International Journal of Advanced Biotechnology and Research
ISSN 0976-2612, Online ISSN 2278–599X,
Vol 4, Issue 3, 2013, pp 404-414
http://www.bipublication.com

IN-SILICO IDENTIFICATION AND SEQUENCE ANNOTATIONS OF POTENTIAL VACCINE CANDIDATE IN *NEISSERIA GONORRHOEAE*

Rajashri N. Bhairamadgi and Anil Kumar S. Katti*
Dept. of Bioinformatics, Walchand College of Arts and Science, Solapur, MH, India
*Corresponding author: Email: anisk09@gmail.com, Tel: 0217-2651863; Fax: 0217-2391849

[Received-29/07/2013, Accepted-21/09/2013]

ABSTRACT:

Gonorrhea is one of the most common sexually transmitted diseases (STD) caused by *Neisseria gonorrhoeae*. The transmission and prevention of gonococcal infection is a global health problem, till date no effective vaccine or specific drug could be developed and only antibiotic treatment is in use. Perhaps, due to excess use of antibiotics, several resistant strains have been found. In the present study, using an in-silico proteomic approach, candidate drug and vaccine targets have identified in *N. gonorrhoeae* virulent strain FA 1090. Proteome set was retrieved from *N. gonorrhoeae* virulent strain FA 1090, which contains 2002 proteins. Out of 2002 proteins, 70 transmembrane proteins were screened using SOSUI-GramN, among these 52 potent antigenic proteins were identified by VaxiJen v 2.0. Finally, a candidate peptide "VRMGIPVTE" of hypothetical protein YP_208831.1 identified for both T-cell and B-cell by querying the sequence of non-homologous antigenic protein. Hypothetical Protein YP_208831.1 is antigenic and has potentiality to induce both the T-cell and B-cell mediated immunity in *N. gonorrhoeae* but requires proper design and subsequent validation of the peptide target.

Keywords: *Gonorrhoeae, vaccine designing, STD, N. gonorrhoeae FA1090, Sequence annotation*

[I] NTRODUCTION

Neisseria Gonorrhoeae is the Gram-negative diplococcus causative agent of gonorrhoeae, a sexually transmitted disease [4,5,16]. In these days STDs are major public health concern both in industrialized and developing countries. The only host for *N. gonorrhoeae* is human, as it could not survive outside human body. Gonorrhoea is transmitted by human sexual contact of an infected tissue surface with susceptible mucosa epithelium surface and also vertically from mother to child during delivery [30,35]. These Cocci first adhere to the columnar epithelium cells by pilli and opa protein then penetrate, multiply on basement membrane and then produce lipopolysaccharide endotoxin [5,30,32].

According to the World Health Organization's report published in 2001 estimates the world's STD epidemics were in 1999. There were about 340 million new cases estimated including gonorrhoeae throughout the world among South & Southeast Asia was having largest number of case followed by Sub-Saharan Africa & Latin America and Caribbean and this was in 15 - 49 age group [9,24,38]. *N. gonorrhoeae* infection symptoms are distinct in men and women. Those

might be serious if remained untreated. Men will experience lower urinary tract infection as urethritis, epididymitis, proctitis while women are asymptomatic frequently get symptoms like vaginal and pelvic discomfort and it may lead to pelvic inflammatory disease (PID) which may cause permanent fallopian tube blockage with subsequent infertility and ectopic pregnancies [7,11,16].

Antibiotic resistance is only effective and essential method to control the gonorrhea. Over the last ten years *N. gonorrhoeae* strains developed elevated resistance to many antimicrobial agent as penicillin, tetracycline, ciprofloxacin, gemifloxacin, sulfonamides, cephalosporin [13,19,20,29,36] giving rise to problem of managing gonorrhea. Due to the unavailability of vaccine and less effectiveness of the antibiotics the gonococcal infection and transmission still have a global health problem [19,36]. Numerous virulence causes have been finding and researchers are in advancement to discover new improved methods and procedure to foster effective therapeutics against pathogen while barrier contraceptives and antibiotic therapies are prescribed to control and prevalence [9]. *N. gonorrhoeae* infection still remained important and unsolved problem in front of scientist because of its adaptability and resistance to broad range of antimicrobial agents. It will not be wrong if we expect in few decades that the gonococci might become more resistant to the antibiotic. That time tetracycline and cephalosporin will just useless. So now increasing resistance of *N. gonorrhoeae* to antibiotics raises the need of Vaccine development.

There was relative work done before on the same gonorrhoeae vaccine. tspA (T-cell stimulating protein-A) and tspB (T-cell stimulating protein-B) (US patent: 6861507) and Pili protein (US patent: 4443431) have reported and patented as a potent candidate against *Neisseria* [33] . There are already reporting potent vaccine candidates like Pili [25], PorB phospholipase A (pldA) [6], transferrin-binding proteins (tbpA and tbpB) , [22,23,31] competence lipoprotein and ddl [4] not found that much effective and promising in practice.

