J. Bangladesh Acad. Sci. 49(2); 227-242: December 2025 DOI:https://doi.org/10.3329/jbas.v49i2.79820

Journal of Bangladesh Academy of Sciences

Journal homepage: http://www.bas.org.bd/publications/jbas.html

Research Article

A computational analysis of a hypothetical protein from *Acinetobacter nosocomialis* identifies it as a multicopper oxidase linked to the copper resistance system

Nizam Uddin and Anika Tasnim*

Bangladesh Council of Scientific and Industrial Research, Dhaka, Bangladesh

ARTICLE INFO

Article History

Received: 12 Feburay 2025
Revised: 21 August 2025
Accepted: 25 August 205

Keywords: *Acinetobacter nosocomialis*, Hypothetical protein, Homology modeling, Virtual screening.

ABSTRACT

Acinetobacter nosocomialis, a nosocomial pathogen, particularly affecting immunocompromised individuals. This species is retaining multidrug resistance and it is increasing over the time. The organism is dependent on human hosts for survival and exerts several methods to circumvent both innate and adaptive immune response. Some of the defence systems are intrigued by extracellular or intracellular proteins of this pathogen. Hence, characterize the unannotated proteins, also called hypothetical proteins, could play a big role in identifying novel treatment targets. In this study, we focused on the characterization of a hypothetical protein (HP) (accession No. KDM58575.1) derived from *Acinetobacter nosocomialis.* In-silico tools was employed to examine the hypothetical proteins, focusing on physicochemical parameters, subcellular localisation, secondary structure, three-dimensional structure, and functional annotation. To study the active site, protein-protein interactions and molecular docking, the bioinformatics tools CASTp, the STRING server, and PyRx were utilized, respectively. The findings identified this protein as a multicopper oxidase involved in detoxification, transport, and metal binding. The 3D structure derived from the SWISS Model server, and it was matched 89.92% similarity to the AlphaFold DB model A0A0M3ADD5.1, a copper oxidase derived from *Acinetobacter sp*. AG1 (gene: A0A0M3ADD5_9GAMM). This protein also showed considerable biological activity and consists of five functional domains, like cupredoxin. The protein-protein interaction also delineates the essential partners required for bacterial viability, whereas KEGG suggested the protein associated with the bacterial two-component signalling system. In docking analysis, the ligand exhibited significant binding activities and suggesting a potential target for treating nosocomials infection. This study highlights the bioinformatics tools for protein characterization and understanding molecular pathways for *Acinetobacter nosocomialis.* Furthermore, it offers significant insights into the development of innovative therapeutic procedures.

Introduction

Acinetobacter spp., a nosocomial pathogen, is accountable for a variety of infections, especially in Immunocompromised individuals or those who have experienced prolonged hospitalisation. These infections are linked to considerable morbidity and mortality rates. (Ainsworth et al., 2017; McConnell and Pachón, 2010). Most Acinetobacter species are widely distributed in the environment and can survive and persist in harsh, desiccated environments (Williams et al., 2016). Acinetobacter species are presently widespread in healthcare environments, especially within intensive care units (ICUs). Anton et al. (2008) say Acinetobacter species can cause pneumonia, urinary tract infections, bacteremia, skin

*Corresponding author:<anika-2016514146@bmb.du.ac>

and soft tissue infections, osteomyelitis, and meningitis. Antibiotic-resistant bacteria can cause more serious sickness and death without proper treatment (Metan et al., 2009). The incidence of extensively drug-resistant (XDR) and multidrug-resistant (MDR) strains makes Acinetobacter species a "serious" concern for the CDC.

The National Cancer Institute of Mexico conducted a study of ICU infections. Acinetobacter species caused 6% of nosocomial infections. These infections were strongly associated with pneumonia (60% of cases) and bloodstream (25% of cases) (Cornejo-Juárez et al., 2015). The rise of drug-resistant Acinetobacter species and other highly resistant bacteria has highlighted the need for alternative antibacterial methods to address broad antibiotic resistance (Williams et al., 2016). Maintaining metal homeostasis is essential and offers many unique targets for bacterial reduction. Copper is needed for cellular processes such as redox balance regulation and enzyme cofactor function (Williams et al., 2016). However, large copper ion concentrations can be hazardous. Fenton chemistry produces hydroxyl radicals that damage important biomolecules when these ions participate (Liochev and Fridovich, 2002).

