



Epitope-Based Vaccine Designing of *Nocardia asteroides* Targeting the Virulence Factor Mce-Family Protein by Immunoinformatics Approach

Prasanta Patra¹ · Niladri Mondal¹ · Bidhan Chandra Patra^{1,2} · Manojit Bhattacharya^{1,2} 

Accepted: 16 August 2019 / Published online: 21 August 2019
© Springer Nature B.V. 2019

Abstract

Nocardia asteroides is the main causative agent responsible for nocardiosis disease in immunocompromised patient viz. Acquired Immunodeficiency Syndrome (AIDS), malignancy, diabetic, organ recipient and genetic disorders. The virulence factor and outer membrane protein pertains immense contribution towards the designing of epitopic vaccine and limiting the robust outbreak of diseases. While epitopic based vaccine element carrying B and T cell epitope along with adjuvant is highly immunoprophylactic in nature. Present research equips immunoinformatics to figure out the suitable epitopes for effective vaccine designing. The selected epitopes VLGSSVQTA, VNIELKPEF and VVPSNLFAV amino acids sequence are identified by HLA-DRB alleles of both MHC class (MHC-I and II) molecules. Simultaneously, these also accessible to B-cell, confirmed through the ABCPred server. Antigenic property expression is validated by the Vaxijen antigenic prediction web portal. Molecular docking between the epitopes and T cell receptor delta chain authenticate the accurate interaction between epitope and receptor with significantly low binding energy. Easy access of epitopes to immune system also be concluded as transmembrane nature of the protein verified by using of TMHMM server. Appropriate structural identity of the virulence factor Mce-family protein generated through Phyre2 server and subsequently validated by ProSA and PROCHECK program suite. The structural configuration of these epitopes also shaped using DISTILL web server. Both the structure of epitopes and protein will contribute a significant step in designing of epitopic vaccine against *N. asteroides*. Therefore, such immunoinformatics based computational drive definitely provides a conspicuous impel towards the development of epitopic vaccine as a promising remedy of nocardiosis

Keywords Immunocompromised · Nocardiosis · Immunoinformatics · Protein · Epitope · Vaccine

Abbreviations

AIDS	Acquired immune deficiency syndrome
MHC	Major histocompatibility complex
HLA	Human leukocyte antigen
NCBI	National centre for biotechnology information
PDB	Protein data bank
ACE	Atomic contact energy
UCSF	University of California, San Francisco

ProSA-web	Protein structure analysis
BCPreds	B-Cell epitope prediction Server
MCC	Matthew correlation coefficient

Introduction

The gram positive bacillus bacteria, *Nocardia asteroides* causing nocardiosis targeting to immunocompromised patients in direct state (Wilson 2012). A total of 225 species have been estimated till now under the genera of *Nocardia* (Bennett et al. 2014). It is an opportunistic pathogen responsible for several diseases in human as well as other vertebrate animal (Moylett et al. 2003). Practically these bacteria serve as predominant causing factors for pulmonary infection. Consequently it shows acute necrotizing pneumonia that also reflects in the inflammation of cutaneous and subcutaneous tissue. While the brain is recognized as most favored site for secondary infection in

✉ Manojit Bhattacharya
pfzf_manojit@mail.vidyasagar.ac.in

¹ Department of Zoology, Vidyasagar University, Midnapore 721 102, West Bengal, India

² Centre For Aquaculture Research, Extension & Livelihood, Department of Aquaculture Management & Technology, Vidyasagar University, Midnapore 721 102, West Bengal, India

25% patients amongst all victims (Ellis and Beaman 2002; Quinones-Hinojosa 2012). Subsequently *N. asteroides* is the prime cause of serious cerebral abscess as reported by researcher (Fleetwood et al. 2000). Nocardiosis is distributed worldwide but it is most prevalence through the United States (Patil et al. 2012; Saubolle and Sussland 2003). Infection of nocardiosis is influenced by the sex and age of the patient. As noted the male female patient ratio 3:1 reflects that men are prone to nocardiosis rather than the female. Recently, around 0.375 nocardia infection for every 100 000 individuals have been reported throughout the world (Palmieri et al. 2014). Exclusively 500–1000 individuals are affected in United States every year in a random mode (Sethy et al. 2016). Numerous endeavors are taken to overcome the fatal disease but till now the disease is prevalence as research data reported. The Nocardiosis disease is predominant to the alcoholic, diabetic, cancer, AIDS and organ transplanted patient so, its need immediate control and monitoring supports. In present scenario records of these patients are quiet in frequent numbers. Therefore, this disease already occupied great concern and requires immediate remedy by promising, cost effective ways. Vaccination against the disease will be the finest and reliable option for controlling the disease outbreak (Nichol et al. 2003). A vaccine would have more effective if it can elicit both B and T-cell dependent immune responses. Vaccine designing against any disease can be performed using two specific approaches, conventional approach and reverse vaccinological approach. Within the present research plan the reverse vaccinological approach has been applied to design epitopic vaccine against *N. asteroides* as the conventional technique has some limitations. The conventional technique requires much time and also very expensive to formulate. Furthermore, the technique applied to killed, attenuated or indolent pathogens for vaccine development which will prove deadly if the pathogen alters to its pathogenic appearance (Khan et al. 2015). Besides from such conventional approach, the reverse vaccinology harnesses the immunoinformatics for developing vaccine against pathogenic diseases (Vivona et al. 2008). The advance approach involves several web portals and computational tools to predict the potential vaccine candidate. The fundamentality of reverse vaccinology showing it targets of protein derived epitopic sequence for novel vaccine generation. The reverse vaccinology provides easy and reliable techniques to find out the accurate B and T-cell epitopes (Srivastava et al. 2019). While the exomembrane and secreted protein of pathogens are the ideal for vaccine generation as these can easily interact with immune system (Gourlay et al. 2013). Considering the bacterial proteins are foreign to host body, it will be further recognized as antigenic determinants. Expression of both B and T-cell epitopes will enhance the immunogenicity driven by B

and T-cell proliferation. The virulence factor of Mce family protein found to be responsible for the virulence of *N. asteroides* and pathogenicity expression. The current research focuses on the computational analysis and identification of potential epitopes present within the bacterial Mce family protein candidate. Therefore, identified epitopes will be served as ideal component for future vaccine development against the infection of *N. asteroides*. In order to execute the subsequent research performance the supreme computational aided techniques has been employed. Consequently the targeted protein sequence has been crucially analyzed through several authentic web prediction portals focus to extract the functional epitopes leading to narrative vaccine generation against Nocardiosis disease.

Materials and Methods

Retrieval of Amino Acids Sequence

The amino acid sequences of a protein is the preliminary requirement for development of computer aided epitopic vaccine component. Amino acid sequence of the virulence factor Mce-family protein has been retrieve from NCBI protein database (Coordinators 2017). Identified amino acids sequence was downloaded and carried out further computer based analysis particularly supported in FASTA file format.