For targeting any protein from pathogen should have little or no similarity with the host organism. However, target must be important in cellular growth and sustain life of pathogenic organism. This concept of differences in proteins of host and pathogen could be efficiently and effectively used in designing the novel drug targets. It is true that prevention and cure can be achieved by preparing antigenic epitope vaccines as these triggers human immune response. So it would work if a single candidate may act as dual target vaccine and drug then that would be more effective and significant to suppress growth of *N. gonorrhoeae.*

Therefore novel target for vaccine and drug must be designed and developed to tackle the *N. gonorrhoeae* infection and transmission too. Present study involves screening of potent antigenic target and its sequence annotation study, which is of non-human homologous proteins and can be used as for both vaccine and drug designing.

[II] MATERIALS AND METHODS

2.1. Retrieval of Proteome Set

Proteome set of target organism *N. gonorrhoeae* FA 1090 retrieved from species search of EBI integr8.

2.2. Prediction of Transmembrane Proteins

Proteome set was screened by SOSUI-GramN to predict their sub-cellular localization. Sosui-GramN [15] predictive software system was developed for assessing subcellular localization of proteins in Gram-negative bacteria. The system does not require the sequence homology data of any known sequences. It uses only

physicochemical parameters of the N- and C-terminal sequences and the total sequence.

2.3. Determination of Antigenic Proteins

All the transmembrane proteins were uploaded as a multiple sequence file in FASTA format to VaxiJen v 2.0 [8] 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.

2.4. Selection of Non-Human Homologue Proteins

To identify proteins that are essential but not human homologue proteins, each protein sequence was subjected to human BLAST-P in NCBI server.

2.5. Identification of Epitope

Each antigenic protein sequence was then subjected for B-cell epitope prediction using BCPred. As the B-cell epitope have significant role in vaccine development. BCPred calculates based on Support Vector Machine (SVM) five different Kernel methods and finally predicts B-cell epitopes, which are highly, appropriate [10]. Top Scored (cut off value for BCPred is 1) antigenic with B-cell binding sequences were selected.

These antigenic B-cell epitope sequences were evaluated in ProPred-I and ProPred for MHC class I and MHC class II binding epitope prediction respectively. As ProPred –I [27]is online server to classify MHC class-I binding regions in epitope while ProPred [28] is to classify MHC class-II binding regions. Both server implements matrix based prediction algorithm.

2.6. Hypothetical Protein Sequence Analysis

2.6.1. Primary Protein Sequence Analysis

To compute various physico-chemical properties of protein sequence, ProtParam tool was used [12]. 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).

2.6.2. Secondary Structure Prediction

The hypothetical protein sequence was subjected in FASTA format to PSIPRED for predicting secondary structure. As PSIPRED [17] is a highly accurate method for protein secondary structure prediction. It is widely used transmembrane topology prediction method and one of GenTHREADER, pGenTHREADER and pDomTHREADER - sequence profile based fold recognition methods.

2.6.3. Functional Analysis

Protein functional analysis was carried out by subjecting the sequence to InterProScan. This is a tool that combines different protein signature recognition methods into one resource. Domain and Motif data from InterProScan are also displayed in the genome browser, grouped under “Protein Domains/Motifs” track. The input data is a bacterial or micro-sporidian protein sequence in FASTA format [37].

2.6.4. Domain Analysis

The antigenic B-cell and T-cell sequence was then subjected to domain analysis tool using SMART (Sequence Modular Architecture Research Tool) [26], for the identification and annotation of protein domains and the analysis of protein domain architectures. Sequences with more than three domains were selected for further analysis.

2.6.5. Protein-Protein Interactions

Protein-protein interaction study for the hypothetical protein NGO1801 was done by using STRING [21] 8.3 database. STRING is a database of known and predicted pr otein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources Genomic

Context, High-throughput Experiments, Conserved Co-expression.

2.6.6. 3D Modeling

Homology modeling of the antigenic protein was done using Swiss Model server [1]by choosing template from Protein Model Portal [2] . Then consequential model was optimized using Swiss-pdb Viewer [14]. Superlative model was selected using ProsaWeb [34]. The model visualized by RasMol viewer and PDBSum [18] was used at last to check the proteins 3D structure by different angles and amino acid positions too in specific foldings.

[III] RESULTS

3.1. Screening of Proteome set

Proteome set of *N. gonorrhoeae FA1090* was retrieved from integr8 of EBI. Retrieved proteome set contains 2002 protein sequences. All these proteins were mainly responsible for cells normal functioning and protection purpose from other host organism. SOSUI-GramN predictive software was used for assessing the subcellular localization of proteins in Gram-negative bacteria. Total transmembrane proteins were screened from 2002 proteins. Total 70 proteins were identified as (Outer membrane) transmembrane proteins present in *N. gonorrhoeae.* While 1930, proteins were located in other subcellular location such as inner membrane, periplasm, cytoplasm, extracellular and unknown (297, 182, 1185, 48, 220 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.

3.2. Selection of antigenic and Non-human homologue proteins

Total 70 transmembrane proteins were determined for their antigenicity using VaxiJen server. The protective antigen or non-antigen, were predicted according to a predefined cutoff. Since more of the models had their highest accuracy at a threshold of 0.5, this threshold value was chosen in our study. Out of 70 transmembrane, 52 proteins were found to be antigenic in nature, 6 proteins were identified as non-antigenic, 12 were less antigenic than 0.5 antigenic value. Proteins those present on outer membrane and antigenic in nature could be useful in potent vaccine and drug designing. We selected protein sequences that did not show any similarity with any of the human sequences by NCBI Blastp. Because these are non-human homologue essential enzymes, involved in pathogen specific metabolic pathways, and localized to the cell membrane or cell wall. Because of the subjected sequence was a non-human homologue essential protein and putative drug target for the *N. gonorrhoeae*; it was presumed that the homologous sequences also would be drug targets for the corresponding pathogens. All the 70 proteins screened by SosuiGramN were non-human homologue that could be useful to design a vaccine for human, as it did not involve in normal metabolic pathway.

