plots are shown in Supplementary Figures 1–6.

The RMSD values depend up on binding interaction and energy between target and ligand. The optimised protein has the lowest RMSD values indicating small fluctuations and constant backbone (Razzaghi-Asl et al., 2018). The RMSD analysis of EP: 90–106 and IgG showed a stable nature across the simulation time ranging from 40 to 100 ns. The B-cell epitopes and IgM(Fab) complexes have shown stable RMSD. The EP: 90–106—IgM(Fab) complex exhibited stable RMSD after 60 ns of production run. The complexes EP: 397–413 and EP: 398–418 with IgM(Fab) showed a somewhat similar RMSD graph with a predominantly stable nature across the simulation time and the overall RMSD values were below 7 Å. However, the EP: 230–250 and IgM(Fab) showed a fairly stable behaviour at initial production run from 15 to 55 ns and after that the residual fluctuation increased with time. The TCR(αβ)—EP: 216–230 and TCR(αβ)—EP: 133–144 exhibited similar stable conformation for time period from 20 ns and
the RMSD values were below 6 Å. The RMSD analysis of EP: 181–189 and EP: 341–358—TCR complex showed a higher deviation during initial time periods and attained equilibrium after 50 ns.

The RMSF plot shows fluctuation of the residues compared to the relative structure across the production trajectory. This can be used to identify the regions of highest and lowest flexibility. Typically terminal residue and surface loop regions have the highest flexibility, while the mobility of the protein core is more restricted. Also, the flexibility of a region can be affected by the interactions that it makes with surrounding residues. The RMSF analysis of EP: 90–106 and IgG showed a stable nature along the simulation time without much deviation. The B-cell epitopes and IgM(Fab) complexes have also exhibited stable RMSF except for EP: 230 with TCR complex showed a higher deviation during initial time periods and attained equilibrium after 50 ns.

### Binding energy analysis

The receptor–ligand binding strength of the entire complexes were calculated using g_mmpbsa analysis after the MD simulation trajectory analysis. The RMSD of the simulated complexes showed some degree of variation initially, while it became stable at later time frames. This time frame of stable nature was different for the complexes analysed. Therefore, the analysis was performed for the time intervals which had displayed a stable RMSD over 100 ns. From each of the RMSD graph generated, four regions in 10ns time intervals showing stable confirmations were randomly selected for binding energy calculations and the results are shown in Table 10. There was no significant difference (p > 0.05) in the binding energy of each of the complexes across the different time intervals indicating energetic stability. Out of the eight complexes analysed, EP: 398–418-IgM complex showed the lowest average binding energy (−3770.1 ± 168.9 KJ/mol) followed by EP: 397–413-

---

**Table 8.** The list of MHC-I and MHC-II epitopes of Lsa46 selected for docking.

| Sl. No. | HLA allele | Start | End   | Sequence      | IC50     | Adjusted rank | Antigenicity | IFN/immunogenicity |
|---------|------------|-------|-------|---------------|----------|---------------|--------------|-------------------|
| 1       | HLA-A*23:01 | 181   | 189   | RLYTRYTPF     | 2.58     | 0.01          | 1.34         | Positive (0.19)    |
| 2       | HLA-A*02:01 | 319   | 327   | SLDSIHFSL     | 8.39     | 0.08          | 1.36         | Positive (0.03)    |
| 3       | HLA-A*02:06 | 133   | 144   | FQFANGKIGV    | 9.96     | 0.09          | 0.81         | Positive (0.003)   |
| 4       | HLA-DRB1*01:01 | 216   | 230   | PINTVSYEIAATSK | 18.43 | 4.90          | 0.40         | Positive (0.07)    |
| 5       | HLA-DRB1*1 3:02 | 341   | 358   | DNLSYLNKLENKIKIT | 9.56 | 0.14          | 1.39         | Positive (0.33)    |

The list of T-cell epitopes selected for site specific docking of Lsa46 with human TCR is given. The first three epitopes are MHC-I epitopes and EP4 and EPS are MHC-II epitopes. The MHC-II epitopes were selected based on low IC50, low adjusted rank, antigenic and positive IFN-gamma inducers. The MHC-I epitopes were chosen based on low IC50, low adjusted rank, antigenic and positive immunogenicity.

