Journal of Computer Science & Systems Biology

The International Open Access Journal of Systems Biology and Applied Computer Science

Subtractive Genomics Approach for *in Silico* Identification and Characterization of Novel Drug Targets in *Neisseria Meningitides Serogroup B*

Aditya Narayan Sarangi[1]*, Rakesh Aggarwal[1], Qamar Rahman[2], Nidhi Trivedi[1]

[1]Sanjay Gandhi Postgraduate Institute of Medical Sciences,
Lucknow, Uttar Pradesh, INDIA

[2]Integral University, Lucknow, Uttar Pradesh, INDIA

Abstract

Meningococcal disease is a life-threatening illness with annual incidence rates varying from 1 to 1000 per 100 000 persons in different parts of the world. Effective polysaccharide and polysaccharide-protein conjugate vaccines that offer protection against infection with meningococcal serogroups A, C, Y and W-135 have been licensed and are available worldwide. Serogroup B remains the most prevalent cause of meningococcal disease responsible for 32% of all meningococcal disease in the United States, 45 to 80% of the cases in Europe, and for the majority of cases in the rest of the world. The development of a vaccine against serogroup B poses the biggest problem due to the similarity between the B capsular polysaccharide structure and a polysialic acid containing glycopeptides that are a part of human brain tissue. Prevention of meningococcal disease will require the development of an effective vaccine to combat serogroup B, which is the cause of most meningococcal cases in developed countries. The availability of the complete sequence information of *Neisseria meningitides* serogroup B proteome has made it possible to carry out the *in silico* analysis of its genome for identification of potential vaccine and drug targets. Our study revealed 1413 proteins which are non-homologous to human genome. Screening these proteins using the Database of Essential Genes (DEG) resulted in the identification of 362 proteins as essential proteins of the bacterium. Analysis of the identified essential proteins, using the KEGG Automated Annotation Server (KAAS) housed at Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database, revealed 35 enzymes of *N. Meningitides* that may be used as potential drug targets, as they belongs to pathways present only in the bacterium and not present in humans. Subcelluler localization prediction of these essential proteins revealed that 9 proteins lie on the outer membrane of the pathogen which could be potential vaccine targets. Screening of the functional inhibitors against these novel targets may result in discovery of novel therapeutic compounds that can be effective against *Neisseria meningitides* Serogroup B.

Keywords: *Neisseria meningitides*; Essential gene; Drug targets; Vaccine candidates; Subtractive genomics; Pathway analysis

Introduction

Despite rapid advances in the diagnosis of bacterial infections and the availability of effective antibiotics, meningococcal disease continues to represent a substantial public health problem for most countries (Tzeng et al., 2000). *N. meningitides* is a Gram-negative bacterium that colonizes asymptomatically the upper nasopharinx tract of about 5–15% of human population. The bacterium can traverse the epithelium and reach the blood stream causing septicemia. From the blood meningococcus is able to cross the blood brain barrier and infect the meninges, causing meningitis potentially leading to shock and death (van et al., 2000; Stephens et al., 2007). 95% of total cases of meningitis are caused by five major serogroups: A, B, C, Y and W135. Vaccines against serogroups A, C, Y and W135 were developed in the 1960s by using the purified capsular polysaccharide as antigen. In the last 40 years a lot of efforts have been directed to the identification of meningococcus B antigens as basis of new vaccines. However, the high variability of these proteins among the different MenB strains represents a serious obstacle to the production of a globally effective anti-MenB vaccine (Stephens et al., 2007).

Till date the complete genome sequence of about 863 bacteria has been determined and 1653 bacterial genome projects are currently in progress. Availability of genome sequences of pathogens has provided a tremendous amount of information that can be useful in drug target and vaccine target identification. Any gene vital to sustain life and cellular growth of pathogenic organism may be a possible drug target, particularly when the sequence of that gene has little or no similarity to those of the host organism. The differences in the proteins of the host and the pathogen can be effectively used for designing a drug specifically targeting the pathogen. The

application of *in silico* subtractive genomics technology significantly shortens the time required both for identification of antigens and their utility as protective immunogen. Subtractive genomics help enables the subtraction dataset between the host and pathogen genome and provides information for a set of genes that are likely to be essential to the pathogen but absent in the host. This approach has been used successfully to identify essential genes in *Pseudomonas aeruginosa* (Sakharkar et al., 2004; Perumal et al., 2007). Same methodology was implemented by us for the analysis of the whole proteome sequence of the human pathogen *Neisseria meningitides* Serogroup B.

