



# In silico vaccine design against *Chlamydia trachomatis* infection

Shilpa Shiragannavar<sup>1</sup> · Shivakumar Madagi<sup>1</sup> · Joy Hosakeri<sup>1</sup> · Vandana Barot<sup>2</sup>

Received: 6 December 2019 / Revised: 30 March 2020 / Accepted: 20 May 2020  
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

## Abstract

*Chlamydia trachomatis*, a gram-negative bacterium known to infect the genital sites mainly columnar epithelial cells of the cervix, urethra and rectum in women and causes acute epididymitis, urinary tract inflammation and DNA damage to the sperms in men, hence considered to be one of the major sexually transmitted infections. The infection is asymptomatic in many people and remains untreated leading to serious health complications, including pelvic inflammatory disease, ectopic pregnancy and infertility. The current treatment options include antibiotics, but the pathogen has gained resistance against many antibiotics. The present work involves an in silico reverse vaccinology approach for identifying the immunogens as vaccine candidates that can be effective against reinfections and should be capable of inducing long-term protective immunity against *Chlamydial* infections. This study identifies the putative vaccine candidates that are membrane bound with high antigenicity properties; antigenicity induces the immunogenicity which involves identification of T-cell and B-cell epitopes that induce both humoral and cell-mediated immunity. The epitopes ‘LSWEMELAY’, ‘LSNTEGYRY’, ‘TSDLGQMEY’, ‘FIDLLQAIY’ and ‘FSNNFSDIY’ were predicted as core sequences for class I MHC molecules. The identified epitopes showed promising ability to interact with the human leukocyte antigens (HLA). These epitopes showed maximum population coverage with epitope conservancy above 80%. Molecular docking was performed to test the binding affinities of the identified epitopes with the HLA molecule to study the binding cleft interactions. The vaccine candidate thus identified from this study showed to possess the potential to activate the B- and T-cell immune responses which are more specific and make the body stronger against infections and effective for reinfections.

**Keywords** *Chlamydia trachomatis* · Infertility · Vaccine · epitope · Sexually transmitted infections · Molecular docking

## 1 Introduction

*Chlamydia trachomatis* commonly known as chlamydia is a pathogenic gram-negative, ovoid-shaped non-motile bacterium. The bacterium is widely recognized as the most common sexually transmitted pathogen with approximately 100 to 150 million new cases occurring each year worldwide among 15- to 49-year-olds (CDC yearly Report 2011). The bacterium is usually transmitted through sexual activity, a

man who is infected with the bacterium has a 25% chance of transmitting the infection to an uninfected female per sexual encounter (CDC yearly Report 2015). Chlamydia can also spread vertically, i.e. from a mother to child, the transmission rate from infected mother to new born is 50–60%, eventually causing conjunctivitis (Betha et al. 2016). Many times, patients with one STD are at high risk of coinfection with another STD (Paavonen and Eggert-Kruse 1999). Chlamydial infections are usually asymptomatic and produce less severe symptoms than other sexually transmitted infections. As there are no symptoms, the pathogen goes unnoticed until secondary or tertiary stages of disease develop (Malik et al. 2006).

The sequelae of undetected and untreated infections include acute salpingitis and cause adverse outcomes in women, like cervicitis, urethritis and pelvic inflammatory disease (PID)—which is an ascending infection of the uterus, fallopian tubes or neighbouring pelvic structures (Luis et al. 2017). Chlamydial infections are associated with

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13721-020-00243-w>) contains supplementary material, which is available to authorized users.

✉ Shilpa Shiragannavar  
shilpa.shiragannavar@gmail.com

<sup>1</sup> Department of Studies and Research in Bioinformatics and Biotechnology, Karnataka State Akkamahadevi Women’s University, Vijayapura, India

<sup>2</sup> Biodiscovery Group for Life Sciences, Bangalore, India

HIV infection and also with cervical cancer (Gewirtzman et al. 2011). In addition, infection with *C. trachomatis* has also been linked with adverse pregnancy outcomes with high risks for miscarriage, stillbirth, preterm birth and ectopic pregnancy (Hafner and Timms 2015). In males, it causes epididymitis, most severe complication that can also lead to infertility (Braxton et al. 2017). Additionally, infections with *C. trachomatis* can cause severe ocular infections by invading conjunctival epithelial cells. If not treated, repeated infections can result in trichiasis, i.e. in-turning of the eye lashes, leading to corneal abrasions, corneal scarring, opacification and ultimately blindness (Finethy and Coers 2016).

The bacteria have developed multidrug resistance and heterotypic resistance. The intracellular pathogen was found to be resistant to potent antibiotics, such as tetracycline, doxycycline, azithromycin, erythromycin and ofloxacin, and thus shows heterotypic resistance (Somani et al. 2000). Due to lack of symptoms and treatment compliance, the infections lead to severe reproductive complications. Although antibiotics are available, the chlamydia vaccine will dramatically reduce the rates of chlamydial infections (Elwell et al. 2016). In spite of many attempts made to develop one, there are no available vaccines yet. The protective immune response to *C. trachomatis* is generated by inducing CD4+ T cells. Also, Th-1 cytokines specifically INF- $\gamma$  and interleukin 12 are essential to induce protective immunity.

Initial efforts to develop the vaccines in 1950 was done to develop whole-organism vaccine in inactivated or live attenuated form. The first ever vaccine to treat the *C. trachomatis* infections was a live attenuated vaccine. The vaccine had a risk of immunopathology and also the production of large-scale pure chlamydia is very difficult. The vaccine was limited to reducing early periods of *C. trachomatis* infections. As the live vaccines are not safe always, killed or inactivated vaccines were studied. Heating and chemical treatment was used for inactivation. The inactivated vaccines could not provide maximum protection because of their inability to replicate and induce the immunity altogether.

Subunit vaccines are parts of antigens and can be the choice to overcome previous vaccines designs. Several studies indicate the Major Outer Membrane Protein (MOMP) shows good results for treating the pathogen. Although many attempts made to develop the vaccine against infections, the failures for success are still unclear. The failure of several attempts may be due to the fact that the protective immune responses developed may result in harmful effects for host. The whole-cell vaccine components could induce either immunopathogenic or protective responses. Subunit vaccines are efficient to overcome the complications with scrupulous design (Banatvala et al. 2009).

The study uses reverse vaccinology approach to evaluate the proteins of the pathogen. Whole proteome of the pathogen was taken from UniProt database and screened

for antigenicity, surface accessibility and allergenicity. Further, the proteins that are membranous and antigenic were screened for epitope-based vaccine design. Past studies have concentrated on the Polymorphic/Probable Membrane Proteins (Pmps), while the current study focuses on all the antigenic membrane bound proteins that are conserved and are immunologically important.