3.3. Potent Antigenic Target

Fifty-two predicted potent antigenic proteins were subjected to BCPred server to identify B-cell binding epitopes with cut off value 1. Out of which, two proteins having accession number (YP_208831.1) and (YP_207267.1), antigenic proteins showed B-cell binding epitopes. This method had generated a stretch of 20-mer sequence of TTTAGGGVRMGIPVTEYDRV from 508 to 527 for YP_208831.1 and IHITGNNKTRDEVVRRELRQ spanning amino acid position from 352 to 371 for YP_207267.1 as shown in Table 1. B-cell epitopes are antigenic determinants that are recognized and bound by receptors (membrane bound antibodies) on the surface of B-lymphocytes. The identification and

characterization of B-cell epitopes play an important role in vaccine design, immunodiagnostic tests, and antibody production. MHC binding T-cell epitopes were identified using ProPred 1 (for MHC I) and ProPred (for MHC II) with default parameters. The 20-mer antigenic B-cell epitope sequence was analyzed with these two servers and the common epitope(s) that could bind both the MHC classes and covers maximum MHC alleles were selected. In this way only one 9-mer sequence (VRMGIPVTE) spanning at amino acid position 508-527 of the hypothetical protein was selected as shown in Table 2. The epitope peptide sequence VRMGIPVTE was taken for further analysis as this can potentially induce both the B-cell and T-cell.

3.4. Searching Similar Epitope in Other Organisms

BlastP analysis was carried out to find out similar epitope in other organism so that it can be used as putative drug target for other organisms. The BlastP result showed that the sequence (VRMGIPVTE) was highly conserved in *S. enterica S. typhimurium*, *E. coli* and moderately conserved in *P. aeruginosa*, *A. baylyi* and *F. novicida* were shown in Fig 1 (a) and (b). Therefore, the identified epitope may be a candidate vaccine against these pathogens also.

3.5. Protein Sequence Analysis

3.5.1. Primary Protein Sequence Analysis

Sequence of hypothetical protein subjected to ProtParam and information retrieved for molecular weight, Theoretical PI, amino acid composition, atomic composition, total number of atoms, extinction coefficient, estimated half life, instability index and details of the GRAVY recorded as in Table 3.

3.5.2. Secondary Structure Predictions

Protein secondary structure was predicted by using PSIPRED Server v3.0 as shown in Fig. 2. The PSIPRED prediction showed the results consisted 12 helices, coils and 33 strands with secondary stricture sequence alignment and prediction confidence in graphical view.

3.5.3. Functional Analysis

Using InterProScan tool did functional analysis of hypothetical protein NGO1801 and result was obtained for protein domain and functional site. The subjected sequences of hypothetical protein gave the bacterial surface antigen (D15) (Bac_surface_Ag), surface antigen variable number (Surf_Ag_VNR), outer membrane assembly protein (OM_assembly_OMP85, OM_YaeT) and un integrated two regions (Bac_surfAg_D15, PTHR12815:SF6) as shown in Fig. 3.

3.5.4. Domain Analysis

Domain analysis was done by SMART database. Out of 52 antigenic proteins, only two proteins were found with more than 3 domains. The domain regions obtained for YP_208831.1 from Pfam were five of Surf_Ag_VNR and one Bac_Surface_Ag domain. Hypothetical protein YP_207267.1 having STN, Pfam_Secretin_N and Pfam_Secretin_N as shown in Fig 4 (a) and (b). The protein YP_208831.1 was uncharacterized while protein YP_207267.1 was highly characterized and well annotated.

3.5.5. Protein-Protein Interactions

Protein-Protein interaction study for hypotheical protein YP_208831.1 was revealed by STRING database and the interacting proteins are NGO 1800, IpxD, NGO1802, fabZ, IpsA, comL, uppS, dxr, NGO 1798, IpxB etc. were shown in Fig.5 as a graphical view which showed interactions in abstractive and attractive way of interested protein with other proteins.

3.5.6. 3-D Modeling of *N. gonorrhoeae* Hypothetical Protein

The 3D structure of *N. gonorrhoeae* hypothetical protein YP_208831.1 was not available, hence the model was generated using Swiss Model Server and the template was selected from

ModBase based on the Blast parameters. The template selected was *E. coli* (PDB id: 2qdzA) having 31% sequence identity and for selected region between 236-792 amino acids. The X-Ray diffraction structure had 3.15 A^o resolutions. Visualization of the built model was done using RasMol as shown in Fig.6. To validate the model, ProSA-web was used, that compares and analyzes the energy distribution in protein structure as a function of sequence position to determine a structure as native like or fault as shown in Fig.7. Local model quality and overall model quality was checked. The overall model quality obtained in *Z*-Score: -6.37. PDBSum was used to analyze the protein 3D structure by different angles and amino acid positions in different specific folding.

