Functional Annotation and Classification of the Hypothetical Proteins of *Neisseria meningitidis* H44/76

Archana Singh[1], #, Bharti Singal[2], #, Onkar Nath[2], Indrakant Kumar Singh[2], *

[1]Department of Botany, Hans Raj College, University of Delhi, New Delhi, India
[2]Molecular Biology Research Laboratory, Department of Zoology, Deshbandhu College, (University of Delhi), Kalkaji, New Delhi India

Email address:
iksingh@db.du.ac.in (I. K. Singh)

Abstract: *Neisseria meningitidis* is a parasitic gram-negative bacterium of the family Neisseriaceae (Proteobacteria) and it causes many human diseases including meningitidis and septicemia. One of its strains, H44/76, has natural transformation capacity, thus it is important to identify possible novel drug targets and to develop serogroup B vaccines against this opportunist pathogen. In the complete genome of *N. meningitidis* strain H44/76, there are 1961 coding genes out of which 544 encodes for hypothetical proteins (HPs). Due to their less homology and relatedness to other known proteins, HPs may serve as potential drug targets. We performed extensive functional analysis of these HPs with the help of Bioinformatics tools and assigned functions to 235 HPs, out of which 202 were annotated with high confidence whereas 33 with less confidence. In this study, we have used a combination of latest tools to acquire information about the conserved regions, families, pathways, interactions, localization and virulence related to a particular protein. We also categorized these proteins as transporters, regulators, enzymes, binding proteins, virulent proteins. The outcome of this intensive study may help in the comprehensive understanding of pathogenesis, drug resistance, adaptability to host, epidemic causes and drug discovery for treatment of the diseases.

Keywords: *Neisseria meningitidis*, Hypothetical Proteins, Functional Annotation, Drug Targets

1. Introduction

N. meningitidis is a parasitic bacterium and it is an obligate nasopharyngeal human pathogen, which leads to severe diseases like septicemia and meningitis [1, 2]. Among children and infants, it can get its way to the brain by invading the respiratory epithelial tissues and then crossing the blood brain barrier. The common symptoms are high fever, lethargy, confusion, nausea, neck stiffness, vomiting, and petechial rash [3][4]. Surveillance is of utmost importance for getting better grasp of meningococcal diseases as they may lead to epidemics and outbreaks [5][6]. The strain H44/76 is very much related to strain MC58 that also belongs to serogroup B. There are 13 serogroups in which *N. meningitidis* can be classified on the basis of the immunological reactivity of their capsular polysaccharides out of which 5 (A, B, C, Y and W) are found to be the most common causes of diseases [3].

The efficiency of this strain to transform itself naturally in

a favorable manner makes it important for the development of the serogroup B vaccines [2]. The vaccine for serogroup B has been developed recently which has the potential to minimize mortality and morbidity associated to the diseases caused by serogroup B strains [7][8]. *N. meningitidis* has enormous capability to change its surface structures thus enabling it to escape the defense mechanism of the host.

The whole genome sequencing has taken its pace with the use of high-throughput techniques. It is becoming necessary to give meaningful direction to this web reservoir of genomic information. The computational sequence analysis tools are playing crucial role in annotating the novel genes. There is a huge repertoire of Hypothetical proteins (HPs, proteins which are derived from translating nucleic acid sequences and yet not characterized functionally and biochemically), which need to be identified to gain ample knowledge about the complete genomic and proteomic content of an organism [9]. The functional annotation of these HPs not only helps in understanding the unknown metabolic pathways in which

# These authors contributed equally to the paper.

they are involved, but also helps in identifying the unfamiliar functions of the previously annotated proteins. The functionally annotated HPs can be used as drug targets for novel drug discovery process. These HPs can also be used as potential biological markers [10, 11].

The genome size of H44/76 is 2.18 Mb with 2,480 reads [2]. There are 1961 proteins expressed under certain conditions out of which 544 are putative HPs. Here, in this study, analysis of the HPs found in *N. meningitidis* strain H44/76 was done using advance bioinformatics tools. A wide range of tools are used to predict physicochemical properties, subcellular localization, domains, motifs, presence of helices and family of the HPs. In strain H44/76, we have

successfully annotated 235 HPs. Our study has taken account of approximately 27% of the whole proteome of strain H44/76 of *N. meningitidis* which is remarkably the huge percentage of what is unknown. This shows that we cannot underestimate the annotation of HPs in an organism to identify probable targets for pharmacological studies.

