International Journal of Advanced Biotechnology and Research (IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol-7, Issue-3, 2016, pp1103-1110
http://www.bipublication.com

Research Article

In silico analysis of clumping factors ClfA and ClfB of multidrug-resistant *Staphylococcus aureus* MRSA 252 strain

Shilpa M. Deodurg[1]., Anil Kumar S. Katti[2] and Sulochana M. B[1]*

[1]Dept. of PG Studies and Research in Biotechnology,
Gulbarga University, Kalaburagi-585106, India.
[2]Dept. of Biotechnology, Walchand Centre for Biotechnology, Solapur-413006, India.
*Corresponding author: Email: mbsulochanabt@gmail.com, Mob: 09449618676.

ABSTRACT:
Infectious diseases cause millions of deaths every year and several gram-positive bacteria such as *Staphylococcus aureus*, *Streptococcus pyogenes* and *Mycobacterium tuberculosis* are the main cause of severe illnesses. *S. aureus* is a major cause of infection in both healthcare and community settings. *S. aureus* is equipped with a great variety of virulence factors and carries numerous surface proteins named "microbial surface components recognizing adhesive matrix molecules" (MSCRAMMs) that mediate attachment to host tissues exclusively to the gamma-chain of human fibrinogen which induces formation of bacterial clumps and initiate colonization leading to an infection. In the present investigation, *in silico* analysis of adhesins such as clumping factors A & B (ClfA and ClfB) of MRSA 252 was carried out. Primary protein sequence analysis reveals that Clf A & B proteins were acidic in nature having pI value 3.47 & 3.89 respectively. The secondary structure analysis by SOPMA revealed that Clf A & B have less percent of alpha helix than beta sheets and more number of random coils and these proteins belong to serine-aspartate repeat containing protein (SDr) family. The domain analysis by SMART showed that ClfA contains YSIRK type signal peptide and SdrG_C_C domain and ClfB contains YSIRK type signal peptide, SdrG_C_C and Gram_pos_anchor domains. Antigenicity of ClfA and ClfB were 1.2023 and 1.0931 respectively was predicted by Vaxijen 2.0 server. The homology modeling of ClfA & B built by SWISS-MODEL was validated using PROCHECK that showed 92.9% and 91.0% model quality respectively. The detailed analysis of function and mechanisms of action of each virulence factor could open the way to control infectious diseases by using specific inhibitors and may be helpful for the development of novel therapies for *S. aureus* diseases.

Key words: *In silico*, ClfA and ClfB, clumping factors, multidrug-resistant, *Staphylococcus aureus.*

[1] INTRODUCTION

Staphylococcus aureus, especially Methicillin-resistant *S. aureus*, causes a wide range of opportunistic infections that range from superficial skin infections to deadly diseases such as endocarditis, pneumonia, and septicemia. *S. aureus* infections cause extensive morbidity, mortality, and medical expenses [12]. Antibiotic resistance for *S. aureus* isolates continues to increase and the rapid worldwide spread of community-associated methicillin-resistant *S. aureus* infection is particularly alarming. Identification and analysis of new targets to prevent and treat *S. aureus* infections are required, that could serve as a great prospective for the development of novel therapeutics against this pathogen.

Adherence of bacteria to host matrix components that is mediated by bacterial surface adhesins is the initial critical event in the pathogenesis of most infections. *S. aureus* utilize the extracellular matrix

(ECM) as substrata for their adhesion by way of a family of adhesins called MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) [13] that specifically recognize host matrix components. One important component of the ECM, also occurring in soluble form in blood plasma, is fibrinogen. This protein is recognized by several integrins including the platelet integrin aIIbb3. Activation of platelets and integrin aIIbb3 result in fibrinogen-dependent aggregation *in vitro* and the formation of platelet-fibrin thrombi *in vivo*. *S. aureus* contains several fibrinogen-binding proteins, one of which is clumping factor A (ClfA) and is primarily responsible for the clumping of bacteria in fibrinogen solutions and bacterial adherence to fibrinogen substrata [10].