B-Cell Epitope Prediction

B-cell epitope identification is one of the most crucial steps intended to development of epitopic vaccine (Jamal et al. 2017). Potential B-cell epitope has been predicted employing the ABCpred web server (Saha and Raghava 2006). The server utilizes recurrent neural network technique to locate the specific B-cell epitopes present in targeted protein sequence. The server also uses five parameters to validate the predicted outputs.

These factors are:

$$Q_{\text{sens}}(\text{sensitivity}) = \frac{TP}{TP + FN} \times 100\%$$

$$Q_{\text{spec}}(\text{specificity}) = \frac{TN}{TN + FP} \times 100\%$$

$$Q_{\text{acc}}(\text{accuracy}) = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%$$

$$Q_{\text{ppv}}(\text{positive prediction value}) = \frac{TP}{TP + FP}$$

MCC (Matthew Correlation Coefficient)

$$= \frac{(TP)(TN) - (FP)(FN)}{\sqrt{[TP + FP][TP + FN][TN + FP][TN + FN]}}$$

TP, FN, TN and FP represents true positive, false negative, true negative and false positive output respectively (Saha and Raghava 2006). The collective FASTA sequence of targeted protein virulence factor Mce-family protein has been submitted to the ABCpred server. Further, for the B-cell epitope identification, threshold score is set to 0.75 and window length is set to 20 in order to clarify the result. Moreover an overlapping filter has been used to eliminate the overlapped epitopic prediction.

T-Cell Epitope Prediction Within the B-Cell Epitope

In order to generate the immense immune response, an antigenic epitope necessarily to be accessible for the both MHC class type molecules (Naz et al. 2015) along with the B-cell (Barh et al. 2010; Bhattacharya et al. 2019). In support to the prediction of potent T-cell epitopes, the ProPred (Singh and Raghava 2001) and ProPred-I (Singh and Raghava 2003) servers are appointed for MHC-II and MHC-I molecules correspondingly. The ProPred and ProPred-I web servers predict epitopes that can be recognized by 51 MHC-II and 47 MHC-I alleles (Mustafa and Shaban 2006). Both the server applied matrices prediction algorithm method (Lafuente and Reche 2009; Lin et al. 2008) to find out the potent T-cell epitope. Current research work predicted epitopes for B-cell, are submitted to the ProPred and ProPred-I server with default parameters to identify the common epitopes anticipate by both B and T-cell.

Analysis of Antigenic Property of Epitopes

Predicted common epitopes for both the B and t-cell are further investigated for their antigenic property through the VaxiJen (v. 2.0) server (Doytchinova and Flower 2007). The server requires a query sequence of amino acids to evaluate the antigenic propensity with 70–97% accuracy level (Dimitrov et al. 2016). The server also specifies the field of antigenic prediction in five subsequent target group like- bacteria, virus, tumor, parasite and fungal (Zaharieva et al. 2017). The auto cross covariance (ACC) method has been also incorporated to convert the submitted sequence into uniform length of amino acids chain (Doytchinova and Flower 2007).

The ACC is estimated by the formulas

$$A_{jj}(l) = \sum_i^{n-l} \frac{z_{j,i} \times z_{j,i+1}}{n-l} \tag{1}$$

$$C_{jk}(l) = \sum_i^{n-l} \frac{z_{j,i} \times z_{k,i+1}}{n-l} \tag{2}$$

Where ‘j’ is the z-scale (j= 1, 2, 3), ‘n’ is the amino acids present in sequence (i = 1, 2, …n) and ‘l’ is the lag (l = 1, 2, …L). The Eq. (2) is used when there are two different z-scales. The query sequence has been submitted to the server by selecting bacteria as target organism. In order to obtain more precise result, a threshold value of 1.0 has been set instead of default threshold value (0.4).

Epitope Selection

The common epitopes of B and T-cell having antigenic property are promising candidate for the epitopic vaccine designing. The epitopes predicted by ABCpred, ProPred and ProPred-I server also having Vaxijen predicted antigenic nature are selected for the vaccine development.

3D Conformation of the Epitopes

Three dimensional conformation of the epitope have its significant role leading to vaccine development. It is the most vibrant object for investigation of epitope-antibody docking system (Alam et al. 2016). Due to short amino acid length of the epitope conventional modeling servers are not in use so, DISTILL 2.0(Baú et al. 2006) server have been introduced for the process. The server uses two sets of bidirectional recurrent neural networks technique (Pollastri et al. 2002). For the result output, filtering is applied in two successive stages. In first stage network input has been amplified through prophecies and averaged over several adjacent windows. If $\sigma_j = (\alpha_j, \beta_j, \gamma_j)$ signifies output of j, subsequent to helix, strand and coil prediction reflects the inputs. In second stage network, I_j is the input instead of j.

$$I_j = \left(\sigma_j, \sum_{h=k-p-\omega}^{k-p+\omega} \sigma_h, \dots, \sum_{h=k-p-\omega}^{k_p+\omega} \sigma_h \right)$$

In which $k_f = j + f(2\omega + 1)$, $2\omega + 1$ represents window size ($\omega = 7$) and $2p + 1$ represents window numbers ($p = 7$) taken into account (Pollastri and Mclysaght 2004). PBD file of the experimental epitopic sequence is provided by the server as result element.

Structural Configuration of the Protein

The three dimensional (3D) architecture of a protein plays essential role in protein function and stability (Roy et al. 2012; Willard et al. 2003). So, unrevealing the protein

structural configuration is quite necessary and purposeful for the study. The virulence factor Mce-family protein structure of *N. asteroides* is unavailable in PDB database (Berman et al. 2000). Because of the unavailability of PDB structure, the Phyre2 (Kelley et al. 2015) web server has been introduced to generate the protein structural data. The Phyre2 server applies the Hidden Markov model alignment to detect structure of target protein through HHsearch, open-source software package (Kumar and Jena 2014; Nema and Pal 2013). This server also uses Poing folding simulation to figure out the non allied part of the protein sequence (Nema and Pal 2013). Amino acid sequence of virulence factor Mce-family protein has been submitted to the server and a zip file containing predictions is comes out as systematic result component.

Successive Justification of Protein Model

Justification of a predicted protein model is very much fundamental aspect for establishing the specified protein 3D configuration (Laskowski et al. 2006; Rodriguez et al. 1998). To properly justify the model quality, two web portals are methodologically assigned. The PROCHECK web server (Laskowski et al. 1993), to analyze stereochemical properties of the protein model emphasizing on the torsion angle of C_{α} atoms of amino acids. A PDB file of the targeted protein model has to provide to the online web server. The server offered a Ramachandran plot of all available amino acids, that also vital for the validation of targeted protein model (Laskowski et al. 1993). The ProSA-web is introduce to calculate the Z score of the protein model and furnishes a plot of the protein model within all known protein structure (Wiederstein and Sippl 2007). This server also furnishes energy plot for perfect assessment of the model quality. While the negative energy value of the amino acid residues indicates a good model quality of examined protein (Belkina et al. 2001; Wiederstein and Sippl 2007).