Vaccination could be more effective than any other biomedical interventions in controlling the epidemics of chlamydial infections. Unfortunately, yet there are no protective vaccines, either fully or partially available, though many attempts are made to develop a vaccine for chlamydial infections (Schautteet et al. 2011). An epitope is an antigenic determinant part that is important in bringing immunity of an organism. Reverse vaccinology is a computational analysis of genome that predicts the surface epitopes which are important in the development of a candidate vaccine (Kamampalliwar et al. 2013). B- and T-cell lymphocytes play a major role in the immune system which are responsible for the humoral and cell-mediated immunity (De Temmerman et al. 2011). T and B Cells in mammals originate from haematopoietic stem cells that are involved in adaptive or antigen-specific immune responses. The lymphocytes have receptors T-cell receptor (TCR) and B-cell receptor (BCR) on their surface that presents the antigen to the cells, which recognize and react to the antigenic epitopes specially. The antigen presenting cells (APC) carry the fragment antigen on special molecules called MHC proteins (Rees et al. 1999). CD4 cells or helper T cells recognize MHC class II proteins that coordinate the immune responses by stimulating the other immune cells, such as macrophages and monocytes. CD8 cells or cytotoxic T cells recognizes MHC class I proteins. The cytotoxic T cell attacks cells directly and destroys the antigens.

The T-cell activation as well as differentiation will possible only if the following three signals are present: (a) interaction of the peptide with the HLA molecule, (b) participation of cytokines that initiate clonal expansion and (c) signalling through co-stimulatory molecules (Rangarajan et al. 2018).

In the present study, an attempt is made to recognize major immunogenic epitopes on *C. trachomatis* proteins that can be vaccine candidates using various bioinformatics tools. Thus, the results offer novel epitopes for vaccine development against *C. trachomatis* infections.

## 2 Materials and methods

### 2.1 Protein sequence retrieval and evaluation of evolutionary analysis

The study protein sequence set included 895 protein sequences of *C. trachomatis* which were downloaded in

FASTA format from UniProt proteome database [www.uniprot.org](http://www.uniprot.org) (The UniProt Consortium 2016). The phylogenetic tree was constructed for all the sequences of *C. trachomatis* for the analysis of evolutionary divergence in various proteins using Clustal Omega tool (Sievers et al. 2011).

## 2.2 Antigenic protein identification and prediction of sub-cellular localization

Proteins were subjected to VaxiJen server v2.0 (Doytchinova and Flower 2007) for effective antigen prediction. The antigenic and non-antigenic proteins were identified, and antigenic proteins with high antigenic scores were considered for further analysis. *C. trachomatis* is a gram-negative bacterium and the sub-cellular localization of the pathogen proteins was done using PSORTb 2.0 (Gardy et al. 2004).

## 2.3 Allergy prediction

Allergenicity, i.e. whether the antigenic protein is allergen or non-allergen, was predicted using Allertop<sup>24</sup> (Dimitrov et al. 2014). The prediction is based on the physicochemical properties of proteins along with allergen prediction utilizing amino acid E-descriptors in the protein, auto-cross covariance transformation and together with several machine learning methods (Awad Elkareem et al. 2017).

## 2.4 Protein physical–chemical and functional analysis

Different physicochemical properties for the proteins were estimated using ProtParam (Gasteiger et al. 2005) server. The parameters including molecular weight (MW), stability index, isoelectric point (IP), half-life in vitro and in vivo, aliphatic index and grand average of hydropathicity (GRAVY) were estimated, while InterPro (Finn et al. 2016) identifies the conserved domains or important sites in the proteins.

## 2.5 Identification of T-cell epitopes, conservancy analysis and population coverage

Epitopes based on conserved regions were identified from cytotoxic T lymphocytes (CTL) using NetCTL v1.2 (Larsen et al. 2007). Prior to the run, the peptide length was set to 9.0 and the threshold was fixed at 0.5, while the sensitivity and specificity were set at 0.89 and 0.94. To initiate the immune response, the binding of the antigenic peptide to the MHC class I molecules is important. Hence, the binding of epitopes to the MHC I analysis was done using immune epitope database IEDB tools (Hoof et al. 2009). The tool calculates the half-maximal inhibitory concentration value (IC<sub>50</sub>) of the binding epitope to human leukocyte antigen (HLA) molecules. The tool predicts the binding of MHC

class I peptides to 12 different HLA supertypes using the stabilized matrix base method (Peters and Sette 2005). The conservancy of each epitope was also calculated using the conservancy analysis tools at IEDB (Bui et al. 2007). The stabilized matrix base method (SMM) tool at IEDB was used to calculate the total score which includes parameters, like processing score, TAP score, proteasomal cleavage score and binding affinity for the MHC I (Tenzer et al. 2005). Population coverage generally plays a crucial role in vaccine design which was calculated using the IEDB population coverage tool in this study (Bui et al. 2006).

## 2.6 Molecular docking studies

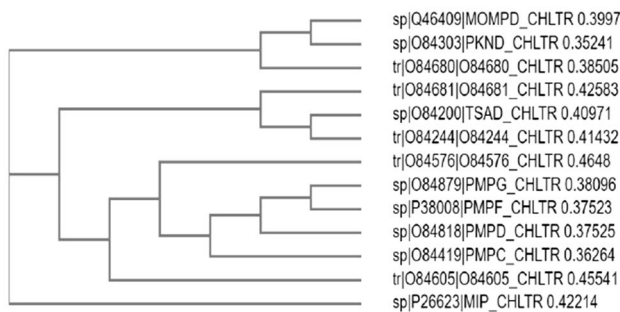
The 3D structures of epitopes were predicted by PEP-FOLD web-based server.2.0 (Thevenet et al. 2012). The server predicts the five most probable structures, and the best structure was taken for docking analysis, i.e. the lowest energy model. HLA-A\*01:01 allele showed the highest binding scores among all the other alleles. To validate the binding of identified epitopes with HLA molecule, the X-ray structure of HLA-A\*01:01 with 2.3 Å resolution was taken from PDB. Before performing the docking, the NP44 influenza epitope (Quiñones-Parra et al. 2014) was removed using UCSF Chimera (Pettersen et al. 2004). Further, the docking was performed with the DINC webserver, DINC is a meta-docking algorithm, in the sense that it relies on a standard docking tool; currently, AutoDock Vina was used to perform the sampling and scoring at each docking round (Dhanik et al. 2013).

## 2.7 Identification of the B-cell epitope

B-cell epitopes were identified using various tools from IEDB. The Bepipred linear epitope prediction analysis and Emini surface accessibility prediction (Emini et al. 1985), Kolaskar and Tongaonkar antigenicity scale (Kolaskar and Tongaonkar 1990), Karplus and Schulz flexibility prediction (Karplus and Schulz 1985), Parker Hydrophilicity Prediction (Parker et al. 1986) and Chou and Fasman (Chou and Fasman 2009) beta-turn prediction tools were used.

## 3 Results

The present investigation was focused on predicting the vaccine candidates for *Chlamydial* infections by a computational reverse vaccinology approach for which a total of 895 protein sequences of *C. trachomatis* were collected from the UniProt Proteome database in FASTA format. These protein sequences were analysed for their antigenic properties that potentially induce immunogenicity. Figure 1 shows a cladogram for the selected proteins.



**Fig. 1** Phylogenetic analysis of the selected outer membrane proteins

### 3.1 Antigenic protein identification and prediction of sub-cellular localization

Protein screening was done using the Vaxijen server which identifies the antigenic and non-antigenic proteins. 0.5 was kept as a threshold and the proteins with antigenic score >0.5 were considered for future analysis. Out of all 895 proteins, 477 proteins were found to be probable antigens. Since outer membrane proteins are important in therapeutic interventions, the proteins were screened for their localization using PSORTb 2.0 server. 15 proteins were found to be outer membrane proteins which were considered for further analysis.