Iron-sulfur cluster proteins are damaged by copper-sulfur atom interactions and iron ion displacement (Macomber and Imlay, 2009). Copper surfaces damage Escherichia coli and Salmonella spp. outer membranes, according to recent investigations. This process produces hydroxyl radicals, suppresses respiration, and degrades DNA (Warnes et al., 2012). Copper as an antibacterial agent has attracted increased study due to the demand for therapies against antibiotic-resistant organisms. Many diseases have been tested for copper susceptibility in labs. The use of different growth media and experimental techniques makes it impossible to compare these findings.

Hypothetical proteins have been found in *Acinetobacter sp*. Hypothetical proteins (HP) are expressed in organisms but require empirical and chemical validation (Ashrafi et al., 2019; Rahman et al., 2020; Paul et al., 2015). An ORF can predict the creation of these proteins even while there is no evidence for their existence (Paul et al., 2015; Rahman et al., 2020). Putative proteins account for 50% of protein-coding regions in most genomes, and their functions are unknown (Paul et al., 2015; Mazumder et al., 2021). In hypothetical proteins, there are two types: uncharacterized protein families (UPF) and domains of unknown function (DUF) (Rahman et al., 2020). Uncharacterized protein families (UPF) lack a clear relationship with genes but are identified through experimentation. On the other hand, proteins identified empirically lacking functional or structural domains are called DUFs (Rahman et al., 2020). Studying the structural and functional UPF and DUF can reveal domains and motifs, pathways and cascades, and protein networks. This study will identify pharmaceutical targets and help in diagnosis (Ashrafi et al., 2019; Paul et al., 2015). Recently, a number of bioinformatics tools have been used to characterize proteins, including three-dimensional structural conformations, the discovery of novel domains and pathways, phylogenetic profiling, and the provision of functional annotations (Ashrafi et al., 2019; Rahman et al., 2020). This research was used to characterize a hypothetical protein (KDM58575.1) identified in *Acinetobacter nosocomialis*. Multiple bioinformatics tools and databases were used to gain a thorough understanding of the protein's physical and structural characteristics, as well as its possible functions. Furthermore, molecular docking was performed to elucidate the structural features and potential functions of this protein.

Materials and Methods

Sequence Retrieval

At first, the term "Hypothetical proteins AND *Acinetobacter nosocomialis*" was searched in NCBI, and the accession number KDM58575.1 (hypothetical protein) was retrieved in FASTA format. A sequence-based peptide search was also conducted using the UniProt database (https://www.uniprot.org/peptidesearch/) to assess protein redundancy. The sequence was subsequently analyzed using various

computational tools for in-silico characterization and functional analysis (Table 1).

Physicochemical properties

The ProtParam tool on the ExPASy server application (http://web.expasy.org/protparam/) (Gasteiger et al., 2005) was used to analyze various physical and chemical properties of the protein. Among these were the grand average of hydropathicity (GRAVY), instability index, aliphatic index, extinction coefficient, estimated half-life, molecular weight, theoretical pI, amino acid, and atomic composition, and the total number of negatively and positively charged residues (Asp+Glu and Arg+Lys) (Gasteiger et al., 2005).

Identification of Subcellular Localization

Accurate determination of subcellular localization is critical for proper protein function and genetic analysis. Since drug and vaccine targets are situated within cells, understanding this information is crucial for discovering and developing new pharmaceuticals. Subcellular localization and protein topology were analyzed using CELLO v.2.5, which employs a two-level support vector machine (SVM) prediction system (Yu et al., 2006). The localization predictions from CELLO were further cross-validated with results from HMMTOP v.2.0 (Tusnády and Simon, 2001), TMHMM v.2.0 (Krogh et al., 2001; Möller et al., 2001), and PSORTb v3.0 (Yu et al., 2010).