2. Materials and Methods

The URLs to all tools/servers/databases that were used in the functional annotation of the HPs found in the Genome of *N. meningitidis* H44/76 from NCBI Genome Bio-project: PRJNA61079 given in Table 1.

Table 1. List of Tools, servers and databases used for the functional annotation of 544 HPs in N. meningitidis.

S.NO.	TOOL/ SERVERS/ DATABASES	URL
Sequence Homology Search		
1	BLAST: Basic Local Alignment Search Tool	http://www.ncbi.nlm.nih.gov/BLAST/
Physiochemical Characterization		
2	ExPASy – ProtParam tool	http://web.expasy.org/protparam/
Sub-cellular Localization		
3	PSORT B v3.0	http://www.psort.org/psortb/
4	PSLpred	http://www.imtech.res.in/raghava/pslpred/
5	CELLO	http://cello.life.nctu.edu.tw/
6	SignalP 4.1	http://www.cbs.dtu.dk/services/SignalP/
7	SecretomeP	http://www.cbs.dtu.dk/services/SecretomeP/
8	SOSUI	http://harrier.nagahama-i-bio.ac.jp/sosui/
9	HMMTOP	http://www.enzim.hu/hmmtop
10	TMHMM	http://www.cbs.dtu.dk/services/TMHMM/
Protein categorization		
11	SMART	http://smart.embl-heidelberg.de/
12	INTERPRO	http://www.ebi.ac.uk/interpro/
13	MOTIF	http://www.genome.jp/tools/motif/
14	CATH	http://www.cathdb.info/search/
15	SUPERFAMILY	http://supfam.cs.bris.ac.uk/SUPERFAMILY/
16	PANTHER	http://www.pantherdb.org/
17	Pfam	http://pfam.xfam.org/
18	Conserved domain database	http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
19	ScanProsite	http://prosite.expasy.org/scanprosite/
20	HAMAP	http://hamap.expasy.org/hamap_scan.html
21	ProtoNet	http://www.protonet.cs.huji.ac.il/index.php
22	SVMprot	http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi
Virulence Prediction		
23	VICMpred	http://www.imtech.res.in/raghava/vicmpred/
24	VIRULENTpred	http://203.92.44.117/virulent/
Protein-Protein Identification		
25	STRING 9.1	http://string-db.org/

2.1. Sequence Retrieval

We have found that out of 1961 proteins of this strain, 544 proteins have been designated as HPs but upon analysis of these proteins, it was observed that only 525 HPs are unique. So, further analysis was performed for these 525 unique HPs only. The Gene IDs for these 525 HPs of H44/76 were retrieved from the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome). These IDs were queried against UniProt database (http://www.uniprot.org/uniprot) to retrieve their UniProt

IDs, primary accession number and protein sequences in FASTA format.

2.2. Homology Search

The functional annotation of HPs can be performed by using homology search i.e. searching for functions by looking at the conserved regions in the sequences of various organisms. Here, we have used BLASTp for the sequence similarity in which protein sequences are queried against database to find out homology among various organisms [12]. BLASTp was queried at 0.005 e-value against non-

redundant (nr) protein sequences database. The homologs for each HP were analyzed and best hits were taken for the function prediction depending upon their percent query coverage and identity. The results are presented in Supplementary Table A.3.

2.3. Physicochemical Characterization

Physicochemical characterization was carried out by using ProtParam server on Expasy [13]. It was used to measure theoretical physicochemical properties such as Molecular weight (Mw), pI, Extinction coefficient (M^{-1} cm^{-1}), Instability index classification, aliphatic index and Grand Average of Hydropathicity (GRAVY). The results of the above analysis are summarized in Supplementary Table A.1.