Staphylococcus aureus can express up to 20 different potential MSCRAMMs that are anchored to peptidoglycan cell wall following sortase mediated cleavage of the C-terminal LPXTG motif [8]. However, only the fibrinogen-binding proteins ClfA and ClfB [10], the fibronectin-binding proteins FnBPA and FnBPB (which also bind to fibrinogen) [20], the collagen binding protein Cna [14] and protein A, which binds to IgG and to von Willebrand factor [5] have been characterized in detail. ClfB is a major determinant in *S. aureus* nasal colonization and binds to the α-chain of fibrinogen [15]. *S. aureus* expresses several different proteins that can bind specifically to fibrinogen (Fg), including clumping factors A and B (ClfA and ClfB) and the bifunctional fibronectin binding proteins A and B, FnBPA and FnBPB [15], bind to the extreme C terminus of the γ-chain protruding from domain D of Fg.

In this report, we propose *in silico* structure and functional analysis of ClfA and ClfB, which could help in the designing of drug and vaccine for the treatment of multidrug-resistant *S. aureus* involved in various diseases.

[2] MATERIALS AND METHODS

2.1. Primary protein sequence analysis

Protein sequences of ClfA (Q6GIK4) and ClfB (Q6GDH2) were retrieved from UniProt database. To compute various physico-chemical properties of protein sequence, ProtParam tool was used. The parameters computed by ProtParam include molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

2.2. Secondary structure prediction

Secondary structure prediction of ClfA and B were carried out using SOPMA tool [4]. Default parameters were selected to increase reliability and accuracy of the query sequences. The secondary structures such as α-helix, β-sheet and coil in the given protein were predicted using SOPMA. Solvent accessibility of secondary structures was prediction using PredictProtein [17].

2.3. Protein function prediction

Protein function prediction (PFP) version 2.0 beta release was used for the prediction of biological function of ClfA and B. PFP algorithm searched conventional databases with relative probability of gene ontologies (GO) to predict the most probable GO annotations in three biological processes (BP), molecular function (MF) and cellular component (CC) categories [7].

2.4. Protein functional sites prediction

InterProScan and Fingerprint scan were the tools used to predict the signatures and the motif regions in the sequence. InterProScan is a tool to automatically annotate proteins with families and domains and functional sites [22].

2.5. Protein motif identification

FingerPRINTScan was used to identify the closest matching PRINTS sequence motif fingerprints in a protein sequence. It identifies a queried protein sequence as a member of a known family, inferring a wealth of known information pertinent to that family and its members [2].

2.6. Predicting transmembrane topology

Transmembrane helices and topological domains were predicted using TMpred [6]. This tool was originally meant for analyzing the properties of transmembrane proteins with both specificity and sensitivity. Finds positional preferences of certain

amino acids and prediction of transmembrane domains.

2.7. Prediction of antigenicity

Proteins were uploaded as a multiple sequence file in FASTA format to VaxiJen v 2.0 [3] server to predict antigenicity. VaxiJen v 2.0 is the server for alignment independent prediction of protective antigens, solely based on the physicochemical properties of proteins without recourse to sequence alignment.

2.8. Identification of domains

The domains and signal peptides of ClfA and ClfB were analyzed using SMART (Simple Modular Architecture Research Tool) [18]. SMART is an online resource for the identification and annotation of protein domains and analysis of protein domain architecture

2.9. Homology modeling

Homology modeling of the antigenic proteins was done using Swiss Model server [1] by choosing template. The model was visualized by RasMol viewer and model quality was checked by PROCHECK.

2.10. Identification of Non-Human Homologue

To identify ClfA and ClfB proteins were human homologue proteins or not, each protein sequence was subjected to human BLAST-P in NCBI server.

