***Original Research Article***

**Computational Structure Analysis,** **Characterization and Functional Annotation of a Hypothetical Protein from** ***Mycobacterium marinum*** **(****WP\_020728386.1): In Silico Approach**

**ABSTRACT**

**Aims:** The objective of this investigation was to assess the intended function of the hypothetical protein WP\_020728386.1.

**Study design:** It is a computer based in silico dry lab study based on different software use and manipulating databases.

**Place and duration of study:** The study is conducted on dry lab from 2024 to 2025.

**Methodology:** The projected tertiary structure was assessed using servers such as Swiss Model. The Swiss-Model Interactive Workplace, ProSA-web, PROCHEK, Ramachandran Plot, Z scores, and structural evaluations are used to choose the most suitable materials. For finding out homologous protein, different bioinformatics tools were used for searching sequence similarity.

**Results:** Fish dwelling in both fresh and saltwater are susceptible to infection from the opportunistic pathogen *Mycobacterium marinum.* This infection causes morbidity and mortality in fish by causing necrotising granuloma similar to tuberculosis. Again, in human, it causes nodular lymphangitis, skin nodules, and ulcers as signs of preliminary cutaneous infections that can progress to osteomyelitis, tenosynovitis, and arthritis leading to serious health hazards. Due to the existence of extradiol ring cleavage dioxygenase family protein, the protein WP\_020728386.1 present in *Mycobacterium marinum* has an enormous effect in catalysing the incorporation activity of molecular oxygen atoms into substrates, resulting in ring cleavage in aromatics. However, the protein has not yet been fully explicit. As a result, an in-silico method was developed in this study to record the structure and function of the unidentified protein WP\_020728386.1. The analysis of physicochemical parameters, functional annotation, predicted active site of ligand binding, Ramachandran plot analysis as well as overall fundamental features of the hypothetical protein provide an insight into the studied protein that it might be a very functional and potential protein possessing crucial role.

**Conclusions:** As a result of this research, our understanding of pathogenesis and pathophysiology will increase, and we will be able to concentrate on the protein complex which might assist in the development of effective treatment drugs and vaccines against *Mycobacterium marinum* infections.

KEYWORDS: *Mycobacterium marinum, in silico*, functional annotation, homology modeling, Ramachandran plot.

**1. INTRODUCTION**

*Mycobacterium marinum*, a macrophage pathogen similar to *Mycobacterium tuberculosis*, causes a persistent, systemic disease in ectotherms that resembles tuberculosis. In 1926, Aronson isolated and identified this slow-growing nontuberculous Mycobacterium (NTM) from fish for the first time. (Haenen et al. 2013; Aroson 1926; Collins 1985). Later, in 1954, Collins et al. characterised M. marinum as an opportunistic pathogen that affects humans (Collins 1985; Linell, 1954). These bacteria are ubiquitous in aquatic environment and are mostly spread by coming into touch with fresh and salt water, such as from swimming pools, aquariums, and marine life (Wayne 1992). It is considered as the most serious ﬁsh pathogen, known for triggering a variety of symptoms at a time, including uncoordinated swimming, abdominal swelling, weight loss, skin ulceration, the growth of white nodules in the liver, kidney, and spleen in both fresh and marine water ﬁsh (Ferguson [2006](#_bookmark159); El Amrani et al. [2010](#_bookmark156)). In fish, infection by M. marinum has long been referred to as fish tuberculosis. In human, Mycobacterium marinum infections may cause treatment refractory skin and soft tissue lesions, cutaneous manifestations like solitary papules or nodules, in extreme case develop tenosynovitis.

Although *M*. *tuberculosis* is most common macrophage, it’s slow growth hinders the discovery of new drugs. On the other contrary, fast-growing nature of *M*. *marinum* is suitable for it to use as a surrogate model for studyingthe pathogenesis of *M*. *tuberculosis*. Additionally, the feasibility of employing M. marinum as a model for assessing anti-TB activity has been explained (Dharra et al. 2019).