2.4. Sub-cellular Localization

Protein sub-cellular localization prediction is important as it helps in understanding the location of the protein in the cell. This further aids in predicting the function of a particular protein in an organism. We can also utilize this information in refining the list of drug targets in a cell. Surface proteins can be of great importance for this purpose. The features present in the primary structure of the protein such as presence of trans-membrane helices or signal peptides determines the sub-cellular localization of the protein [13]. We have used PSORT B v3.0 [14], PSLpred [15], and CELLO [16] for sub-cellular localization prediction. SignalP 4.1 server was used for predicting the availability as well as location of signal peptides which was based on artificial neural network analysis [17]. SecretomeP was used to predict non-classical (not signal peptide triggered) proteins by considering post-translational and localization information of the sequence [18]. SOSUI [19] was used to classify protein into membrane and soluble proteins and also predicting the number of trans-membrane helices. HMMTOP [20] and TMHMM [21, 22] were used for predicting trans-membrane helices in the HPs. The analyzed results of above tools are presented in Supplementary Table T2.

2.5. Functional Prediction

We have used variety of tools in order to do precise and accurate functional prediction of 544 HPs. The databases and tools used for this purpose are SMART [23], INTERPRO [24], MOTIF [25- 28], CATH [29], SUPERFAMILY [30], PANTHER [31], Pfam [32], Conserved domain database [33, 34], ScanProsite [35], HAMAP [36], ProtoNet [37] and SVMprot [38]. SMART search was used to detect presence of domains. It could be used to identify signal peptides, coiled coil regions trans-membrane helices and compositionally biased regions [39]. The results are summarized in Supplementary Table T3.2.

2.6. Virulence Analysis

Virulence factors are gene products that facilitate the microorganism to launch itself as a pathogen. These factors enable its interaction with the cells inside a host. There is huge list of virulence factors, which includes cell surface protein and carbohydrates for attachment and protection respectively, bacterial toxins and hydrolytic enzymes. For identification of these factors in *N. meningitidis* H44/76, we had used VICMpred [40] and VIRULENTpred [41]. These methods are based on Support Vector Machine (SVM). The accuracy rate of VICMpred and VIRULENTpred is 70.75% and 81.8% respectively. The results are produced in Supplementary Table A.6.

2.7. Protein-Protein Interaction Prediction

It is crucial to gather system level information of the cellular functions. It requires correct annotation of functional interactions among proteins. In bioinformatics, we have several computational tools to detect these interacting partners. Here, we have used STRING 9.1 database to search known and predicted protein interacting partners (Supplementary Table A.5). These relations can be physical or functional interactions. STRING is an integrated tool that uses data from genomic context, high throughput experiments, conserved/co-expressed and from previous knowledge available in the text [42].

3. Results and Discussion

3.1. Sequence Analysis

The BLASTp was used to analyze HP sequences on the basis of homology. We had analyzed 525 HPs out of which 334 had been predicted to have their homologs (Supplementary Table A.3). These results were used for precise functional annotation of HPs in *N. meningitidis* along with the results obtained from other analysis. For example, HP E6MU35 was found similar to alanine racemase protein. HP E6MU60 was found to be homologous to cyanate hydratase.

3.2. Physiochemical Characterization

On the basis of instability index computed by ProtParam, 311 proteins were found stable (Supplementary Table A.1). The instability index for HPs was ranging from 40.01 to 82.07. The HPs belonging to this range were found to be unstable. Other analysis performed by ProtParam for the physiochemical characterization of HPs. The pI prediction helps in developing the buffer systems which are crucial for Isoelectric Focusing. Extinction coefficient was predicted on the basis of the concentration of Cys, Trp and Tyr amino acids within the protein sequence. It is important to know about extinction coefficient of the proteins in the drug development process. The Extinction coefficient helps in studying the protein-ligand and protein-protein interactions. Aliphatic index depends on the presence of aliphatic groups in the sequence. The number of aliphatic residues in the protein sequences increases the incidence of thermal stability in the case of globular proteins. The higher the aliphatic index, greater will be the protein's thermal stability. GRAVY value helps in determining the extent of protein-water interaction. The HPs, predicted with low GRAVY score are

likely to interact better with water [13].

3.3. Subcellular Localization Analysis

We have characterized 525 HPs into five subcellular localizations namely cytoplasmic, inner-membrane, outer-membrane, periplasmic and extracellular proteins. These five locations are characteristic to gram-negative bacteria. Out of these 525, 219 HPs are predicted to be in cytoplasm and 84 HPs as membrane bound proteins including inner membrane and outer membrane proteins. 56 HPs are predicted as periplasmic proteins and rest found to be extracellular (Supplementary Table A.2). We have found 150 HPs, which are having transmembrane helices (TMH). The prediction of membrane proteins is important, as they are vital for survival. TMH constitute these membrane proteins, which are key component in cell-cell signaling, ion and solute transportation and self-recognition. The membrane bound receptors are very significant in pharmaceutical industry [43].