### 3.2 Allergy prediction

Final protein screening was done based on their probable allergenic properties. Allertop was used for the analysis out of 15 outer membrane antigenic proteins 12 were non-allergenic. The proteins with negligible scores were selected to avoid the allergenic reactions of the vaccine. These 12 proteins then were subjected to epitope-based analysis.

### 3.3 Protein physical–chemical and functional analysis

The primary structure analysis of the proteins from ProtParam tools is given in Table 1. Amino acid composition,

**Table 1** ProtParam analysis for the proteins

Protein	No. of amino acids	Molecular weight	Theoretical pI	Extinction coefficients	Estimated half-life (h)	Instability index	Aliphatic index	GRAVY
PmpD	1531	160,748.22	4.75	105,795, 104,170	30	33.6	83.75	−0.056
PmpF	1034	112,392.82	8.88	118,900, 118,150	30	33.85	76.29	−0.284
PknD	934	107,666.41	6.24	143,895, 143,020	30	48.63	97.98	−0.113
Yop trans-location protein	921	100,580.4	5.55	56,185, 55,810	30	48.97	96.22	−0.186

molecular weight, isoelectric point (pI), grand average hydropathicity (GRAVY), instability index, aliphatic index, estimated half-life and extinction coefficient of the proteins were predicted using default parameters through ProtParam. The results suggest lower GRAVY and higher aliphatic index with acceptable aliphatic index values. The conserved domain analysis was performed using InterPro. The proteins have domains ranging from two to five.

### 3.4 Identification of T-cell epitopes and conservancy analysis

Since the selected 12 proteins are immunologically important, surface bound, conserved and non-allergenic, these proteins were then subjected to epitope-based peptide vaccine design analysis. NetCTL server was utilized for prediction of T-cell epitopes in the proteins. The server gives the highest combinatorial score of the epitopes from the proteins based on the IEDB MHC class I binding prediction. IC50 values less than 200 nM were considered this ensured selection of the MHC-I molecules for which the selected epitopes showed higher affinity. The list is given in Table 2.

**Table 2** T-cell epitopes identified according to the overall score predicted by the NetCTL server

Sl. no	Epitope	Total score (nM)
	TSDLGQMEY	3.5638
	PTDKTVFWY	3.4502
	CSDGEAVYY	3.3661
	FSNNFSDIY	3.3349
	ASDNANLLF	3.3127
	FIDLLQAIY	3.0324
	CSSTHPNKY	3.0243
	PTSEEFYY	2.9427
	LSNTEGYRY	2.8921
	LSWEMELAY	1.3958
	LSDSANSGL	1.2867
	NFDALRTLY	1.1811

The NetCTL server-predicted peptides were further predicted for effective designing of the T-cell epitopes, in combination with methods, like Proteasomal cleavage, Transporter Associated with antigen Processing (TAP) score, MHC-I binding affinity and processing. Proteasomal cleavage score gives the cleavage site in the peptide at C terminus. TAP score estimates the effective log values of the peptides binding to TAP. MHC-I processing prediction was done for analysing the binding efficiencies of the predicted peptides. It provides access to the prediction of antigen processing through the MHC-I antigen presenting pathway and ability to bind to a specific MHC-I molecule. The peptides with a higher total score represent higher processing capabilities. Based on the overall scores, the 12 epitopes were selected. The epitopes were subjected to MHC-I binding predictions using the stabilized matrix-based method. The epitopes that showed higher affinity, i.e.  $IC_{50} < 200$  nM, was considered for further analysis the epitopes were selected based on their  $IC_{50}$  values. The epitope conservancy was performed: the higher the conservancy the more the immunogenic are the proteins. Table 3 gives the scores for the interaction and conservancy analysis.

### 3.5 Population coverage

Each epitope that was recognized as optimum for the MHC-I binders was then subjected to the population coverage analysis. The epitopes indicated 100% coverage in India, East Asia, South Asia, Southeast Asia, North America, South America and 99.64% in Europe. The population coverage analysis in other areas is listed in Table 4.

**Table 4** Population coverage analysis for the proposed epitopes against *Chlamydia trachomatis*

Population	Coverage (%) <sup>a</sup>	Average hit <sup>b</sup>	PC90 <sup>c</sup>
India	100	6.38	1.31
East Asia	100	5.08	1.43
Europe	99.64	13.67	2.3
Northeast Asia	44.96	2.87	0.18
South Asia	100	6.41	1.31
Southeast Asia	100	3.99	1.33
Southwest Asia	58.14	6.8	0.24
Central Africa	49.95	5.0	0.2
East Africa	51.56	5.95	0.21
North Africa	57.84	7.04	0.24
North America	100	10.2	1.72
Oceania	56.01	3.93	0.23
South Africa	45.65	4.14	0.18
South America	100	6.09	1.35
West Africa	55.11	6.0	0.22
West Indies	66.66	7.87	0.3

<sup>a</sup>Projected population coverage

<sup>b</sup>An average number of epitope hits/HLA combinations recognized by the population

<sup>c</sup>A minimum number of epitope hits/HLA combinations recognized by 90% of the population

### 3.6 Molecular docking studies

To validate the epitopic potential of the identified peptides molecules, molecular docking studies were performed. As all the selected epitopes showed a good binding affinity with allele HLA-A\*01:01, its structure was downloaded from PDB with ID 4nqv. The epitopes structure was predicted by PEPFold, and the molecular docking was performed for

**Table 3** Interaction and conservancy results of identified T-cell epitopes using tools for binding and conservancy analysis in IEDB

Epitope	Interacting MHC-I allele (proteasome score, TAP score, MHC-I score and processing score)	Epitope conservancy (%)
TSDLGQMEY	HLA-A*01:01(1.83) HLA-A*29:02(-0.02) HLA-B*35:01(0.19)	86.58
FIDLLQAIY	HLA-A*01:01(1.26) HLA-A*29:02(0.88) HLA-B*35:01(0.68) HLA-C*03:02(0.14)	86.24
FSNNFSDIY	HLA-A*01:01(1.58) HLA-A*29:02(0.85) HLA-B*15:01(0.46) HLA-B*35:01(1.12) HLA-B*58:01(0.13) HLA-C*02:02(0.99) HLA-C*03:02(1.27) HLA-C*12:02(0.83) HLA-C*14:02(0.1) HLA-C*16:01(0.89)	84.45
LSNTEGYRY	HLA-A*01:01(0.7) HLA-A*29:02(0.34) HLA-B*58:01(0.28)	89.71
LSWEMELAY	HLA-A*01:01(0.75) HLA-A*29:02(1.45) HLA-B*15:01(0.7) HLA-B*35:01(1.58) HLA-B*53:01(0.22) HLA-B*58:01(1.07) HLA-C*03:02(1.21) HLA-C*16:01(0.69)	91.95

the HLA-A\*01:01 against the predicted epitopes as ligands using DINC web server that relies on Autodock vina. The XYZ coordinates were set to  $x=60.2$ ,  $y=2.2$  and  $z=40.4$ . The structures of the predicted epitopes and the PDB structure of HLA-A\*01:01 is shown in Fig. 2. The interaction of the epitopes and HLA-A\*01:01 is visualized using Chimera; the interactions are shown in Fig. 3. The HLA-A\*01:01 amino acids Ser, His, Val, Arg and Asp are involved in the interactions with LSWEMELAY as shown in Fig. 4. The binding energies and the RMSD values for the epitopes are given in Table 5.