The optimized final 3-D model of *N. gonorrhoeae* hypothetical protein with main, bottom and right view is as shown in Fig. 8 (a), (b) (c) respectively and Fig 8 (D) consist of 4 sheets (A, B, C, D mixed-2, antiparrallel-2), 18 beta hairpins, 1psi loop, 4 beta bulges, 25 strands, 4 helices, 2 helix-helix interactions, 51 beta turns, 9 gamma turns.

[IV] DISCUSSION

The present investigation identifies a most probable candidate peptide vaccine from fifty-two outer membrane antigenic drug targets of *N. gonorrhoeae.* This outer membrane protein therefore may be useful in developing drug as-well-as vaccines against the pathogen. According to Shivkumar *et. al., (*2011) membrane localized proteins have 70% chances of effective drug target in any organism. Predicted epitope of Hypothetical Protein YP_208831.1 is antigenic and potential to induce both the T-cell and B-cell mediated immunity. The hypothetical protein NGO1801 protein is potent drug target. VRMGIPVTE sequence of Hypothetical protein NGO1801 from *Neisseria gonorrhoeae* strain *FA 1090.* (Accession no. YP_208831.1) can be used as potent target to develop drug as well as peptide vaccine. As per the records in NCBI Hypothetical protein NGO1801 protein is predicted as outer membrane, which we also proved by using the tools. There are also NCBI records for the protein with accession no. YP_208831.1 is antigenic in nature at the position 505-525 and is "Bac-surface_Ag". This epitope VRMGIPVTE is produces both T-cell and B-cell mediated immunity, which is novel characteristic of a peptide to design an epitope for target-based vaccine designing [3]. So this candidate and predicted epitope may be superior and safe to those previously identified vaccine candidates and epitope. Experimental validation of this candidate epitopes is required.

[V] CONCLUSION

To conclude, hypothetical protein NGO1801 (YP_208831.1) could be a good target for developing effective antibiotic and vaccine for pathogenic *Neisseria, P. aeruginosa, Acinetobacter* sp., *S. typhi, E. coli s*tr. K-12 substr. and *F. novicida.* Effective inhibitor screening for hypothetical protein YP_208831.1 is required. Similarly, the identified epitope (VRMGIPVTE) require proper design and successive validation for their uses as peptide vaccine against these human pathogens.

REFERENCES

1] Arnold, K., *et al.* (2006), The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling, *Bioinformatics*. 22, 2, 195-201

2] Arnold, K., *et al*. (2009), The Protein Model Portal, *Journal of Structural and Functional Genomics*. 10, 1, 1-8

3] Barh D, *et al*. (2010), A novel strategy of epitope design in Neisseria gonorrhoeae, *Bioinformation*. 5, 2, 77-85

4] Barh, D. and Kumar, A. (2009), In silico Identification of Candidate Drug and Vaccine Targets from Various Pathways in Neisseria gonorrhoeae, *In Silico Biology*. 9, 4, 225-231

5] Barh, D., Misra, A.N. and Kumar, A. (2010), In silico identification of dual ability of n. gonorrhoeae

ddl for developing drug and vaccine against pathogenic neisseria and other human pathogens, *Journal of Proteomics and Bioinformatics*. 3, 3, 082-090

6] Bos, M.P., *et al*. (2005), Function of neisserial outer membrane phospholipase A in autolysis and assessment of its vaccine potential, *Infection and Immunity*. 73, 4, 2222-2231

7] Densen, P. (1989), Interaction of complement with Neisseria meningitides and Neisseria gonorrhoeae, *Clinical Microbiology Reviews*. 2, S11–S17

8] Doytchinova, I.A. and Flower, D.R. (2007), VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinformatics*. 8, 4, 7

9] Edwards, J.L. and Apicella, M.A. (2004), The molecular mechanisms used by Neisseria gonorrhoeae to initiate infection differ between men and women, *Clinical Microbiology Reviews*. 17, 4, 965-981

10] El-Manzalawy, Y., Dobbs, D. and Honavar, V. (2008), Predicting linear B-cell epitopes using string kernels, *Journal of Molecular Recognition*. 21, 4, 243-255

11] Furuya, R. and Tanaka, M. (2009), Neisseria gonorrhoeae infections, *Nippon Rinsho*. 67, 129-135

12] Gasteiger, E., *et al.* (2005), *The Proteomics Protocols Handbook*

13] Golparian, D., *et al.* (2010), Emergence, spread and characteristics of Neisseria gonorrhoeae isolates with in vitro decreased susceptibility and resistance to extended-spectrum cephalosporins in Sweden, *Sexually Transmitted Infections*. 86, 6, 454-460

14] Guex, N. and Peitsch, M.C. (1997), SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis*. 18, 15, 2714-2723

15] Imai K, *et al.* (2008), SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram negative bacteria, *Bioinformation*. 2, 9, 417-421

16] Johnson, M.B. and Criss, A.K. (2011), Resistance of Neisseria gonorrhoeae to neutrophils, *Frontiers in Microbiology*. 2, APR

17] Jones, D.T. (1999), Protein secondary structure prediction based on position-specific scoring matrices, *Journal of Molecular Biology*. 292, 2, 195-202