**Table 9.** Docking parameters of Lsa46 epitopes with TCRs(β).

| Pose | Z rank score | H-bond interacting residues | Bond distance (Å) |
|------|--------------|-----------------------------|-------------------|
| Pose2 | −118.397     | Lsa46:LYSTSFQ54—TCR:ASP126 | 2.74             |
| Pose3 | −104.505     | Lsa46:LYSTSFQ54—TCR:ASP126 | 2.50             |
| Pose1 | −129.732     | Lsa46:TYR221—TCR:GLU77     | 2.82             |
|       |              | Lsa46:SER222—TCR:ASP63     | 1.83             |
|       |              | Lsa46:TYR221—TCR:ARG77     | 2.09             |
| Pose8 | −96.241      | Lsa46:ARG181—TCR:ASN199    | 3.05             |
|       |              | Lsa46:ARG181—TCR:THR206    | 2.79             |
|       |              | Lsa46:THR185—TCR:GLU204    | 2.67             |
| Pose10| −94.652      | Lsa46:ASN137—TCR:GLY24     | 2.56             |
|       |              | Lsa46:ASN137—TCR:ASP87     | 2.87             |

List of desired poses and their corresponding z-rank scores, hydrogen bond interactions and their bond distance of Lsa46 T-cell Epitopes with human TCR(β) is given. The pose 1 of EP: 216–230, pose2 and 3 of EP: 341–358 and pose 8 of EP: 181–189 and pose10 of EP: 133–144 with ligand have desired interactions and best z-rank scores.

**Table 10.** Binding energy calculations.

| Receptor-ligand complex | Time interval 1 (KJ/mol) | Time interval 2 (KJ/mol) | Time interval 3 (KJ/mol) | Time interval 4 (KJ/mol) | Average binding energy (KJ/mol) |
|-------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------------------------|
| EP:90–106-IgG           | −392.6 ± 122.9           | −436.4 ± 137.3           | −362.1 ± 145.9           | −235.7 ± 63.1            | −356.7 ± 117.3                  |
| EP:90–106-IgM           | −1365.6 ± 256.6          | −1792.5 ± 99.8           | −1706.8 ± 323.8          | −1509.7 ± 148.6          | −1593.6 ± 207.2                 |
| EP: 397–413-IgM         | −3342.4 ± 217.7          | −3668.2 ± 189.9          | −3869.9 ± 184.2          | −3515.2 ± 206.5          | −3623.9 ± 199.6                 |
| EP:398–418-IgM          | −3525.8 ± 155.1          | −3630.1 ± 152.4          | −3983.2 ± 189.5          | −3941.1 ± 178.5          | −3770.1 ± 168.9                 |
| EP:216–230-TCR          | −1039.4 ± 45.6           | −1126.8 ± 192.4          | −818.9 ± 217.9           | −1084.2 ± 65.7           | −1017.3 ± 130.4                 |
| EP:341–358-TCR          | −647.4 ± 161.7           | −769.8 ± 122.4           | −867.1 ± 175.1           | −1142.4 ± 238.1          | −856.45 ± 174.3                 |
| EP:133–144-TCR          | −669.4 ± 155.9           | −825.1 ± 173.3           | −688.6 ± 237.9           | −910.6 ± 99.1            | −773.4 ± 166.5                  |
| EP:181–189-TCR          | −1215.5 ± 221.2          | −1295.7 ± 156.8          | −1142.4 ± 138.1          | −978.3 ± 68.5            | −1147.3 ± 146.1                 |

The binding energy of receptor-ligand complexes was calculated at different time intervals. The time interval 1, 2, 3 and 4 (10ns intervals) mentioned are different time frames considered for energy calculations which were structurally stable. The comparison of binding free energies using ANOVA showed no significant difference (p > 0.05).
Figure 7. H-bond interaction of docked Lsa46 epitopes poses with TCR(αβ). Dock result of four T-cell epitopes of Lsa46 with TCR(αβ) showing the hydrogen bond interactions of best poses. (a) Pose 2 interactions of TCR(αβ) with EP 341–358. (b) Pose 10 H-bond interactions of TCR(αβ) and EP 133–144. (c) Pose 8 interactions of TCR(αβ) and EP 181–189 and (d and e) represents pose 1 H-bond interactions of TCR(αβ) and EP 216–230. The H-Bond is represented as bright green cylinders. The Lsa46 is represented as red lines and IgM(Fab) is represented as blue lines.
IgM (−3623.2 ± 199.6 KJ/mol), EP: 90–106-IgM (−1593.6 ± 207.2 KJ/mol), EP: 181–189-TCR (−1147.7 ± 146.1 KJ/mol), EP: 216–230-TCR (−1017.3 ± 130.4 KJ/mol), EP: 341–358-TCR (−856.45 ± 174.3 KJ/mol), and EP: 90–106-IgG (−356.7 ± 117.3 KJ/mol). The one way ANOVA of binding energy of each complex at different time intervals (Table 10), showed that they are not significantly different with \( p \) values greater than 0.05.