[3] RESULTS AND DISCUSSION

3.1. Primary protein sequence analysis

The sequence of ClfA (Q6GIK4) and ClfB (Q6GDH2) contains 1029 and 873 amino acids with molecular weight 106716.3 Da, pI 3.47 and 92995.9 Da, pI 3.89 respectively. The extinction coefficient of ClfA and ClfB were 43320 and 42290 respectively. The estimated half life was 30 h for both the proteins. The instability index and predicted aliphatic index of proteins was 52.73 & 55.24, 43.03 & 44.93 for ClfA and ClfB respectively. The grand average of hydropathicity (GRAVY) of ClfA and ClfB were -1.149 and -1.157 respectively.

Molecular weight provides information not only on full-length of the protein expression, but also on the expression of modified, splice variant, cleavage product and processed proteins. However, any protein modification may lead to a change in the overall protein charge and/or molecular weight.

A protein isoelectric point (pI) forms the basis for its isolation, with this criterion the predicted pI values of ClfA and ClfB were 3.47 and pI 3.89 respectively. Since at this pI, the proteins would be stable and compact and will be of use for developing buffer system for purification by isoelectric focusing method [16].

The computed protein concentration and extinction coefficient could help in the quantitative study of protein-protein and protein-ligand interactions in solution. The bio-computed half life of both the proteins were estimated to be 30 h which indicated stability.

Both the proteins were stable and greater than 30 (II>30) instability index. Perhaps, the higher aliphatic index of the protein with relatively high volume of aliphatic side chains (alanine, valine, isoleucine and leucine), provided the thermostability of the proteins.

The GRAVY score reflected the different hydropathy indices with negative values representing the presence of more hydrophilic residues [21]. Similar tool for the analysis of GlmU protein of *M. tuberculosis* strain CDC1551 was used [11].

3.2 Protein secondary structure prediction by SOPMA

The parameters considered were window size as 17 residues and similarity threshold as 8. SOPMA results confirm that ClfA protein contains less alpha helix (10.01%) than ClfB (10.88%).

ClfB has alanine rich α-helices than ClfA and extended strands were 13.80% & 16.49% for ClfA and ClfB respectively. Very high coil structural content of ClfA (71.91%) than ClfB (67.70%) and the details of secondary structure of ClfA and ClfB were shown in Fig.1a & 1b respectively.

PredictProtein predicts that ClfA and ClfB loop regions are more exposed to the solvent, strand regions buried and helix regions are intermediate as shown in Fig. 2.

Similar tool was used for the prediction of secondary structure of GlmU antigenic protein of *Mycobacterium tuberculosis* CDC1551 [11] and SOPMA was also used to predict the secondary structure Gp41 envelope glycoprotein of human [19].

3.3. Protein function prediction

PFP version 2.0 revealed ClfA & ClfB protein functions such as molecular function, biological processes and cellular components as shown in Table. 1 and Table. 2 respectively.

3.4. Protein functional sites prediction

InterProScan predicted common functional sites in ClfA and ClfB were YSIRK Gram-positive signal peptide domain, adhesion domain, fibrinogen-binding domain 1 and fibrinogen-binding domain 2, LPXTG cell wall anchor domain and Gram-positive anchor as shown in Fig. 3.

3.5. Protein motif identification

FingerPRINTScan results of ClfA and ClfB were as tabulated in Table.3 and Table.4 respectively.

The result shows 10 FingerPRINTScan for ClfA & ClfB with number of motifs as 2 for each but in ClfA having FingerPrint–LVDCCALPHA1C has 3 motifs.

FingerPRINTScan classifies sequences using familial definitions from the PRINTS database, allowing progress to be made with the identification of distant evolutionary relationships. The approach makes use of the contextual information inherent in a multiple-motif method, and has the power to identify hitherto unidentified relationships in mass genome data.

3.6. Predicting transmembrane topology

ClfA and ClfB both having one transmembrane helix each predicted by TMpred server as shown in Fig. 4a and 4b respectively. Subcellular localization is one of the most important characteristics of proteins, which is central to understand their function and the constitution of biological systems.