18] Laskowski, R.A., *et al*. (1997), PDBsum: a web-based database of summaries and analyses of all PDB structures, *Trends in Biochemical Sciences*. 22, 12, 488-490

19] Lee, H.S.W., Ostrowski, M.A. and Gray-Owen, S.D. (2008), CEACAM1 dynamics during Neisseria gonorrhoeae suppression of CD4 + T lymphocyte activation, *Journal of Immunology*. 180, 10, 6827-6835

20] Newman, L.M., Moran, J.S. and Workowski, K.A. (2007), Update on the management of gonorrhea in adults in the United States, *Clinical Infectious Diseases*. 44, SUPPL. 3, S84-S101

21] Perng, C.Y.J., Smith, J. and Rossing, T. (2010), Physical modeling of the harpsichord plectrum-string interaction. *13th International Conference on Digital Audio Effects, DAFx 2010 Proceedings*

22] Price, G.A., *et al.* (2007), Gonococcal transferrin binding protein chimeras induce bactericidal and growth inhibitory antibodies in mice, *Vaccine*. 25, 41, 7247-7260

23] Price, G.A., Russell, M.W. and Cornelissen, C.N. (2005), Intranasal administration of recombinant Neisseria gonorrhoeae transferrin binding proteins A and B conjugated to the cholera toxin B subunit induces systemic and vaginal antibodies in mice, *Infection and Immunity*. 73, 7, 3945-3953

24] Report, W.H.O. (2001), STD statistics worldwide

25] Rothbard, J.B., *et al.* (1985), Antibodies to peptides corresponding to a conserved sequence of gonococcal pilins block bacterial adhesion, *Proceedings of the National Academy of Sciences of the United States of America*. 82, 3, 915-919

26] Schultz, J., *et al.* (2000), SMART: A web-based tool for the study of genetically mobile domains, *Nucleic Acids Research*. 28, 1, 231-234

27] Singh, H. and Raghava, G.P.S. (2003), ProPred1: Prediction of promiscuous MHC class-I binding sites, *Bioinformatics*. 19, 8, 1009-1014

28] Singh, H. and Raghava, G.P.S. (2002), ProPred: Prediction of HLA-DR binding sites, *Bioinformatics*. 17, 12, 1236-1237

29] Tapsall, J. (2006), Antibiotic resistance in Neisseria gonorrhoeae is diminishing available treatment options for gonorrhea: Some possible remedies, *Expert Review of Anti-Infective Therapy*. 4, 4, 619-628

30] Tapsall, J. (2001), Antimicrobial resistance in Neisseria gonorrhoeae. World Health Organization (WHO)

31] Thomas, C.E., *et al.* (2006), Vaccination of mice with gonococcal TbpB expressed in vivo from Venezuelan equine encephalitis viral replicon particles, *Infection and Immunity*. 74, 3, 1612-1620

32] Todar, K. (2008-2012), Todar's Online Textbook of Bacteriology. *Neisseria: Gonorrhea and Meningitis*. University of Wisconsin

33] Van Ulsen, P., *et al.* (2008), Two-partner secretion systems of Neisseria meningitidis associated with invasive clonal complexes, *Infection and Immunity*. 76, 10, 4649-4658

34] Wiederstein, M. and Sippl, M.J. (2007), ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Research*. 35, SUPPL.2, W407-W410

35] Wiesner P J and Thompson S E 3rd (1980), Gonococcal diseases, *Dis Mon*. 26, 5, 1-44

36] Workowski, K.A., Berman, S.M. and Douglas Jr, J.M. (2008), Emerging antimicrobial resistance in Neisseria gonorrhoeae: Urgent need to strengthen prevention strategies, *Annals of Internal Medicine*. 148, 8, 606-613

37] Zdobnov, E.M. and Apweiler, R. (2001), InterProScan - An integration platform for the signature-recognition methods in InterPro, *Bioinformatics*. 17, 9, 847-848

38] Zhu, W., *et al.* (2011), Vaccines for gonorrhea: can we rise to the challenge?, *Frontiers in Microbiology*. 2

(a)

***Pseudomonas aeruginosa* UCBPP-PA14**

```
Query  481  DGVSLGYDIYGKAFDPRKASTSVKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLTVNT
540
            DGVSLGY+ + +  D  +    V  Y    + G G+ +G P++E  R+ +GL+ +   ++T
Sbjct  480  DGVSLGYNAFYRKTDYDELDVDVASYSVNSLGAGMSIGYPISETSRLTYGLSVQRDQIDT
539
```

***Acinetobacter* sp. ADP1**

```
Query  503  VKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLTVNTYNKAPKRYADFIRKYG------
556
            V  Y T + GG +  G P+ E   ++  L  +   V T          D++   G
Sbjct  506  VNNYVTDSFGGNLSFGYPIDENQSLSASLGIDSTKVKTGPYVSTYVRDYLLANGGKATGT
565
```

***Salmonella enterica serovar Typhi,* yaeT**

```
Query  479  TADGVSLGYDIYGKAFDPRKASTSVKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLTV
538
            T DGVSLG  I+   F+   A  S   Y   + G  V +G P+ EY+ +  GL   H  +
Sbjct  479  TVDGVSLGGRIFYNDFEADDADLS--DYTNKSYGTDVTLGFPINEYNTLRAGLGYVHNKL
536
```