The binding energy analysis could not be done for the B-cell—EP: 230–250-IgM complex as its RMSD and RMSF structural analysis had shown high degree of fluctuations.

**Discussion**

The currently available mode of disease management of leptospirosis is not adequate as it does not provide a complete protection from infection and re-infection. When considering the risk of leptospirosis such as occupational exposures, recreational activities and extreme weather events, prevention is quite difficult without vaccination. Administration of antibiotics, which is the common mode of treatment is only effective within seven days after infection and should be immediately taken on suspicion. In the present study, different *in silico* analyses were done to discover a suitable candidate which can be developed into a potential vaccine later.

One of the failures of modest numbers of leptospiral vaccines against the infections is the lack of cross protection among the different serovars (Yang et al., 2006). The similarity search of Lsa46 showed it as a conserved protein among the pathogenic species of leptospires and thus can mediate cross protection against various serovars. Most of the surface outer membrane proteins of leptospires are immunogenic (Nally et al., 2005). Lsa46 and Lsa77 were predicted as outer membrane proteins when analysed with PSORTb which is regarded as the most accurate predicting tool for surface localisation (Yu et al., 2010). Also the domain prediction of Lsa46 shows that it belongs to the OmpA super family. The antigenic proteins can trigger the immune response in the cell, thus the antibody-mediated immune response. IgG and IgM are the two types of immunoglobulins involved in the
leptospiral immune responses (Lessa-Aquino et al., 2017). The antigenicity and virulence prediction of Lsa46 showed it as virulent and antigenic, thus it can be used as an effective vaccinogen. Antigenic epitopes are identified based on the amino acid sequences and the tertiary structures. The primary predictions of selected Lsa proteins with the respective sequence revealed Lsa46 as a better antigen for vaccine designing. The 3D structures or the structural features of Lsa46 including its active sites and pockets were not available in the public databases or scientific articles. I-TASSER is ranked as the best method in the server section for the protein 3D structure prediction (Roy et al., 2010; Zhang, 2008). The modelled 3D structure of Lsa46 was of good quality when computed with Ramachandran plot and the whole protein was used for the molecular interaction studies with the immunoglobulins and TCR. Out of a total of 15 Leptospiral surface adhesion proteins used for the study, the computational prediction of Lsa46 protein shows conserved, outer membrane, virulent, antigenic epitopes with stable binding interaction with IgG, IgM and TCR, making it an ideal candidate for peptide as a vaccine.

The principles governing structural vaccinology are rooted in the observation that; arousing an effective immune response does not require recognition of the entire antigenic protein, but can be achieved with the recognition of a single or multiple selected epitopes and be sufficient to induce protective immunity (Rappuoli et al., 2016). Selection of antigenic protein as a vaccine candidate is crucial. The proteins which are not exposed on the surface of the pathogen cannot induce a strong and fast immune response as the antibodies produced against these proteins would not be able to bind due to the presence of subsurface proteins. This makes the vaccine ineffective (Adler, 2015b). One of the pathogenic serovars of *Leptospira interrogans* i.e. *L. interrogans* serovar Copenhageni strain L1-130 genome reveals that there are many proteins, both surface-exposed and outer membrane proteins, which show a high degree of conservation among the other pathogenic species of leptospires (Grassmann et al., 2017). This study targets the reverse vaccinology approach for predicting the surface exposed adhesion proteins, as they are regarded most likely to be a vaccine candidate (Serruto & Rappuoli, 2006).