Transmembrane Helix Prediction

Transmembrane helix prediction of the protein model was performed through TMHMM (ver. 2.0) server, the stand-alone software package (Sonnhammer et al. 1998). The server predicts transmembrane helices based on a Hidden Markov model method with 97–98% accuracy of result (Krogh et al. 2001). FASTA sequence of the protein has been submitted to the server and the server provides lists of transmembrane prediction along with an apparent graphical interpretation.

Molecular Docking Analysis

Molecular docking between epitope and antibody component is the critical for proper functioning of vaccine candidate. In order to carry out molecular docking the PatchDock server have been appointed in present research (Duhovny et al. 2002). The geometric contour complementarity method has been applied for docking the peptide and protein sequences (Schneidman-Duhovny et al. 2003). Hence, the PDB file or PDB code of the protein and epitope has to be submitted to the server for molecular docking analysis. The accurate docked complexes ranked according to geometric shape complimentary score and presented along with ACE value, interface area and PDB data (Schneidman-Duhovny et al. 2005). In molecular docking the aggregated desolvation energy of atom pairs termed as ACE. Atom set inside S_1 and S_2 with threshold distance d , the ACE is:

$$E_{ACE} = \sum_{s \in S_1, t \in S_2, \|s-t\| \leq d} T[s, t]$$

Here $\|s-t\|$ signifies Euclidean distance amid s and t , $T[s, t]$ signifies the prearranged score of s and t atom pair.

$T[s, t]$ is estimated using the subsequent formula:

$$T[s, t] = -\ln \frac{N_{s,t}/C_{s,t}}{\left(\frac{N_{s,0}}{C_{s,0}}\right) \times \left(\frac{N_{t,0}}{C_{t,0}}\right)}$$

Here 0 signifies the solvent. ($N_{s,t}$) is number of s,t connection and the number of $s-0$ connection ($N_{t,0}$) are suitable connection numbers of recognized complexes. Moreover, $C_{s,t}$ and $C_{s,0}$ are signifies as the possible numbers of s,t connection and $s-0$ connection (Guo et al. 2012).

Results

Retrieving of Sequence

The amino acid sequence of the protein virulence factor Mce-family protein has been retrieved from the NCBI protein database. The virulence factor Mce-family protein consisting of 493 amino acids and the sequence was downloaded in FASTA format (GenBank: SFL64340.1) (Benson et al. 2012). Afterwards, this sequence processed through several web servers to find out the potential epitopes for promising vaccine development.

B-Cell Epitope Prediction

ABCpred server predicts 20 linear epitopes (Table 1) within the virulence factor Mce-family protein of 20 amino

acids length (window length). The sensitivity, specificity and accuracy are 57.14%, 71.57% and 64.26% correspondingly at window length of 20 (Saha and Raghava 2006). A score is also assigned against predicted epitopes and are ranked accordingly (Han et al. 2015). Predicted sequence with the higher score secures better chance to be an potent

epitope (Jones and Carter 2014). The predicted epitopes are superior because of high threshold score of 0.75.

T-Cell Epitope Prediction Within the B-Cell Epitope

The identified B-cell epitope were further investigated for detection of prospective T-cell epitope through ProPred and ProPred-I to improve immunogenicity (Oprea and Antohe 2013). The B-cell epitopes those are recognized by both MHC molecules are taken under additional consideration. The epitopes common to both MHC classes and B-cell are listed in the Table 2 which may be recognized for ideal vaccine candidate. Following MHC alleles are also enlisted in the supplementary table. The ProPred server predicts epitopes of only nine amino acids moreover, the common epitopes comprised of 9mer sequence.

Table 1 Predicted B-cell epitopes of virulence factor Mce-family protein by ABCpred server

Serial no.	Sequence position in protein	Epitopic sequence	Score
1	377–396	AIPGLTTPGIAPMQTANGTE	0.94
2	329–338	PRYGDLLGSPCFAPSEVSL	0.93
3	144–163	VRDILNEVDPVQLGRVLGTL	0.89
4	273–292	LGALAADPGSITKTFDVFPG	0.85
5	419–438	GLGGGAPAPAAGPAAAPAD	0.83
6	358–377	KALASAAGLPPLVAIPGMPA	0.83
7	215–234	ELLDVLGSSVQTANTIATKR	0.83
8	394–413	GTEVKPFAGTPLEGLVPIIQ	0.82
9	52–71	DVKFRGVLVGVVDVEIVDK	0.81
10	178–197	IQRDLHWLTTVADAGPNLGG	0.81
11	112–131	FNGASNQYLHEGSQIPEDTS	0.81
12	68–87	IVDKGTVQKVNIELKPEFAG	0.80
13	461–480	QLLGRKPTASEFLLLGPIMK	0.79
14	94–113	TARVVPSNLFVAVTSVELVFN	0.77
15	76–95	KVNIELKPEFAGDIPANVTA	0.77
16	59–78	LVGVVDVEIVDKGTVQKVN	0.76
17	408–427	LVPPIQLPGLPGLGGGAPA	0.75
18	387–406	APMQTANGTEVKPFAGTPLE	0.75
19	16–35	ALILGFLMARYQGYFVPKVN	0.75
20	133–152	GTVALQDTLTTVRDILNEVD	0.75

Analysis of Antigenic Property of Epitopes

Manifestation of antigenicity is the prime criterion of a novel epitope in immunobiological aspect (Chen et al. 2007). The common B and T-cell epitopes are successively validated against antigenic propensity via Vaxijen server; Table 3 presents the epitopes along with relevant antigenic score. The 9mer epitopes availing antigenic score beyond the threshold value 1.0 secures antigenic characteristics. In that consequence 10 out of 13 9meric epitopes proved futile to express antigenicity. Three 9mers VLGSSVQTA, VNIELKPEF and VVPSNLFVAV having antigenic score 1.1110, 2.4569 and 1.0810 respectively proved to secure antigenic propensity.

Table 2 Common MHC epitopes within the B-cell epitope

Serial no.	Common epitopic sequence of ABCpred, ProPred and ProPred-I	Number of ProPred predicted MHC-II allele recognize the epitope	Number of ProPred-I predicted MHC-I allele recognize the epitope	Sequence start point
1	IAPMQTANG	9	5	386
2	VQLGRVLGT	14	9	154
3	LVAIPGMPA	25	4	369
4	VLGSSVQTA	13	8	219
5	VKPFAGTPL	7	8	397
6	FRGVLVGVV	21	11	55
7	FNGASNQYL	4	7	112
8	VNIELKPEF	6	7	77
9	FLLLGPIMK	27	10	472
10	VVPSNLFVAV	14	14	97
11	VEIVDKGTV	11	11	65
12	YQGYFVPKV	14	11	26
13	VALQDTLTT	7	10	135

Table 3 Vaxijen predicted antigenic and non antigenic common B and T-cell epitopes

Serial no.	Sequence	Antigenic prediction	Score
1	IAPMQTANG	Non antigen	0.7176
2	VQLGRVLGT	Non antigen	-0.1922
3	LVAIPGMPA	Non antigen	0.3552
4	VLGSSVQTA	Antigen	1.1110
5	VKPFAGTPL	Non antigen	0.1380
6	FRGVLVGVV	Non antigen	-0.4871
7	FNGASNQYL	Non antigen	0.7069
8	VNIELKPEF	Antigen	2.4569
9	FLLLGPIMK	Non antigen	0.6999
10	VVPSNLFAV	Antigen	1.0810
11	VEIVDKGTV	Non antigen	-0.3870
12	YQGYFVPKV	Non antigen	-0.3972
13	VALQDTLTT	Non antigen	0.6264