Determining the domain, coil, folding pattern, protein family, and superfamily

The conserved domain was predicted using the NCBI CD tool (Marchler-Bauer et al., 2011) via Reverse Position-Specific BLAST (RPS-BLAST). Evolutionary relationships were analyzed using GenomeNet (Kanehisa et al., 2002), Pfam (Finn et al., 2014), the SuperFamily program (Wilson et al., 2007), and the ScanProsite tool (Sigrist et al., 2012).

Pfam, a database of protein families, uses hidden Markov models (HMMs) for annotation and multiple sequence alignment. Protein sequence motifs were detected using the MOTIF search tool (https://www.genome.jp/tools/motif/) and the InterProScan tool.

Analysis of Multiple Sequence Alignment and Phylogeny

Protein sequences with similar functionalities were retrieved by querying the NCBI non-redundant protein sequence (nr) database using BLASTp with default settings. Multiple sequence alignment was performed using EBI's MUSCLE server (https://www.ebi.ac.uk/Tools/msa/muscle/) (Edgar, 2004) and visualized using CLC Sequence Viewer 7.0.2 (http://www.clcbio.com). Phylogenetic analysis was performed utilizing the Phylogeny online platform (http://phylogeny.lirmm.fr/).

Secondary structural assessment

The SOPMA program was used with default parameters, including a window width of 17, a similarity threshold of 8, and 4 states (Combet et al., 2000). The PSIPRED version 4.0 (Buchan and Jones, 2019) is used to determine additional secondary structure and protein topology.

Three-dimensional structure prediction and assessment of Model quality

The tertiary structure of the uncharacterized protein was modelled to examine its three-dimensional configuration and folding characteristics. The Swiss Model program (Bitencourt-Ferreira and de Azevedo, 2019) and the D-I-TASSER servers were used to generate the protein's tertiary structure. The HHpred tool was used to search for the best template, including the SWISS model template, for further analysis (Biegert et al., 2006; Gabler et al., 2020; Zimmermann et al., 2018).

PROCHECK and Verify3D programs within the SAVES (v.6.0) platform were used to validate the protein structure (Bowie et al., 1991). Using the ProSA-web application, Z-score and modelled 3D-structure validation were performed (Wiederstein and Sippl, 2007).

Table 1. Sequence-Based Function Annotation Databases and Bioinformatics Tools.

Sl no.	Software	Function	Reference
1.	BlastP	Identifies similar sequences within protein databases	Altschul et al., 1997
2.	MUSCLE	Performs multiple sequence alignment	Edgar, 2004
3.	ExPASy-Protparam tool	Calculates physical and chemical properties of proteins	Gasteiger et al., 2003
4.	CELLO	Predicts localization of proteins in prokaryotic and eukaryotic cells	Yu et al., 2004
5.	PSLpred	Predicts subcellular localization for proteins in Gram-negative bacteria	Bhasin et al., 2005
6.	PSORTb	Predicts the subcellular localization of bacterial proteins	Yu et al., 2010
7.	SOPMA	Predicts the secondary structure of proteins	Mazumder et al., 2021
8.	PSIPRED	Analyzes protein secondary structure using PSI-blast	Jones, 1999
9.	HHpred	Detects protein homology using HMM-HMM comparison	Zimmermann et al., 2018
10.	PyMOL	Facilitates structural analysis and generation of model figures	Mazumder et al., 2021
11.	YASARA	Enhances the stability of 3D protein models	Krieger et al., 2009
12.	PROCHECK's Ramachandran plot analysis	Evaluates the quality and accuracy of predicted 3D protein structures	Laskowski et al., 1993
13.	Verify3D	Assesses protein models using 3D profiles	Eisenberg et al., 1997
14.	ERRAT	Analyzes nonbonded atomic interactions to validate protein structures	Colovos andYeates, 1993
15.	CD Search	Identifies conserved structural and functional domains within sequences	Marchler-Bauer et al., 2011
16.	InterProScan	Searches InterPro for motif and domain discovery	Jones et al., 2014
17.	STRING	Predicts protein-protein interactions	Szklarczyk et al., 2019
18.	CASTp	Identifies and measures surface regions within 3D protein structures	Tian et al., 2018

Energy minimization of the model structure

The 3D structure was energy-minimized using YASARA (Krieger et al., 2009). The PDB format of the hypothetical protein was submitted to the server, which optimizes the protein structure and provides an accurate, stable 3D conformation.

