Pak. J. Biochem. Mol. Biol., 2023, 56(1):12-20

Research paper

Structural Identification of Formate Tetrahydrofolate Synthetase from *Neisseria meningitidis*

Ayesha Habib, Sana Aurangzeb, Mehwish Hamid, Yasmeen Rashid*

Department of Biochemistry, University of Karachi, Karachi-75270, Pakistan.

*Corresponding author: Dr. Yasmeen Rashid (yasmeen.rashid@uok.edu.pk)

ABSTRACT

Neisseria meningitidis, a gram-negative diplococcus, is a life-threatening pathogen that is responsible for causing meningitis and severe sepsis in humans. Over 2.5 million cases of *N. meningitidis* are reported yearly, with an approximately 10% mortality rate. Structural bioinformatics has been utilized in this research to predict the three-dimensional model of Formate tetrahydrofolate synthetase (FTHFS) of *N. meningitidis*. The FTHFS of N. *meningitidis* serogroup B (strain MC58) is a gene product of NMB1839 and is known to play a role in the one-carbon metabolism and transfer one-carbon-containing units to a variety of biosynthetic pathways. The enzyme FTHFS has an alpha/ beta fold and is mainly composed of three domains, in which one is comparatively larger than the other two. The crystal structure of FTHFS from *Morella thermoacetica* (PDB ID; 1FP7) was used as a template to predict the homology model of *N. meningitidis* FTFHS. The structural comparison of both enzymes has suggested that the amino acid residues interacting with the active site and potassium moiety were well conserved. This high conservation among crucial residues showed the shared catalytic mechanism by both enzymes. This study provided additional data for research on novel anti-meningitis drug target.

KEYWORDS Neisseria meningitidis; Formate tetrahydrofolate synthetase; Homology Modeling, NMB1839; Structural Bioinformatics

INTRODUCTION

Neisseria meningitidis (meningococcus) is the lethal pathogenic bacterium that colonizes the nasopharyngeal region of its human host. The deadly pathogen is capable of causing a variety of life-threatening diseases, including sepsis [1]. Based on its capsular polysaccharide, *N. meningitidis* is divided into thirteen serogroups, five of which, i.e., A, B, C, W-135, and Y, are the ones that result in deadly infectious diseases in humans throughout the world. Vaccines against all the serogroups are available except serogroup B as it is poorly immunogenic and has a resemblance with human antigens [2]. Therefore, it is crucial to develop a safer drugs against serogroup B. As *N. meningitides* inhabits the throat of its host and causes high fever, nausea, confusion, difficulty concentrating, seizures, and vomiting [3]. Meningococcal diseases are transmitted through respiratory droplets in a healthy individual, where the pathogen uses IgA protease to escape IgA antibodies which results in the invasion of the bloodstream. In the bloodstream, the microorganisms evade the immune system and then result in the disruption of the meninges [4-6]. The whole genome analysis of *N. meningitidis* indicated that seventy-three genes were mainly involved in its pathogenesis. Approximately 50% of these genes have been identified in encoding proteins and participate in metabolic pathways. Likewise, eight genes

were determined to encode known pathogenic factors [7]. Several structural studies of *N. meningitidis* pathogenic factors have been reported indicating the importance of structural analysis in anti-meningococcal drug discovery [8-13]. Formate tetrahydrofolate synthetase (FTHFS) is encoded by the gene NMB1839 and it is one of the significant factors involved in *N. meningitidis* pathogenesis [7].

FTFHS is an essential enzyme that contributes to the synthesis of various macromolecules like DNA, polyamine, amino acids, creatinine, and phospholipids by adding a one-carbon-containing unit, which is the key element in distinct biosynthetic pathways [14]. This enzyme catalyzes the ATP-dependent activation of the formate ion via its addition to the N10 position of tetrahydrofolate [15]. It is a highly expressed enzyme in the acetogenesis pathway of carbon fixation, as well as the glycine reductase/synthase pathway of purinolysis for cellular biosynthesis [16]. In acetogens, this enzyme catalyzes the formylation of tetrahydrofolate, and in purinolytic organisms, this reaction is reversed in which it yields formate from 10-formyltetrahydrofolate with the simultaneous generation of ATP [16-19].