The current studies makes use of the subtractive genomics approach, database of essential gene (DEG), differential pathway analysis and sub cellular localization prediction to analyze the complete proteome of *N. meningitides* serogroup B to search for potential drug and vaccine targets.

Methodology

The essential proteins in *N. meningitides* serogroup B were identified using subtractive genomics approach and further analyzed for identification of potential drug targets. The flow chart describing the detailed methodology for identifying of pathogen specific essential proteins using subtractive genomics approach is mentioned in figure 1.

Retrieval of Proteomes of Host and Pathogen

The complete proteome of *N. meningitides* serogroup B was retrieved from SwissProt (Bairoch et al., 2000) and complete *Homo sapiens* proteome was downloaded from NCBI (Maglott et al., 2007). Prokaryotic essential protein sequences were retrieved from the database of essential genes (DEG) (Zhang et al., 2004) from its location http://tubic.tju.edu.cn/deg/ by using an in-house designed perl script.

Identification of Essential Proteins in *N. Meningitidis*

N. Meningitides Serogroup B proteome was purged at 60% threshold using CD-HIT (Li et al., 2006) to exclude paralogs or duplicates for further analysis. The set of proteins obtained were subjected to BlastP against *Homo sapiens* proteome with the expectation value (E-value) cutoff of 10^{-4}. An in-house designed perl script was used to retrieve the protein sequences of *N. Meningitides* serogroup B showing no significant similarity with *Homo sapiens* proteome. BlastP analysis was performed for the non homologous protein sequences of *N. meningitides* serogroup B against database of essential proteins retrieved from DEG with E-value cutoff score of 10^{-10} and 30% identity. A minimum bit-score cut-off of 100 was used to screen out proteins that appeared to represent essential genes. The protein sequences obtained are non homologous essential proteins of *N. meningitides* serogroup B.

Stand-alone BLAST software (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/blast-2.2.21-ia32-win32.exe) was used for the BlastP analysis. E-value cutoff threshold for BlastP and DEG search were set by using the command line parameters option in stand-alone BLAST.

Functional Classification of the Uncharacterized Essential Proteins

Functional family prediction of the putative uncharacterized essential proteins was done by using the SVMProt web server (http:/
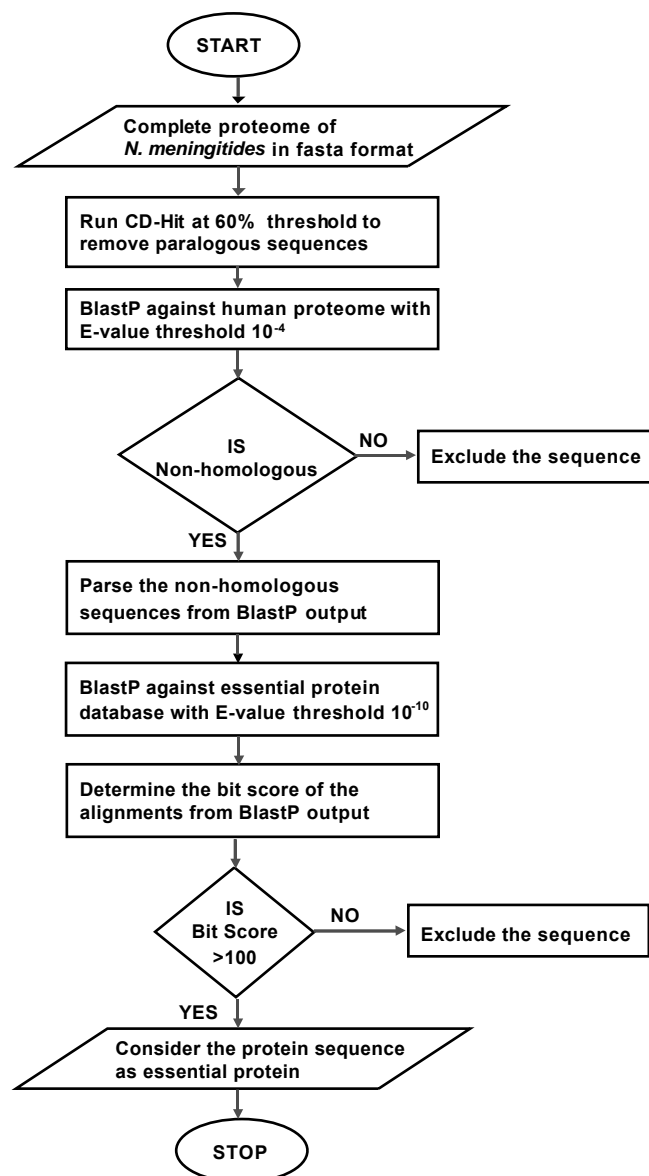

Figure 1: Flowchart for identification of essential proteins in *N. meningitides* Serogroup B.

/jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) (Cai et al., 2003). SVMProt utilizes Support Vector Machine for classification of a protein into functional family from its primary sequence.

Metabolic Pathway Analysis

Metabolic pathway analysis of the essential proteins of N. *meningitides* was done by KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007). Comparative analysis of the metabolic pathways of the host and pathogen was performed by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa et al., 2000) to trace out essential proteins involved in pathogen specific metabolic pathways for the identification of potential drug targets.

Sub Cellular Localization Prediction

Sub-cellular localization analysis of the essential proteins has been done by Proteome Analyst Specialized Sub Cellular Localization Server v2.5 (PA-SUB) (Lu et al., 2004) to identify the sur-

face membrane proteins which could be probable vaccine targets.

Results and Discussion

N. meningitides serogroup B proteome was found to contain 2001 proteins. Out of these, 1939 non-paralogs sequences were retrieved by using CD-HIT program at 60% identity. All the proteins of *N. meningitides* that are non-homologous to the human proteome could not be taken directly as targets as these also include a large number of proteins which are not essential for the viability of the organism. Therefore, out of 1939 proteins, 1413 proteins were identified as non-homologues to the human proteome and were subjected to BLASTP against DEG with E-value cutoff score of 10^{-10} and 30% identity. 362 proteins comprising 18% of the total number of protein coding sequences in *N. meningitides* serogroup B found to be essential. Detailed analysis of *N. meningitides* serogroup B proteome is mentioned in supplementary table 1.

Functional classification of the 36 putative uncharacterized essential proteins was performed by using the SVMProt web server based on P value, which is the expected classification accuracy in terms of percentage. 10 proteins were classified as transmembrane proteins , 6 as zinc binding, 4 as lipid-binding, 4 as iron binding, 3 as transferases, 1 as manganese binding, 1 as oxidoreductases, 1 as carbon-carbon lyases, 1 as carbon-sulfur Lyases, 1 as electrochemical potential driven Transporter, 1 as DNA-binding protein (Supplementary Table 2) . The predicted transmembrane proteins and transporter protein may be considered as potential drug targets.

The sub cellular localization prediction for the 362 essential proteins of *N. meningitides* was done by using PA-SUB server to locate the outer membrane proteins which could be probable vaccine targets. 9 outer membrane proteins and 10 periplasmic proteins were identified, which are listed in Supplementary Table 3.

Metabolic pathway analysis of the essential proteins of *N. meningitides* was done by KEGG Automatic Annotation Server (KAAS). Result of comparative analysis of the metabolic pathways of the host and pathogen by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database reveals 11 pathways are unique to *N. meningitides* serogroup B (supplementary table 4). Thereafter, each selected pathway was screened for the presence of pathogen specific essential enzymes. The C5-branched dibasic acid metabolism pathway did not show involvement of any meningococcal essential protein. As shown in supplementary table 4, 33 non-human homologous essential proteins identified from the remaining 10 pathogen specific pathways. This small group of proteins is required to be further verified for their role in meningococcal survival and virulence by mutagenesis study.

In the present study, phosphoenolpyruvate carboxylase (ppc) has been identified as a possible drug target due to its involvement in carbon fixation in photosynthetic organisms, reductive carboxylate cycle and pyruvate metabolism.

Lipopolysaccharides (LPS) and peptidoglycans are the main constituents of the outer cell wall of gram negative bacteria and play an important role in pathogenesis. Out of the thirteen enzymes involved in LPS biosynthetic pathways, eleven enzymes found to be essential for the viability of bacteria, were identified as drug targets as these did not show homology to any human gene. Two proteins (lipid A biosynthesis lauroyl acyl transferase and D-beta-D-heptose 7-phosphate kinase) are not essential meningococcal enzymes.

Type II secretion pathway supports the translocation of proteins associated with the virulence factors, across the outer membrane. Five virulent factors i.e type IV pilus assembly protein (pilF), pilus assembly protein (pilG), twitching motility protein1 (pilT-1), twitching motility protein2 (pilT-2), competence protein (comA) are found to be involved in pathogen specific Type II secretion pathway. Type IV pilus assembly proteins are the major components in type II secretion pathway. Type IV pili are associated with bacterial adhesion, aggregation, invasion, host cell signaling, surface motility, and natural transformation (Rothbard et al., 1985).In this study, it has been found that type IV pilus assembly protein (PilF) may be considered for effective anti-meningococcal drug development.