In bacteria information related to the subcellular location of pathogen proteins can facilitate the development of drugs and vaccines for treatment.

3.7. Prediction of antigenicity

The protective antigen or non-antigen, were predicted according to a predefined cut off. Since more of the models had their highest accuracy at a threshold of 0.5, this threshold value was chosen in our study. VaxiJen server 2.0 revealed that ClfA and ClfB were potent antigenic proteins having antigenic score 1.2023 and 1.0931 respectively as shown in Fig. 5a & 5b.

Proteins those present on outer membrane and antigenic in nature could be useful in potent vaccine and drug designing. We have predicted that both ClfA and ClfB were non-human homologues by NCBI Blastp, proving pathogen specific proteins and localized to the cell membrane or cell wall served as drug and vaccine targets for *S. aureus* MRSA 252.

3.8. Identification of domains

SMART revealed that ClfA and ClfB contain similar domains such as Pfam: YSIRK_signal, Pfam: SdrG_C_C and Pfam: Gram_pos_anchor as shown in Fig. 6. Many surface proteins found in *Streptococcus*, *Staphylococcus*, and related lineages share apparently homologous YSIRK_signal signal sequences. A motif resembling [YF]SIRKxxxGxxS[VIA] appears at the start of the transmembrane domain. The GxxS motif appears perfectly conserved, suggesting a specific function and not just homology. There is a strong correlation between proteins carrying this region at the N-terminus and those carrying the Gram-positive anchor domain with the LPXTG sortase processing site at the C-terminus. The protein domain SdrG_C_C terminal refers to the C terminus domain of an adhesin found only on the cell walls of bacteria. More specifically, SdrG is only found in gram-positive bacteria. This particular domain binds to a glycoprotein named fibrinogen. SdrG stands for serine-aspartate dipeptide repeats, which as its name suggests, contains repeats of two amino acids, serine and aspartate [12].

3.9. Homology modeling

Homology modeling of ClfA and ClfB was carried out by SWISS-MODEL automated protein

structure homology-modeling server using template 1n67.1A and 4F27A having resolution 1.9A^{o} and 1.92A^{o} respectively. The PROCHECK revealed the model quality of 92.3% and 91.0% for ClfA and ClfB as shown in Fig. 7a and 7b respectively.

3.10. Identification of Non-Human Homologue

Both ClfA and ClfB were predicted to be non human homologues by BLASTp tool as these proteins were not having any homology with human proteins.

[4] CONCLUSIONS

In the dearth of experimental data, model-building on the basis of the known three dimensional structure of a homologous protein is the only dependable method to obtain structural information which is based on sequential information. Computational methods play a vital role in accelerating the drug development process, since the present study focused on the *in silico* analysis of clumping factors ClfA and ClfB of *S. aureus* MRSA252 and provided a new insight into the molecular biology and functional properties for rational designing of inhibitors in its battle against *S. aureus* infections.