Structural bioinformatics approaches were used to predict the three-dimensional homology model of the *N. meningitidis* FTHFS, which is extensively significant in structure-based drug discovery against meningococcus. This study plays a significant role in increasing the understanding of the prediction of the structural aspects of important factors of *N. meningitidis* pathogenesis.

MATERIALS AND METHODS

The sequence of the target protein was retrieved in FASTA format from the UniProt database [20], and the selection of the best template was done by subjecting the target protein sequence against PDB [21] using PSI-BLAST [22]. The secondary structural information of the target protein was predicted using the PSIPRED web server [23], while the secondary structure information of the template was obtained through PDBsum [24]. Multiple sequence alignment (MSA) was performed using CLUSTALX software [25] which showed the evolutionary relationship and conservation among the protein sequence. Using the secondary structure and MSA results the structure-based sequence alignments of the target and the template were made and used to predict the refined homology model by the computational program MODELLER 10.1 [26]. The standalone software, i.e., PROSA [27-28], and PROCHECK [29] were used to validate the stereochemistry of the predicted models. The visualizing software. i.e., Discovery Studio Visualizer (DSV) [30] was utilized to examine the three-dimensional predicted model's structural features, including the active site residues, the domains, the ion binding sites, etc. The iMods server was used to validate the overall stability of the predicted model in terms of normal mode analysis (NMA) [31].

RESULTS

Sequence Alignment

Sequence analysis of identified *N. meningitidis* gene product NMB1839, i.e., FTHFS has a length of 557 amino acid residues. The secondary structure prediction of this sequence revealed that the protein consists of twenty beta sheets and twenty-five alpha-helices. The crystal structure of *M. thermoactica* FTHFS (PDB ID: 1FP7) was selected as a sequence homolog because the Psiblast results of the amino acid sequence of *N. meningitidis* FTHFS and *M. thermoacetica* FTHFS showed 61% similarity. The secondary structures of both the FTFHS proteins were also conserved (Figure 1).

Homology Modeling

The predicted model was shown to have an alpha/beta fold and consist of three major domains. The first which is the largest domain has seven beta sheets covered by alpha-helices on both sides. The second domain has five beta-sheets and two alpha-helices, whereas the third has four beta-sheets making a half-barrel. The concave side of the barrel is covered with two alpha-helices, whereas the convex side is another wall of the large cavity (Figure 2). The lower Root mean square deviation (RMSD) value of 0.222 Å between the C-alpha backbones of the crystal and the model structures indicated higher structural conservation.

The Prosa plot showed a stable internal energy profile of the *N. meningitidis* FTHFS model with a z-score of -11.50, which was comparable with the computed z-score of the template, i.e., -12.44. The PROCECK results, i.e., Ramachandran plot, indicated that the model has only 2 residues in the disallowed region, i.e., LYS101 and ALA516, 409 residues in the most favored region, 53 residues in the additionally allowed region, and 6 in the generously allowed region. For authenticating the model, the template Ramachandran plot was also obtained that showed 3 residues in the disallowed region, i.e., LYS1108, LYS1015, and PHE1129.

Active site analysis

The active site residues including E1009, R1097, K1074, T1075, T1076, and W1412 were found to be conserved in both *N. meningitidis* FTHFS and *M. thermoacetica* FTFHS proteins. The high conservation among these active residues suggested that both proteins share the same catalytic mechanism (Figure 3A). The calculated RMSD of active site residues between the model and crystal was 0.4 Å which is an indication of shared protein fold between the two proteins.