3.4. Functional Analysis

We have analyzed 525 hypothetical protein sequences of *N. meningitidis* strain H44/76 for further characterization by searching functional motifs, domains, families, superfamilies and clusters present in them. The results are deduced on the basis of predictions made by latest bioinformatics tools (Supplementary Table A.3, Table A.4). We have successfully assigned functions to 202 proteins with high confidence. We have categorized these HPs into 91 enzymes, 25 binding proteins, 28 transporters, 9 immunity proteins, 6 bacteriophage related proteins and 43 regulatory proteins illustrated in Fig. 1. We have found well-conserved domains in 167 HPs. Many HPs contain domains with enzymatic activities and predicted as hydrolases, oxidoreductases, transferases, isomerases, lyases, ligases, kinases, phosphatases etc.
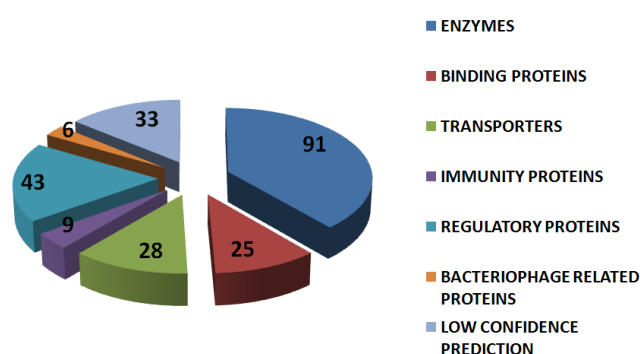

Figure 1. *Functional Annotation of 235 Hypothetical Proteins of Neisseria meningitidis.*

3.4.1. Enzymes

In every organism, enzymes affect the metabolic processes directly or indirectly. Enzymes produced by bacterial species are crucial for their sustenance in the host organism. These pathogens utilize the nutrients taken from the host for triggering pathogenicity, survival, maintenance and growth in the host. Information of these enzymes provides broad understanding of the interactions between pathogen and host [44].

Table 2. *Classification of enzymes into sub-classes.*

S.No.	Sub-category	Predicted Hypothetical protein
1.	Oxidoreductases	HP E6MWI9, HP E6N024, HP E6MTY1, HP F0MKA2, HP F0MKA2 etc.
2.	Transferases	HP E6MY14, HP F0MJ56, HP E6MZG2 etc.
3.	Hydrolases	HP E6MZ19, HP F0MM28, HP F0MKI8, HP E6MUL4, HP E6MX06, HP F0MIS3 etc.
4.	Lyases	HP E6MU60, HP E6MWH8, HP E6MXD6, HP F0MNU9
5.	Isomerases	HP E6MYS2, HP E6MYS3, HP F0ML13, HP E6MWT7, HP E6MZY1, HP E6MU35
6.	Ligases	HP E6MV50, HP E6MY93

Here, we have identified 91 HPs as enzymes, which are classified further into sub categories. Few are listed in the Table 2.

3.4.2. Transporters

The survival of the pathogen, *N. meningitidis* H44/76, also depends on the ability of the pathogen to extract essential nutrients from the host [45] (http://mmbr.asm.org/content/68/1/154.full). Import and export of these substances takes place with the help of transporter proteins. In our thorough analysis of H44/76 strain, we have annotated 28 HPs as transporters, which are functioning as membrane transporters, carriers and receptors. HP E6MWB7 and HP E6MXV5 were predicted as a divalent cation transporter. HP E6MX75, HP E6MUK8, HP E6MX75 and HP E6MZ77 were identified as the member of sulfite exporter TauE/SafE family.