Active Sites Determination

Active Sites were determined by the Computed Atlas of Surface Topography of Proteins (CASTp) algorithm (http://sts.bioengr.uic.edu/castp/) (Tian et al., 2018). The CASTp results were subsequently visualised using PyMOL software (Mazumder et al., 2021).

Analysis of Protein-protein interaction

A computational software, STRING (version 11.5), was used to elucidate protein-protein interactions (Szklarczyk et al., 2018).

Analysis of Molecular Docking

Molecular Docking was performed to investigate the interactions between the protein and two distinct ligands. Previous research has not identified the ligands association with this potential protein. Ligand selection was based on the most structurally similar protein in the Protein Data Bank (PDB). This study investigated the compounds that are agonists or antagonists of multicopper oxidases (MCOs). MCOs, a redox enzymes, helps convert molecular oxygen into water by transferring four electrons.

Based on their expected high-affinity binding to the protein, two ligands were chosen: NAG (2-acetamido-2-deoxy-beta-D-glucopyranose) and beta-D-Mannopyranose. The ligands structure were obtained from PubChem and later converted to PDB format using Discovery Studio software. Docking was performed using PyRx, and interactions were visualized using PyMOL and Discovery Studio. Chimaera was also used to examine hydrophobic and hydrogen-bond interactions within 5 Å of the receptor and ligands. Fig.1. illustrates the workflow of the study.

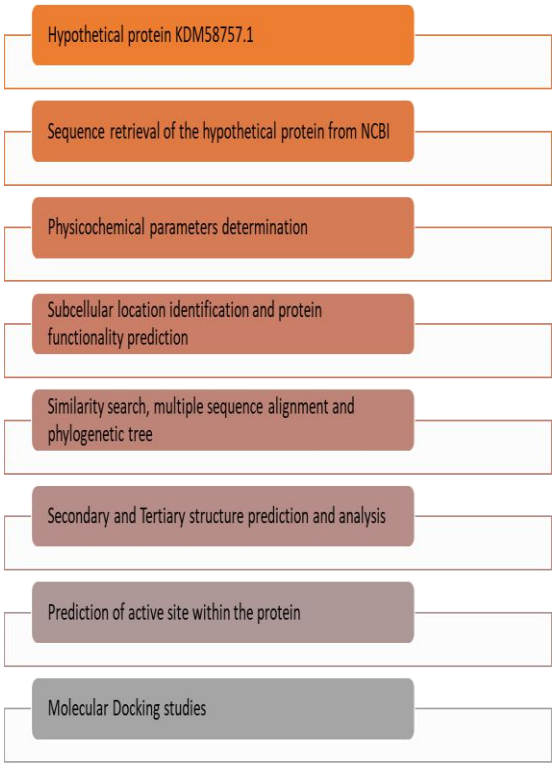

Fig. 1. A comprehensive study workflow.

Results and Discussion

Information about sequence retrieval and similarity

The protein's amino acid sequence of 645 residues was obtained in FASTA format from the National Centre for Biotechnology Information (NCBI) database. The sequence was subsequently analyzed using UniProt. The UniProt database entries confirmed that the protein is non-redundant, which suggests its potential significance. Additional details retrieved from the NCBI database are presented in Table 2.

The ProtParam tool was used to obtain a full picture of its physicochemical properties. There are 645 amino acids in the protein. The most common ones are serine (n=56, 8.7%) and methionine (n=55, 8.5%). It is important to note that the sequence does not include selenocysteine or pyrrolysine.

Table 2. Protein distinctiveness and protein data.

Parameters	Protein Information
Locus	KDM58575
Amino Acid	645 aa
Definition	hypothetical protein AE32_00579 [*Acinetobacter nosocomialis*]
Accession	KDM58575
Version	KDM58575.1
GenBank ID	KDM58575.1
Source	*Acinetobacter nosocomialis*
Organism	*Acinetobacter nosocomialis*

Determination of Physicochemical Parameters

The half-life of the protein, which is the time it takes for the concentration of a radiolabeled protein to drop by 50% during the chase phase (Zhou, 2004), was about 30 hours in vitro in mammalian reticulocytes, over 20 hours in vivo in yeast, and more than 10 hours in vivo in Escherichia coli. The isoelectric point (pI), total atom count, and molecular weight were found to be 6.34, 10,041, and 72,740.88 Daltons, respectively (Table 3).