Potassium Binding site

The protein requires a monovalent K+ to stabilize the structure at higher temperatures. The interacting residues with K+, i.e., E1098, G1272, D1132, T1130, N1258, and F1129 were observed to be conserved. However, I1270 of *M. thermoacetica* FTFHS was substituted V1270 in *N. meningitidis* FTHFS, and 1133 *M. thermoacetica* FTFHS was substituted F1133 in *N. meningitidis* FTHFS. Isoleucine and valine both residues shared non-polar side chains. Similarly, isoleucine and phenylalanine also shared non-polar side chains that suggested a similar cation-binding mechanism (Figure 3B).

Normal Mode Analysis

Deformability graphs can be useful in assessing the quality of protein structures by providing information on the flexibility and deformability of the protein backbone. Deformability graphs are plots of the root mean square fluctuations (RMSF) of each residue in a protein structure, which can be used to identify regions of the protein that are more flexible (Figure 4A). B-factor graphs show the comparison between NMA (Normal Mode Analysis) and PDB (Protein Data Bank) fields. These graphs highlight the areas of flexibility and dynamics of the protein structure by calculating the approximate root mean square (Figure 4B). Eigenvalues can be used in protein structure prediction to analyze the low-frequency vibrational modes of a protein structure. In the case of a predicted structure, a lower eigenvalue is an indication of a good-quality protein. This is because lower eigenvalues correspond to low-frequency vibrational modes of the protein structure, which are typically associated with large-scale, collective motions of the protein backbone. These motions are thought to be important for protein function and stability and are often disrupted in misfolded or unstable proteins (Figure 4C). The variance graph has an inverse relationship with eigenvalues, the green bars depict the individual variance while the purple bars

show the cumulative variance. This graph also provides information about the flexibility and rigidity of the predicted structure (Figure 4D). Covariance graphs can be used in protein structure prediction to identify pairs of residues that are likely to be in close proximity to the folded protein structure. The colors red, white, and blue depict correlated, uncorrelated, and anti-correlated motions, respectively (Figure 4E). Elastic network maps can be used to identify regions of the protein that are particularly flexible or unstable, which may be associated with regions of the protein that are less important for structure. The lighter gray color in the predicted structure elastic map indicates less stiff regions in the structure (Figure 4F).

DISCUSSION

N. meningitidis is precisely an adaptable organism, capable of adapting to various environmental conditions. It comes into contact and maintains the host alive to allow transmission, just like many other bacterial pathogens [1]. A study reported that seventy-three genes in the *N. meningitidis* genome are crucially involved in its pathogenesis; 11 of which encode proteins that are involved in amino acid biosynthesis [7]. The protein encoded by the NMB1839 gene is formate tetrahydrofolate synthetase (FTFHS). FTHFS is an essential enzyme in one-carbon metabolism that is known to transfer one carbon-containing unit (methyl groups) in various biosynthetic pathways [7].

The primary goal of this study was to ascertain the three-dimensional structure of the *N. meningitidis* FTHFS enzyme due to its importance in *N. meningitidis* pathogenicity [21]. The target protein was modeled in three dimensions using computational methods. Pharmaceutical companies frequently employ computational approaches for predicting 3D protein models, and much effort has been made to increase model accuracy as well as to improve the applicability of these techniques [32].

The homology model of the targeted protein, i.e., *N. meningitidis* FTHFS was constructed utilizing three-dimensional coordinates of the crystal structure of *M. thermoacetica* FTHFS.

The *FTHFS* from *M. thermoacetica* was considered the best template among other homologs based on the same enzyme family, high, i.e., 61%, sequence identity, maximum query cover, minimum gaps, and a lower E-value. Secondary structure analysis of the model and the crystal structure showed high conservation. Multiple sequence alignment significantly helped in determining the structural and functional relationship between the protein-sequence families. The MSA results also showed conservations among all the active site residues across the distinct species. PROSA [27-28] and PROCHECK [29] software was utilized to evaluate the overall quality of the predicted model. The Ramachandran plot of the selected model has shown lesser residues, i.e., 2 in the disallowed region compared to the residues of the template protein, i.e., 3. These residues were found to be a part of the loop area, which was far from the active site. Therefore, we believe that the overall architecture of the active site would not be affected by these residues with disallowed stereochemistry. Prosa software also provided a good energy profile and a z-score within the allowed range.