3.4.3. Binding Proteins

We have identified four HPs namely HP E6MYN4, HP E6MYR5, HP E6MUN8 and HP E6MYY5 containing tetratricopeptide repeat. These repeats are responsible for protein-protein interactions and assembly of multiprotein complexes. Proteins carrying such repeats are identified to participate in cell cycle regulation, mitochondrial and peroxisomal protein transport, transcriptional control and protein folding [46]. HP FOMNK9, HP FOMJK0, HP FOMJK3, HP FOMJW3, HP FOMJW5 and HP FOMJW8 have been identified to have Hemagglutinin repeat. Such proteins are found to act as adhesins, filamentous haemagglutinins or Haem/haemopexin-binding proteins [47].

3.4.4. Regulatory Proteins

Some proteins are involved in various cellular processes like cell cycle, signaling pathway, replication, transcription, translation etc. These proteins are of crucial in the pathogenesis of the organism. HP E6MX02, HP FOMXL4 and HP E6N0E7 were identified to have Hedgehog/intein (Hint) domain. These proteins are required for embryonic cell differentiation. These proteins are released as inactive precursors having N terminal signaling domain and C

terminal auto-processing domain [48].

There are four HPs that are identified as transcriptional regulator. HP E6MUE0 belongs to Ic1R family of transcriptional regulator as predicted by using various tools. The proteins of this family share winged helix-turn-helix DNA-binding domain which is responsible for their activity. It works as a repressor of the acetate operon in *Escherichia coli* and *Salmonella typhimurium* [49].

HP E6MWQ9 was predicted as lipoate regulatory protein YbeD. YbeD protein is homologous to 3-phosphoglycerate dehydrogenase and the function of this enzyme is to regulate the synthesis of lipoic acid [50].

3.4.5. Virulent Proteins

We have identified 26 HPs as virulence causing factors (Table 3). These are concluded on the basis of VIRULENTpred and VICMpred predictions (Supplementary Table A.6). Virulence causing proteins can serve as potent drug targets for the drug discovery process. Lipooligosaccharides are found in the outer membrane of *N. meningitidis* and responsible for septic shock and hemorrhage. This is due to the destruction of red blood cells by this endotoxin [51]. Its polysaccharide capsule and fimbriae also helps the pathogen in spreading the virulence [52, 53].

Table 3. *Virulence analysis of 525 HPs using VIRULENTpred and VICMpred predicted 26 virulence causing proteins.*

S.No.	UNIPROT ID	VIRULENTpred			PSI-BLAST PSSM profiles	Cascade of SVMs and PSI BLAST	VICMpred
		Amino acid composition	Dipeptide Composition	Higher order Dipeptide composition			
1	E6MU90	Virulent	Virulent	Non-Virulent	Virulent	Virulent	Virulence factors
2	F0MNJ4	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
3	E6MWD8	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
4	E6MVZ3	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
5	E6MZG2	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
6	E6MWS1	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
7	E6MXM5	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
8	E6MUH7	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
9	E6N0E8	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
10	E6MYC1	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
11	E6MW11	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
12	F0MKG3	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
13	E6MVE2	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
14	E6MVG0	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
15	F0MLM9	Virulent	Virulent	Virulent	Virulent	Virulent	Virulence factors
16	E6MWC3	Virulent	Virulent	Virulent	Virulent	Virulent	Metabolism Molecule
17	F0MJ74	Virulent	Virulent	Virulent	Virulent	Virulent	Metabolism Molecule
18	F0MJ88	Virulent	Virulent	Virulent	Virulent	Virulent	Cellular process
19	F0MJ91	Virulent	Virulent	Virulent	Virulent	Virulent	Cellular process
20	F0MJ94	Virulent	Virulent	Non-Virulent	Virulent	Virulent	Cellular process
21	F0MJ95	Virulent	Non-Virulent	Non-Virulent	Virulent	Virulent	Information and storage
22	F0MM93	Virulent	Virulent	Virulent	Virulent	Virulent	Information and storage
23	F0MLM6	Virulent	Virulent	Virulent	Virulent	Virulent	Cellular process
24	E6MWY8	Virulent	Virulent	Virulent	Virulent	Virulent	Cellular process
25	E6MUE1	Non-Virulent	Virulent	Virulent	Virulent	Virulent	Cellular process
26	E6MUE2	Non-Virulent	Virulent	Virulent	Virulent	Virulent	Information and storage