Leptospires enter and colonize the host tissue by binding to ECM components mainly laminin, plasminogen and fibronectin. The surface adhesion proteins usually help in the tissue penetration and thus help in the immune invasion and disease progression. The affinity and potentiality of binding may differ with respect to other interacting proteins and other factors (Vieira et al., 2014). The Leptospiral surface adhesion proteins, Lsa27 (Longhi et al., 2009), Lsa20 (Mendes et al., 2011), Lsa25 and Lsa33 (Domingos et al., 2012), Lsa26 (Siqueira et al., 2013), Lsa21 (Atzingen et al., 2008; Oliveira et al., 2010), Lsa63 (Vieira et al., 2010), Lsa30 and Lsa23 (Souza et al., 2012; Vieira et al., 2014) are regarded as laminin binding adhesins when assessed by immunofluorescence assay and protease k accessibility. Some of the Lsa proteins like Lsa33 (Domingos et al., 2012), Lsa21 (Atzingen et al., 2008; Oliveira et al., 2010), Lsa23, Lsa30 (Souza et al., 2012) show broader range for binding as they can bind with laminin, plasminogen, plasma fibronectin etc. The Lsa25, Lsa33, Lsa30 and Lsa23 are also reactive with complement pathway regulator protein C4BP (Domingos et al., 2012; Vieira et al., 2014).

Studies have shown that anti-leptospiral antibodies produced against LPS antigen (Evangelista & Coburn, 2010), but there is no evidence of involvement of cell-mediated immune responses in humans against leptospires. Indeed, administration of killed *L. borgpetersenii* as vaccine in cattlees has shown proliferation of γδ T cells and CD4+γδ T cells and INF γ production (Fraga et al., 2011; Klimpel et al., 2003). Clinical studies on both human and animals only give an indirect evidence of involvement of TCRγδ T cells against bacterial, viral and parasitic infections. TCRγδ T cells can respond to non-peptide molecules like alkylamines, small organic phosphate molecules, un-processed antigens apart from peptide molecules (Bukowski et al., 1999), whereas TCR alpha beta cells specifically bind to MHC restricted peptides. The involvement of different T-cell (TCRγδ vs TCRβ+) population together with immunoglobulins can determine the severity or level of septicemia or leptospermia. Experimental models suggest that macrophage and neutrophil-mediated phagocytosis is only possible if the leptospiral antigen is opsonized by specific immunoglobulins. Thus the Class-I and Class-II MHC epitopes and their corresponding antigenicity and immunogenicity were predicted and used in docking studies against human TCR(αβ). IFNγ is an important MHC-II molecule inducer and macrophage activator thereby helping in complete elimination of antigens. Thus for the identification of better subunit vaccine candidates, prediction of IFN-gamma inducer peptides from a list of multiple peptides is necessary. IFNγ prediction was done only for the antigenic peptides.

B-cell epitopes play a vital role in the development of peptide vaccines, in the diagnosis of diseases, and also for allergy research. Antigenicity of the proteins as well as the small peptide region is vital as they can induce immunogenicity. Antigenicity of all the selected epitopes are predicted, even though epitopes are generally antigenic determinant. Top five epitopes which are antigenic and having highest epitope prediction score were used in the antibody interaction study. The epitopes 398–418 and 397–413 have overlapping regions. When these two epitopes were docked individually they have shown stable interactions with the IgM(Fab), so the two epitopes can be combined to a 22-mer long peptide for eliciting stronger immunogenicity (Meza et al., 2017; Nezafat et al., 2016; Rahman et al., 2020). The epitope 90–106 is found to be interactive to IgG and IgM, and this is the only peptide which has shown interaction with both the immunoglobulin. The non-antigenic epitope EP: 230–250 has shown H-bond interactions with IgM(Fab) when docked, but when complex was subjected to MD simulation very high structural deviation was observed. On the MD simulation animation analysis, it was observed that two chains of IgM(Fab) were moving apart from it. This could be the reason for unstable and highly fluctuated RMSD behaviour. The binding of the receptor protein might cause the
disruption of intra-molecular bonding of IgM. Thus this non-antigenic region can be eliminated when the peptide as a vaccine is considered. The two MHC-II epitopes and two MHC-I epitopes also showed stable binding interactions and energy with TCR(αβ). These small peptides can be combined together and can be subjected to further in vitro and in vivo study to identify the cellular toxicity, immunogenicity and antigenicity, and used as efficient vaccinogen against leptospirosis.