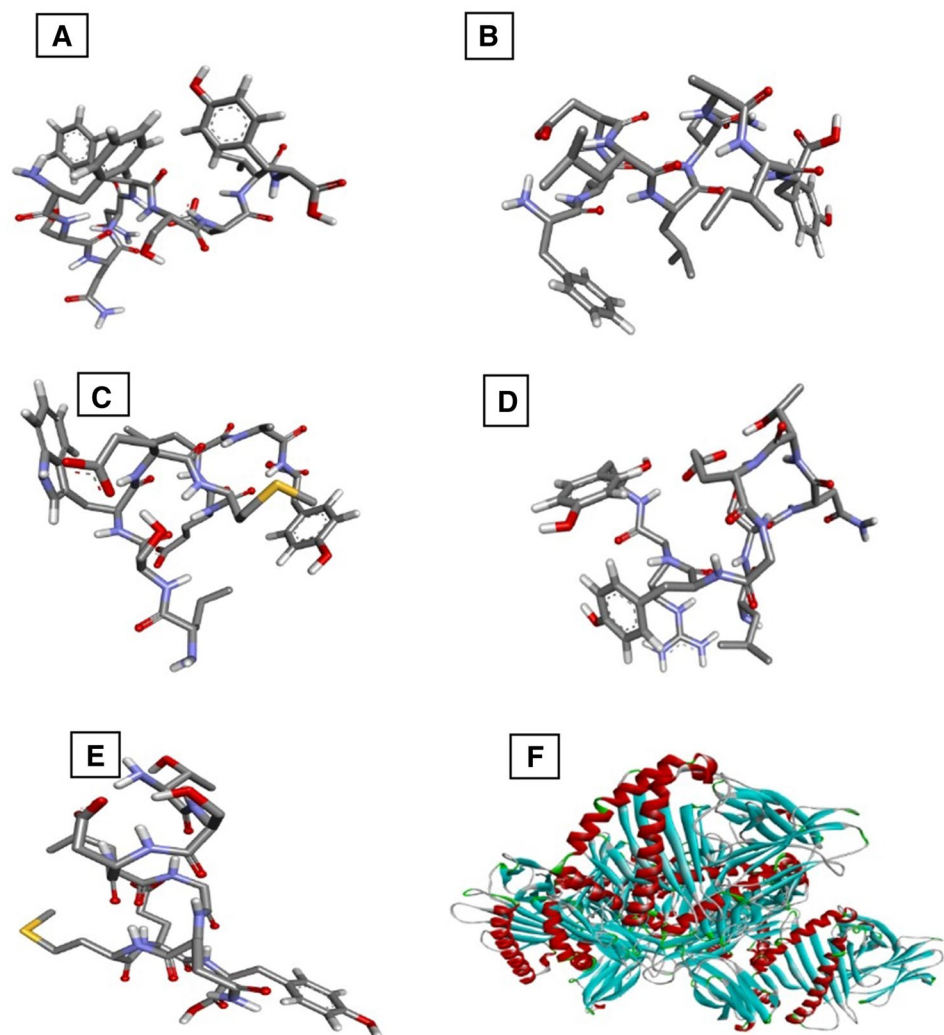
### 3.7 B-cell epitope prediction

The protein epitopes were further evaluated to identify for B cell-specific epitopic nature by various online tools IEDB. To evaluate the physicochemical properties of the amino acids in the protein and their ability to identify the B-cell epitopes was performed by Kolaskar and Tongaonkar antigenicity prediction tool. The peptides FSNNFSDIY,

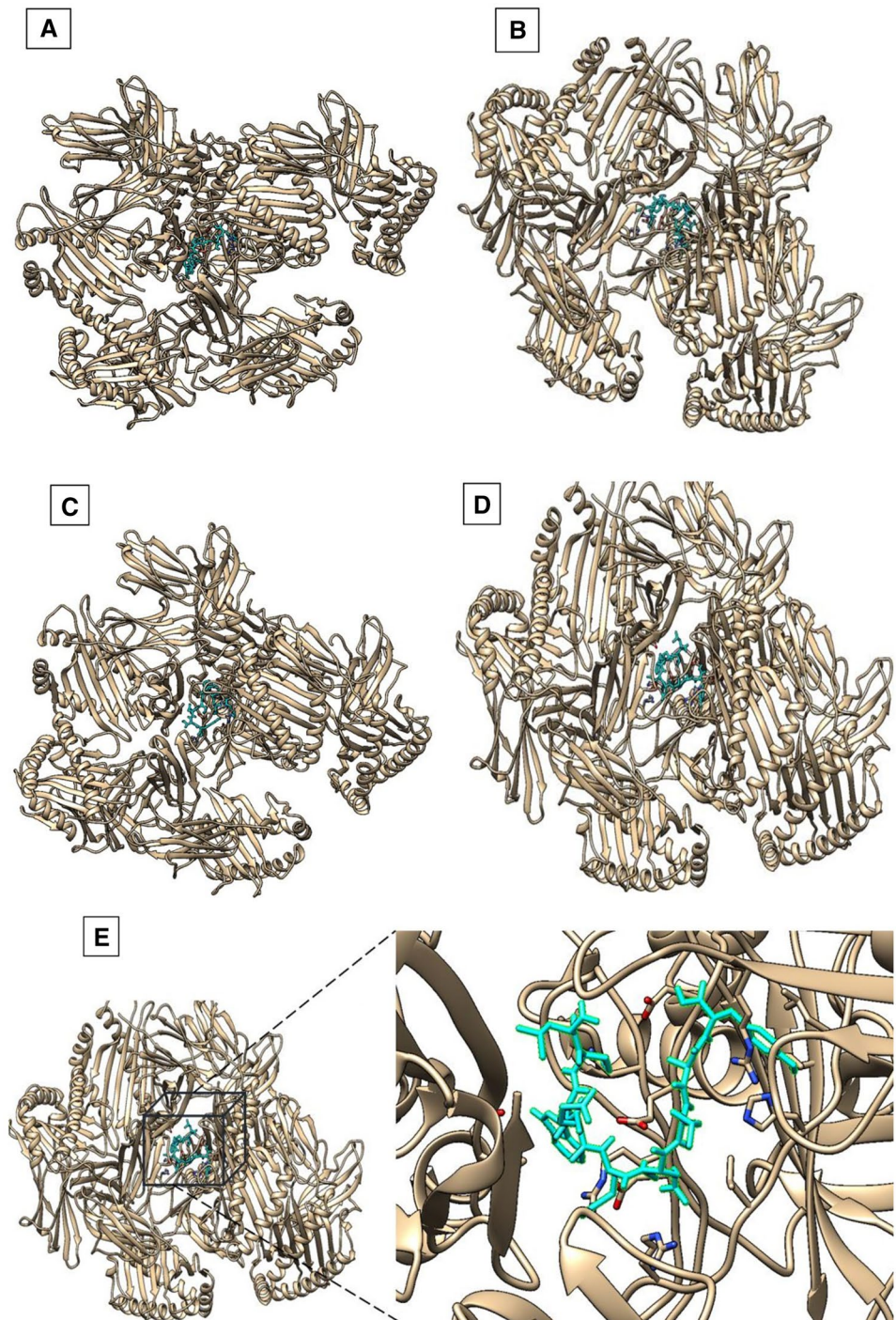
FIDLLQAIY, LSNTGYRY, LSWEMELAY and TSDLGQMEY showed the score ranging from 1.004, 1.608, 0.976, 0.99 and 0.965, respectively. The Emini surface accessibility prediction was carried out for checking the surface accessibility of the epitopes. The Emini surface accessibility prediction scores for the selected peptides were ranging from 0.796 to 2.542, indicating good surface accessibility. Parker hydrophilicity prediction tool was utilized to check the hydrophilic regions of the proteins. The identification of surface accessible hydrophilic regions is essential as these regions are likely to evoke the B-cell immune responses. Results indicated that the selected peptides are hydrophilic.