We have found 6 HPs being associated with Mu-like prophage FluMu protein which are still functionally uncharacterized. Out of these 6, 3 HPs namely HP E6MWY8, HP E6MUE2 and HP E6MUE1 are found to be virulent. We have also found 9 HPs working as immunity proteins. These 9 proteins namely E6MWC3, F0MJ74, F0MJ88, F0MJ91, F0MJ94, F0MJ95, F0MM93, E6MVL2 and F0MLM6 have been found to participate in bacterial polymorphic toxin system and present next to toxin gene. These immunity proteins are characterized on the basis of the all alpha helical fold and a conserved proline residue. They usually contain Tox-REAse-1 or Tox-REase-6 family domain [54]. VIRULENTpred predicted HP E6MVL2 as non-virulent.

4. Conclusion

With the advent of new genomic data, it becomes mandatory to annotate the HPs sequenced in the genome of parasitic pathogens. The structural and functional characterization will allow us to understand the role of these HPs in causing pathogenicity and thereby finding new drug targets. We precisely analyzed 525 HPs of *N. meningitidis* H44/76 strain which might help in prioritizing the targets for experimentation among the reservoir of proteins. 202 HPs for which the functional prediction is performed with high confidence can be used in development of pathogen-system targeted drugs. Other 33HPs need to be analyzed further to get that confidence level in prediction. Characterization done on the basis of subcellular localization and physiochemical properties is going to help in distinguishing the membrane proteins and transporters specifically. By using docking studies, the structure of these proteins can aid in finding the possible inhibitors that act on these targets. Further, 26 Virulence proteins that are being found are of utmost importance in

understanding the pathways in which they are involved and to understand how they are helping the pathogen to survive. All these findings and annotations will definitely help in further characterization studies.

Acknowledgements

References

[1] Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Moule, S., Mungall, K., Quail, M. A, Rajandream, M. A, Rutherford, K. M., Simmonds, M., Skelton, J., Whitehead, S., Spratt, B. G., Barrell, B. G., 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature 404, 502–506. doi:10.1038/35006655.

[2] Piet, J. R., Huis in 't Veld, R. A. G., van Schaik, B. D. C., van Kampen, a H. C., Baas, F., van de Beek, D., Pannekoek, Y., van der Ende, A, 2011. Genome sequence of *Neisseria meningitidis* serogroup B strain H44/76. J. Bacteriol. 193, 2371–2372. doi:10.1128/JB.01331-10.

[3] MacNeil, J., Cohn, A., 2011. Meningococcal Disease, in: Roush, S.W., McIntyre, L., Baldy, L. M. (Eds.), Manual for the Surveillance of Vaccine-Preventable Diseases, Centers for Disease Control and Prevention, Atlanta, GA. http://www.cdc.gov/vaccines/pubs/surv-manual/chpt08-mening.html#f5.

[4] Brandtzaeg, P., van Deuren, M., 2012. Classification and pathogenesis of meningococcal infections. In *Neisseria meningitidis*, Humana Press, pp. 21-35.

[5] Harrison, L. H., Trotter, C. L., Ramsay, M. E., 2009. Global epidemiology of meningococcal disease. Vaccine 27. doi:10.1016/j.vaccine.2009.04.063.

[6] Whelan, J., Bambini, S., Biolchi, A., Brunelli, B., Robert-Du Ry van Beest Holle, M., 2015. Outbreaks of meningococcal B infection and the 4CMenB vaccine: historical and future perspectives. Expert review of vaccines, 14(5), 713-736.

[7] Andrews, S. M., Pollard, A. J., 2014. A vaccine against serogroup B *Neisseria meningitidis*: dealing with uncertainty. Lancet Infect. Dis. 14, 426-434. doi: 10.1016/S1473-3099(13)70341-4.

[8] Seib, K. L., Scarselli, M., Comanducci, M., Toneatto, D., Masignani, V., 2015. *Neisseria meningitidis* factor H-binding protein fHbp: a key virulence factor and vaccine antigen. Expert review of vaccines 14(6), 841-859.

[9] Nimrod, G., Schushan, M., Steinberg, D. M., Ben-Tal, N., 2008. Detection of Functionally Important Regions in "Hypothetical Proteins" of Known Structure. Structure 16, 1755–1763. doi:10.1016/j.str.2008.10.017.