The protein comprises 72 positively charged residues (Arg + Lys) and 81 negatively charged residues (Asp + Glu). The instability index (40.61) indicates the protein's stability, whereas the aliphatic index (63.60) denotes its structural robustness over a wide spectrum of temperatures. Furthermore, the overall average hydropathicity (GRAVY) value of -0.576 indicates an improvement in thermostability (Pawel et al., 2007).

Subcellular location determination and protein topology prediction

The classification of the proteome, identification of potential therapeutic targets, and development of vaccines all depend on automated prediction of bacterial protein subcellular localization. In recent years, researchers have developed seversl subcellular localisation predictors, including both generalized and feature-based predictors (Restrepo-Montoya et al., 2009; Schneider and Fechner, 2004). Using the CELLO (v.2.5) and PSORTb (v3.0) computational methods, it was anticipated that the protein would be found outside of the cell (Yu et al., 2004).

Table 3. Protein Physicochemical parameters determination.

Parameter	Value
Molecular Weight	72740.88
Formula	C3159H4966N884O976S56
Total number of atoms	10041
Theoretical pI	6.34
The total number of positively charged residues (Arg + Lys)	72
The total number of negatively charged residues (Asp + Glu)	81
The measured half-life	30 hours (mammalian reticulocytes, in vitro) over 20 hours (yeast, in vivo) over 10 hours (Escherichia coli, in vivo)
Aliphatic index	63.60
Instability index (II)	40.61
Grand average of hydropathicity	-0.576

Moreover, one of the most important aspects of bioinformatics is the detection of transmembrane helices in integral membrane proteins. Furthermore, the most successful current methodologies aim to predict the comprehensive topology of proteins, encompassing the total number of transmembrane helices and their orientation with respect to the membrane (Krogh et al., 2001; Möller et al., 2001). A single transmembrane helix was predicted to be present in the protein by the TMHMM version 2.0 and HMMTOP version 2.0 software programs (Table 4).

Forecasting the domain, coil, folding pattern, protein family, and superfamily

The preliminary objective of the Conserved Domain Database (CDD) is to provide annotations for

Table 4. Subcellular location Identification.

Web Tools Analysis	Result
CELLO v2.5	Periplasmic
PSORTb v3.0	Periplasmic
HMMTOP v.2.0	one transmembrane helix
TMHMM v.2.0	one transmembrane helix
CCTOP	one transmembrane helix

molecular sequences by identifying and assigning evolutionarily conserved protein domains. NCBI's Entrez database maintains a collection of pre-computed domain identification for proteins, along with real-time search capabilities. The protein was categorized by the CDD tools as a multicopper oxidase belonging to the copper resistance system. This protein, with a domain architectural ID of 1001039, exhibits significant similarity to the copper resistance protein A found in Escherichia coli. The latter is known to play a crucial role in the copper-inducible production of copper resistance and is believed to possess oxidase function. The GenomeNet program has successfully identified five motifs in the given dataset. These motifs include multicopper oxidase, which is located between positions 32 and 146, with an independent E-value of 1.1*10-34. Another instance of multicopper oxidase is found between positions 159 and 317, with an independent E-value of 3.8*10-26. Additionally, there is a multicopper oxidase motif between positions 505 and 617, with an independent E-value of 2.9*10-27. Another motif identified is the cupredoxin-like domain, spanning positions 529-610, with an independent E-value of 0.92. Lastly, a motif corresponding to a protein of unknown function is located between positions 206 and 233, with an independent E-value of 0.2. The five motifs, namely multicopper oxidase (accession id PF07732), multicopper oxidase (accession id PF00394), multicopper oxidase (accession id PF07731), cupredoxin-like domain (accession id PF13473), and protein of unknown function (accession id PF05339), were validated by both the Pfam program (Finn et al., 2014) and the ScanProsite tool.

