Bioinformation
www.bioinformation.net
Volume 5
Issue 2
open access
Hypothesis

A novel strategy of epitope design in *Neisseria gonorrhoeae*

Debmalya Barh[1]*, Amarendra Narayan Misra[2], Anil Kumar[3], Vasco Azevedo[4]

[1]Centre for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, WB-721172, India. [2]Department of Biosciences and Biotechnology, School of Biotechnology, Fakir Mohan University, Jnan Bigyan Vihar, Balasore-756020, Orissa, India. [3]School of Biotechnology, Devi Ahilya University, Khandwa Road, Indore, MP-452001, India; [4]Laboratorio de Genetica Celular e Molecular, Departmento de Biologia Geral, Instituto de Ciencias Biologics, Universidade Federal de Minas Gerais CP 486, CEP 31270-901 Belo Horizonte, Minas Gerais, Brazil, Debmalya Barh - Email: dr.barh@gmail.com; Phone: +91-944 955 0032,*Corresponding author

Received March 23, 2010; revised May 17, 2010; accepted June 08, 2010; published July 06, 2010

Abstract:
In spite of genome sequences of both human and *N. gonorrhoeae* in hand, vaccine for gonorrhea is yet not available. Due to availability of several host and pathogen genomes and numerous tools for *in silico* prediction of effective B-cell and T-cell epitopes; recent trend of vaccine designing has been shifted to peptide or epitope based vaccines that are more specific, safe, and easy to produce. In order to design and develop such a peptide vaccine against the pathogen, we adopted a novel computational approach based on sequence, structure, QSAR, and simulation methods along with fold level analysis to predict potential antigenic B-cell epitope derived T-cell epitopes from four vaccine targets of *N. gonorrhoeae* previously identified by us [Barh and Kumar (2009) In Silico Biology 9, 1-7]. Four epitopes, one from each protein, have been designed in such a way that each epitope is highly likely to bind maximum number of HLA molecules (comprising of both the MHC-I and II) and interacts with most frequent HLA alleles (A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401) in human population. Therefore our selected epitopes are highly potential to induce both the B-cell and T-cell mediated immune responses. Of course, these selected epitopes require further experimental validation.

Key words: gonorrhea, vaccine designing, epitope mapping, antigenicity HLA alleles, immune response

Background:
The disease, gonorrhea is a wide spread sexually transmitted disease caused by infection and transmission of *N. gonorrhoeae.* It is a drastic disease responsible for developing ***acquired immune deficiency syndrome*** (AIDS). Several attempts have been made for the development of vaccines for drastic diseases like AIDS **[1]** and flue viruses **[2]. V**accine development for gonorrhea, an age old disease is not in sight. However, several vaccine targets viz. protein-I B (porB) **[3]**, opacity protein (opa), lipo-oligosaccharides, protein-I, lactoferrin-1 and -2 (lbpl, lbp2), Immunoglobulin A1 (IgA1) proteases **[4]**, phospholipase A (pldA) **[5]**, transferrin-binding protein -A and -B (tbpA and tbpB) **[6-8]**, 2C7 oligosaccharide **[9]** have been identified for *N. gonorrhoeae* and few of them viz. pili (US4443431), T-cell stimulating protein-A and -B (tspA and tspB) (US6861507) have been patented. However, the problem still remains unsolved. Candidate vaccines identified and tested from these targets are only partially successful perhaps due to wide adoptability of the gonococcus, lack of proper animal models for the disease, cell lines to produce the vaccine, and inadequate immune response of used epitopes derived from these target proteins.

Secreted and surface proteins of any given pathogen are mostly antigenic and are responsible for pathogenicity **[10]**. Hence, these are good vaccine candidates. B-cell epitopes are parts of proteins that are antigenic and therefore are recognized by the antibodies. Identification of B-cell epitopes is useful development of diagnostic tests and can be employed as the first step in vaccine designing **[11]**. Similarly, effective immune response depends on specificity and diversity of the antigen binding to the human leukocyte antigen (HLA) alleles **[12]** class I (recognizes CD8+ T-cells) and class II (recognizes CD4+ T-cells) **[13, 14]**. The HLA molecules are highly polymorphic and according to IMGT/HLA Database (http://www.ebi.ac.uk/imgt/hla/stats.html), there are 4,633 HLA alleles (Class-I =3,411 and Class-II = 1,222) **[15]**. Among these, DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*1801, DRB1*1101 and DRB1*1501 of Class-II HLA alleles are most frequent (20-50%) **[16]**. Evidences are that gonococcal infection decreases the CD4+ and CD8+ T-cell counts **[17, 18]** and Pili proteins modulate initial CD4+ T-cell proliferation and regulatory T-cell activation thereby may evoke immune responses **[19]**. Therefore both these T-cells mediated immunities should be considered in designing peptide vaccines against *N. gonorrhoeae*. A list of previously reported T-epitopes is given in **Table 1 (see supplementary material)**.

Compared to the conventional vaccines, peptide or epitope based vaccines are easy to produce, more specific, and also safe. The advent of human genome project, sequencing and functional annotation of several pathogenic bacteria has boosted peptide based vaccine designing. Efficacy of computational prediction of effective B-cell and T-cell epitopes forms pathogenic genome as the key success to develop such vaccines. It is recommended to get an overview of various computational approaches in this regard **[20]**.

Previously, we have reported several membrane associated essential proteins of the gonococcus using computational approaches that can be used in designing new vaccine (s) against the pathogen **[21]** In the present study, four such membrane associated essential gonococcal proteins are explored for designing peptide vaccine (s) using a novel *in silico* strategy combined with simulation and fold level verification to identify best possible and most effective epitopes that can produce both the B-cell and T-cell mediated immunity.

Bioinformation Volume 5
www.bioinformation.net Issue 2 Hypothesis
open access

Methodology:

Prediction of antigenic B-cell epitopes:

Four essential membrane proteins namely D-alanine--D-alanine ligase (ddl), Sulfate transport permease protein C (cysW), Competence lipoprotein (comL), and Type IV pilin protein (pilV) of *N. gonorrhoeae* virulent strain *FA 1090* earlier identified as best vaccine candidates **[21]** were selected for the current study and a novel approach of epitope designing was adopted where an epitope should produce both the B-cell and T-cell mediated immunity.