[10] Lubec, G., Afjehi-Sadat, L., Yang, J. W., John, J. P. P., 2005. Searching for hypothetical proteins: Theory and practice based upon original data and literature. Prog. Neurobiol. 77, 90–127. doi:10.1016/j.pneurobio.2005.10.001.

[11] Galperin, M. Y., Koonin, E. V., 2004. "Conserved hypothetical" proteins: Prioritization of targets for experimental study. Nucleic Acids Res. 32, 5452–5463. doi:10.1093/nar/gkh885.

[12] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

[13] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server, in: Walker, J. M. (Ed.), The Proteomics Protocols Handbook-2005, Humana Press Inc. Totowa, NJ, pp. 571-607.

[14] Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L.J., Brinkman, F. S. L., 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615. doi:10.1093/bioinformatics/btq249.

[15] Bhasin, M., Garg, A., Raghava, G. P. S., 2005. PSLpred: Prediction of subcellular localization of bacterial proteins. Bioinformatics 21, 2522–2524. doi:10.1093/bioinformatics/bti309.

[16] Yu, C. S., Lin, C. J., Hwang, J. K., 2004. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 13, 1402–1406. doi:10.1110/ps.03479604.calization.

[17] Petersen, T. N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786. doi:10.1038/nmeth.1701.

[18] Bendtsen, J. D., Kiemer, L., Fausbøll, A., Brunak, S., 2005. Non-classical protein secretion in bacteria. BMC Microbiol. 5, 58. doi:10.1186/1471-2180-5-58.

[19] Mitaku, S., Hirokawa, T., Tsuji, T., 2002. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. Bioinformatics 18, 608–616. doi:10.1093/bioinformatics/18.4.608.

[20] Tusnády, G. E., Simon, I., 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J. Mol. Biol. 283, 489–506. doi:10.1006/jmbi.1998.2107.

[21] Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–80. doi:10.1006/jmbi.2000.4315.

[22] Sonnhammer, E. L., von Heijne, G., Krogh, A, 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 175–182.

[23] Letunic, I., Doerks, T., Bork, P., 2012. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40, D302–5. doi:10.1093/nar/gkr931.

[24] Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A.F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, C. J. A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C., Yong, S.-Y., 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 40, D306–D312. doi:10.1093/nar/gkr948.

[25] Kanehisa, M., 1997. Linking databases and organisms: GenomeNet resources in Japan. Trends Biochem. Sci. 22, 442–444. doi: 10.1016/S0968-0004(97)01130-4.

[26] Henikoff, J. G., 2000. Increased coverage of protein families with the Blocks Database servers. Nucleic Acids Res. 28, 228–230. doi:10.1093/nar/28.1.228.

[27] Corpet, F., Gouzy, J., Kahn, D., 1999. Recent improvements of the ProDom database of protein domain families. Nucleic Acids Res. 27, 263–267. doi:10.1093/nar/27.1.263.

[28] Attwood, T. K., 2002. PRINTS and PRINTS-S shed light on protein ancestry. Nucleic Acids Res. 30, 239–241. doi:10.1093/nar/30.1.239.

[29] Orengo, C., Michie, A., Jones, S., Jones, D., Swindells, M., Thornton, J., 1997. CATH − a hierarchic classification of protein domain structures. Structure 5, 1093–1109. doi: 10.1016/S0969-2126(97)00260-8.

[30] Gough, J., Karplus, K., Hughey, R., Chothia, C., 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. 313, 903–19. doi:10.1006/jmbi.2001.5080.

[31] Thomas, P. D., Kejariwal, A., Guo, N., Mi, H., Campbell, M. J., Muruganujan, A., Lazareva-Ulitsky, B., 2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. 34, W645–W650. doi:10.1093/nar/gkl229.

[32] Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–30. doi:10.1093/nar/gkt1223.

[33] Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V, Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., Bryant, S. H. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, D225–9. doi:10.1093/nar/gkq1189.

[34] Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Tasneem, A., Thanki, N., Yamashita, R. A., Zhang, D., Zhang,

N., Bryant, S. H., 2009. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 37, D205–10. doi:10.1093/nar/gkn845.

[35] De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–5. doi:10.1093/nar/gkl124.