Recent reports have demonstrated that MODELLER has provided lower RMSD values in predicting protein structures compared to AI-based tools such as AlphaFold and Rosetta. The differences in these values are due to the large variation in loop prediction of structures by AI-based methods, whereas MODELLER provides impressive accuracy in loop predictions [33].

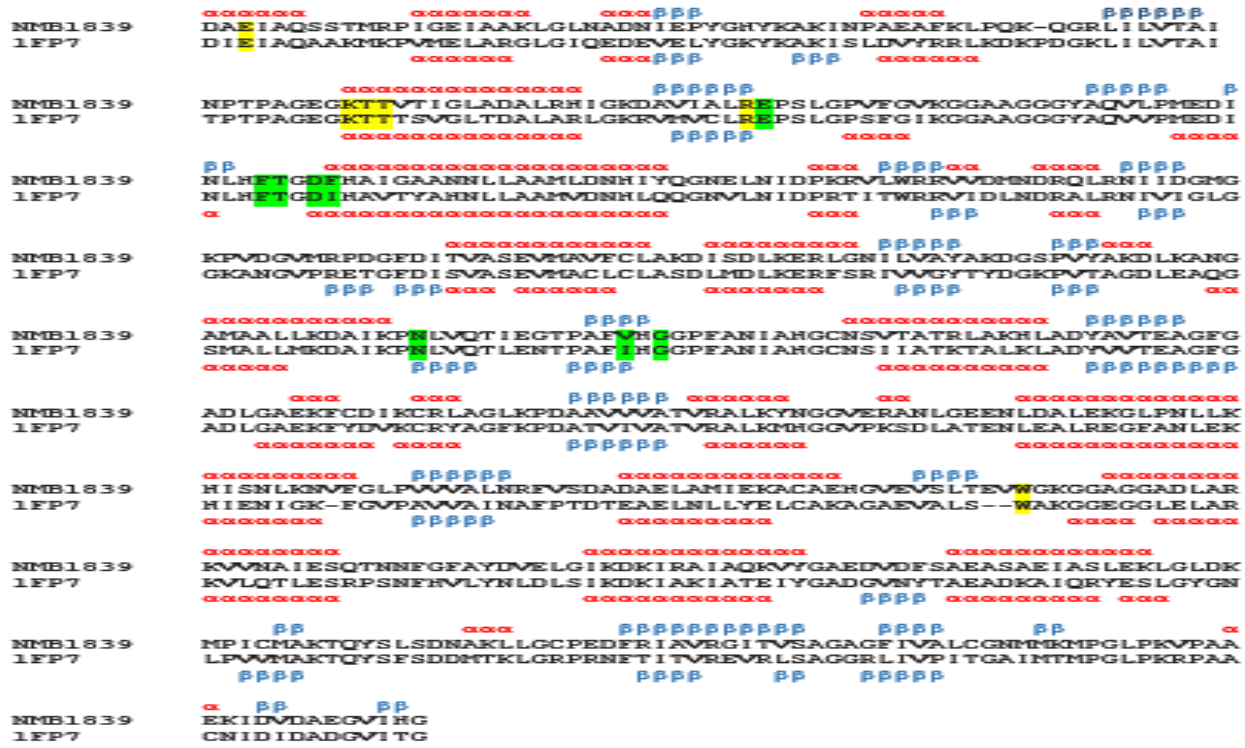

Figure 1: Structure-based pair-wise sequence alignments between *N. meningitidis* (NMB1839) and *M. thermoacetica* (PDB ID: 1FP7) proteins. Alpha helices and beta sheets are shown using α and β characters respectively. The active site residues are highlighted in yellow potassium binding residues in green.