The complete amino acid sequence of each protein was retrieved from Swiss-Prot protein database (http://us.expasy.org/sprot) and analyzed using VaxiJen v2.0 antigen prediction server **[22]**. The default parameters (threshold=0.4, ACC out put) were used against bacterial species to check the antigenicity of each full length protein. Proteins having antigenic score >0.5 were selected. Each selected full length amino acid sequence was then subjected to transmembrane topology analysis using TMHMM v0.2 prediction server **[23]** in order to identify exo-membrane (surface exposed) amino acid sequences of each protein.

For prediction of B-cell epitopes, each full length protein sequence was subjected to BCPreds analysis **[24]** and all predicted B-cell epitopes (20 mers) having a BCPreds cutoff score >0.8 were selected. Selected B-cell epitopes were then subsequently checked for membrane topology by comparing with TMHMM results for exo-membrane amino acid sequences. Surface exposed B-cell epitope sequences having the cutoff value for BCPreds (>0.8) were selected and further analyzed using VaxiJen (threshold=0.4, ACC out put) to check the antigenicity. Finally 2-3 epitopes with top VaxiJen scores were selected for use in prediction of T-cell epitopes.

Prediction of T-cell epitopes from B-cell epitopes:

T-cell epitopes were predicted from the selected B-cell epitopes. Both the sequence based and structure based QSAR simulation approaches were taken into account to predict T-cell epitopes and two screening steps were adopted. In the first screening, the selection criteria were: i) the sequence should bind to both the MHC class-I and class-II molecules and minimum number of total interacting MHC molecules should be >15, ii) the sequence must interact with HLA-DRB1*0101 of MHC class-II, and iii) should be antigenic based on VaxiJen score. Propred-1 (47 MHC Class-I alleles) **[25]** and Propred (51 MHC Class-II alleles) **[26]** servers that utilize amino acid position coefficients inferred from literature employing linear prediction model **[27],** were used to identify common epitopes that bind to both the MHC class molecules as well as to count total numbers of interacting MHC alleles. For QSAR simulation approach, the half maximal (50%) inhibitory concentration (IC_{50}) and antigenicity of common epitopes predicted by Propred-1 and Propred was calculated using MHCPred v.2 **[28]** server (selecting DRB1*0101) and VaxiJen, respectively. Epitopes with highest antigenicity and those bind more than 15 MHC molecules comprising of both the MHC class I and II alleles and less than 100 nM IC_{50} scores for DRB1*0101 were selected. The second screening was based on structure and QSAR simulation methods using T-Epitope Designer (http://www.bioinformation.net/ted/) **[29]** and MHCPred, respectively. T-epitope Designer predicts HLA-peptide binding based on virtual binding pockets using X-ray crystal structures of HLA-peptide complexes followed by calculation of peptide binding to binding pockets. The server can screen peptides for >1000 HLA alleles. In the second screening, the criteria were: i) binding prediction with large number of HLA alleles (>1000), ii) the peptide should bind > 75% of total HLA molecules, iii) must bind with high scores to A*0201, A*0204, and B*2705, and iv) must bind to DRB1*0101 and DRB1*0401. T-epitope Designer was used for criteria i), ii), and iii) and MHCPred was used for selection of DRB1*0101 and DRB1*0401 binding peptides. The final list of epitopes was made with non overlapping peptide sequences that pass these above mentioned criteria and VaxiJen and IC_{50} scores. Selected epitopes were further analyzed for fold level topology.

Epitope analysis:

Homology modeling for each full length protein was carried out using Phyre version 2.0 Web-server **[30]** and best models were selected based on super families and E-values of templates. The 3D folding and clusters of epitopes in folded protein were analyzed to confirm the exo-membrane topology of these epitopes using Pepitope server **[31]**. Pepitope was fed with Phyre derived 3D structure of each protein and all identified epitopes from the same protein to analyze the linear alignment of epitopes on the corresponding protein and to determine the epitope clusters and exo-membrane position of epitopes in the folded proteins. The detailed method has been represented in **Figure. 1.**

Results:

Antigenicity and topology of selected proteins

VaxiJen analysis of exo-membrane full length proteins selected for this study exhibited various degree of antigenicity ranging from 0.3986 to 0.6091 (**Table 2,** column 1 **see supplementary material**). Out of the four essential membrane proteins, cysW showed highest antigenicity (0.6091). Although comL exhibited the lowest score (0.3986) and predicted to be non-antigenic by VaxiJen, it was also considered for further analysis. The basic criterion of a good epitope is that it must be exposed to cell outside. The transmembrane topology analyses of these proteins were done using TMHMM and the result revealed that ddl (1- 304 amino acids) and comL (1- 267 amino acids) are fully exposed to outside of the membrane. The exo-membrane amino acid sequences of cysW are 42 – 66 and 129 – 137, and for pilV, the sequence is 30 - 129 (**Table 2,** column 6) **(see supplementary material)**.

Antigenic B-cell epitopes:

A peptide should be hydrophilic and produce both the B-cell and T-cell mediated immunity for becoming a good vaccine candidate **[26]**. Therefore to identify such epitopes, full length proteins were first subjected to B-cell epitope prediction using BCpreds and all B-cell epitopes were listed from each protein (**Table 3 see supplementary material**). Best epitopes were selected based on the criteria as mentioned in methods. In general, epitopes having BCpreds and VaxiJen cutoff values respectively >0.8 and >0.4 were selected except epitope from cysW where only one 20 mers epitope was found exposed to cell out side that has the BCpred and VaxiJen scores respectively 0.056 and 0.6316 (**Tables 3 see supplementary material**). Therefore, two epitopes each from ddl, comL, and pilV and one from cysW were finally selected for further analysis (**Table 3 see supplementary material**).