***Escherichia coli* str. K-12 substr. MG1655**

```
Query  479  TADGVSLGYDIYGKAFDPRKASTSVKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLTV
538
            T DGVSLG  ++   F    A  S   Y   + G  V +G P+ EY+ +  GL   H ++
Sbjct  479  TVDGVSLGGRLFYNDFQADDADLS--DYTNKSYGTDVTLGFPINEYNSLRAGLGYVHNSL
536
```

***Francisella novicida* U112**

```
Query  478  FTADGVSLGYDIYGKAFDPRKASTSVKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLT
537
            FT  GVS     Y     +  K + +V  Y+  T G  +  G+P++ +  V+ G+    +
Sbjct  485  FTTSGVSQSISAYINRSNFAK-TNAVAAYQLDTIGARLMYGVPISTFSNVSGGITFAN--
541
```

(b)

```
gi|116051645|ref|YP_789516.1|       YRKTDYDELDVDVASYSVNSLGAGMSIGYPISETSRLTYGLSVQRDQIDT
gi|118423959|gb|ABK90349.1|         INRSNFA-KTNAVAAYQLDTIGARLMYGVPISTFSNVSGGITFANNTVKQ
gi|16128170|ref|NP_414719.1|        YNDFQA--DDADLSDYTNKSYGTDVTLGFPINEYNSLRAGLGYVHNSLSN
gi|16763614|ref|NP_459229.1|        YNDFQA--DDADLSDYTNKSYGTDVTLGFPINEYNTLRAGLGYVHNKLSN
gi|50084556|ref|YP_046066.1|        YRKTKLSS-SYTVNNYVTDSFGGNLSFGYPIDENQSLSASLGIDSTKVKT
gi|59802119|ref|YP_208831.1|        GKAFDPRKASTSVKQYKTTTAGGGVRMGIPVTEYDRVNFGLAAEHLTVNT
gi|81515728|sp|Q8Z9A3.1|YAET_SALTI  YNDFEA--DDADLSDYTNKSYGTDVTLGFPINEYNTLRAGLGYVHNKLSN
```

Figure. 1. (a) NCBI blastp with *N. gonorrhoeae FA1090*, (hypothetical protein) **YP_208831.1** showing conserved epitope sequences "VRMGIPVTE" in other human pathogens. **(b)** multiple sequence alignment of **YP_208831.1** (hypothetical protein) from identified other human pathogens using T-Coffee.