The bacterial two component system is crucial for the growth and survival under extreme conditions. Eight essential enzymes found to be potential drug targets in this pathway. Among those identified enzymes, tryptophan synthase subunit a (trpA), tryptophan synthase subunit b (trpB), indole-3-glycerol-phosphate synthase (trpC), anthranilate phosphoribosyltransferase (trpD) and anthranilate synthase component I (trpE), are also found to be key components in tyrosine, and tryptophan biosynthesis pathways. Therefore, targeting of these three enzymes may disrupt pathways essential for *N. meningitidis* survival and virulence and therefore might be a potential antibacterial therapeutic strategy.

Thus, further investigations on the predicted essential proteins are required to verify the reliability of the data. The complete list of the proteins identified via this *in silico* approach is available as supplementary data.

Conclusion

The large scale genome sequencing projects have increased the availability of completely sequenced genomic and proteomic data in public domain. Use of the DEG database is more efficient than conventional methods for identification of essential genes and facilitates the exploratory identification of the most relevant drug targets in the pathogen. The present study has thus led to the identification of several proteins that can be targeted for effective drug design and vaccine development against *N. meningitides* serogroup B. The drugs developed against these will be specific to the pathogen, and therefore less or not toxic to the host. Since the number of essential genes in the metabolic pathways of *N. meningitides* serogroup B, identified in the present study, is relatively small (only 33), these can be further characterized and their role in the survival of the bacteria can be verified. Homology modelling of these targets will help identify the best possible sites that can be targeted for drug design by simulation modelling. Virtual screening against these novel targets might be useful in the discovery of novel therapeutic compounds against *N. meningitides* serogroup B.

Acknowledgements

References

1. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28: 45-48. » CrossRef » Pubmed » Google Scholar

2. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: Web-Based Support Vector Machine Software for Functional Classification of a Protein from Its Primary Sequence. Nucleic Acids Res 31: 3692-3697. » CrossRef » Pubmed » Google Scholar

3. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30. » CrossRef » Pubmed » Google Scholar

4. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658-1659. » CrossRef » Pubmed » Google Scholar

5. Maglott D, Ostell J, Pruitt KD, Tatsuova T (2007) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 35: D26-D31. » CrossRef » Pubmed » Google Scholar

6. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35: W182-W185. » CrossRef » Pubmed » Google Scholar

7. Perumal D, Lim CS, Sakharkar KR, Sakharkar MK (2007) Differential genome analyses of metabolic enzymes in Pseudomonas aeruginosa for drug target identification. In Silico Biol 7: 0032. » CrossRef » Pubmed » Google Scholar

8. Rothbard JB, Fernandez R, Wang L, Teng NNH, Schoolnik GK (1985) Antibodies to peptides corresponding to a conserved sequence of gonococcal pilins block bacterial adhesion. Proc Natl Acad Sci USA 82: 915-919. » CrossRef » Pubmed » Google Scholar

9. Sakharkar KR, Sakharkar MK, Chow VTK (2004) A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol 4: 0028. » CrossRef » Pubmed » Google Scholar

10. Stephens DS (2007) Conquering the meningococcus. FEMS Microbiol Rev 31: 3-14. » CrossRef » Pubmed » Google Scholar

11. Stephens DS, Greenwood B, Brandtzaeg P (2007) Epidemic meningitis, meningococcaemia, and Neisseria meningitidis. Lancet 369: 2196-210. » CrossRef » Pubmed » Google Scholar

12. Tzeng Y, Stephens DS (2000) Epidemiology and pathogenesis of Neisseria meningitidis. Microbes. Infect 2: 687- 700. » CrossRef » Pubmed » Google Scholar

13. van Deuren M, Brandtzaeg P, van der Meer JW (2000) Update on meningococcal disease with emphasis on pathogenesis and clinical management. Clin Microbiol Rev 13: 144-166 . » CrossRef » Pubmed » Google Scholar

14. Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, et al. (2004) Predicting Subcellular Localization of Proteins using Machine-Learned Classifiers. Bioinformatics 20: 547-556. » CrossRef » Pubmed » Google Scholar

15. Zhang R, Ou HY, Zhang CT (2004) DEG: A database of essential genes. Nucleic Acids Res 32: D271-D272. » CrossRef » Pubmed » Google Scholar