There is a correlation between the location of antigenic sites and the prediction of turns in proteins; hence, Beta-turns prediction in proteins was performed as these regions are involved in initiating antigenic properties (Pellequer et al. 1993). Chou–Fasman beta-turn prediction was performed to find the beta-turn regions in the proteins. Karplus Schulz's flexibility prediction tool was used to know the flexible regions in the proteins. The regions of

**Fig. 2** Structures of epitopes **a** 'FSNNFSDIY', **b** 'FIDLLQAIY', **c** 'LSWEMELAY', **d** 'LSNTEGYRY', **e** 'TSDLGQMEY' predicted by PepFold and structure of **f** HLA-A\*01:01 retrieved from PDB



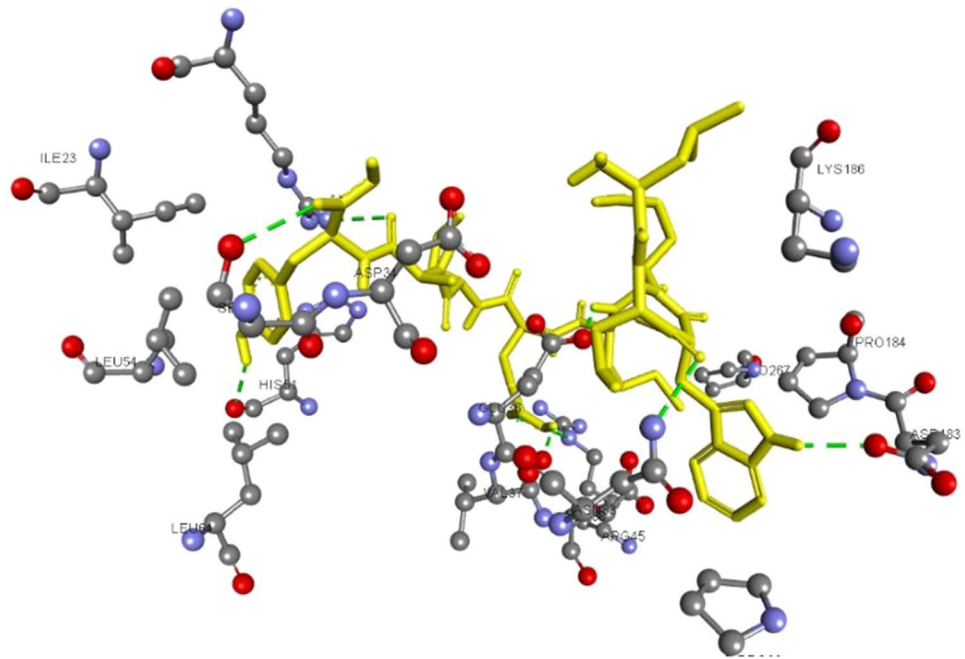
**Fig. 3** Visualization of docking results of epitopes **a** 'TSDLGQMEY', **b** 'FIDLLQAIY', **c** 'FSNNFSDIY', **d** 'LSNTEGYRY', **e** 'LSWEMELAY' with the HLA-A\*01:01 using Chimera. The epitopes interacting at the centre of the HLA-A\*01:01 protein



the peptide locations were considered in the most favourable regions of the flexibility prediction analysis. The peptides showed good flexibility scores ranging from 0.936 to 1.064. Further, the linear B-cell epitopes were predicted using Bepipred, hidden Markov model-based machine learning process. Almost all the regions in the proteins showed the most favorable regions in BepiPred linear epitope predictions. By analysing overall data obtained from the B-cell epitope prediction tools, the peptides as

selected as T-cell epitopes were found to satisfy the requisites required for inducing B-cell response. The overall result of B-cell epitope analysis is represented in Fig. 5. Molecular dynamic simulations give a time-dependent evolution of the complex interaction networks which govern some of the fundamental aspects as molecular recognition, binding strength, and mechanical and structural stability. Figure 6 gives the distance between masses of amino acid residues through 1000 ns simulations

**Fig. 4** The green dotted line indicate the hydrogen bond interactions between epitope ‘LSWEMELAY’ and HLA-A\*01:01 chains (colour figure online)



**Table 5** Binding energy and RMSD values for the epitopes with the HLA-A\*01:01

Epitope	Binding energy (kcal/mol)	RMSD Å	Vaxijen score
TSDLGQMEY	-8.20	0.00	0.5772
FIDLLQAIY	-8.40	0.00	0.501
FSNNFSDIY	-8.60	0.00	0.5727
LSNTEGYRY	-8.40	0.00	1.0623
LSWEMELAY	-8.60	0.00	1.6240

performed using NAMD and VMD molecular dynamics simulator and visualizer programs.