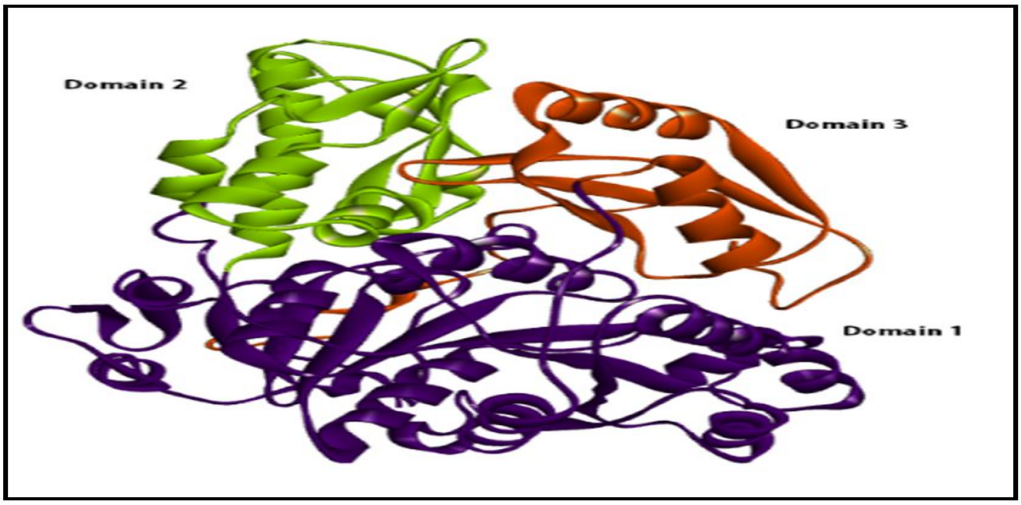

Figure 2: Ribbon representation of the three functional domains in the predicted homology model. The larger Domain 1 is purple, Domain 2 is green, and Domain 3 is orange.

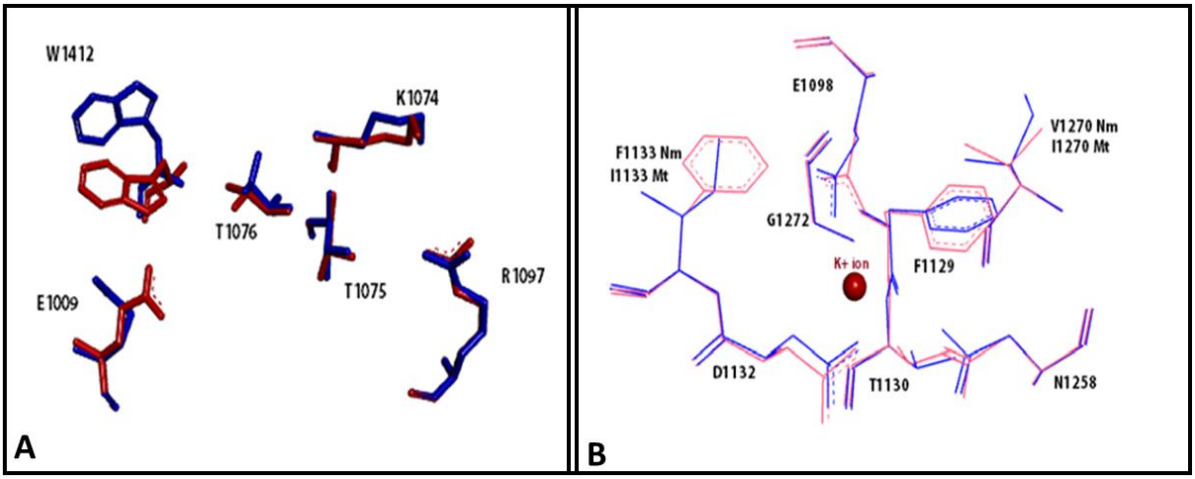

Figure 3: Superposition of the (A) active residues and (B) cation-binding residues of *N. meningitidis* FTHFS (blue) and *M. thermoacetica* FTFHS proteins(red).