[36] Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuche, B. A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I., Bridge, A., 2013. HAMAP in 2013, new developments in the protein family classification and annotation system. Nucleic Acids Res. 41, D584–9. doi:10.1093/nar/gks1157.

[37] Rappoport, N., Karsenty, S., Stern, A., Linial, N., Linial, M., 2012. ProtoNet 6.0: organizing 10 million protein sequences in a compact hierarchical family tree. Nucleic Acids Res. 40, D313–20. doi:10.1093/nar/gkr1027.

[38] Cai, C. Z., 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. doi:10.1093/nar/gkg600.

[39] Schultz, J., 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231–234. doi:10.1093/nar/28.1.231.

[40] Saha, S., Raghava, G. P. S., 2006. VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. Genomics. Proteomics Bioinformatics 4, 42–7. doi: 10.1016/S1672-0229(06)60015-6.

[41] Garg, A., Gupta, D., 2008. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinformatics 9, 62. doi: 10.1186/1471-2105-9-62.

[42] Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L. J., von Mering, C., 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39, D561–8. doi:10.1093/nar/gkq973.

[43] Chen, C. P., Kernytsky, A., Rost, B., 2002. Transmembrane helix predictions revisited. Protein Sci. 11, 2774–91. doi:10.1110/ps.0214502.

[44] Bjornson, H. S., 1984. Enzymes Associated with the Survival and Virulence of Gram-Negative Anaerobes. Clin. Infect. Dis. 6, S21–S24. doi:10.1093/clinids/6.Supplement_1.S21.

[45] Perkins-Balding, D., Ratliff-Griffin, M., Stojiljkovic, I., 2004. Iron Transport Systems in Neisseria meningitidis. Microbiol. Mol. Biol. Rev. 68, 154–171. doi:10.1128/MMBR.68.1.154-171.2004.

[46] D'Andrea, L. D., Regan, L., 2003. TPR proteins: the versatile helix. Trends Biochem. Sci. 28, 655–62. doi:10.1016/j.tibs.2003.10.007.

[47] Kajava, A. V., Cheng, N., Cleaver, R., Kessel, M., Simon, M. N., Willery, E., Jacob-Dubuisson, F., Locht, C., Steven, A. C., 2001. Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins. Mol. Microbiol. 42, 279–292. doi:10.1046/j.1365-2958.2001.02598.x.

[48] Perler, F. B., 1998. Protein Splicing of Inteins and Hedgehog Autoproteolysis: Structure, Function, and Evolution. Cell 92, 1–4. doi: 10.1016/S0092-8674(00)80892-2.

[49] Reverchon, S., Nasser, W., Robert-Baudouy, J., 1991. Characterization of kdgR, a gene of Erwinia chrysanthemi that regulates pectin degradation. Mol. Microbiol. 5, 2203–2216. doi:10.1111/j.1365-2958.1991.tb02150.x.

[50] Kozlov, G., Elias, D., Semesi, A., Yee, A., Cygler, M., Gehring, K., 2004. Structural similarity of YbeD protein from Escherichia coli to allosteric regulatory domains. J. Bacteriol. 186, 8083–8. doi:10.1128/JB.186.23.8083-8088.2004.

[51] Griffiss, J. M., Schneider, H., Mandrell, R. E., Yamasaki, R., Jarvis, G. A., Kim, J. J., Gibson, B. W., Hamadeh, R., Apicella, M. A., 1988. Lipooligosaccharides: The Principal Glycolipids of the Neisserial Outer Membrane. Clin. Infect. Dis. 10, S287–S295. doi:10.1093/cid/10.Supplement_2.S287.

[52] Koomey, M., 2009. Type IV Pilus Biogenesis, Structure and Function: Lessons from Type IVa Pilin Systems, in: Jarrell, K. F. (Ed.), Pili and Flagella: Current Research and Future Trends, Caister Academic Press, Norfolk, UK, pp. 19-40.

[53] Corbett, D., Roberts, I. S., 2009. Genetics and Regulation of Bacterial Polysaccharide Expression in Human Pathogens Bacteria, in: Ullrich, M. (Ed.), Bacterial Polysaccharides: Current Innovations and Future Trends, Caister Academic Press, Norfolk, UK, pp. 69-86.

[54] Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M., Aravind, L., 2012. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biol. Direct 7, 18. doi: 10.1186/1745-6150-7-18.