## 4 Discussion

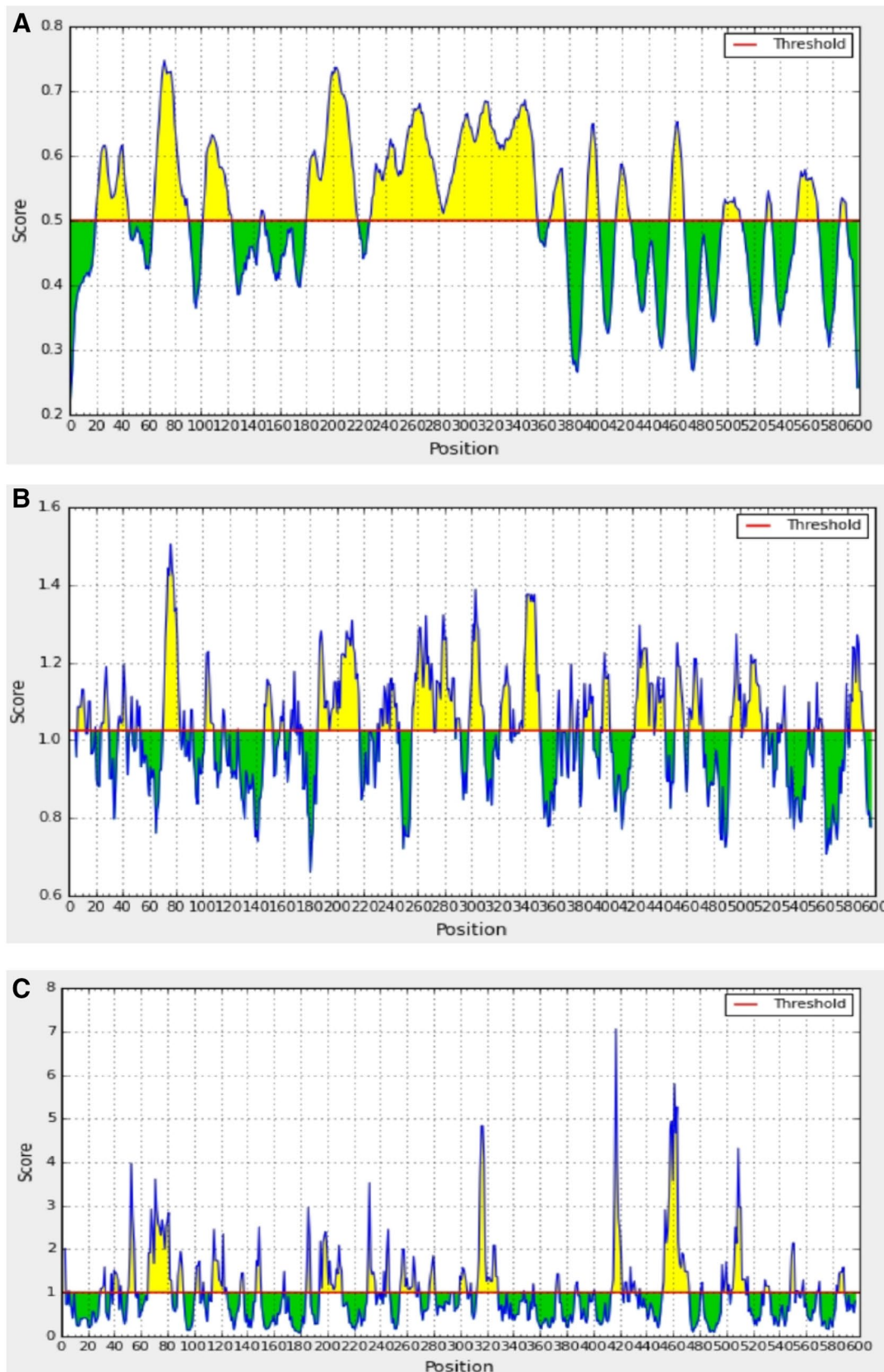
Although many attempts made, Chlamydia genital vaccines are still under discussion. The priority in developing a vaccine is safety. Therefore, the negative effects, delayed hypersensitive reactions and susceptibility to the infections have to be avoided. Inactivated or live attenuated vaccine could treat trachoma but how effective the vaccine could be for genital infections is a question. A distinctive T-cell and B-cell vaccination initiates both humoral and cell-mediated immunity effective on controlling the reinfection, engender protection and minimizing the host pathology. Studies confirm the release of CD4+ T cells and IFN  $\gamma$  cells to protect from the infections. Thus, suggesting the importance of identification of T-cell epitopes epitope-based vaccine approach has been

reported for *Francisella tularensis* (Whelan et al. 2018), *Haemophilus influenzae* (Zahroh et al. 2016), *Dengue virus* (Chakraborty et al. 2010), *Human coronaviruses* (Oany et al. 2014), *Chikungunya virus* (Islam et al. 2012), and *Ebola* (Srivastava et al. 2016).

The proteome of *C. trachomatis* was retrieved from UniProt proteome database. The proteins were further analysed for antigenic properties using vaxijen; the antigenic proteins with antigenic score >0.5 were then subjected for localization prediction using Psortb. The outer membrane proteins with higher antigenicity scores were then evaluated for allergic properties using Allertop. The conserved domain analysis was done using InterPro. The membrane-bound proteins with higher antigenic scores, non-allergens with maximum conserved domains, were selected for the further immunoinformatic study. The peptidyl-prolyl *cis-trans* isomerase Mip, Major outer membrane porin, Polymorphic/Probable outer membrane protein C (PmpC), Polymorphic/Probable outer membrane protein G (PmpG), Polymorphic/Probable outer membrane protein D (PmpD), Polymorphic/Probable outer membrane protein F (PmpF), Serine/threonine-protein kinase (PknD), tRNA N6-adenosine threonylcarbamoyltransferase, Yop proteins translocation protein, Omp85 Analog protein, Peptidoglycan-associated protein and Gen. Secretion Protein D were the proteins finally considered for further analysis. The proteins were subjected to NetCTL T-cell epitope analysis; around 12 epitopes were selected from these proteins based on IC50 values. Peptides with the highest scores have the highest processing capabilities.

An effective peptide epitope should be conserved among the host proteins, good processing capabilities and binding





**Fig. 5** Bepipred linear epitope prediction of the most antigenic protein PmpF. The threshold is 0.5. The yellow color indicates flexible regions in polypeptide while green color indicates the region that could not satisfy the threshold margin (colour figure online)

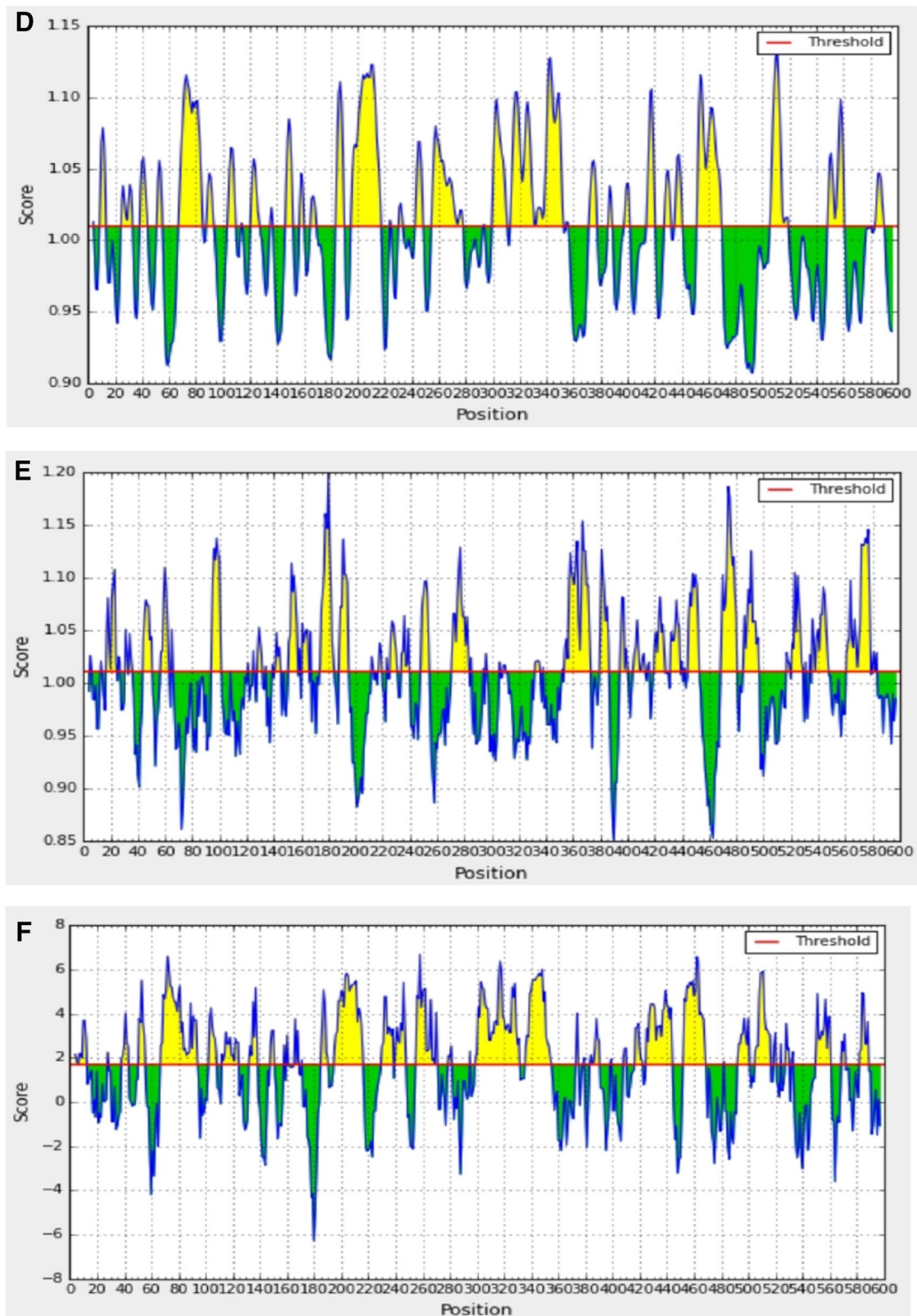
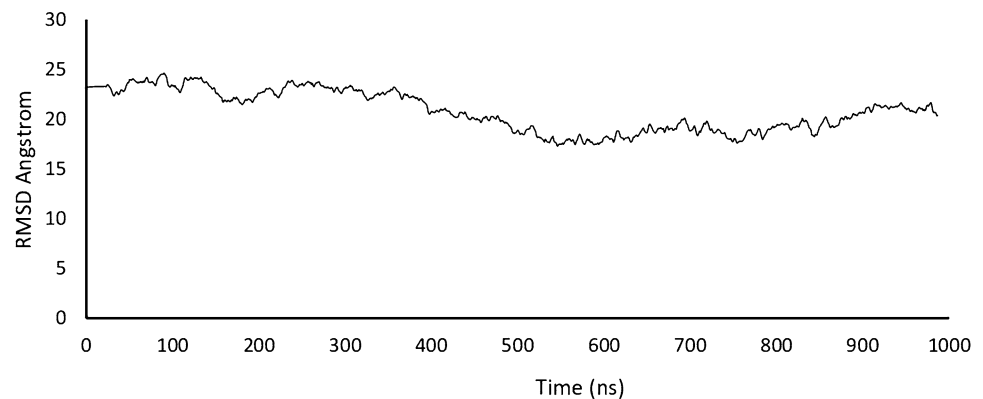