REFERENCES

1. Arnold, K, Bordoli, L, Kopp, J, and Schwede, T, (2006), The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling, Bioinformatics. 22, 195-201.
2. Attwood, T.K, Croning, M.D, Flower D.R, Lewis, A.P, Mabey, J.E, Scordis, P, Selley, J.N, Wright, W, (2001), PRINTS-S: the database formerly known as PRINTS, Nucleic Acids Research. 28(1), 225-227.
3. Doytchinova, I.A, and Flower, D.R, (2007), VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, BMC Bioinformatics. 5(8) 4.
4. Geourjon, C, and Deleage, G, (1995), SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput Appl Biosci. 11(6), 681-684.
5. Hartlieb, J, Koehler, N, Dickinson, R.S, Chhatwal, S, Sixma, J.J, Hartford, O, et al., (2000), Protein A is the von Willebrand factor binding protein on *Staphylococcus aureus*, Blood. 96, 2149–2156.
6. Hofmann, K, and Stoffel, W, (1993), TMbase - A database of membrane spanning proteins segments, Biol. Chem. Hoppe-Seyler, 374, 166.
7. Lakshmi, P.T.V, and Subhashini, S, (2010), *In silico* sequence specific analysis of ERBB2 RTK alterations responsible for neuroectodermal tumors of Homo sapiens, Journal of Bioinformatics and Sequence Analysis. 2(6), 75-84.
8. Mazmanian, S.K, Liu, G, Ton-That, H, and Schneewind, O, (1999), *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall, Science. 285, 760–763.
9. McCrea, K.W, Hartford, O, Davis, S, Eidhin D.N, Lina, G, Speziale, P, et al., (2000), The serine-aspartate repeat (Sdr) protein family in *Staphylococcus epidermidis*, Microbiology. 146 (7), 1535–46.
10. McDevitt, D, Francois, P, Vaudaux, P, and Foster, T.J, (1994), Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*, Mol. Microbiol. 11, 237–248**.**
11. Mubashshira Pathan, A, Aasma Jamkhandi, S, and Anil Kumar Katti S, (2014), *In silico* characterization of GlmU antigenic mycobacterium tuberculosis protein, International journal of pharmaceutical science and research. 5(4), 1493-1499.
12. Noskin, G.A, Rubin, R.J, Schentag, J.J, et al., (2005), The burden of *Staphylococcus aureus* infections on hospitals in the United States: an analysis of the 2000 and 2001 Nationwide Inpatient Sample database, Arch Intern Med. 165, 1756 – 61.

13. Patti, J.M, Allen, B.L, McGavin, M.J, and Hook, M, (1994), MSCRAMM-mediated adherence of microorganisms to host tissues, Annu. Rev. Microbiol. 48: 585–617.
14. Patti, J.M, Jonsson, H, Guss B, Switalski L.M, Wilberg, K, Lindberg, M, and Hook, M, (1992), Molecular characterization and expression of a gene encoding *Staphylococcus aureus* collegen adhesin, J Biol Chem. 267, 4766–4772.
15. Perkins, S, Walsh, E.J, Deivanayagam, C.C, Narayana, S.V, Foster, T.J, and Hook, M, (2001), Structural Organization of the Fibrinogen-binding Region of the Clumping Factor B MSCRAMM of *Staphylococcus aureus*, J Biol Chem. 276, 44721–44728.
16. Rosenberg, I, and Ian, M, (2005), Protein Analysis And Purification: Benchtop Techniques, Birkhauser Boston book, 2nd ed., 25, 526 pp.
17. Rost, B, Yachdav, G, Liu, J, (2004), The PredictProtein server, Nucleic Acid Res. 32, 321-326.
18. Schultz, J, et al., (2000), SMART: A web-based tool for the study of genetically mobile domains, Nucleic Acids Research. 28 (1), 231-234.
19. Syyada, S.J, Saliha, K, Syed, B.J, Masaud, S, (2012) Structure based sequence analysis & epitope prediction of gp41 HIV1 Envelope glycoprotein isolated in Pakistan, Genet Vaccines Ther. 23(10), 10.
20. Wann, E.R, Gurusiddappa, S, and Hook, M, (2000), The fibronectin-binding MSCRAMM FnBPA of *Staphylococcus aureus* is a bifunctional protein that also binds to fibrinogen, J Biol Chem. 275, 13863–13871.
21. Xiong, X, Huang, S, Zhang, H, Jianjun, L, Jianying, S, Jixian, X, Yong, L, Liping, J, Xianchun, W, and Sonping, L, (2009), Enrichment and proteomic analysis of plasma membrane from rat dorsal root ganglions, Proteome Sci. 7, 41-1477-5956.
22. Zdobnov, E.M, and Apweiler, R, (2001), InterProScan - An integration platform for the signature-recognition methods in InterPro, Bioinformatics. 17(9), 847-848.