Threshold score = 1.0

Values in bold are beyond the threshold score = 1.0, these are Antigenic

Epitope Selection

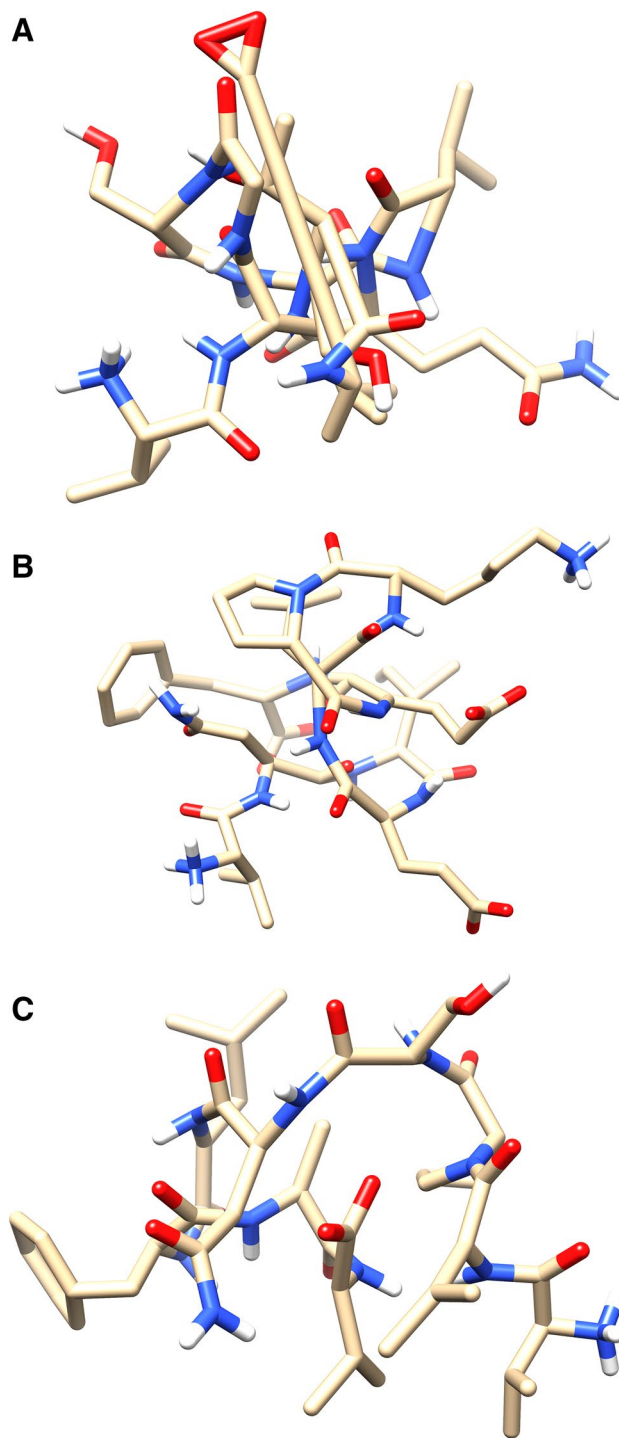
The three 9mer B and T-cell epitopes VLGSSVQTA, VNIELKPEF and VVPSNLFAV having antigenicity (Table 4) will be easily accessible to the immune system (Comerford et al. 1991). These epitopes can be utilized for the future prospective of epitopic vaccine development. The epitopes will elicit strong immune response when administrated in the body as potent vaccine element.

3D Conformation of the Epitopes

The selected epitopes VLGSSVQTA, VNIELKPEF and VVPSNLFAV are submitted to DISTILL server (2.0). The DISTILL server contributes five models of each of the epitope in PDB format. Only the top ranked model of epitopes is taken under consideration. UCSF Chimera (ver. 1.13.1) program has been implemented for visualization of the selected PDB files of the targeted epitopes (Pettersen et al. 2004). Subsequently the generated images are retrieved and presented in the Fig. 1a, b, c in order of VLGSSVQTA, VNIELKPEF and VVPSNLFAV protein.

Table 4 Selected common B and T-cell antigenic epitope

9mer epitopes	Antigenic score
VLGSSVQTA	1.1110
VNIELKPEF	2.4569
VVPSNLFAV	1.0810

**Fig. 1** a 3D structure of VLGSSVQTA epitope. b 3D structure of VNIELKPEF epitope. c 3D structure of VVPSNLFAV epitope

Structural Configuration of the Protein

The virulence factor Mce-family protein lacking PDB entry, for that reason the structure of protein has been depicted by homology modeling. Modeling of virulence factor

Mce-family protein has been performed via Hidden Markov Model in Phyre2 server. Amongst 493 amino acids of virulence factor Mce-family protein 41%, 11%, and 4% amino acid construct α -helix, β -strand and transmembrane helix respectively. Phyre2 predicted PDB data of the protein processed through UCSF Chimera program in order to generate structural image. Figure 2 showing the three dimensional conformation of virulence factor Mce-family protein in *N. asteroides*.

Rational Justification of Protein Model

The topology of virulence factor Mce-family protein is justified using ProSA and PROCHECK web servers. According to the prediction of PROCHECK generated Ramachandran plot (Fig. 3), 79.3% of residues exist in most favoured region and only 1.7% exists in disallowed region (Table 5). Existence of only two non-glycine and non-proline residues in disallowed region established the model quality. ProSA estimated Z score (-2.38) of the model resides within the range of experimentally proved protein structures (Fig. 4) (Sharma and Jaiswal 2009). As presented in Fig. 5, most of the residues of the model possess negative energy value only few have positive value then the model is justified (Bodade et al. 2010).

Transmembrane Helix Prediction

TMHMM server (ver. 2.0) predicts the existence of one transmembrane helix in virulence factor Mce family protein. The transmembrane helix comprises with 20 amino acids of virulence factor Mce family protein. Four of 493 amino acids present inside the membrane while 469 secure outer membrane location and rest of the other 20 form transmembrane helix. The graphical output of transmembrane helix prediction is interpreted in Fig. 6. Transmembrane nature of a protein is very crucial to express antigenicity and