Fig. 5 (continued)

affinities with the MHC alleles with higher population coverage. Five peptides were selected from 12 epitopes based on the epitope conservancy, and binding with the

MHC alleles. All the five epitopes showed good conservancy and interactions with the HLA alleles. However, the epitope LSWEMELAY showed maximum conservancy

**Fig. 6** Distance between masses of amino acid residues through 1000 ns simulations performed using NAMD and VMD molecular dynamics simulator and visualizer programs



with 91.95% and interacting with eight HLA variants. Further population coverage analysis showed maximum coverage in most regions. It calculates the percentage of people showing potential responses to the query epitopes among the people living in that region. The selected five epitopes were subjected for molecular docking analysis, and the results showed good binding interactions with the HLA-A allele. The epitope LSWEMELAY showed the lowest binding energy of  $-8.60$  and RMSD  $0.00 \text{ \AA}$  with the highest antigenicity score of 1.6240. B-cell epitope identification was also performed using various tools of IEBD; these epitopes induce primary and secondary immunity. After analysing the results, it was found that the selected T-cell epitopes satisfied the potential to B-cell epitopes also. IEDB B-cell epitope prediction is a collection of methods to identify B-cell epitopes in the given antigenic sequence. The results of the analysis are given in Supplementary Table 1. The BepiPred method results indicate the predicted T-cell epitopes fall above the threshold.

The nine mere epitopes TSDLGQMEY, FSNNFSDIY, LSWEMELAY, FIDLLQAIY and LSNTEGYRY were identified in proteins with positions 465 and 808 in proteins PmpD, 952 in PmpF, 787 in PknD and 874 in Yop translocation protein, respectively.

The predicted B- and T-cell epitopes were analysed for IFN-gamma reactivity using IFNepitope (Ghahremanifard et al. 2019). IFNepitope is a module designed for the prediction of IFN-gamma-inducing regions in a protein or antigen. The findings suggests that the predicted epitopes have good probability to release IFN- $\gamma$  with positive scores except FIDLLQAIY. The epitope FSNNFSDIY showed the highest score of 0.774, epitope LSWEMELAY 0.5702, epitope TSDLGQMEY 0.4093 and epitope LSNTEGYRY showed positive score of 0.4072.

Many studies infer major outer membrane proteins important in designing a subunit vaccine and can protect against infertility. In the present study, outer membrane proteins were screened for immunologically important proteins. The identified epitope sequences were evaluated for conservancy

among the other chlamydia strains. The epitopes are conserved among 27 Polymorphic/Probable outer membrane protein sequences in different chlamydia strains. The multiple sequence alignment of the epitopes with proteins from different strains is given in Supplementary Fig. 1. As the epitopes were conserved among the other chlamydia species, the vaccine developed may elicit cross-serovar protective responses.

A chimeric peptide was generated by joining the five peptides, and the B-cell analysis was performed for the chimeric peptide. The results of the IEDB B-cell epitope prediction on physicochemical properties of the chimeric protein are shown in Supplementary Fig. 2. BepiPred server predicts the sequential epitopes based on Random Forest algorithm, and the results are shown in Supplementary Fig. 3. SYFPEITHI was used for the prediction of T-cell epitopes. The analyses show that the chimeric peptide has potent epitopes that can be detected by immune cells and molecules, and the results are shown in Supplementary Table 2. Further population coverage for the chimeric peptide was performed, and the results are given in Supplementary Table 3.

## 5 Conclusion

The untreated chlamydial infections continue to have a negative impact on the reproductive health of humans due to a lack of adequate treatment options. In the present study, an attempt was made in designing a peptide-based vaccine for the pathogen *C. trachomatis* which causes infertility in both men and women. Through in silico approach using bioinformatics tools, novel therapeutic epitope vaccine candidates were identified. Both T-cell and B-cell epitopes are offered for stimulating long-term immunity. The identified peptides show both B- and T-cell selectivity, wide population coverage, good conservancy and significant binding interactions with the MHC-1 HLA allele. Studies not only of B- and T-cell epitopes but also of INF- $\gamma$  induction were done. The predicted epitopes

supposed to offer protective immunity for the long term against *C. trachomatis*. The chimeric peptide with multiple antigenic regions from different proteins is interacting with different HLA molecules; hence, the combined epitope vaccine will be more efficacious. However, in vitro studies have to be performed in order to validate the predicted vaccine candidates. Further studies to experimentally validate the obtained data from study are in progress. The predicted epitopes supposed to offer protective immunity for the long term against *C. trachomatis* infections. However, in vitro studies have to be performed in order to validate the predicted vaccine candidates.