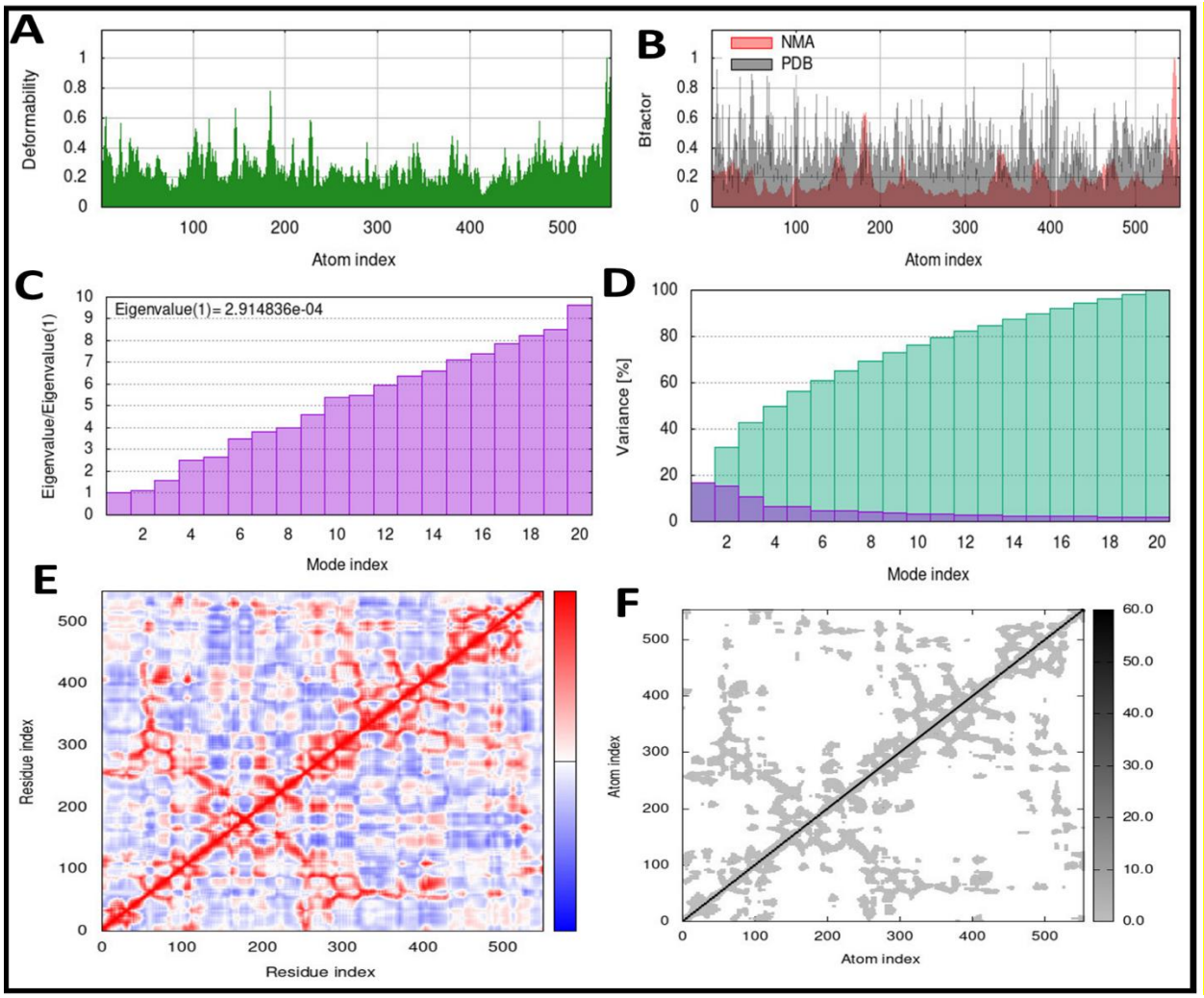

Figure 4: (A) deformability (B) B-factor (C) eigenvalue (D) variance, (E) co-variance (F) Elastic networks maps of the predicted structure.