Though the in silico findings are satisfactory as a vaccine candidate and helps narrow down the datasets, still the size of the databases, conformational sampling and uncertainties in the prediction, limits it to be rational. Further, in vitro and in vivo studies will reveal the exact biological effects.

Conclusion

The initial analyses of Lsa46 with computational prediction approaches give a better understanding of the protein. The immunoinformatics analyses revealed Lsa46 as stable, conserved and non-secreted outer membrane protein. The Lsa46 is also predicted as antigenic peptide having B-cell and T-cell epitopes with good acceptable scores. Both 2D and 3D structure of Lsa46 were studied and validated using standard Ramachandran plot analysis, indicating it as a reliable model for further studies. The docking, dynamic simulation and binding energy studies with epitope region and the human IgG/IgM(Fab)/TCR(αβ) revealed it as a stable stereochemically ideal structure for developing as a vaccine candidate.

Though the data generated for Lsa46 can be used for immunotherapies, further wet lab studies including experimental designing of the vaccine candidate are needed. The effectiveness and potency of designed vaccine studies both in vitro and in vivo are essential to substantiate the prediction.

Authors’ contributions

J.M.: conceived the study, designed the study, data analysis, writing and drafting the manuscript. S.A.: coordinated the study. P.R.S., A.S.N. and O.V.O.: designed the study, data analysis, coordinated the study, revised and proof read the draft manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors acknowledge the SIUCEB support at the Department of Computational Biology and Bioinformatics, University of Kerala, India for providing the necessary facilities to carry out this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by University of Kerala, by way of a fellowship to JMI. PRS was supported by ISCA, Kolkata by Asutosh Mukerjee Fellowship.

References

Adler, B. (2015a). History of leptospirosis and leptospira. Current Topics in Microbiology and Immunology, 387, 1–9. https://doi.org/10.1007/978-3-662-45059-8_1
Adler, B. (2015b). Vaccines against leptospirosis. Current Topics in Microbiology and Immunology, 387, 251–272. https://doi.org/10.1007/978-3-662-45059-8_10
Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389
Atzingen, M. V., Barbosa, A. S., De Brito, T., Vasconcellos, S. A., de Morais, Z. M., Lima, D. M., Abreu, P. A., & Nascimento, A. L. (2008). Lsa21, a novel leptospirotic protein binding adhesive matrix molecules and present during human infection. BMC Microbiology, 8(1), 70. https://doi.org/10.1186/1471-2180-8-70
Barbosa, A. S., Abreu, P. A. E., Neves, F. O., Atzingen, M. V., Watanabe, M. M., Vieira, M. L., Morais, Z. M., Vasconcellos, S. A., & Nascimento, A. L. T. O. (2006). A newly identified leptospiral adhesin mediates attachment to laminin. Infection and Immunity, 74(11), 6356–6364. https://doi.org/10.1128/IAI.00460-06
Berman, H. M., Westbrook, J., Feng, Z., Gililand, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. Nucleic Acids Research, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235
Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., Levet, P. N., Gilman, R. H., Willig, M. R., Gotuzzo, E., & Vinetz, J. M. (2003). Leptospirosis: A zoonotic disease of global importance. The Lancet. Infectious Diseases, 3(12), 757–771. https://doi.org/10.1016/S1473-3099(03)00830-2
Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Research, 47(W1), W402–7. https://doi.org/10.1093/nar/gkj297
Bukowski, J. F., Morita, C. T., & Brenner, M. B. (1999). Human γδ T cells recognize aliphamines derived from microbes, edible plants, and tea: Implications for innate immunity. Immunity, 11(1), 57–65. https://doi.org/10.1016/S1074-7613(00)80081-3
Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Computational Biology, 9(10), e1003266. https://doi.org/10.1371/journal.pcbi.1003266
DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. {CCP4} Newsletter on Protein Crystallography, 40, 1–8.
Dellagnostin, O. A., Grassmann, A. A., Hartwig, D. D., Félix, S. R., da Silva, É. F., & McBride, A. J. A. (2011). Recombinant vaccines against leptospirosis. Human Vaccines, 7(11), 1215–1224. https://doi.org/10.4161/hv.7.11.17944
Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-γ-gamma inducing MHC class-II binders. Biology Direct, 8(1), 30. https://doi.org/10.1186/1745-6150-8-30
Domingos, R. F., Vieira, M. L., Romero, E. C., Gonçalves, A. P., de Morais, Z. M., Vasconcellos, S. A., & Nascimento, A. L. T. O. (2012). Features of two proteins of Leptospira interrogans with potential role in host-pathogen interactions. BMC Microbiology, 12(1), 50. https://doi.org/10.1186/1471-2180-12-50
Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics, 8(1), 4. https://doi.org/10.1186/1471-2105-8-4
El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting flexible length linear B-cell epitopes [Paper presentation]. Comput Syst Bioinformatics Comput Syst Bioinforma Conf., 7:121–32. https://doi.org/10.1142/S17933876100011
Evangelista, K. V., & Coburn, J. (2010). Leptospirosis as an emerging pathogen: A review of its biology, pathogenesis and host immune responses. Future Microbiology, 5(9), 1413–1425. https://doi.org/10.2217/fmb.10.102
Fraga, T. R., Barbosa, A. S., & Isaac, L. (2011). Leptospirosis: Aspects of innate immunity, immunopathogenesis and immune evasion from
Pappas, G., Papadimitriou, P., Siozopoulou, V., Christou, L., & Akritidis, N. (2008). The globalization of leptospirosis: Worldwide incidence trends. *International Journal of Infectious Diseases: IJD*, 12(4), 351–357. https://doi.org/10.1016/j.ijid.2007.09.011