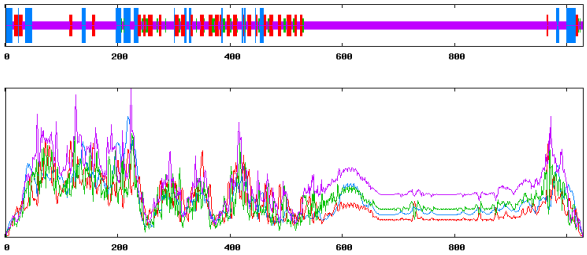

Fig: 1a. Secondary structure of ClfA by SOPMA

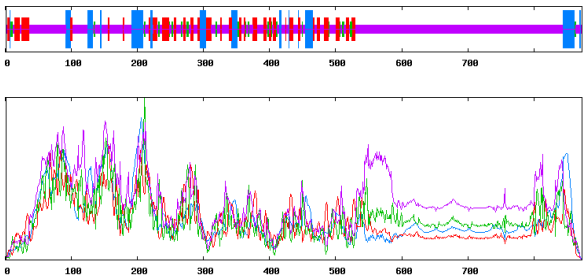

Fig: 1b. Secondary structure of ClfB by SOPMA.

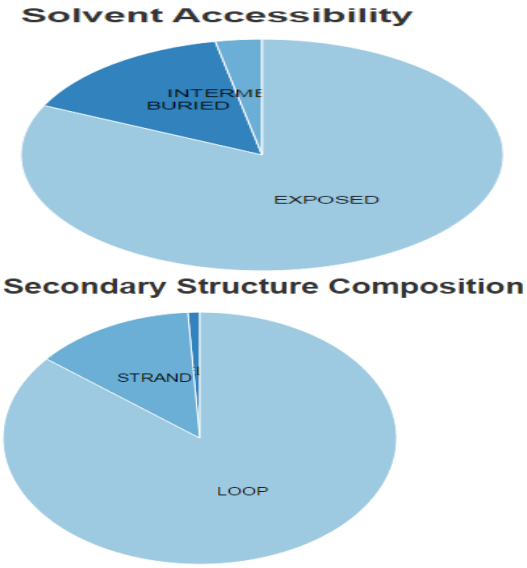

Fig: 2. Prediction of Secondary structure composition and solvent accessibility of ClfA and ClfB by PredictProtein.

Table: 1. Prediction of ClfA function using PFP.

PFP Score	Description
Molecular Function	
777.18	Protein binding
759.12	Catalytic activity
649.85	Cation binding
693.64	Ion binding
599.39	Metal ion binding

Biological processes	
3912.20	Cell adhesion
3220.49	Pathogenesis
Cellular components	
5753.31	Membrane
5394.17	Extracellular region
4315.90	Cell wall

Table: 2. Prediction of ClfB function using PFP.

PFP Score	Description
Molecular Function	
1043.56	Catalytic activity
741.14	Protein binding
648.34	Cation binding
647.22	Ion binding
599.39	Metal ion binding
Biological processes	
3438.67	Cell adhesion
3220.49	Pathogenesis
Cellular components	
5296.40	Membrane
4838.20	Extracellular region
3756.50	Cell wall

Fig: 3. Functional sites predicted in ClfA and ClfB by InterProScan.