Additionally, superposition of the template and the model proteins showed a lower RMSD value of 0.22Å suggesting overall similar backbone architectures of the *N. meningitidis* FTFHS and *M. thermoacetica* FTFHS. All the active site residues were conserved in both *NmFTHFS* and MtFTHFS proteins in addition to the K+ ion binding site residues and suggested a shared catalytic mechanism. The predicted model has shown good quality with improved stereochemistry, energy profiling, and a lower RMSD, indicating high structural similarity to the crystal structure and overall accuracy. These predictions are valuable because the protein's structure is closely related to its function. Knowing the structure can provide insights into its biological role, interactions with other molecules, and potential drug targets [8-13]. To verify the stability of the predicted model in this research, the iMODS server was utilized. NMA analysis is an alternative and more efficient approach than expensive atomistic simulations [31]. This server provides normal mode analysis of the predicted structures through deformability, B-factors, eigenvalues, covariance graphs, and elastic network maps. These analyses are very useful for gaining insight and enhancing understanding regarding protein stabilities, reactivity, and mechanical response [31].

CONCLUSION

N. meningitidis is a lethal pathogen known to inhabit the nasopharyngeal region of its host. It is responsible for causing hazardous diseases like sepsis and meningococcemia and can lead to the death of its host (humans) in severe conditions. The prediction of the three-dimensional structure of the protein *Formate tetrahydrofolate synthetase* of *N. meningitidis* was beneficial as it is involved in the pathogenesis. The overall fold of the predicted model was quite similar to that of the template with the conserved active site and cation-binding residues, suggesting a similar catalytic and cation-binding mechanism for both enzymes. These interpretations will help us to grease the wheels toward the discovery of many potent drug candidates against *N. meningitidis* that will, in turn, aid in the treatment of meningococcal diseases.

REFERENCES

1. Yazdankhah, S.P. and Dominique A.C, Neisseria meningitidis: an overview of the carriage state. *J. Med. Microbiol.*, 2004 **53**(9): p. 821-832.
2. Rosenstein, N.E., et al., Meningococcal disease. *N. Engl. J. Med.*, 2001. **344**(18): p. 1378-1388.
3. Tzeng, Y.L. and David S.S, Epidemiology and pathogenesis of Neisseria meningitidis. *Microbes Infect.*, 2000. **2**(6): p. 687-700.
4. Vidarsson, G., et al., Activity of human IgG and IgA subclasses in immune defense against Neisseria meningitidis serogroup B. *J. Immunol.*, 2001. **166**(10): p. 6250-6256.
5. Ladhani, S.N., et al., Meningococcal disease and sexual transmission: urogenital and anorectal infections and invasive disease due to Neisseria meningitidis. *The Lancet*, 2020. **395**(10240): p. 1865-1877.
6. Spoerry, C., et al., Neisseria meningitidis IgA1-specific serine protease exhibits novel cleavage activity against IgG3. *Virulence*, 2021 **12**(1): p. 389-403.
7. Sun, Y.H., et al., Functional genomics of Neisseria meningitidis pathogenesis. *Nat. Med.*, 2000. **6**(11): p. 1269-1273.
8. Aurangzeb, S., et al., Three Dimensional Structural Investigation of Lead Molecules

against Neisseria meningitis Pathogenic Factors; A step towards drug designing.PJBMB, 2018. **51**(1-2): p. 31-56.