Although there exists an uncountable amount of biological data, for instance, genomic sequences, functional genomic data etc. gathered from different high-throughput experiments, there is still lacking of functional annotation for most of the recently sequenced proteins. For instance, about half of the *Mycobacterium tuberculosis* genome is made up of proteins of unknown functions (Mazandu et al. 2012**)**. It is now crying need to extract information from this big data pool. For exploitation of these data, there is need for effective computational procedure/methods to get genome annotation to predict functional annotations of uncharacterized protein. Hence, the experiment is conducted intended to characterize and annotate a hypothetical Protein from *Mycobacterium marinum* (WP\_020728386.1).

**2. MATERIALS AND METHODS**

**2.1 Sequence retrieval and Homology searching**

The sequence information of the selected hypothetical protein obtained from the protein database of National Center for Biotechnology Information (NCBI) (Sayers et al. 2022) having Accession no.WP\_020728386 (Version WP\_020728386.1)which contains 231 amino acids. After that, the sequence of amino acid was extracted in FASTA format for further analysis. To locate the protein's homologues, homology searching was done using the NCBI's BLASTp search tools (http://www.ncbi.nlm.nih.gov/) against the non-redundant database with default parameters (Altschul et al. 1990).

**2.2 Physiochemical properties analysis**

The physiochemical properties of the hypothetical protein such as molecular weight, theoretical pI, amino acid composition, atomic composition, instability index, and grand average of hydropathicity (GRAVY) were evaluated using web-based server named Protparam tool of ExPaSy server (Gasteiger et al. 2003). Furthermore, the theoretical isoelectric point (pI) of the hypothetical protein was measured using SMS Suite (v2.0) (Stothard. 2000).

**2.3 Subcellular location determination**

The subcellular location of the hypothetical protein was predicted by CELLO v.2.5 (Yu et al. 2006). The protein's location, such as its membrane, extracellular space, cytoplasm, or cell wall, is primarily predicted by the CELLO Prediction. The location is appropriately indicated by a higher reliability value.

**2.4 Functional annotation prediction**

We used NCBI Conserved Domains Database (NCBI-CDD) (Wang et al. 2023) for functional annotation of the hypothetical protein from *Mycobacterium marinum.*

**2.5 Secondary structure prediction**

The secondary structure features for selected hypothetical protein were predicted by the reliable PSIPRED server (McGuffin et al. 2000). For element prediction, the self-optimized prediction method with alignment (SOPMA) (Combet et al. 2000) was also used.

**2.6 Three-dimensional (3D) structure prediction and model quality validation**

The three-dimensional (tertiary) structure of the hypothetical protein under study was predicted using the Swiss-Model server (Waterhouse et al. 2018). Finally, the anticipated 3D structure of the hypothetical protein was assessed for quality using the PROCHECK service available in the SAVES (v6.0) server (Laskowski et al. 1996). Furthermore, the Z-score of the modeled structure for structural evaluation was calculated using the ProSA-Web server (Wiederstein& ProSA-Web, 2007)

**2.7 Active site detection**

CASTp server (Computed Atlas of Surface Topography of Protein) (Tian et al. 2018) was used to determine the active sites of the modelled protein. The CASTp web server delivers an online service for identifying, characterising, and measuring concave surface areas on three-dimensional structures of proteins.