B-cell epitopes derived T-cell epitopes:

Each selected B-cell epitope was analyzed for identification of T-cell epitopes within the B-cell epitope sequence. For the first level screening, Propred-I (47 MHC Class-I alleles), Propred (51 MHC Class-II alleles), and MHCPred (DRB1*0101 allele) were used to identify common T-cell epitopes that share B-cell epitope sequence, can interact with both the MHC classes with highest number, and specifically interact with DRB1*0101 (as the DRB1*0101 is commonest bound allele, therefore the interaction epitope should produce better antigenic response) **[32] (Table 4)**. At the second level of screening, identified peptides in the first screen were used to predict their binding abilities to >1000 MHC alleles using T-Epitope Designer and epitopes that bind to >75% alleles were selected **(Table 5)**. Similarly, as A*0201, A*0204, and B*2705 alleles are mostly used in various prediction methods **[20]**, we set the cut off that selected peptides must bind to these three HLA molecules and T-epitope Designer was also used for this purpose **(Table 5 see supplementary material)**. Since the frequency of DRB1*0101 and DRB1*0401 alleles of MHC class-II is 20-50% **[16]**, we selected T-epitopes that interact with these two HLA molecules using MHCPred as described in methods **(Table 6 see supplementary material)**. The final list of epitopes was made with non overlapping peptide sequences that confirm the above mentioned criteria and VaxiJen and IC_{50} scores **(Table 6 see supplementary material)**. Finally, one epitope from each ddl, comL, cysW, and pilV are found to be prospective candidates for vaccine design (**Table 6**, in bold **see supplementary material**). Selected epitopes were further analyzed for fold level topology.

Clusters and folding of epitopes:

There is every possibility of the identified epitopes getting folded inside the tertiary structure of the corresponding protein. Therefore, a fold level analysis of all identified epitopes was carried out to determine the position of epitopes in folded proteins and to confirm their exo-membrane

Bioinformation **Volume 5**
www.bioinformation.net Issue 2 **Hypothesis**
open access

topology. Homology based 3D structures of all four proteins were carried out using Phyre server as described under Methods. The **Table 7 (see supplementary material)** represents the summary of Phyre results. Analysis of the epitope clusters and positions of epitopes with in the folded protein by Pepitope showed that identified all epitopes are situated within clusters having acceptable scores, located at the surface of the corresponding protein, and exposed to cell out side (**Table 4 see supplementary material,** bold highlighted in last column**,** and **Figure 2**). Therefore, these results confirmed the suitability of all identified epitopes as prospective vaccine candidates.

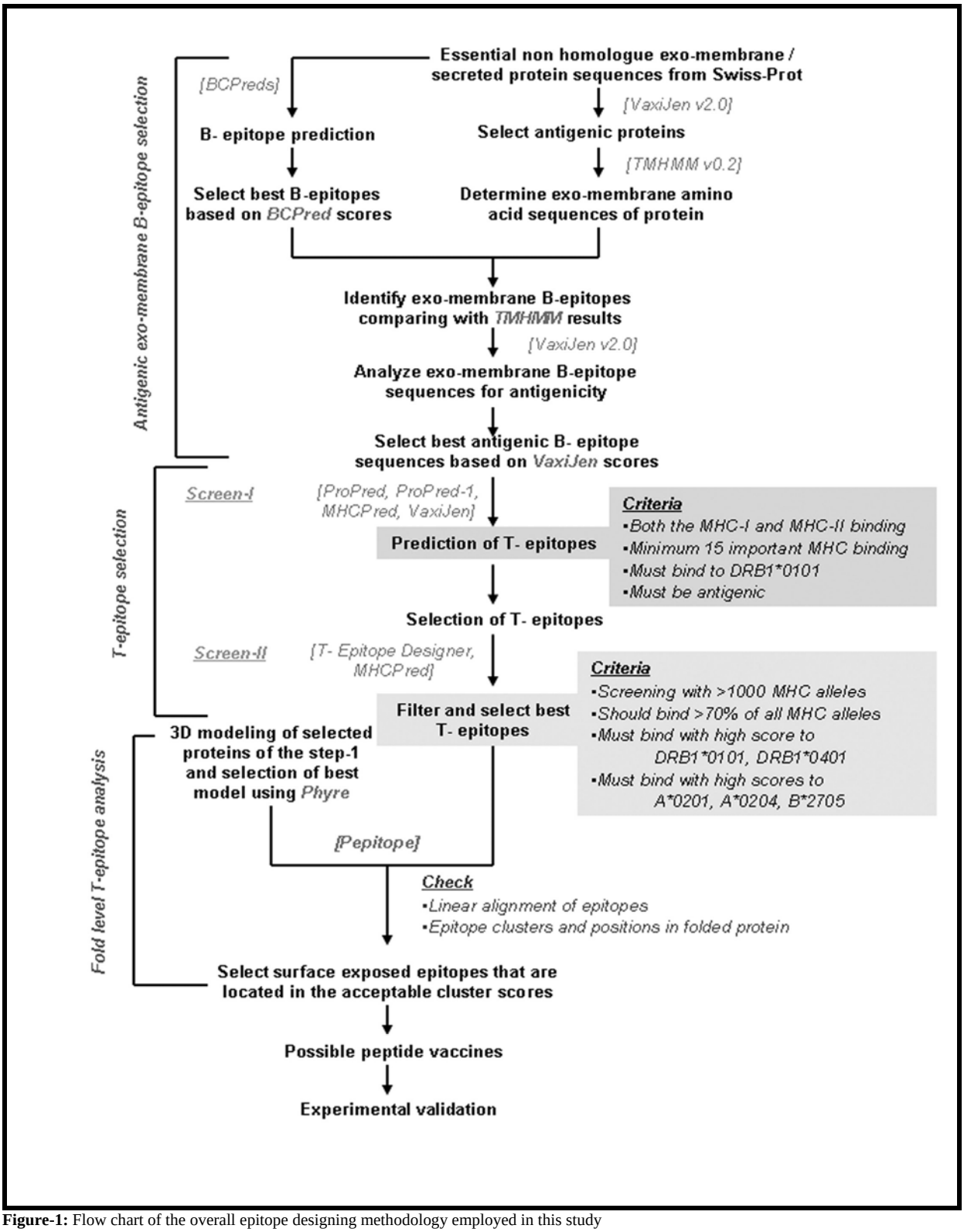

Figure-1: Flow chart of the overall epitope designing methodology employed in this study