9. Safdar, R. and Y. Rashid, Structural Prediction of Pathogenic Factors of Neisseria Meningitidis using Comparative Modeling.PJBMB, 2018. **51**(1-2): p. 15-30.
10. Shakil, U., S. Sadia, and Y. Rashid, Comparative modeling of potential drug targets of Neisseria meningitidis. PJBMB, 2019. **52**(1): p. 37-51.
11. Ahrar, M. and Y. Rashid, Homology modeling-based structure prediction of pathogenic factors of Neisseria meningitidis. PJBMB, 2019. **52**(1).
12. Ali, M., M. Aurongzeb, and Y. Rashid, In-silico three dimensional structure prediction of important Neisseria meningitidis proteins. Pak J Pharm Sci, 2021. **34**(2).
13. Hamid, M., et al., Identification and structural investigation of potential novel drug candidates against lethal human pathogen.Pak J Pharm Sci, 2021. **34**(1).
14. MacKenzie, R.E., Biogenesis and interconversion of substituted tetrahydrofolates. *Folates and pterins*, 1984. 1: p. 255-306.
15. Mejillano, M.R., et al., Formation and utilization of formyl phosphate by N10-formyltetrahydrofolate synthetase: evidence for formyl phosphate as an intermediate in the reaction. *Biochemistry*, 1989 **28**(12): p. 5136-5145.
16. Ljungdhal, L.G., The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annu. Rev. Microbiol*, 1986 **40**(1): p. 415-450.
17. Ljungdahl, L.G., The acetyl-CoA pathway and the chemiosmotic generation of ATP during acetogenesis. *Acetogenesis*, 1994: p. 63-87.
18. Radfar, R., et al., The crystal structure of N 10-formyltetrahydrofolate synthetase from Moorella thermoacetica. *Biochemistry*, 2000. **39**(14): p. 3920-3926.
19. Radfar, R., et al., Cation binding and thermostability of FTHFS monovalent cation binding sites and thermostability of N 10-formyltetrahydrofolate synthetase from Moorella thermoacetica. *Biochemistry*, 2000 **39**(47): p. 14481-14486.
20. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*, 2021. **49**(D1): p. D480-D489.
21. Burley, S.K., et al., Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. *Methods mol. biol.*, 2017.160: p. 627-641.
22. Altschul, S.F., et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997.**25**(17):p. 3389-3402.
23. McGuffin, L.J., Kevin B, and David T.J., The PSIPRED protein structure prediction server. *Bioinformatics*, 2000. **16**(4): p. 404-405.
24. Laskowski, R.A., PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*, 2001. **29**(1): p. 221-222.
25. Thompson, J.D., et al., The CLUSTAL_X windows interface: flexible strategies for multiple

sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997. **25**(24): p. 4876-4882.

26. Webb, B. and Andrej S. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinform.*,2016. **54**(1): p.5-6.
27. Wiederstein, M. and Manfred J.S., ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res,* 2007 **35**(suppl_2): p. W407-W410.
28. Sippl, M.J., Recognition of errors in three-dimensional structures of proteins. Proteins: Structure, Function, and Bioinformatics, 1993. **17**(4): p. 355-362.
29. Laskowski, R.A., et al., PROCHECK: a program to check the stereochemical quality of protein structures *J. Appl. Crystallogr.*, 1993. **26**(2): p. 283-291.
30. Visualizer, D.S., Accelrys software inc. Discovery Studio Visualizer, 2005. **2**.
31. López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., and Chacón, P. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res*, 2014. **42**(Web Server issue): p. W271–W276.
32. Schmidt, T., Andreas B., and Torsten S. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today,* 2014.**1 9**(7): p. 890-897.
33. Lee, C., Su, B. H., and Tseng, Y. J. (2022). Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Brief. Bioinformatics*, 2022. **23**(5): p. bbac308.