**3. RESULTS AND DISCUSSION**

**3.1 Protein Sequence retrieval and homology Information**

The basic determinants of biological function and structure are protein sequences which are the ultimate instrument in both cases of drug discovery and development. The NCBI protein sequence databank is a reservoir of protein sequences collecting from different sources such as RefSeq, Genebank, TPA etc. Here, the studied protein obtained from the NCBI protein database is present in the locus WP\_020728386.1 containing 231 amino acids (aa). However, additional information of the respected protein is shown in Table 1.

|  |  |
| --- | --- |
| **Parameters** |  **Information of Protein** |
| Locus | WP\_020728386.1 |
| Definition | Hypothetical protein WP\_020728386.1(*Mycobacterium marinum)* |
| Amino acid | 231 aa |
| Accession | WP\_020728386 |
| Version | WP\_020728386.1 |
| Source | *Mycobacterium marinum* |
| Keywords | RefSeq |
| Organism | [*Mycobacterium marinum*](https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&id=1781&lvl=3&keep=1&srchmode=1&unlock) |
| FASTA sequence | MLSAIAIIPSAPVLVPELTGLAGAEVADLRSAVVAAASSLPNHWVGIGVGPEDQVAGPDAVGTFAGFGVDVQVRLAPALGEQPGPPSDLPLCALLAGWVRGQIRPAATAKIQVYAHGRESEDALVQGKLLRAEIDRTADPVGVLVVADGVNTLTPSAPGGYDPTGAPAQLLLDDALAGGDVAALARLPERVLGRPAFQVLAGLCGAGPRSATELYRGAPFGVGCFVGVWQP |

**Table 1. Protein retrieval**

Well known NCBI BLASTp tools were used for analysing the homology search of our hypothetical protein. According to the homology analysis, our query protein shows structural similarities with other dioxygenase domain containing proteins from different Mycobacterium species. 5 results of homology searching is presented in the Table 2.

**Table 2. Protein similarity is shown by the BLASTp result.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Accession no.** | **Organism** | **Protein Name** | **Score** | **Percent Identity** | **E-Value** |
| WP\_198965783.1 | Mycobacterium shottsii | Hypothetical Protein | 434 | 99.13% | 8e-153 |
| WP\_011741159.1 | Mycobacterium ulcerans | Hypothetical Protein | 431 | 97.84% | 2e-151 |
| WP\_036411733.1 | Mycobacterium gastri | Hypothetical Protein | 290 | 67.10% | 8e-96 |
| WP\_103802134.1 | Mycobacterium kansasii | Hypothetical Protein | 288 | 69.26% | 8e-95 |
| WP\_083111484.1 | Mycobacterium Angelicum | Hypothetical Protein | 281 | 65.80% | 4e-92 |

**3.2 Physicochemical Characterization**

The amino acid sequence of WP\_020728386.1 present in *Mycobacterium marinum* was retrieved in FASTA format for using as a query sequence to determinate physicochemical parameters. The instability index of WP\_020728386 is 33.61(<40) indicates the stable nature of the protein (Guruprasad et al. 1990). The protein has a molecular weight of 23253.66 Da and is naturally acidic (pI 4.66\*). Cystine is present according to extinction coefficient values (21095). Higher aliphatic index values (106.49) suggest as an eventual factor of increased thermo stability of the query protein for a wide temperature range. Since the protein is hydrophobic, the higher grand average of hydropathicity (GRAVY) index value (0.367) suggests that there may be less interaction with water (Table 3). The ExPASy ProtParam tool was put to use to determine the amino acid composition where Alanine (16.5%) was the most prevalent amino acid, and methionine (0.4%) was the least abundant amino acid. The amino acid composition can help us to reveal the active amino acid pocket for drug and vaccine targeting against the protein (Saikat and Ripon 2020).