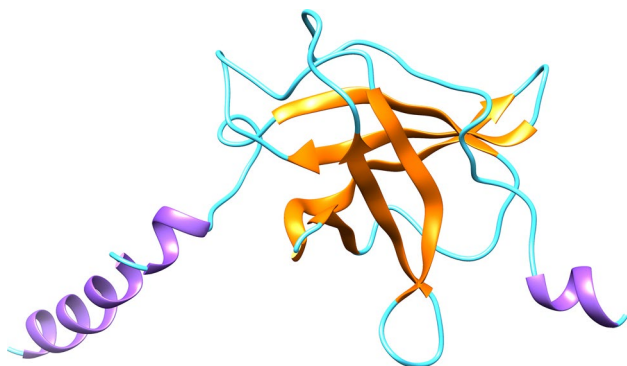


Fig. 2 Tertiary structure of the protein virulence factor Mce family protein

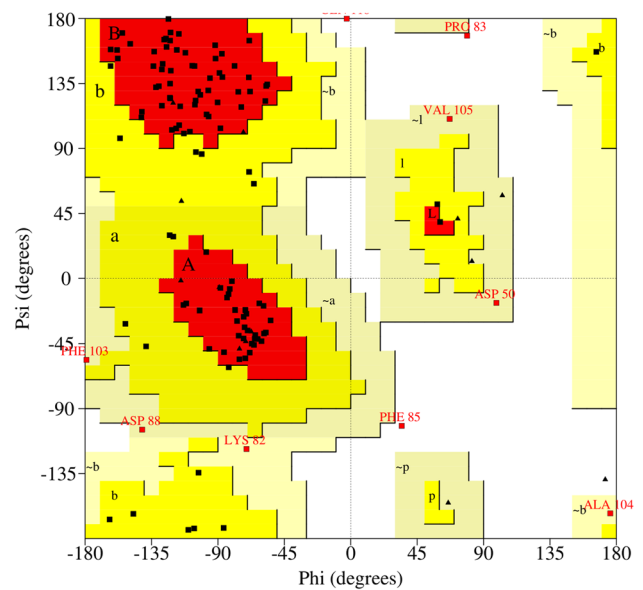


Fig. 3 All atom presentation of virulence factor Mce family protein model in Ramachandran plot by PROCHECK server

accessible to the immune system (Sánchez-Martínez et al. 2006; Xu et al. 1991).

Molecular Docking Analysis

PDB file of VLGSSVQTA, VNIELKPEF and VVPSNL-FAV epitopes and variable domain of T cell receptor delta chain (PDB id:1TV D) are submitted to PatchDock algorithm based server (Li et al. 1998). The server provides 20 docking complex for each epitope ranking them against geometric shape complimentary score. Only the top ranked complex of each epitopes is picked for computational analysis. Epitope along with score, area of interface and ACE value are listed in Table 6. The significantly low ACE value of docking complexes indicated elevated reactivity between epitope and

Table 5 Plot statistics of all atoms of virulence factor Mce family protein

Type of amino acids			Total no
Non Glycine and non Proline residues			
Amino acids position	Number	Percentage (%)	
Favoured region	92	79.3	116
Additional allowed region	16	13.8	
Generously allowed region	6	5.2	
Disallowed region	2	1.7	
End-residues			2
Glycine residues			13
Proline residues			6
Total residues			137

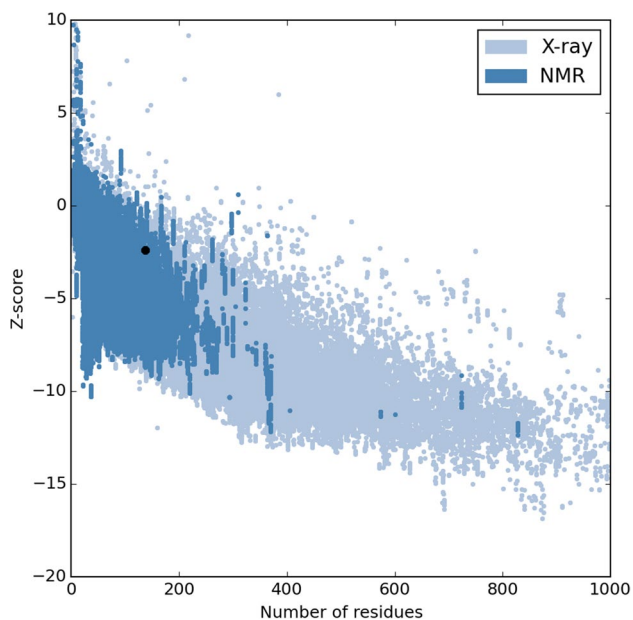


Fig. 4 Predicted model (black dot) of virulence factor Mce family protein within the Z-score range of ProSA plot

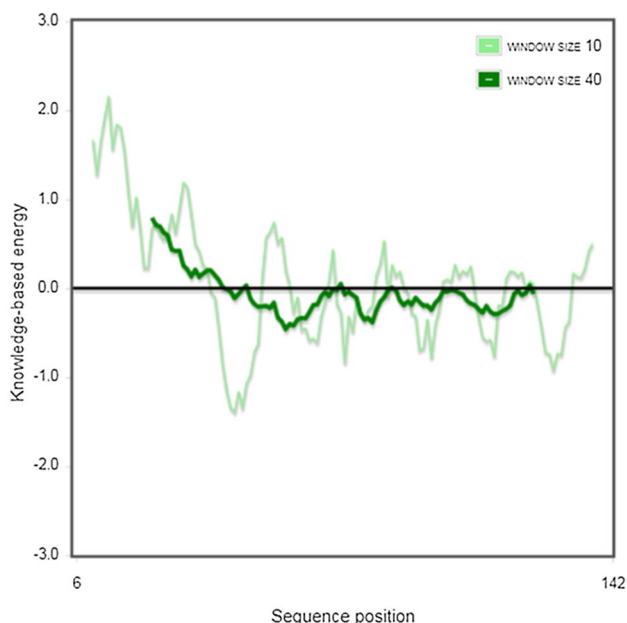


Fig. 5 Energy plot of amino acids of virulence factor Mce family protein by ProSA prediction server

receptor molecules (Ramanathan et al. 2009). The PDB file of docking complex are viewed aiding UCSF Chimera and PyMOL (v1.7.4) graphic system (DeLano 2002). The docking complexes are presented in Figs. 7a, b, 8a, b and 9a, b along with molecular surface interaction. Docking of protein and peptide performs the essential role maintaining cellular activities and linked regulation (Guo et al. 2012; Lavi et al.

2013). The docking of selected epitopes and T cell receptor confirms that the epitopes will accessible to immune system and generate specific immunogenicity.