## References

- Awad Elkareem MA, Ahmed Osman S, Mohamed HA (2017) Prediction and conservancy analysis of multi-epitope based peptide vaccine against merkel cell polyomavirus: an immunoinformatics approach. *Immun Res* 13(134):2
- Banatvala JE, Griffiths P, Schoub B, Mortimer P (eds) (2009) Principles and practice of clinical virology, Wiley
- Betha K, Robertson JM, Tang G, Haggerty CL (2016) Prevalence of Chlamydia trachomatis among childbearing age women in India: a systematic review. *Infect Dis Obstet Gynecol* 2016:8561645
- Braxton J, Davis DW, Emerson B, Flagg EW, Grey J, Grier L et al (2017) Sexually transmitted disease surveillance. Department of Health and Human Services: National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Atlanta, GA
- Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 7(1):153
- Bui HH, Sidney J, Li W, Fusseder N, Sette A (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform* 8(1):361
- Centers for Disease Control and Prevention (CDC) (2011) CDC Grand Rounds: Chlamydia prevention: challenges and strategies for reducing disease burden and sequelae. *MMWR Surveill Summ* 60(12):370
- Centers for Disease Control and Prevention (2015) Sexually transmitted diseases treatment guidelines. *MMWR Recomm Rep* 64(RR-3):1–137
- Chakraborty S, Chakravorty R, Ahmed M, Rahman A, Waise TM, Hassan F, Rahman M, Shamsuzzaman S (2010) A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. *Silico Biol* 10(5–6):235–246
- Chou PE, Fasman GD (2009) Amino acid sequence. *Adv Enzymol Relat Areas Molec Biol* 47:45
- De Temmerman ML, Rejman J, Demeester J, Irvine DJ, Gander B, De Smedt SC (2011) Particulate vaccines: on the quest for optimal delivery and immune response. *Drug Discov Today* 16(13–14):569–582
- Dhanik A, McMurray JS, Kaviraki LE (2013) DINC: a new Auto-Dock-based protocol for docking large ligands. *BMC Struct Biol* 13(1):S11
- Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) AllerTOP v. 2—a server for in silico prediction of allergens. *J Mol Model* 20(6):2278
- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumor antigens and subunit vaccines. *BMC Bioinform* 8(1):4
- Elwell C, Mirrashidi K, Engel J (2016) Chlamydia cell biology and pathogenesis. *Nat Rev Microbiol* 14(6):385
- Emini EA, Hughes JV, Perlow D, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55(3):836–839
- Finethy R, Coers J (2016) Sensing the enemy, containing the threat: cell-autonomous immunity to Chlamydia trachomatis. *FEMS Microbiol Rev* 40(6):875–893
- Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J (2016) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res Suppl* 45(D1):D190–D199
- Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS (2004) PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 21(5):617–623
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. Humana Press, Totowa, pp 571–607
- Gewirtzman A, Bobrick L, Conner K, Tying SK (2011) Epidemiology of sexually transmitted infections. Sexually transmitted infections and sexually transmitted diseases. Springer, Berlin, pp 13–34
- Ghahremanifard P, Afzali F, Rostami A, Nayeri Z, Bamba B, Minuchehr Z (2019) Designing a novel multi-epitope T vaccine for “targeting protein for Xklp-2” (TPX2) in hepatocellular carcinoma based on immunoinformatics approach. *International Journal of Peptide Research and Therapeutics*, pp 1–10
- Hafner LM, Timms P (2015) Development of a vaccine for *Chlamydia trachomatis*: challenges and current progress. *Vaccine Dev Therapy* 5:45–58
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61(1):1
- Islam R, Sakib MS, Zaman A (2012) A computational assay to design an epitope-based peptide vaccine against chikungunya virus. *Future Virol* 7(10):1029–1042
- Kanampalliwar AM, Rajkumar S, Girdhar A, Archana T (2013) Reverse vaccinology: basics and applications. *J Vaccines Vaccin* 4:194. <https://doi.org/10.4172/2157-7560.1000194>
- Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. *Naturwissenschaften* 72(4):212–213
- Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276(1–2):172–174
- Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform* 8(1):424
- Luis M, Zhong G, Brunham RC (2017) Update on *Chlamydia trachomatis* vaccinology. *Clin Vaccine Immunol* 24(4):e00543–e616
- Malik A, Jain S, Hakim S, Shukla I, Rizvi M (2006) *Chlamydia trachomatis* infection and female infertility. *Indian J Med Res* 123(6):770
- Oany AR, Emran AA, Jyoti TP (2014) Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in-silico approach. *Drug Des Dev Ther* 8:1139
- Paavonen J, Eggert-Kruse W (1999) *Chlamydia trachomatis*: impact on human reproduction. *Infect Dis Obstet Gynecol* 5(5):433–447
- Parker JM, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25(19):5425–5432

- Pellequer JL, Westhof E, Van Regenmortel MH (1993) Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunol Lett* 36(1):83–99
- Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinform* 6(1):132
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612
- Quiñones-Parra S, Grant E, Loh L, Nguyen TH, Campbell KA, Tong SY, Miller A, Doherty PC, Vijaykrishna D, Rossjohn J, Gras S (2014) Pre-existing CD8+ T-cell immunity to the H7N9 influenza A virus varies across ethnicities. *Proc Natl Acad Sci* 111(3):1049–1054
- Rangarajan S, He Y, Chen Y, Kerzic MC, Ma B, Gowthaman R, Pierce BG, Nussinov R, Mariuzza RA, Orban J (2018) Peptide–MHC (pMHC) binding to a human antiviral T cell receptor induces long-range allosteric communication between pMHC- and CD3-binding sites. *J Biol Chem* 293(41):15991–16005
- Rees W, Bender J, Teague TK, Kedl RM, Crawford F, Marrack P, Kappler J (1999) An inverse relationship between T cell receptor affinity and antigen dose during CD4+ T cell responses in vivo and in vitro. *Proc Natl Acad Sci* 96(17):9781–9786
- Schautteet K, De Clercq E, Vanrompay D (2011) *Chlamydia trachomatis* vaccine research through the years. *Infect Dis Obstet Gynecol* 2011:963513
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(1):539
- Somani J, Bhullar VB, Workowski KA, Farshy CE, Black CM (2000) Multiple drug-resistant *Chlamydia trachomatis* associated with clinical treatment failure. *J Infect Dis* 181(4):1421–1427
- Srivastava PN, Jain R, Dubey SD, Bhatnagar S, Ahmad N (2016) Prediction of epitope-based peptides for vaccine development from coat proteins GP2 and VP24 of Ebola virus using immunoinformatics. *Int J Pept Res Ther* 22(1):119–133
- Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhütter HG (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci CMLS* 62(9):1025–1037
- UniProt: the universal protein knowledgebase (2017) *Nucleic acids research*. 45(D1):D158–D169
- Thevenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tuffery P (2012) PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res* 40(W1):W288–W293
- Whelan AO, Flick-Smith HC, Homan J, Shen ZT, Carpenter Z, Khoshkenar P, Abraham A, Walker NJ, Levitz SM, Ostroff GR, Oyston PC (2018) Protection induced by a *Francisella tularensis* subunit vaccine delivered by glucan particles. *PLoS ONE* 13(10):e0200213
- Zahroh H, Ma'rup A, Tambunan US, Parikesit AA (2016) Immunoinformatics approach in designing an epitope-based vaccine against meningitis-inducing bacteria (*Streptococcus pneumoniae*, *Neisseria meningitidis*, and *Haemophilus influenzae* type b). *Drug Target Insights* 10:19–29

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.