**Table 3. Physiochemical properties**

|  |  |
| --- | --- |
| **Characteristics** | **Value** |
| Molecular weight | 23253.66 |
| Theoretical pI | 4.66 |
| Total number of negatively charged residues (Asp + Glu) | 22 |
| Total number of positively charged residues (Arg + Lys) | 14 |
| Formula | C1042H1665N287O307S4 |
| Total number of atoms | 3305 |
| The estimated half-life | 1. 30 hours (mammalian reticulocytes, *in vitro*)
2. 20 hours (yeast, *in vivo*)
3. 10 hours (*Escherichia coli, in vivo*)
 |
| Instability index (II) | 33.61 |
| Aliphatic index | 106.49 |
| Grand average of hydropathicity (GRAVY) | 0.367 |

**3.3 Subcellular location determination**

Subcellular localizations dictate the circumstances in which proteins are operated. This subcellular localization is important as it can influence protein function through controlling a range of molecular interaction partners' accessibility and availability. A protein's subcellular localization can be used to identify it as a therapeutic or vaccination target (Shahbaaz et al. 2013). The studied WP\_020728386.1 hypothetical protein is a membrane protein predicted by CELLO v.2.5. Also this serverpredicted the localization of physico-chemical comp., partitioned seq. comp., neighboring seq. comp., amino acid comp. and N-peptide comp. values of 0.933, 0.743, 0.527, 0.634 and 0.570, respectively (Table 4). Proteins found in surface membranes may serve as potential vaccination targets, while cytoplasmic proteins are thought to be powerful pharmacological targets (Vetrivel et al. 2011). This type of subcellular identification research suggest that the hypothetical protein might be helpful in disease recovering through introducing various novel medicines.

**Table 4. Analysis of subcellular localization**

|  |  |  |
| --- | --- | --- |
| **Support Vector Machine (SVM)** | **Localization** | **Reliability** |
| Amino Acid Comp. | Cytoplasmic | 0.634 |
| N-peptide Comp. | Extracellular | 0.570 |
| Partitioned seq. Comp. | Membrane | 0.743 |
| Physico-chemical Comp. | Membrane | 0.933 |
| Neighboring seq. Comp. | Membrane | 0.527 |
| Subcellular localization Predictor (CELLO) value | Membrane | 2.543 \* |
| Cytoplasmic | 1.353 |
| Extracellular | 1.066 |
| CellWall | 0.038 |

\*CELLO predicted the subcellular location of the hypothetical protein as membrane

**3.4 Functional Annotation assessment**

A component of NCBI's Entrez query and retrieval system, CDD (the Conserved Domain Database) annotates protein sequences with the location of conserved domain footprints and functional locations assumed from these footprints (Marchler-Bauer et al. 2012). According to the conserved domain search (CD-search), the chosen hypothetical protein had one domain. The anticipated domain is Extradiol\_Dioxygenase\_3B\_like super family; indicating class III extradiol ring-cleavage dioxygenase family protein (pfam accession: pfam cl00599, Interval: 7-229, E-value: 3.84\*1031). These family proteins can catalyze the incorporation activity of the atoms of molecular oxygen into substrates leading to ring cleavage in aromatics by using different reaction mechanisms. Two main categories of dioxygenases exist as per the cleavage pattern of aromatic ring: Intradiol enzymes cleave the aromatic ring between two hydroxyl groups by utilizing non-heme Fe (III), while extradiol enzymes cleave the aromatic ring between a hydroxylated carbon and an adjacent non-hydroxylated carbon by utilizing non-heme Fe (II). Again, Extradiol dioxygenases are of three classes. Class I and II enzymes have sequence similarity while they are evolutionary related, class II enzymes having two domains evolving from class I enzyme through the process of gene duplication. Class III enzymes have two subunits (A and B) and have difference in sequence and structure from the other two classes; I and II. This model includes the catalytic subunit B of extradiol dioxygenase class III enzymes. However, Class III enzymes belonging to this family include 2'-aminobiphenyl-2,3-diol 1,2-dioxygenase (CarB), Protocatechuate 4,5-dioxygenase (LigAB), 3,4-dihydroxyphenylacetate (homoprotocatechuate) 2,3-dioxygenase (HPCD) and 4,5-DOPA Dioxygenase, 2,3-dihydroxyphenylpropionate 1,2-dioxygenase.