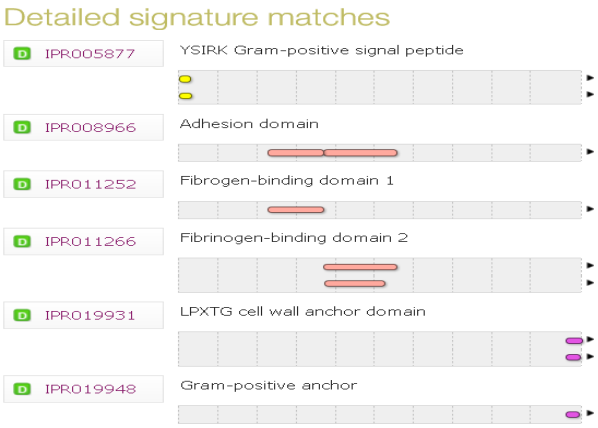

Table: 3. FingerPRINTScan result of ClfA

FingerPrint	No. of Motifs
FLGHOOKAP1	2
LVDCCALPHA1C	3
EDTRNSPORT	2
INTEIN	2
FRAGILYSIN	2
ANTIFREEZEI	2
FIMBRILLIN	2
DENSEGRNULE6	2
ADENOVSFIBRE	2
RVDCCALPHA1	2

Table: 4. FingerPRINTScan result of ClfB

FingerPrint	No. of Motifs
HISTAMINEH3R	2
PRICHEXTENSN	2
KV33CHANNEL	2
BINARYTOXINB	2
VINCULIN	2
TCOMPLEXTCP1	2
5HT6RECEPTR	2
FAM175PLANT	2
PERTACTIN	2
LVDCCALPHA1D	2

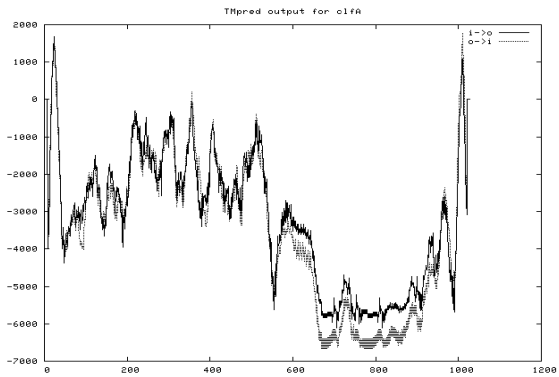

Fig: 4a. Transmembrane topology of ClfA.

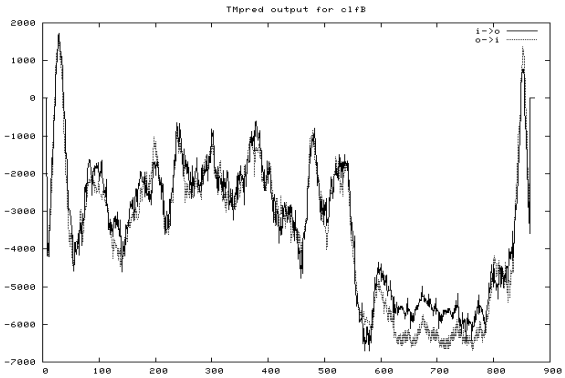

Fig: 4b. Transmembrane topology of ClfB.

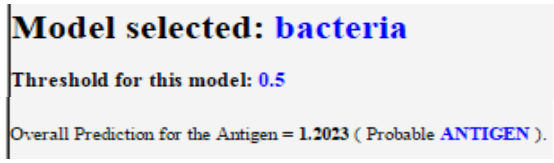

Fig: 5a. Antigenicity of ClfA by VaxiJen 2.0.

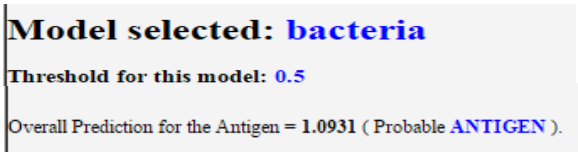

Fig: 5b. Antigenicity of ClfB by VaxiJen 2.0.

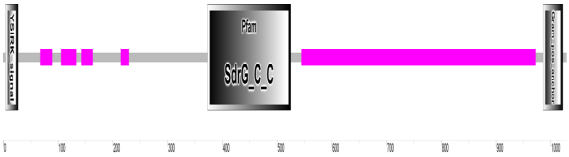

Fig: 6. Domains of ClfA and ClfB by SMART.