Discussion

Rational identification, authentication and *in silico* analysis of epitopic components facilitates the successful generation of novel vaccines. As noted the vaccine element performs a key role not only recovering from the infections but also controlling the future disease outbreak. Hence, the nocardiosis has carried a dreadful influence over the immunocompromised patient as they manifest high level of disease susceptibility. Particularly the organ transplant patients encounter the greater chances of nocardiosis. The casualty of organ recipient patients for nocardia infection often reflects in higher degree (Husain et al. 2002). Consequently, the causing pathogen also highly resistance to several well known, market available antibiotics (Husain et al. 2002). Greater survivability, infections and antibiotic resistance nature of the bacterial pathogen is one of the greatest apprehensions for medical biotechnologist. Present advanced, computational analysis assisted research emphasizes much on the discovery of effective vaccine against to *N. asteroides*. The virulence factor Mce family protein of *N. asteroides* is responsible for the pathogenesis of the bacteria to other organism. Hence this protein served as the supreme component for designing epitopic vaccine against nocardiosis. The virulence factor Mce family protein is processed through several bioinformatic tools to execute the suitable epitopes having antigenicity. The transmembrane localization of the protein permits it to intermingle with exact immune system. The transmembrane helix prediction server TMHMM server (ver. 2.0) attests the exomembrane localization of the protein component. Both the B and T-cell epitopes are considerate for obtaining the maximum immune response through humoral and cell mediated immunity. The epitopes were identified through sequence based prediction method using the several web prediction servers. After retrieving the protein sequence of virulence factor Mce family protein (GenBank: SFL64340.1) from NCBI database, it is processed through the ABCpred server for B-cell epitope motif. The 20 numbers of B-cell epitopes having considering threshold value are again submitted to ProPred and ProPred-I servers simultaneously for MHC-II and MHC-I binding allele (Barh et al. 2010). The common epitopes are shortened to only 9mers because of ProPred prediction module provides only 9meric epitopes. The epitopes VLGSSVQTA, VNIELKPEF and VVPSNLFVAV were selected for vaccine designing after accurate validation against antigenic property through Vaxijen server. These are highly antigenic (VLGSSVQTA = 1.1110, VNIELKPEF = 2.4569

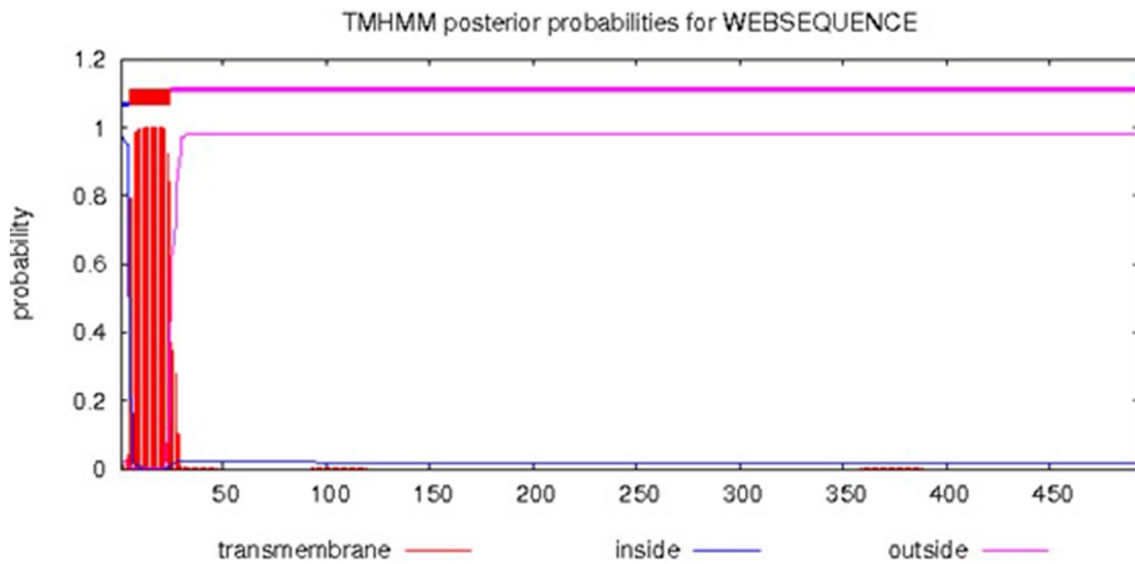


Fig. 6 Predicted transmembrane helix localization by TMHMM server

Table 6 Representation of geometric shape complimentary score, area of interaction and ACE value of the epitope receptor docking complexes

Epitope	Target receptor	Score	Area	ACE value
VLGSSVQTA	1TVD	5574	616.10	-134.28
VNIELKPEF	1TVD	6508	760.20	-151.62
VVPSNLFAV	1TVD	6508	736.30	-203.37

and VVPSNLFAV = 1.0810) being laid over the customized threshold antigenic score. This research design revealed that, three epitopes are the finest vaccine components as identified by B-cell and MHC molecules. The motif of the particular epitopes was mapped out through DISTILL 2.0 for conformational uniqueness and molecular docking. Intend to proper binding of epitopes with the T cell receptor delta chain reflects through the significantly lower ACE values. The structural profile of virulence factor Mce family protein also been mapped out via Phyre2 server that will be fundamental for therapeutics to design the vaccine. The epitopes VLGSSVQTA, VNIELKPEF and VVPSNLFAV will be much more effectual for perfect designing of epitopic vaccine against *N. asteroides* limiting nocardiosis and subsequent casualties linked with it.

Conclusions

Specialized immunoinformatic studies focuses the virulence factor Mce-family protein of *N. asteroides* established its significances lead to expression of bacterial

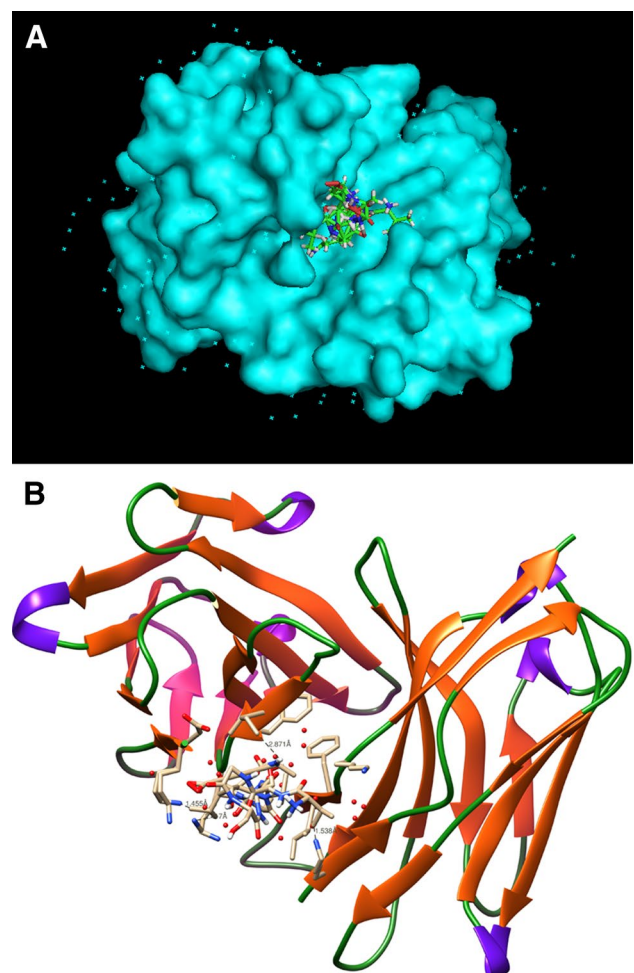


Fig. 7 Docking complex with surface interaction (a) and bond length representation (b) of VLGSSVQTA epitope with 1TVD