Dioxygenases are required for the aerobic breakdown of aromatic compounds by bacteria as they catalyse the oxygenolytic fission of catecholic chemicals. Again, extradiol dioxygenases seem to be more adaptable than their intradiol equivalents. Nevertheless, Extradiol dioxygenases can cleave a greater range of substrates, have evolved on more structural scaffolds, and are found in a a greater range of passageways, such as those that break down non-aromatic substances and biosynthetic pathways. (Vaillancourt et al. 2006).

Research conducted by Senda et al. (1996) and Han et al. (1995) revealed X-ray crystal structures of extradiol dioxygenases which has come to known as the 2-His-1-carboxylate facial triad motif for metal binding at the active site. More than 20 families of enzymes with a wide range of catalytic activity can use the motif.

**3.5 Secondary structure detection**

The structure and function of proteins are closely associated. Secondary structural components, namely helices, coils, sheets, and turns, exhibit a remarkable correlation with protein function, structure, and engagement. (Uchôa et al. 2004; Padjaseket al. 2020; Zhang et al. 2018; Rademaker et al. 2020; Wardah et al. 2019). The secondary structure prediction server SOPMA analyzed the proportions of alpha helix, extended strand, and the random coil of the studied protein as 36.36%, 17.32%, and 46.32%, respectively (Fig. 1). Again, The PSIPRED software estimates the helix, strand, and coil of the matrix protein (WP\_020728386.1) with more assurance (Fig. 2).



**Fig. 1. Predicted secondary structure from multiple alignments using SOPMA server.**



**Fig. 2. Target protein’s Secondary structure was determined using PSIPRED server.**

**3.6 Tertiary structure prediction and validation:**

Homology modeling is an efficient technique for getting the three-dimensional structure (3D) of a hypothetical protein based on identity between template and target protein sequences. The sequence of WP\_020728386.1 in FASTA format was inserted into Swiss-modelserver as input for homology modeling. The server completed the BLASTP search against each protein sequence for identifying homologous templates, and the most suitable template (TK2203) was picked with 100% probability rate. GMQE (Global Model Quality Estimation) is 0.60 for this template. The chosen model's GMQE and QMEAN values indicated that it might be a higher-quality and more dependable model. Fig. 3 represents the structure generated through SWISS\_MODEL.



**Fig. 3. Structure in three dimensions that is generated by the SWISS-Model tool.**

The tertiary structure of matrix protein was assessed employing the Ramachandran plot by PROCHECK program (Fig. 4a), which showed that 91.0 percent of the total residues (462) resided in the core [A, B, L]; 8.1 percent of residues were present in the additional allowed regions [a,b,l,p]; and 0.6 percent of residues were in the generously allowed regions [a,b,l,p]. There were 60 and 44 glycine and proline residues, respectively (Table 5). However, it can be said a valid model as 91.0% residues were found in the most favored region. Hooda *et al* suggested that it is acceptable to have more than 90% of the residues in the most favored regions which is likely to be a valid 3D model (Hooda et al. 2012).

The Prosa-eb server was utilised for standard bond angles identification in the predicted tertiary structures of the studied protein WP\_020728386.1. Here, The constructed tertiary structures from the Swiss-Model server had a Z-score of -7.65 (Fig. 4b). Saikat et. al., reported that Z-scores obtained from the ProSA-web reveal the projected tertiary structures' "degree of nativeness." (Saikat et al. 2020).