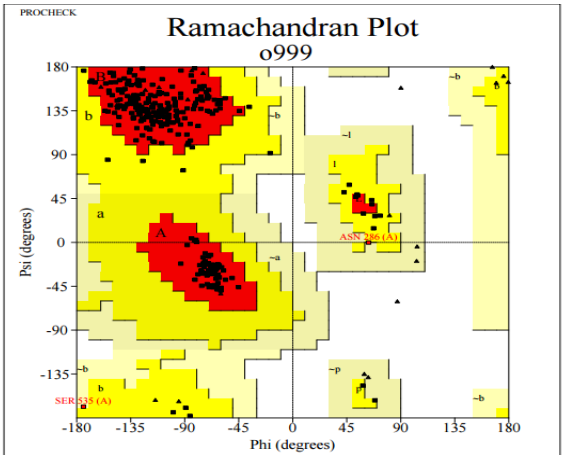

Fig: 7a. Ramachandran plot of ClfA

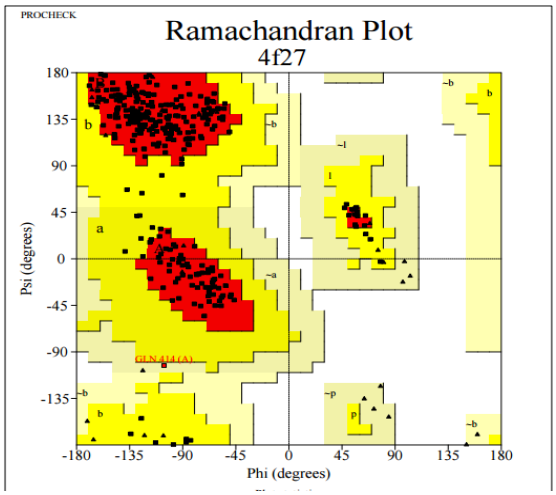

Fig: 7b. Ramachandran plot of ClfB

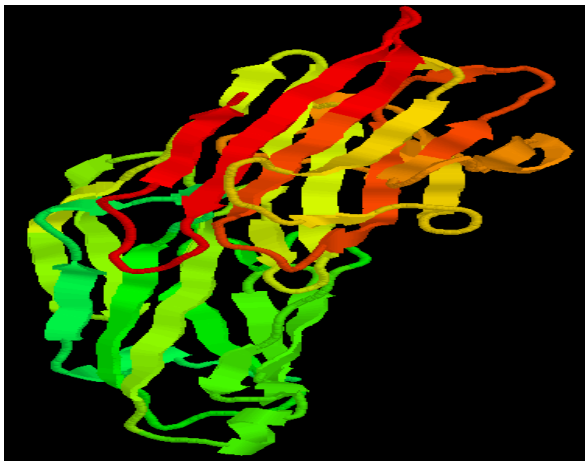

Fig: 8a. Structure of ClfA

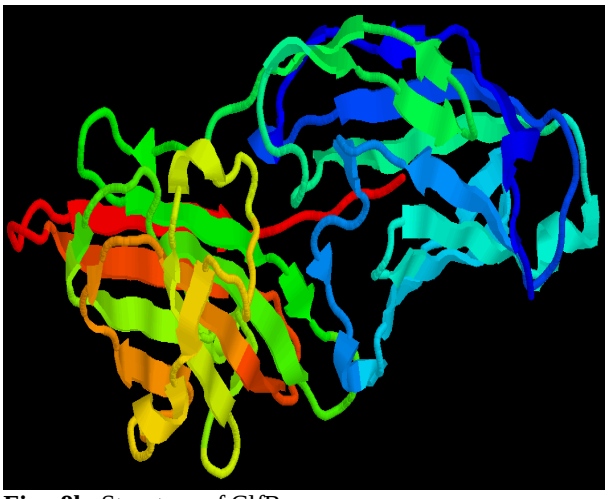

Fig: 8b. Structure of ClfB