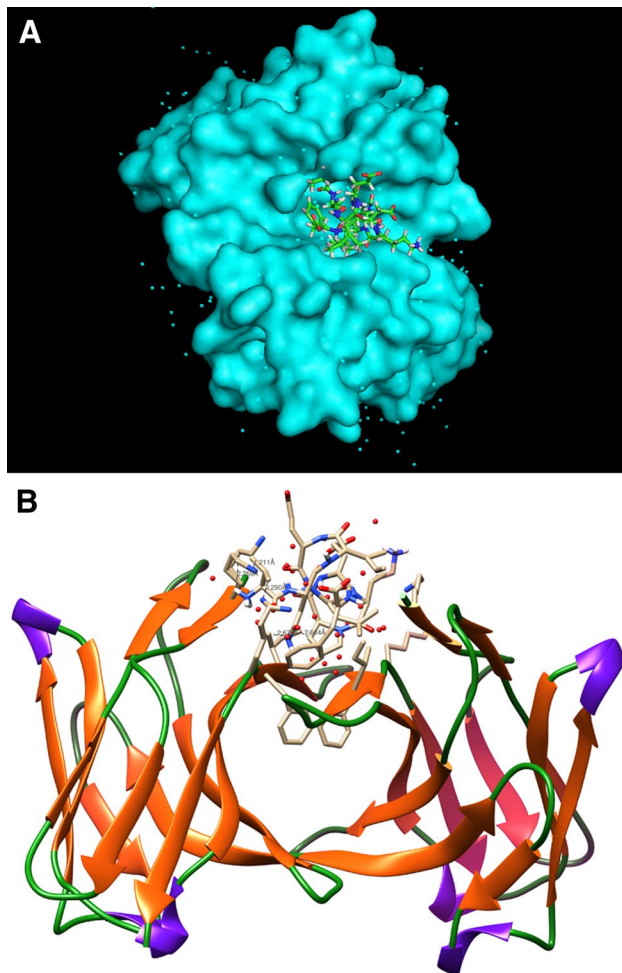


Fig. 8 Docking complex with surface interaction (a) and bond length representation (b) of VNIELKPEF epitope with 1TVD

pathogenicity. Existing research will be surely valuable in modern therapeutics purposes, to resist the nocardiosis outbreak. Particular manifestation of antigenicity by the common B and T-cell epitopes (VLGSSVQTA, VNIELKPEF and VVPSNLFAV) substantiates the critical aptitude to generate humoral and cell mediated immunity. Consequently, the targeted epitopes assist for easy interaction with the immune receptors favor to transmembrane localization of protein element. Literal structural signature of considerable protein along with its epitopes served as decisive factor for novel vaccine development. However, the epitopes requires substantial *in vivo* and *in vitro* justification for accurate refinement to generate finest vaccine component restricting the nocardia infection. Such computer aided research techniques are also highly influential and efficient for designing of desired epitopic vaccine against several associated diseases in light of immunoinformatics.

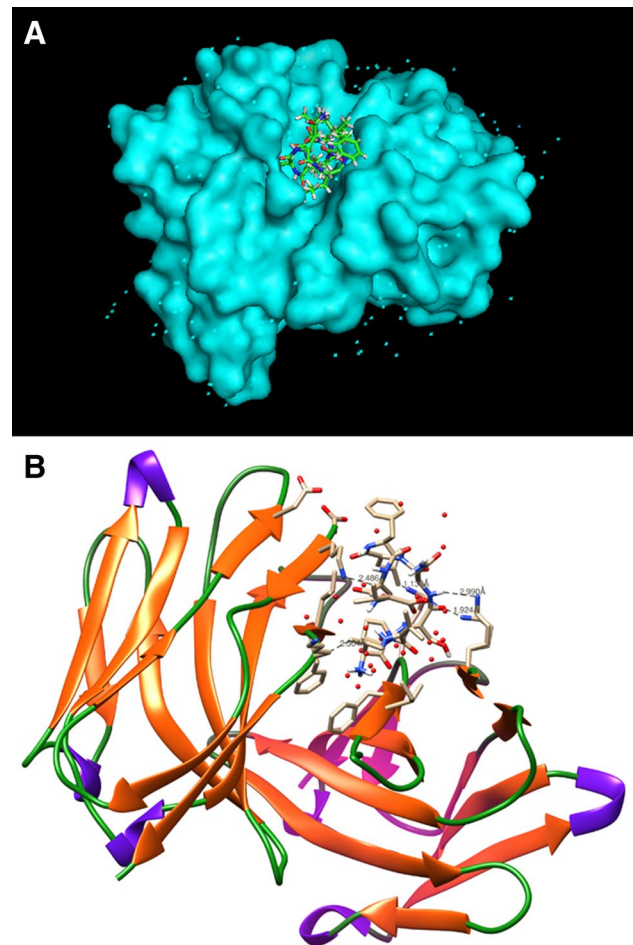


Fig. 9 Docking complex with surface interaction (a) and bond length representation (b) of VVPSNLFAV epitope with 1TVD

Acknowledgements This research work is supported by research project of Science for Equity Empowerment and Development (SEED), Department of Science and Technology (DST), Govt. of India (No. SEED/WTP/059/2014/G).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Research Involving Human and Animal Rights This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R (2016) From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* 149:386–399