Bioinformation **Volume 5**
www.bioinformation.net Issue 2
open access
Hypothesis

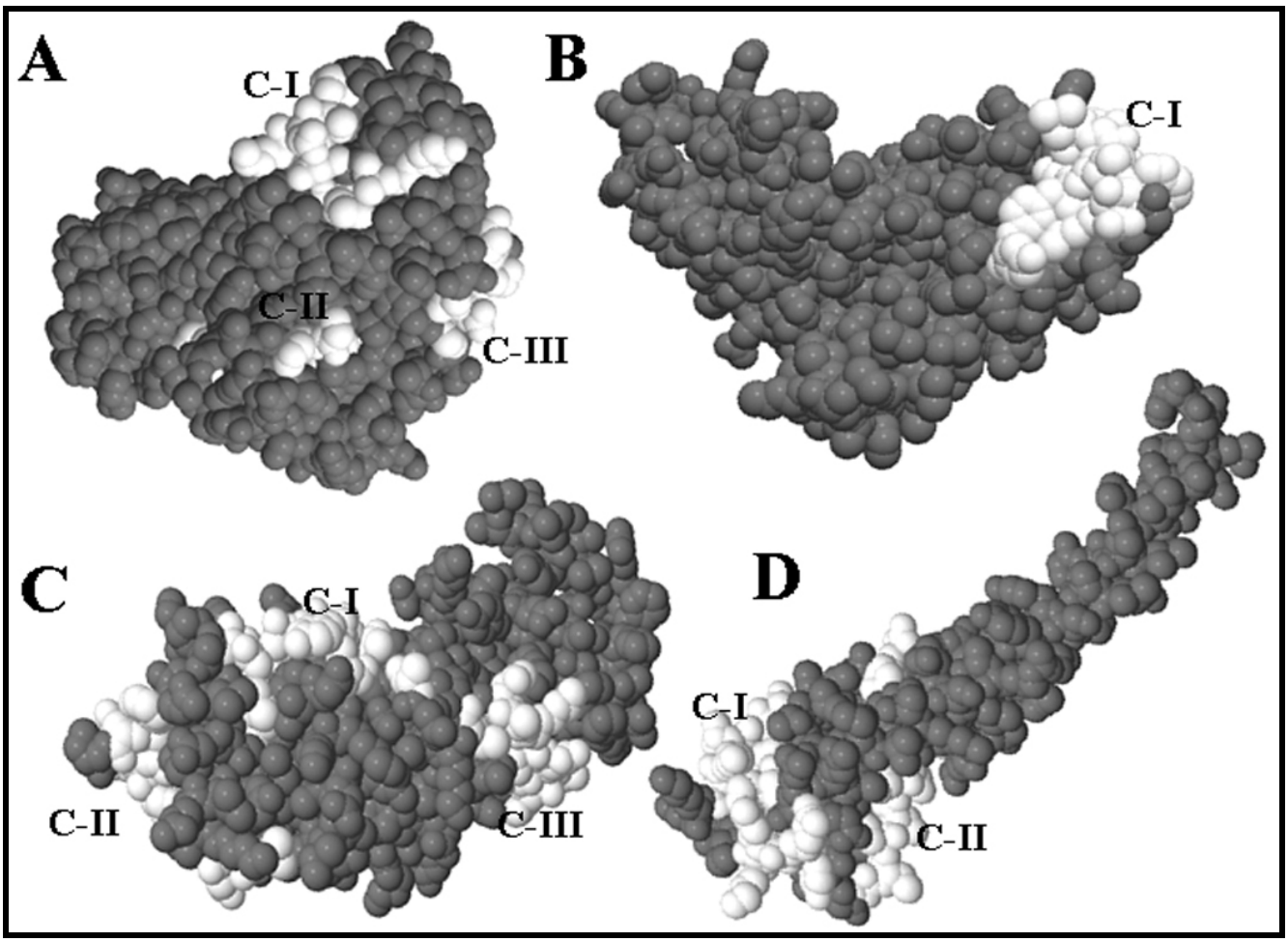

Figure-2: Fold level characterization of cluster and topoloty of best epitopes using Pepitope (in white). Cluster numbers are also represented. A) ddl, B) cysW, C) ComL, and D) pilV.

Discussion:

Bioinformatics approaches have been successfully used for identification of vaccine candidates in several pathogenic bacteria viz. *Bacillus anthracis* **[33]**, *Helicobacter pylori* **[34]** and computer assisted epitope designing and peptide vaccine have been reported for many bacterial and viral pathogens including HIV **[35]**, malaria **[36]**. In the present study, various bioinformatics tools have been used to identify and characterize potential epitopes from four vaccine targets of *N. gonorrhoeae* that might be effective to produce both the B-cell and T-cell mediated immunity. The identified epitopes are highly expected to bind both the class of MHC molecules specifically, A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401, that are most frequent MHC alleles in human population.