**Fig. 4. a) Ramachandran plot of the structural model, verified by the PROCHECK program b) The overall model quality by Z-score (-7.65) obtained from the ProSA-web.**

**Table 5. Analysis of the hypothetical protein's Ramachandran plot statistics**

|  |  |  |
| --- | --- | --- |
| **Statistics** | **Residues** | **Percentage (%)** |
| Residues in the most favored regions [A, B, L] | 324 | 91 |
| Residues in the additional allowed regions [a, b, l, p] | 29 | 8.1 |
| Residues in the generously allowed regions [~a, ~b, ~l,~p] | 2 | 0.6 |
| Residues in disallowed regions | 1 | 0.3 |
| Number of nonglycine and nonproline residues | 356 | 100 |
| Number of end-residues (excl, Gly and Pro) | 2 |  |
| Number of glycine residues (shown as triangles) | 60 |  |
| Number of proline residues | 44 |  |
| Total number of residues | 462 |  |

**3.7 Active site determination:**

The active sites were identified for the hypothetical protein's amino acid residues through the Computer Atlas of Surface Topography of Protein (CASTp) website. According to the CASTp calculation result, total 141 {69(chain A), 72 (chain B)} amino acids are included in the potent active site. Based on the predicted active site of the protein, areas covered the best active site 3242.393 which occupied a volume of amino acids 5258.349. Residues of amino acids in the ligand binding site are presented in Fig. 5.



**Fig.5. Active site of the hypothetical protein WP\_020728386.1 (the active sites are highlighted by the red sphere).**

**4. CONCLUSION**

For manipulating biological processes, proteins functional annotation is indispensable. By predicting physicochemical parameters, predicted active site of ligand binding, overall fundamental features of the hypothetical protein provide an insight into the studied protein, identifying it as functional and potential. It is expected that a variety of useful therapeutic goods as well as nutritious food items could be produced from the selected hypothetical protein. Ultimately, this study may booster and broaden our knowledge of pathophysiology which would help to develop potent therapeutic drugs and vaccines against the infection by *Mycobacterium marinum.*

**REFERENCE**

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.

Aronson, J. D. (1926). Spontaneous tuberculosis in salt water fish. *The journal of infectious diseases*, 315-320.

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, 2(9), 1511-1519.

Collins, C. H., Grange, J. M., Noble, W. C., & Yates, M. D. (1985). Mycobacterium marinum infections in man. *Epidemiology & Infection*, 94(2), 135-149.

Combet, C., Blanchet, C., Geourjon, C., & Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends in biochemical sciences*, 25(3), 147-150.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, et al. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic acids research*, 36(suppl\_2), W465-W469.

Dharra, R., Radhakrishnan, V. S., Prasad, T., Thakur, Z., Cirillo, J. D., Sheoran, et. al. (2019). Evaluation of in silico designed inhibitors targeting MelF (Rv1936) against Mycobacterium marinum within macrophages. *Scientific Reports*, 9(1), 10084.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic acids research*, 34(suppl\_2), W116-W118.

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in enzymology,* Academic Press, (277), 396-404.

Amrani, E. M., Adoui, M., Patey, O., & Asselineau, A. (2010). Upper extremity Mycobacterium marinum infection. *Orthopaedics & Traumatology: Surgery & Research*, 96(6), 706-711.

Ferguson, H. W. (2006). Systemic pathology of fish: a text and atlas of normal tissues in teleosts and their responses in disease.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31(13), 3784-3788.

Guruprasad, K., Reddy, B. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2), 155-161.

Haenen, O. L., Evans, J. J., & Berthe, F. (2013). Bacterial infections from aquatic species: potential for and prevention of contact zoonoses. *Revue scientifique et technique (International Office of Epizootics)*, 32(2), 497-507.

Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W., & Bolin, J. T. (1995). Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. *Science*, 270(5238), 976-980.

## Hooda, V., Gundala, P. B., & Chinthala, P. (2012). Sequence analysis and homology modeling of peroxidase from Medicago sativa. *Bioinformation,* 8,974–979.

doi:10.6026/97320630008974

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8, 477-486.

Linell, F., & Nordén, A. (1954). Mycobacterium balnei. A New Acid-Fast Bacillus occurring in Swimming Pools and Capable of producing Skin Lesions in Humans.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, et al. (2012). CDD: conserved domains and protein three-dimensional structure. *Nucleic acids research*, 41(D1), D348-D352.