Ponomarenko, J., Bui, H.-H., Li, W., Fusseeder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ELiPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9, 514. https://doi.org/10.1186/1471-2105-9-514

Prasad, M., Bothammal, P., Akino Mercy, C. S., Sumaiya, K., Saranya, P., Muralitharan, G., & Natarajaseenivasan, K. (2020). Leptospiral protein LIC11334 display an immunogenic peptide KNSMP01. *Microbial Pathogenesis*, 149, 104407. https://doi.org/10.1016/j.micpath.2020.104407

Razzaq-Asl, N., Mirzayi, S., Mahnam, K., & Sepehri, S. (2018). Identification of COX-2 inhibitors via structure-based virtual screening and molecular dynamics simulation. *Journal of Molecular Graphics & Modelling*, 83, 138–152. https://doi.org/10.1016/j.jmgm.2018.05.010

Revá, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Proteins: Structure, Function, & Bioinformatics*, 32(3), 410–415. https://doi.org/10.1002/(SICI)1097-0004(199804)32:3<410::AID-PROT410>3.0.CO;2-7

Rappuoli, R., Bottomley, M. J., D’Oro, U., Finco, O., & De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *The Journal of Experimental Medicine*, 213(4), 469–481. https://doi.org/10.1084/jem.20151960

Razgaghi-Asl, N., Mirzayi, S., Mahnam, K., & Sepehri, S. (2018). Prediction of COX-2 inhibitors via structure-based virtual screening and molecular dynamics simulation. *Journal of Molecular Graphics & Modelling*, 83, 138–152. https://doi.org/10.1016/j.jmgm.2018.05.010

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Proteins: Structure, Function, & Bioinformatics*, 32(3), 410–415. https://doi.org/10.1002/(SICI)1097-0004(199804)32:3<410::AID-PROT410>3.0.CO;2-7

Rappuoli, R., Bottomley, M. J., D’Oro, U., Finco, O., & De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *The Journal of Experimental Medicine*, 213(4), 469–481. https://doi.org/10.1084/jem.20151960

Razzaq-Asl, N., Mirzayi, S., Mahnam, K., & Sepehri, S. (2018). Identification of COX-2 inhibitors via structure-based virtual screening and molecular dynamics simulation. *Journal of Molecular Graphics & Modelling*, 83, 138–152. https://doi.org/10.1016/j.jmgm.2018.05.010

Revá, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Proteins: Structure, Function, & Bioinformatics*, 32(3), 410–415. https://doi.org/10.1002/(SICI)1097-0004(199804)32:3<410::AID-PROT410>3.0.CO;2-7

Rappuoli, R., Bottomley, M. J., D’Oro, U., Finco, O., & De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. *The Journal of Experimental Medicine*, 213(4), 469–481. https://doi.org/10.1084/jem.20151960

Razga...