VaxiJen **[22]** that predicts antigenicity based on auto cross covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties with 82% accuracy, 91% sensitivity, and 72% specificity for bacterial species based on leave-one-out cross-validation (LOO-CV) was used in this study to identify antigenicity of proteins and peptides. Strategically, B-cell epitope identification can be adopted as first step in vaccine designing **[11]** and BCpreds **[24]** that uses a novel method of subsequence kernel was used to predict linear B-cell epitopes from each protein. A good peptide vaccine must be exposed to cell out side and to predict the exo-membrane topology of amino acid sequences, TMHMM v0.2 prediction server **[23]** was used. *In silico* quantitative structure-activity relationship (QSAR) **[37]** based T-cell epitopes have been reported recently. MHCPred **[28]** that uses additive method for binding affinity prediction of MHC molecules and the Transporter Associated with Processing (TAP) generates allele specific QSAR models using partial least squares (PLS) were used in the present study. MHCPred has been used to identify T-cell epitopes for malarial merozoite surface protein-1 **[36]** and a combination of SYFPEITHI, NetMHC and MHCPred has been used for Epstein-Barr virus latent membrane protein-2A **[38]**. Similarly, Propred and Propred-1 have been also used for the same purpose and these two server based T-cell epitope prediction have recently been reported for development of dengue virus vaccine **[39]**. To select MHC molecule binding peptides from a large number of MHC pool, we have also used T-Epitope Designer that predicts HLA-peptide binding calculated from structure based virtual binding pockets of HLA-peptide complexes **[29]**. We have also used two screening strategies to select T-epitopes in this study. The identified epitopes have been further characterized with Peptiope server **[31]** that predicts discontinuous epitopes based on a set of peptides that were affinity-selected against a monoclonal antibody or peptides extracted from a phage display library. It also aligns a linear peptide sequence onto a 3D protein structure. Therefore, pepitope is helpful to identify epitopes based on physicochemical properties and spatial or fold level organization.

Considering the highest number of MHC allele binding epitope at the first level of T-epitope screening, "LNSSNYTRA" from comL cluster-I (Score: 18.045, Residue No: 10) is found to bind total 31 MHC alleles. Although it showed an antigenicity score of 0.8605, because of its high IC_{50} value for DRB1*0101 (216.77 nM), it may not be considered as a good vaccine candidate. Another epitope "LNKLASQDW" from comL Cluster-III (Score: 10.899, Residue. No: 10) that binds total 18 MHC molecules has

antigenicity and IC_{50} scores 0.7699 and 36.39 nM, respectively. Therefore, this epitope is a better option than the previous one. The third epitope "DELNSSNYT" from comL cluster-I is found to bind total 11 MHC molecules but does not interact with any MHC-II alleles as per the Propred analysis. But the epitope has antigenic and IC_{50} (DRB1*0101) scores 1.0744 and 52.84 nM, respectively. Therefore, it is evident that epitope prediction differs according to different algorithms and the epitope "DELNSSNYT" may not be a good candidate for use as vaccine (**Table 4 see supplementary material**). Based on the second level screening, it has been found that epitope sequence "LNKLASQDW" can probably interact with all the HLA alleles (>1000) and specifically binds to our selected HLA molecules (A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401) with acceptable scores based on T-Epitope Designer, MHCPreds, and VaxiJen analysis **(Tables 4, 5, and 6 see supplementary material)**. Therefore this "LNKLASQDW" epitope is finally selected from comL.

Three epitopes from pilV are selected based on antigenic scores and IC_{50} values at screening-I. Epitopes "TCTVTLNDG" and "ETCTVTLND" of cluster-II (Score: 24.626, Residue. No: 13) showed antigenic scores 0.8669 and 1.1813, respectively. However, similar to the comL "DELNSSNYT" epitope, these two epitopes do not interact with any MHC-II alleles as per the Propred analysis but have IC_{50} values 5.66nM and 40.64nM, respectively for DRB1*0101 as analyzed with MHCpred. Therefore, these two epitopes may not be considered as vaccine candidates based on the same ground similar to comL. The third epitope "FEKYDSTKL" in cluster-I (Score: 43.692, Residue No: 17) is found to bind total 25 MHC molecules comprising of both the classes. The epitope has antigenicity and IC_{50} values 0.8655 and 8.73, respectively. Therefore, this epitope can be considered as best epitope among all epitopes derived form pilV (**Table 4 see supplementary material**). Results from the second level screening **(Tables 5 and 6 see supplementary material)** also support its usefulness to be potential peptide vaccine.

Although the full-length protein is non-antigenic, one epitope "WDLYLKSLS" from cysW Cluster-I (Score: 46.051, Residue No: 13) is found antigenic (VaxiJen score: 1.3624) and can bind total 20 MHC molecules comprising of both the MHC classes. The IC_{50} value of this epitope for DRB1*0101 is 11.69 nM which indicates good inhibition of a given biological process by half (**Table 4 see supplementary material**). The epitope has also been found to bind selected MHC molecules (A*0201, A*0204, B*2705, and DRB1*0401) and more than 80% HLA molecules selected for second level T-epitope screening in this study **(Tables 5 and 6 see supplementary material)**. Therefore this epitope can be considered for vaccine development.

Among the identified epitopes from ddl, "YGEDGAVQG" of cluster-III having VaxiJen and IC_{50} scores 2.6304 and 20.18, respectively binds total 20 MHC alleles comprising of both the classes. This epitope is found to be best epitope having highest antigenicity among all epitopes identified from four proteins under study. Another epitope "YAFDPKETP" from ddl is also found suitable that has antigenicity and IC_{50} scores 0.6362 and 51.29 nM, respectively. This epitope is located in cluster-I (Score: 12.871, Residue No: 9) and interacts with 13 MHC class-I and 8 MHC class-II alleles (**Table 4 see supplementary material**). However, in the second screening, "YGEDGAVQG" peptide fails to bind A*0201, and A*0204 with positive scores **(Table 6 see supplementary material)** and is also capable to bind only 70% HLA molecules **(Table 5 see supplementary material)** based on T-Epitope Designer analysis. Therefore this epitope may not be suitable for designing a vaccine against the pathogen.

Conclusion:

In this study, by using computational approaches based on sequence, structure, QSAR, simulation, and fold level analysis, we identified four potential T-epitopes derived from antigenic B-cell epitopes of four exo-membrane essential proteins of *N. gonorrhoeae*. Selected T-epitopes ["LNKLASQDW" from ComL., "FEKYDSTKL" from pilV, "WDLYLKSLS" from cysW, and "YAFDPKETP" from ddl] are antigenic and have much potential to interact with most common human HLA alleles (A*0201, A*0204, B*2705, DRB1*0101, and DRB1*0401). These epitopes are also found to interact with >75% of HLA molecules in a binding screening using T-Epitope Designer (that contains >1000 HLA molecules). Therefore these selected epitopes are likely to induce both the B-cell and T-cell mediated immune responses. Homology and simulation results also support the suitability of these epitopes as vaccine candidates. However, there are several pitfalls in developing a good vaccine and moreover there is lack of proper experimental disease model for gonorrhea; suitable animal model should be used for experimental validation of these epitopes to confirm these *in silico* results.