Examination of alignment, phylogenetic tree, and sequence similarity

The hypothetical protein's (HP) BLASTp analysis against non-redundant databases revealed a high degree of homology with other copper oxidase proteins from different species. Table 5 lists the copper oxidase proteins extracted from the multiple sequence alignment (MSA) generated by BLASTp. The hypothetical protein's sequence matched that of other copper oxidases, as indicated by the MSA (Fig. 2). A phylogenetic analysis was performed to further validate the homology and to investigate the evolutionary relationship among the aligned oxidases and the target protein. Based on the MSA and BLASTp results, the phylogenetic tree confirmed the hypothetical protein's homology and provided additional information on evolutionary distance (Figs. 2 and 3).

Table 5. A protein BLASTp search analysis was used to find homologs of KDM58575.1.

Accession	Description	Scientific Name	Total Score	E value	Per. ident
WP_050488919.1	copper resistance system multicopper oxidase	*Acinetobacter nosocomialis*	1340	0	100
WP_086376332.1	copper resistance system multicopper oxidase	*Gammaproteobacteria*	1333	0	99.38
MCT9255433.1	copper resistance system multicopper oxidase	*Acinetobacter baumannii*	1325	0	98.6
WP_006582181.1	multicopper oxidase domain-containing protein	*Acinetobacter nosocomialis*	1324	0	95.56
WP_098715181.1	copper resistance system multicopper oxidase	*Acinetobacter baumannii*	1323	0	98.45

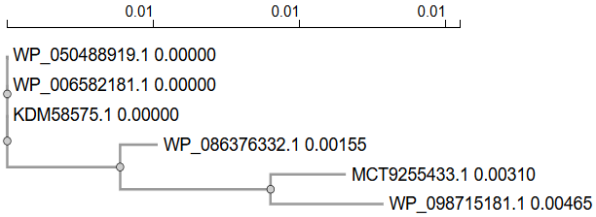

Fig. 2. Phylogenetic trees depicting the true distances between various oxidase proteins are shown. The scale bar in the tree estimates sequence divergence, while the line segments, labeled with the number (0.01), represent the degree of genetic change.

prediction using SOPMA and PSI PRED produced consistent results. The SOPMA analysis identified alpha-helices (n=51, 7.91%), extended strands (n=142, 22.02%), and random coils (n=452, 70.02%) (Fig. 4), emphasizing the predominance of random coil structures.

Tertiary Structure Analysis, Energy minimization, and assessment of Model quality

The Swiss-Model server was used to analyze the query sequence in order to predict the structure and determine protein homology; the Alpha Fold DBmodel of