Mazandu, G. K., & Mulder, N. J. (2012). Function prediction and analysis of Mycobacterium tuberculosis hypothetical proteins. *International journal of molecular sciences*, 13(6), 7283-7302.

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405.

Padjasek, M., Kocyła, A., Kluska, K., Kerber, O., Tran, J. B., & Krężel, A. (2020). Structural zinc binding sites shaped for greater works: Structure-function relations in classical zinc finger, hook and clasp domains. *Journal of Inorganic Biochemistry*, 204, 110955.

Rademaker, D., Van, D. J., Titulaer, W., Lange, J., Vriend, G., & Xue, L. (2020). The Future of Protein Secondary Structure Prediction Was Invented by Oleg Ptitsyn. *Biomolecules*, 10(6), 910.

Saikat, A. S. M., & Ripon, A. (2020). Structure Prediction, Characterization, and Functional Annotation of Uncharacterized Protein BCRIVMBC126\_02492 of Bacillus cereus: An In Silico Approach. *American Journal of Pure and Applied Biosciences*, 2, 104-111.

Saikat, A. S. M., Islam, R., Mahmud, S., Imran, M. A. S., Alam, M. S., Masud, M. H., & Uddin, M. E. (2020). Structural and functional annotation of uncharacterized protein NCGM946K2\_146 of Mycobacterium tuberculosis: an in-silico approach. MDPI, 13(66). (In press).

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic acids research*, 50(D1), D20-D26.

Senda, T., Sugiyama, K., Narita, H., Yamamoto, T., Kimbara, K., Fukuda, M., et al. (1996). Three-dimensional structures of free form and two substrate complexes of an extradiol ring-cleavage type dioxygenase, the BphC enzyme fromPseudomonassp. strain KKS102. *Journal of molecular biology*, 255(5), 735-752.

Shahbaaz, M., Imtaiyaz, H. M. D., & Ahmad, F. (2013). Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. *PloS one*, 8(12), e84263.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6), 1102-1104.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7), 3022-3027.

Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, 46(W1), W363-W367.

Uchôa, H. B., Jorge, G. E., Da Silveira, N. J. F., Camera Jr, J. C., Canduri, F., & De Azevedo Jr, W. F. (2004). Parmodel: a web server for automated comparative modeling of proteins. *Biochemical and biophysical research communications*, 325(4), 1481-1486.

Vaillancourt, F. H., Bolin, J. T., & Eltis, L. D. (2006). The ins and outs of ring-cleaving dioxygenases. *Critical Reviews in Biochemistry and Molecular Biology*, 41(4), 241-267.

Vetrivel, U., Subramanian, G., & Dorairaj, S. (2011). A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *The HUGO journal*, 5, 25-34.

Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., & Marchler-Bauer, A. (2023). The conserved domain database in 2023. *Nucleic Acids Research*, 51(D1), D384-D388.

Wardah, W., Khan, M. G., Sharma, A., & Rashid, M. A. (2019). Protein secondary structure prediction using neural networks and deep learning: A review. *Computational biology and chemistry*, 81, 1-8.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1), W296-W303.

Wayne, L. G., & Sramek, H. A. (1992). Agents of newly recognized or infrequently encountered mycobacterial diseases. *Clinical Microbiology Reviews*, 5(1), 1-25.

Wiederstein, M., & ProSA-Web, M. S. Interactive web service for the recognition of errors in three-dimensional structures of proteins. (2007). 35, pp. *W407-W410. DOI: https://doi. org/10.1093/nar/gkm290*.

Yu, C. S., Chen, Y. C., Lu, C. H., & Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics*, 64(3), 643-651.

Zhang, G. J., Ma, L. F., Wang, X. Q., & Zhou, X. G. (2018). Secondary structure and contact guided differential evolution for protein structure prediction. *IEEE/ACM transactions on computational biology and bioinformatics*, 17(3), 1068-1081.