Financial disclosure:

This work was carried out without any grant or financial support. There is no conflict of interest regarding this work.

Acknowledgement:

We duly acknowledge the motivation and encouragement of all IIOAB members throughout the study

Author contributions:

DB designed the study and analyzed data and wrote the paper. ANM, AK and VA provided input and reviewed analysis.

References:

[1] K Ljungberg *et al. Virolog*, (2002) **302:** 44 [PMID: 12429515]
[2] TW Clark *et al. N Engl J Med* (2009) (in press)
[3] W Zhu *et al. Vaccine* (2004) **22**: 660 [PMID: 14741158]
[4] CT Barbosa-Cesnik *et al. Genitourin. Med* (1997) **73:** 336 [PMID: 9534740]
[5] MP Bos *et al. Infect Immun* (2005) **73:** 2222 [PMID: 15784566]
[6] GA Price *et al. Infect Immun* (2005) **73:** 3945 [PMID: 15972481]
[7] CE Thomas *et al. Infect Immun* (2006) **74:** 1612 [PMID: 16495532]
[8] GA Price *et al. Vaccine* (2007) **25:** 7247 [PMID: 17720283]
[9] S Gulati *et al. Int Rev Immunol* (2001) **20:** 229 [PMID: 11878767]
[10] AM Cerdino-tarraga *et al. Nucleic Acids Res* (2003) **31:** 6516 [PMID: 14602910]
[11] JE Larsen *et al. Immunome Res* (2006) **2:** 2 [PMID: 16635264]
[12] JJ Kuhns *et al. J Biol Chem*, (1999) **274:** 36422 [PMID: 10593938]
[13] C Watts *Immunol*, (1997) **15:** 821 [PMID: 9143708]
[14] RN Germain *Cell* (1994) **76:** 287 [PMID: 8293464]
[15] J Robinson *et al. Nucleic Acids Res* (2009) **37:** D1013 [PMID: 18838392]
[16] M Panigada *et al. Infection and immunity* (2002) **70:** 79 [PMID: 11748166]
[17] AO Anzala *et al. J Infect Dis* (2000) **182:** 459 [PMID: 10915076]
[18] R Kaul *et al. J Infect Dis* (2002) **185:** 1525 [PMID: 11992292]
[19] LJ Plant *et al. Infect Immun* (2006) **74**: 442 [PMID: 16369000]
[20] B Zhao *et al. Int. J. Integrative Bi*, (2007) **2**: 127 [Google Scholar]
[21] D Barh & A Kumar *In silico Bio.*, (2009) **9:** 0019 [PMID: 20109152]
[22] IA Doytchinova & DR Flower *BMC Bioinformatics* (2007) **8:** 4 [PMID: 17207271]
[23] A. Krogh *et al. J Mol Biol* (2001) **305:** 567-580 [PMID: 11152613]
[24] Y EL-Manzalawy *et al. J Mol Recogni*, (2008) **21:** 243 [PMID: 18496882]
[25] H Singh & GP Raghava *Bioinformatics* (2003) **19:** 1009 [PMID: 12761064]
[26] H Singh & GP Raghava *Bioinformatics* (2001) **12:** 1236 [PMID: 11751237]
[27] T Sturniolo *et al. Nat Biotechnology* (1999) **17:** 555 [PMID:10385319]
[28] P Guan *et al. Nucleic Acids Res* (2003) **31:** 3621 [PMID: 12824380]
[29] P Kangueane *et al. Bioinformation* (2005) **1:** 21 [PMID: 17597847]
[30] LA Kelley & MJ Sternberg *Nat Protoc* (2009) **4:** 363 [PMID: 19247286]
[31] I. Mayrose *et al. Bioinformatics* (2007) **23:** 3244 [PMID: 17977889]
[32] BR Eapen *Nature* (2009) hdl:10101/npre.2931.1
[33] N Ariel *et al. Infect Imm* (2003) **71:** 4563 [PMID: 12874336]
[34] A Dutta *et al. In Silico Biol* (2006) **6:** 43 [PMID: 16789912]
[35] AS De Groot *et al. Vaccine* (2005) **23:** 2136 [PMID: 15755584]

[36] V Wiwanitkit *J Microbiol Immunol Infec* (2009) **42:** 19 [PMID: 19424554]
[37] F Tian *et al. Amino Acids* (2009) **36:** 535 [PMID: 18575802]
[38] BWang *et al. Cell Mol Immunol* (2009) **6:** 97 [PMID: 19403058]
[39] P Somvanshi & P K Seth *Indian Journal of Biotechnology* (2009) **8:** 193 [Google Scholar]
[40] A Kondo *et al. Immunogenetics* (1997) **45:** 249 [PMID: 9002445]
[41] JE Beaver *et al.* Immunome Res (2007) **3:** 3 [PMID: 17313688]

Edited by P. Kangueane
Citation: Barh *et al.* Bioinformation 5(2): 77-82 (2010)

Bioinformation Volume 5
www.bioinformation.net Issue 2
open access
Hypothesis

Supplementary material:

Table 1: Reported T-cell epitopes in *Neisseria gonorrhoeae*.