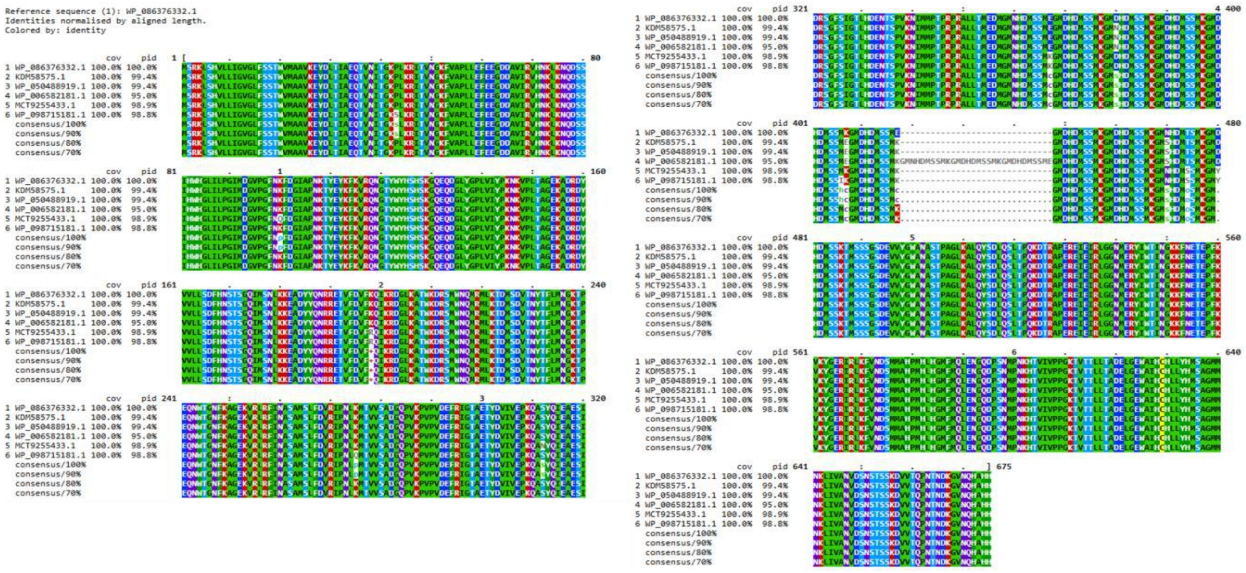

Fig. 3. MSA analysis among the different types of copper resistance system multicopper oxidase with the KDM58575.1 (generated by Mview Version 8).

Assessment of Secondary Structure

Static high-resolution structures have greatly enhanced our understanding of protein architecture and molecular mechanisms. However, developments in structural biology have shown that the molecular basis of protein structure and function in solution cannot be fully understood by high-resolution structures alone (Hodge et al., 2020; Hu et al., 2018; Uversky, 2019). Protein functionality, structural organization, and intermolecular interactions are all significantly influenced by secondary structural elements like helices, sheets, coils, and turns (Hegyi and Gerstein, 1999; Hodge et al., 2020). Secondary structure

A0A0M3ADD5.1, a copper oxidase from *Acinetobacter* sp. AG1 (gene: A0A0M3ADD5_9GAMM), as illustrated in Fig. 5A.

Additionally, 3D structures were generated using the D-I-TASSER server (Fig. 5B).

The quality of the predicted three-dimensional models was assessed using PROCHECK, which analyzed φ and ψ angle distributions using Ramachandran plot. For the Swiss-Model prediction, 85.7% of residues were found in the most favored regions (Table 6, Fig. 6A), indicating a high-quality model. In contrast, the model generated by D-I-TASSER demonstrated lower quality and was excluded from further analysis.