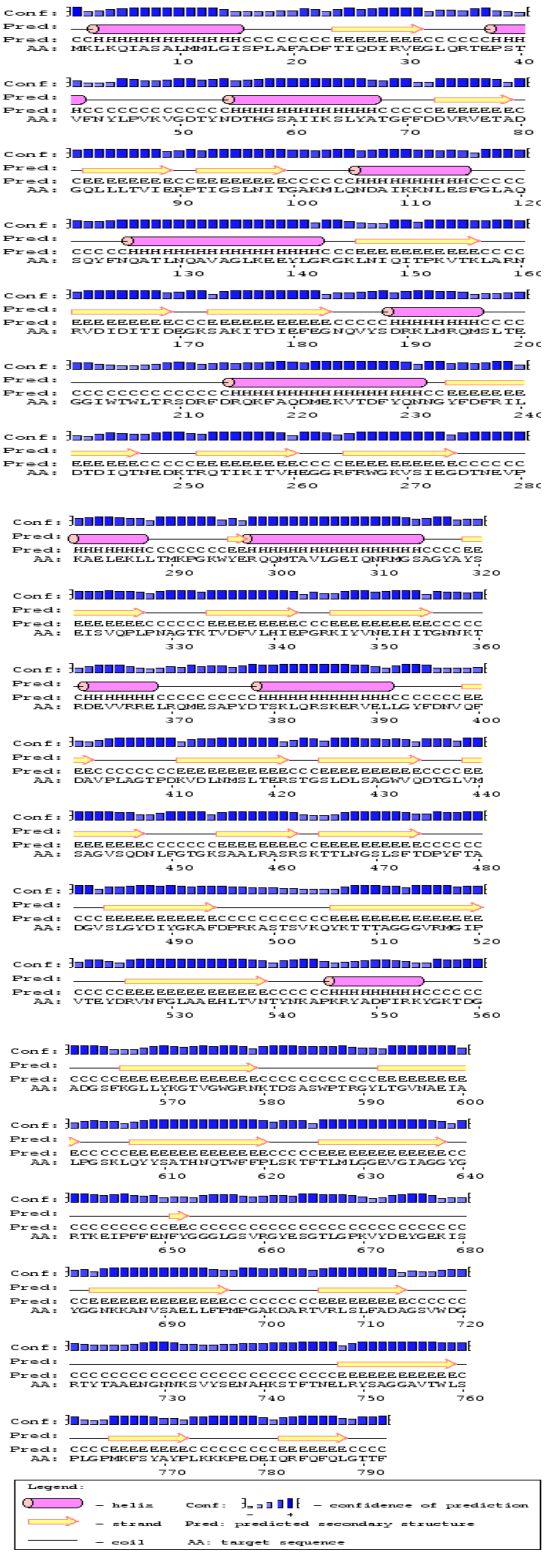


Figure.2. Structure predictions by PsiPred

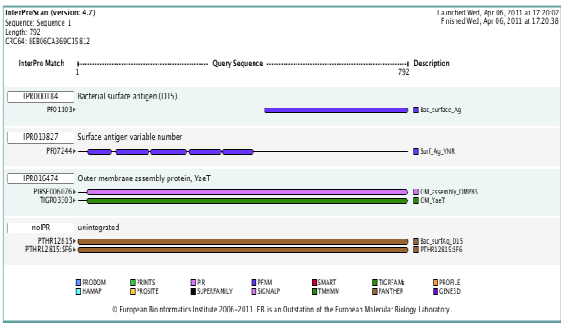


Figure. 3. InterProScan result for domain analysis of hypothetical protein

(a)

(b)

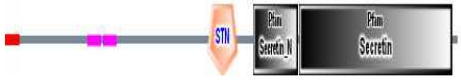


Figure. 4. Domain analysis of (a) Hypothetical Protein (YP_208831.1) (b) Hypothetical Protein (YP_207267.1) by SMART

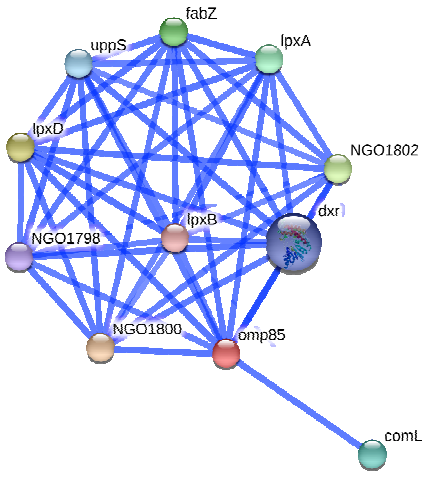


Figure. 5. Protein-Protein interactions by STRING

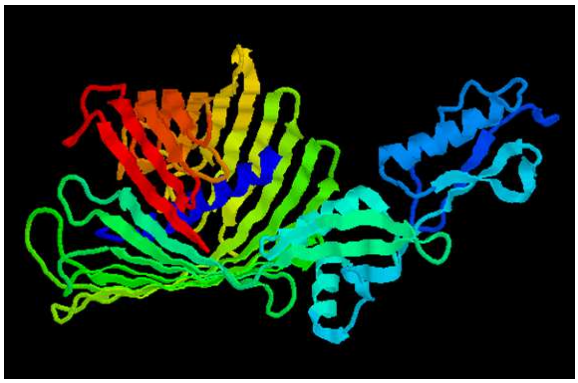
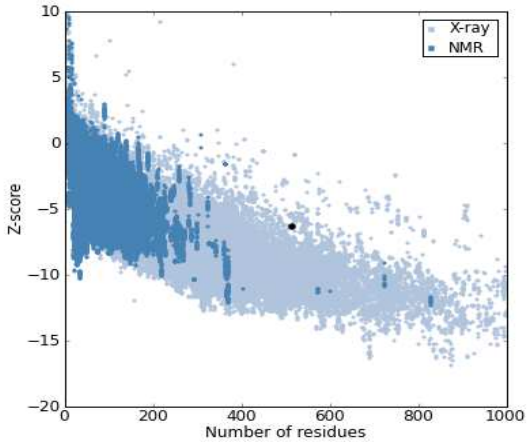


Figure. 6. Model Visualization by RasMol



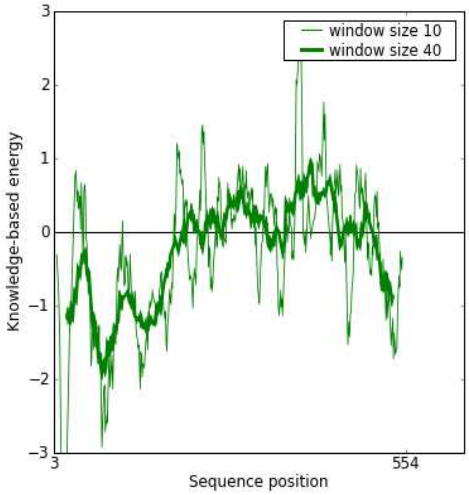


Figure: 7 ProSA-web results to compare & analyze energy distribution.