Proteins	Epitope Sequence	Amino acid positions	MHC Binding	Reference
Lactoferrin receptor precursor	ASADKPYSY	496-504	HLA-A*0101	**[40]**
	FDEKNQDKY	567-575	HLA-A*0101	**[40]**
	IRDMTEKQY	374-382	HLA-A*0101	**[40]**
	LTEIDIRDY	751-759	HLA-A*0101	**[40]**
	RLDAFRQTY	248-256	HLA-A*0101	**[40]**
	RPDLSLRSY	806-814	HLA-A*0101	**[40]**
Carbamoyl-phosphate synthase large chain	DPEMADVTY	57-65	HLA-A*0101	**[40]**
	HSAEALQKY	723-731	HLA-A*0101	**[40]**
	EKEVVPDFY	874-882	HLA-A*0101	**[40]**
IgA-specific serine endopeptidase precursor	GKDLYYKNY	581-589	HLA-A*0101	**[40]**
	GTYDYWAGY	296-304	HLA-A*0101	**[40]**
	YDEDDYSYY	563-571	HLA-A*0101	**[40]**
	KAEVHTFYY	103-111	HLA-A*0101	**[40]**
	KSWQEWNIY	307-315	HLA-A*0101	**[40]**
	LSGNGHFHY	884-892	HLA-A*0101	**[40]**
	LSQDALTNY	264-272	HLA-A*0101	**[40]**
	LTAGLDFAY	1442-1450	HLA-A*0101	**[40]**
	NTFVQANLY	1347-1355	HLA-A*0101	**[40]**
	QNDYDEDDY	560-568	HLA-A*0101	**[40]**
	LRDLSQAYRY	215-224	Mamu-B*17	**[40]**
PilC2	GTATYLPPY	105-113	HLA-A*0101	**[40]**
	KTQNGKKQY	493-501	HLA-A*0101	**[40]**
	QRESTAMAY	252-260	HLA-A*0101	**[40]**
	RSYNLKLSY	555-563	HLA-A*0101	**[40]**
	TTRPGLAGY	136-144	HLA-A*0101	**[40]**
DNA gyrase subunit A	VREFPEDQY	648-656	HLA-A*0101	**[40]**
	VREFPEDQY	648-656	Mamu-B*17	**[41]**
Pilus-associated protein	QRESTAMAY	253-261	Mamu-B*17	**[41]**
	QRESTAMAYY	253-262	Mamu-B*17	**[41]**
50S ribosomal protein L6	GSDKQVVGQ	135-143	HLA-A*0101, HLA-A*0201, HLA-A*0301, HLA-A*1101, HLA-A*2403,HLA-A*2601, HLA-A*3101,HLA-A*6901, HLA-B*0702,HLA-B*0801, HLA-B*1501,HLA-B*2705 HLA-B*4001,HLA-B*5801 HLA-B*4402	**[41]**

Table-2: B-cell epitopes from full length proteins using BCPred (BCPred +AAP). Antigenicity of full length proteins as-well-as all B-cell epitopes derived from each protein was calculated using VaxiJen. TMHMM based surface exposed amino acid sequence of the proteins are given. Selected B-cell epitopes are in bold letters.

Name of protein/ VaxiJen scores	BCPred epitope sequences	Amino acid Positions	BCPred scores	VaxiJen scores for B-cell epitopes	TMHMM based surface exposed amino acid sequence
	BCPred Predictions				
D-alanine--D-alanine ligase (ddl)	KSKGIDAYAFDPKETPLSEL	33	0.961	0.8035	
	NGKGLPGIHIIPATEFYDYE	189	0.938	0.2560	
	DDTIYQCPSEDLTEAEESLM	214	0.893	-0.1183	
	LPVPEFAVLYDDTDFDAVEE	111	0.878	0.4969	
	QTAFNILHGTYGEDGAVQGA	**58**	**0.839**	**1.0836**	
VaxiJen score: 0.5155	KVKEKGRLKSVYEELKHLQG	151	0.834	0.2792	1 — 304
	AAP Predictions				

	VLNGKGLPGIHIIPATEFYD	187	1	0.4356	
	GLPVPEFAVLYDDTDFDAVE	110	1	0.6695	
	YGEDGAVQGALELLGIPYTG	68	1	0.7979	
	LLEINTLPGMTGHSLVPKSA	264	1	0.7017	
	NALKSKGIDAYAFDPKETPL	**30**	**1**	**0.9018**	
	BCPred Predictions				
Sulfate transport permease protein C (cysW)	MKPYSANPNLTEPRRLRMLL	1	0.892	0.6722	
	AAREIIPLMQTQGDSEEQAA	159	0.815	0.6665	
	AAP Predictions				42 — 66
	GAVSVVSGHIRGETNTIPLL	217	1	0.8433	
VaxiJen score: 0.6091	FLGKQLLTTLLDLPFSVSPV	94	1	0.4710	129 — 137
	TQGDSEEQAALVLGASGWQM	169	1	0.9722	
	WSAVKLTLITALIVVPVNAV	61	1	0.5057	
	FYEALKGGWDLYLKSLSDPE	**40**	**0.056**	**0.6316**	
	BCPred Predictions				
	ANRAKKIIGSYQNTRYVEES	198	0.979	-0.0246	
Competence lipoprotein (comL)	**VEKLYAEAQDELNSSNYTRA**	35	**0.936**	**0.5865**	
	RFRRLHPQHPNMDYALYLRG	97	0.825	0.7576	
	QSQLDTAYAYYKDDEKDKAL	73	0.763	0.3129	
	AAP Predictions				1 — 267
	LETNFPKSPFLTHAWQPDDM	241	1	0.1626	
VaxiJen score: 0.3968	**LNKLASQDWSDRDPKANREA**	127	**1**	**0.9646**	
	EKLYAEAQDELNSSNYTRAV	36	1	0.5626	
	RQSQLDTAYAYYKDDEKDKA	72	0.955	0.4956	
	BCPred Predictions				
Type IV pilin protein (pilV)	QAAPDTATNEGETCTVTLND	93	0.987	1.4611	
	QNLERYYRQKGTFEKYDSTK	52	0.866	0.1548	30 — 129
	AAP Predictions				
VaxiJen score: 0.5312	**APDTATNEGETCTVTLNDGG**	95	**1**	**1.4930**	
	YYRQKGTFEKYDSTKLKQNK	57	**1**	**0.5050**	

Table-3: Selected B-cell epitopes using both the modules of BCPreds (BCPred +AAP), TMHMM based topology, and antigenicity using VaxiJen.