- Barh D, Misra AN, Kumar A, Vasco A (2010) A novel strategy of epitope design in *Neisseria gonorrhoeae*. *Bioinformation* 5:77
- Baú D, Martin AJ, Mooney C, Vullo A, Walsh I, Pollastri G (2006) Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. *BMC Bioinform* 7:402
- Belkina N, Lisurek M, Ivanov A, Bernhardt R (2001) Modelling of three-dimensional structures of cytochromes P450 11B1 and 11B2. *J Inorg Biochem* 87:197–207
- Bennett JE, Dolin R, Blaser MJ (2014) Mandell, Douglas, and Bennett's principles and practice of infectious diseases: 2-volume Set vol 1. Elsevier, Amsterdam
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2012) GenBank. *Nucleic Acids Res* 41:D36–D42
- Berman HM et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
- Bhattacharya M, Malick RC, Mondal N, Patra P, Pal BB, Patra BC, Das BK (2019) Computational characterization of epitopic region within the outer membrane protein candidate in *Flavobacterium columnare* for vaccine development. *J Biomol Struct Dyn* 20:1–16
- Bodade R, Beedkar S, Manwar A, Khobragade CN (2010) Homology modeling and docking study of xanthine oxidase of *Arthrobacter* sp. XL26. *Int J Biol Macromol* 47:298–303
- Chen J, Liu H, Yang J, Chou K-C (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33:423–428
- Comerford SA, McCance DJ, Dougan G, Tite JP (1991) Identification of T- and B-cell epitopes of the E7 protein of human papillomavirus type 16. *J Virol* 65:4681–4690
- Coordinators NR (2017) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 45:D12
- DeLano WL (2002) Pymol: an open-source molecular graphics tool. *CCP4 Newslett Protein Crystallogr* 40:82–92
- Dimitrov I, Atanasova M, Patronov A, Flower DR, Doytchinova I (2016) A cohesive and integrated platform for immunogenicity prediction. *Vaccine design*. Springer, New York, pp 761–770
- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform* 8:4
- Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient unbound docking of rigid molecules. *International workshop on algorithms in bioinformatics*. Springer, New York, pp 185–200
- Ellis TN, Beaman BL (2002) Murine polymorphonuclear neutrophils produce interferon- γ in response to pulmonary infection with *Nocardia asteroides*. *J Leukoc Biol* 72:373–381
- Fleetwood IG, Embil JM, Ross IB (2000) *Nocardia asteroides* cerebral abscess in immunocompetent hosts: report of three cases and review of surgical recommendations. *Surg Neurol* 53:605–610
- Gourlay LJ et al (2013) Exploiting the *Burkholderia pseudomallei* acute phase antigen BPSL2765 for structure-based epitope discovery/design in structural vaccinology. *Chem Biol* 20:1147–1156
- Guo F, Li SC, Wang L, Zhu D (2012) Protein-protein binding site identification by enumerating the configurations. *BMC Bioinform* 13:158
- Han J-H et al (2015) Identification of immunodominant B-cell epitope regions of reticulocyte binding proteins in *Plasmodium vivax* by protein microarray based immunoscreening. *Korean J Parasitol* 53:403
- Husain S, McCurry K, Dauber J, Singh N, Kusne S (2002) Nocardia infection in lung transplant recipients. *J Heart Lung Transpl* 21:354–359
- Jamal F et al (2017) Identification of B-cell Epitope of *Leishmania donovani* and its application in diagnosis of visceral leishmaniasis. *J Biomol Struct Dyn* 35:3569–3580
- Jones MS, Carter JM (2014) Prediction of B-cell epitopes in listeriolysin O, a cholesterol dependent cytolysin secreted by *listeria monocytogenes*. *Adv Bioinform* 1:1. <https://doi.org/10.1155/2014/871676>
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845
- Khan M, Hossain M, Rakib-Uz-Zaman S, Morshed M (2015) Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *Scand J Immunol* 82:25–34
- Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580
- Kumar S, Jena L (2014) Understanding rifampicin resistance in tuberculosis through a computational approach. *Genomics Inf* 12:276
- Lafuente EM, Reche PA (2009) Prediction of MHC-peptide binding: a systematic and comprehensive overview. *Curr Pharm Des* 15:3209–3220
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291
- Laskowski R, MacArthur M, Thornton J (2006) PROCHECK: validation of protein-structure coordinates. Kluwer Academic Publishers, Dordrecht
- Lavi A et al (2013) Detection of peptide-binding sites on protein surfaces: the first step toward the modeling and targeting of peptide-mediated interactions. *Proteins* 81:2096–2105
- Li H, Lebedeva MI, Llera AS, Fields BA, Brenner MB, Mariuzza RA (1998) Structure of the V δ domain of a human $\gamma\delta$ T-cell antigen receptor. *Nature* 391:502
- Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusica V (2008) Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunology* 9:8
- Moylett EH et al (2003) Clinical experience with linezolid for the treatment of Nocardia infection. *Clin Infect Dis* 36:313–318
- Mustafa AS, Shaban FA (2006) ProPred analysis and experimental evaluation of promiscuous T-cell epitopes of three major secreted antigens of *Mycobacterium tuberculosis*. *Tuberculosis* 86:115–124
- Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, Ali A (2015) Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach. *Infection, Genetics and Evolution* 32:280–291
- Nema V, Pal SK (2013) Exploration of freely available web-interfaces for comparative homology modelling of microbial proteins. *Bioinformation* 9:796
- Nichol KL, Nordin J, Mullooly J, Lask R, Fillbrandt K, Iwane M (2003) Influenza vaccination and reduction in hospitalizations for cardiac disease and stroke among the elderly. *N Engl J Med* 348:1322–1332
- Oprea M, Antohe F (2013) Reverse-vaccinology strategy for designing T-cell epitope candidates for *Staphylococcus aureus* endocarditis vaccine. *Biologicals* 41:148–153
- Palmieri JR, Santo A, Johnson SE (2014) Soil-acquired cutaneous nocardiosis on the forearm of a healthy male contracted in a swamp in rural eastern Virginia. *International medical case reports journal* 7:41
- Patil SP, Nadkarni NJ, Sharma NR (2012) Nocardiosis: clinical and pathological aspects. In: *Histopathology-reviews and recent advances*. IntechOpen, London
- Petersen 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* 25:1605–1612

- Pollastri G, Mclysaght A (2004) Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21:1719–1720
- Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* 47:228–235
- Quinones-Hinojosa A (2012) Schmidek and sweet: operative neurosurgical techniques 2-volume set: indications, methods and results (Expert Consult-Online and Print), vol 2. Elsevier, Amsterdam
- Ramanathan K, Shanthi V, Sethumadhavan R (2009) In silico identification of catalytic residues in azobenzene reductase from *Bacillus subtilis* and its docking studies with azo dyes. *Interdisciplinary Sciences* 1:290–297
- Rodriguez R, Chinae G, Lopez N, Pons T, Vriend G (1998) Homology modeling, model and software evaluation: three related resources. *Bioinformatics* (Oxford, England) 14:523–528
- Roy A, Yang J, Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 40:W471–W477
- Saha S, Raghava G (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65:40–48
- Sánchez-Martínez S, Lorizate M, Katinger H, Kunert R, Nieva JL (2006) Membrane association and epitope recognition by HIV-1 neutralizing anti-gp41 2F5 and 4E10 antibodies. *AIDS Res Hum Retroviruses* 22:998–1006
- Saubolle MA, Sussland D (2003) Nocardiosis: review of clinical and laboratory experience. *J Clin Microbiol* 41:4497–4501
- Schneidman-Duhovny D et al (2003) Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 52:107–112
- Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33:W363–W367
- Sethy H, Meher B, Pradhan S, Patro M (2016) Pleuropulmonary Nocardiosis in an Immunocompetent Host. *Scholars Journal of Applied Medical Sciences* 2(4):79–83
- Sharma B, Jaiswal MK (2009) EGF domain II of protein Pb28 from *Plasmodium berghei* interacts with monoclonal transmission blocking antibody 13.1. *Journal of Molecular Modeling* 15:369
- Singh H, Raghava G (2001) ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 17:1236–1237
- Singh H, Raghava G (2003) ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 19:1009–1014
- Sonnhammer EL, Von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182
- Srivastava S et al (2019) Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. *J Biomol Struct Dyn* 20:1–16
- Vivona S, Gardy JL, Ramachandran S, Brinkman FS, Raghava G, Flower DR, Filippini F (2008) Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends Biotechnol* 26:190–200
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:W407–W410
- Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* 31:3316–3319
- Wilson JW (2012) Nocardiosis: updates and clinical overview. *Mayo Clinic Proceedings*. Elsevier, Amsterdam, pp 403–407
- Xu J, Gorny M, Palker T, Karwowska S, Zolla-Pazner S (1991) Epitope mapping of two immunodominant domains of gp41, the transmembrane protein of human immunodeficiency virus type 1, using ten human monoclonal antibodies. *J Virol* 65:4832–4838
- Zaharieva N, Dimitrov I, Flower D, Doytchinova I (2017) Immunogenicity prediction by VaxiJen: a ten year overview. *J Proteomics Bioinformatics* 10:298–310

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