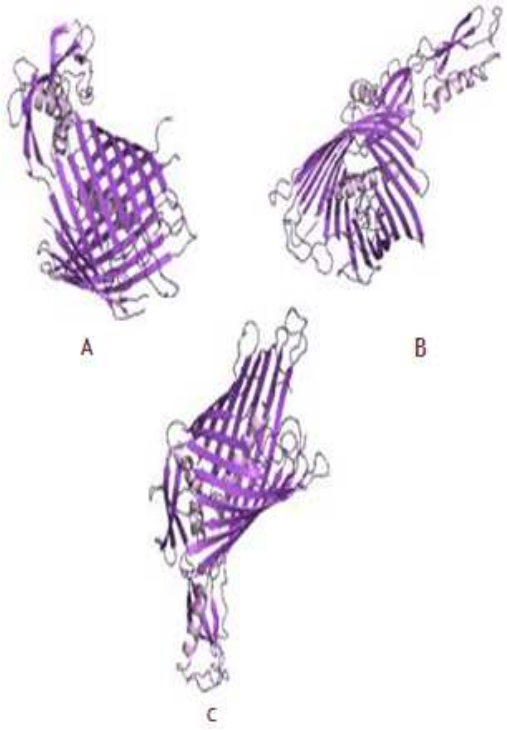


Figure: 8 (A) main (B) bottom and (C) right view of model by PDBSum

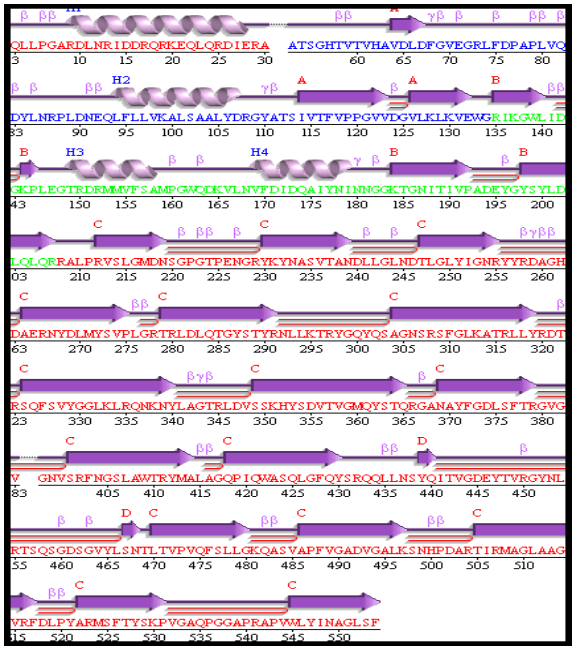


Figure. 9. (D) Amino Acid Positions in Hypothetical protein YP_208831.1 by PDBSum

Table: 1 Antigenic B-cell epitopes of *N. gonorrhoeae* Proteins.

BCPred algorithm		
YP_208831.1		
TTTAGGGVRMGIPVTEYDRV	508	20
YP_207267.1		
PARPAVKAAPAAPAKQQAAA	134	20
AAP Prediction algorithm		
YP_208831.1		
QYKTTTAGGGVRMGIPVTEY	505	20
NKAPKRYADFIRKYGKTDGA	542	20
YP_207267.1		
YEIPFTVTTASGGGNSTNTE	583	20
SAPARPAVKAAPAAPAKQQA	132	20
Common antigenic B-cell epitope		
QYKTTTAGGGVRMGIPVTEYDRV	----	23

Table. 2. T-cell Epitopes of *N. gonorrhoeae* of two hypothetical proteins

Propred (MHC Class II)		
Sequence	**Score**	**Name of binding alleles**
YP_208831.1		
VRMGIPVTE	4.5000	DRB1_0301
MGIPVTEYD	1.0000	DRB1_0301
YP_207267.1		
IHITGNNKT	-0.9000	DRB_0101
ITGNNKTRD	1.0000	DRB_0101
Propred1 (MHC Class II)		
YP_208831.1		
G**VRMGIPVT**	0.17	HLA-A*0205
IPVTEYDRV	0.1	HLA-A*0205
YP_207267.1		
GNNKTRDEV	0.0085	HLA-A*0205
IHITGNNKT	0.0056	HLA-A*0205

Table. 3. Annotation study (Primary, secondary and tertiary structure analysis)

Sr. no.	Parameters	Details	Tool / server references
1	Protein Name	**Hypothetical protein NGO1801**	GenPept
	Gene name	**YP_208831**	GenPept
	Organism	***Neisseria Gonorrhoeae* (strain ATCC 700825 / FA 1090)**	GenPept
	Taxonomy	**Identifier: 242231** **Lineage: Bacteria › Proteobacteria › Betaproteobacteria › Neisseriales › Neisseriaceae › Neisseria**	NCBI Taxonomy Browser
2	Protein Attributes	**Sequence length: 792**	GenPept, ProtParam, InterProScan
3	Physico-chemical properties	**Molecular weight: 87942.2** **Theoretical PI: 8.91** **Amino acid composition: 792 AA** **Negatively charged residues: Asp + Glu= 90** **Positively charged residues: Arg + Lys= 98** **Atomic composition: C=3924, H=6121, N=1067, O=1197, S=17** **Total number of atoms: 12326** **Extinction coefficient: 110130** **Estimated half life: 30 hrs (mammals In vitro), >20hrs (Yeast In vivo), >10hrs (E. coli in vivo)** **Instability index: 29.82 (protein is stable)** **Aliphatic index: 73.65** **Grand average hydropathicity (GRAVY) : -0.467**	ProtParam
4	Secondary structure analysis	**Helix: 12** **Strands: 33**	PsiPred
5	Functional analysis	**Surface antigen: Bac_Surface_Ag** **Surface antigen variable: Surf_Ag_VNR**	InterProScan
6	Domain analysis	**Surf_Ag_VNR** **Bac_Surface_Ag**	SMART InterProScan
7	Protein-Protein Interaction	**Interacting Protein with query sequence: NGO 1800, IpxD, NGO1802, fabZ, IpsA, comL, uppS, dxr, NGO 1798, IpxB**	STRING
8	3-D strucures	**2qdzA** **2qczA**	ModBase