Protein name	SL No. of B-cell epitopes	Amino acid position of epitope	BCPred epitope sequence	BCPred scores	VaxiJen scores
ddl	1	58	QTAFNILHGTYGEDGAVQGA	0.839	1.0836
	2	30	NALKSKGIDAYAFDPKETPL	1	0.9018
cysW	1	40	FYEALKGGWDLYLKSLSDPE	0.056	0.6316
comL	1	35	VEKLYAEAQDELNSSNYTRA	0.936	0.5865
	2	127	LNKLASQDWSDRDPKANREA	1	0.9646
pilV	1	95	APDTATNEGETCTVTLNDGG	1	1.4930
	2	57	YYRQKGTFEKYDSTKLKQNK	1	0.5050

Table-4: Common epitopes from each protein that can produce both the B- and T-cell mediated immunity are representes along with their various parameters. Epitopes selected as vaccine candidates are highlight in bold letters.

Proteins and Uniprot ID	SL No. of B-cell epitopes	Predicted epitopes (Propreds + MHCPred)	Vacxigen score	Amino acid position	IC_{50} value of epitopes for DRB1*0101 (MHCPred)	Number of MHC Class I binding alleles (Propred1)	Number of MHC Class II binding alleles (Propred)	Total number of MHC binding alleles	*Cluster scores (Pepitope)*
ddl	1	**YGEDGAVQG**	**2.6304**	**68**	**20.18**	**17**	**3**	**20**	*Cluster-I Score: 12.871/Res. No: 9*
Q5F6M0		FNILHGTYG	0.3382	61	323.59	0	25	25	
	2	**YAFDPKETP**	**0.6362**	**40**	**51.29**	**13**	**8**	**21**	***YAFDPKETP*** *Cluster-II Score: 8.4077/Res. No: 8* *FNILHGTYG Cluster-III Score: 7.483/Res. No:9* *YGEDGAVQG*
cysW	1	LYLKSLSDP	-0.6286	50	15.56	24	18	42	*Cluster-I Score: 46.051/Res. No:13*
Q5F9F0		**WDLYLKSLS**	**1.3624**	**48**	**11.69**	**2**	**18**	**20**	
		YLKSLSDPE	-0.5090	51	1.05	33	7	40	

									LYLKSLSDP
									WDLYLKSLS
									YLKSLSDPE
	1	**LNSSNYTRA**	**0.8605**	**46**	**216.77**	**27**	**4**	**31**	*Cluster-I*
ComL		VEKLYAEAQ	0.0915	35	156.31	0	11	11	*Score: 18.045/*
		DELNSSNYT	**1.0744**	**44**	**52.48**	**11**	**0**	**11**	*Res. No: 10*
									LNSSNYTRA
Q5F9W0									***DELNSSNYT***
									Cluster-II
	2	LASQDWSDR	0.1974	130	21.68	8	0	8	*Score: 11.057/*
		LNKLASQDW	**0.7699**	**127**	**36.39**	**5**	**13**	**18**	*Res. No: 9*
									LNKLASQDW
									Cluster-III
									Score: 10.899/
									Res. No: 10
									LASQDWSDR
pilV	1	**TCTVTLNDG**	**0.8669**	**105**	**5.66**	**15**	**0**	**15**	*Cluster-I*
		ETCTVTLND	**1.1813**	**104**	**40.64**	**10**	**0**	**10**	*Score: 43.692/*
									Res. No: 17
Q5F6V4									*YYRQKGTFE*
	2	*YYRQKGTFE*	*0.1821*	*57*	*12.19*	*4*	*5*	*9*	***FEKYDSTKL***
		FEKYDSTKL	***0.8655***	***64***	***8.73***	***15***	*10*	***25***	*YRQKGTFEK*
		YRQKGTFEK	*0.1519*	*58*	*175.79*	*7*	*43*	*50*	*Cluster-II*
									Score: 24.626/
									Res. No: 13
									TCTVTLNDG
									ETCTVTLND

Table-5: Screening of T-epitopes using T-epitope Designer.

Proteins	Epitopes	T-Epitope Designer (NO of binders)	Comments	Lowest score	Highest score
ddl	**YAFDPKETP**	99.99%	Almost all alleles	A*2434 [424.64]	A*2607 [2570.92]
	YGEDGAVQG	70%	Mostly positive for almost all B* & C*	A*0234 [41.45]	B*4003 [979.16]
cysW	**WDLYLKSLS**	80%	Mostly positive for almost all B* & C*	A*0221 [10.19]	C*0705 [2197.53]
ComL	**LNKLASQDW**	100%	All alleles	A*0101 [236.75]	C*1802 [2340.96]
pilV	**FEKYDSTKL**	100%	All alleles	A*3306 [170.86]	C*1801 [3465.96]

Table-6: Final selection of T-epitopes from four *N. gonorrhoeae* proteins

Proteins	Epitopes	T-Epitope Designer			MHCPred (IC_{50} Value)	
		A*0201	A*0204	B*2705	DRB1*0101	DRB1*0401
ddl	**YAFDPKETP**	1866.30	1464.79	2261.31	51.29	592.93
	YGEDGAVQG	-103.79	-309.46	714.11	20.18	690.24
cysW	**WDLYLKSLS**	758.14	641.45	1401.35	11.69	223.36
ComL	**LNKLASQDW**	788.83	508.12	1189.16	36.39	959.40
pilV	**FEKYDSTKL**	714.75	248.08	1597.48	8.73	462.38

Table-7: Homology modeling of selected proteins using Phyre. Selected template, E-values, fold descriptor, and super family for each protein are given.

Proteins	SCOP Code/ Protein PDB ID	E-value	Fold/PDB descriptor	Super family
Q5F6M0 **ddl**	c1iowA_	2.1e-32	Ligase	d-ala:d-ala ligase
Q5F9F0 **cysW**	c2onkI_	4.7e-27	Mmembrane protein	Molybdate/tungstate abc transporter permease
Q5F9W0 **ComL**	c2q7fB_	4.2e-31	Protein binding	yrrb protein
Q5F6V4 **pilV**	d2pila_	2.4e-11	Pili subunits	Type IV Pilin Pak from Pseudomonas aeruginosa