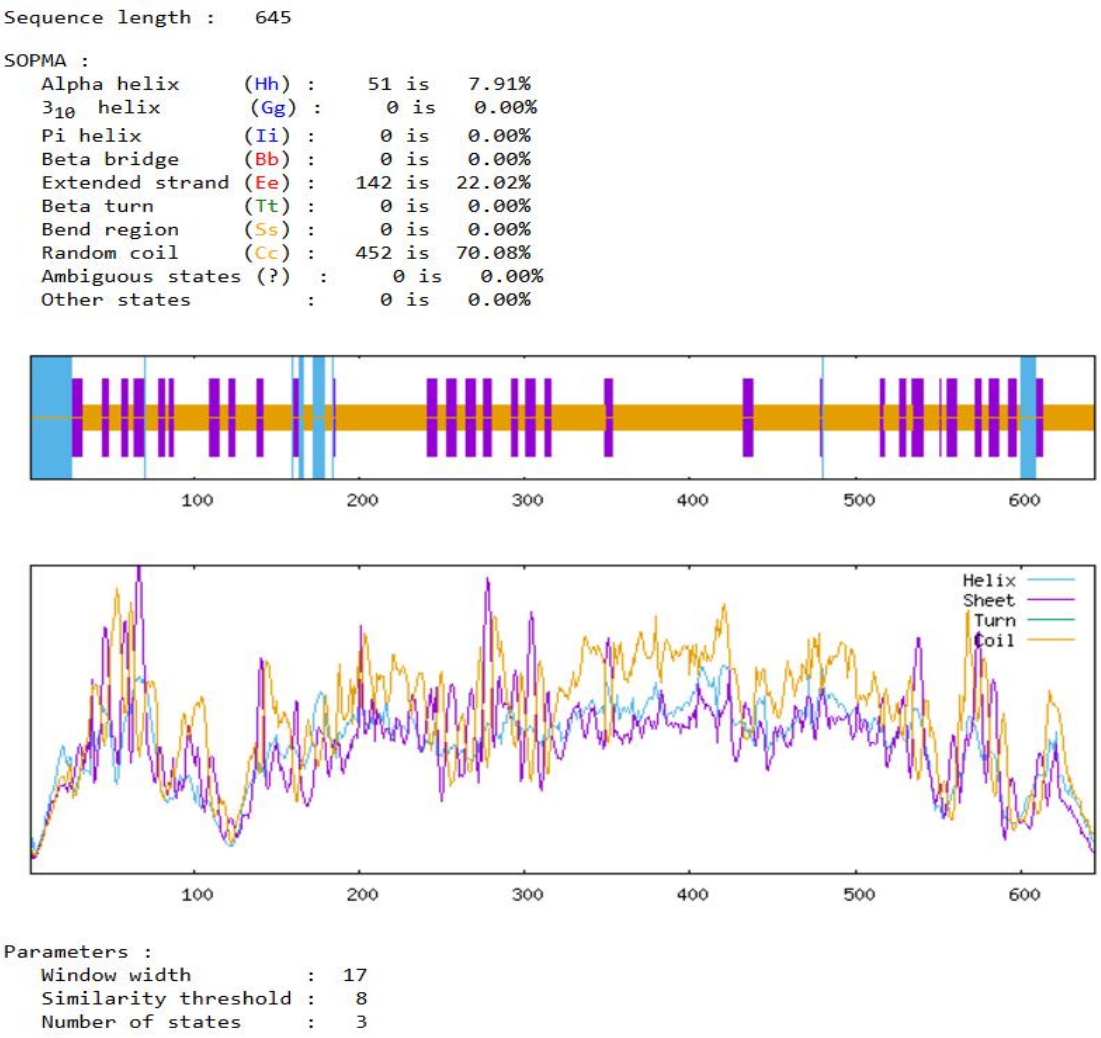

Fig. 4. The predicted secondary structure is visualized by the SOPMA server, which gives information about the coils, sheets, and helices that the protein is expected to contain based on its amino acid sequence.

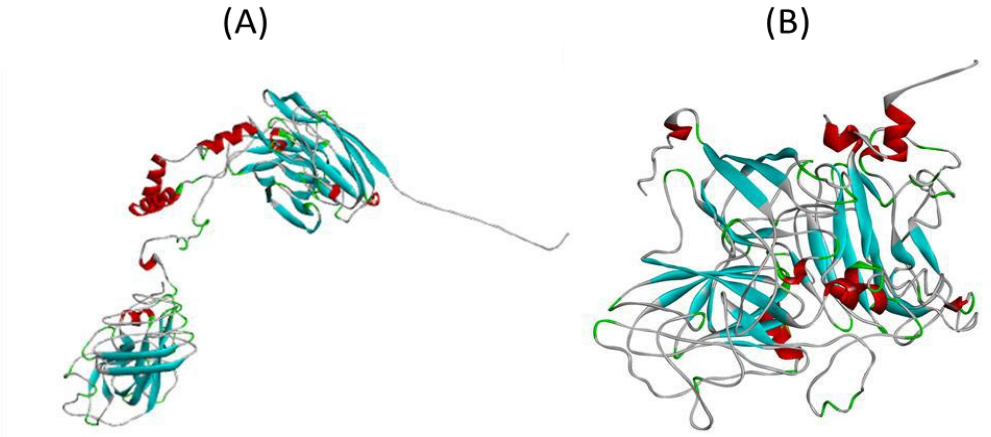

Fig. 5. Predicted three-dimensional structure of the KDM58575.1 by Swiss model (A) and I-TASSER (B) (visualized by Discovery Studio).

Table 6. The hypothetical protein's Ramachandran plot statistics

Statistics for Ramachandran plots	No. (%)
The most preferred regions' residues ⌊A, B, L⌋ 487	85.7
Residues in the additional allowed regions ⌊a, b, l, p⌋ 64	11.3
Residues in the generously allowed regions ⌊a, b, l, p⌋ 9	1.6
residues in the prohibited areas: 8	1.4
Number of non-glycine and non-proline residues: 568	100
Number of end-residues (excl. Gly and Pro): 2	
Number of glycine residues (shown in triangles): 47	
Number of proline residues: 28	
Total Number of residues: 645	

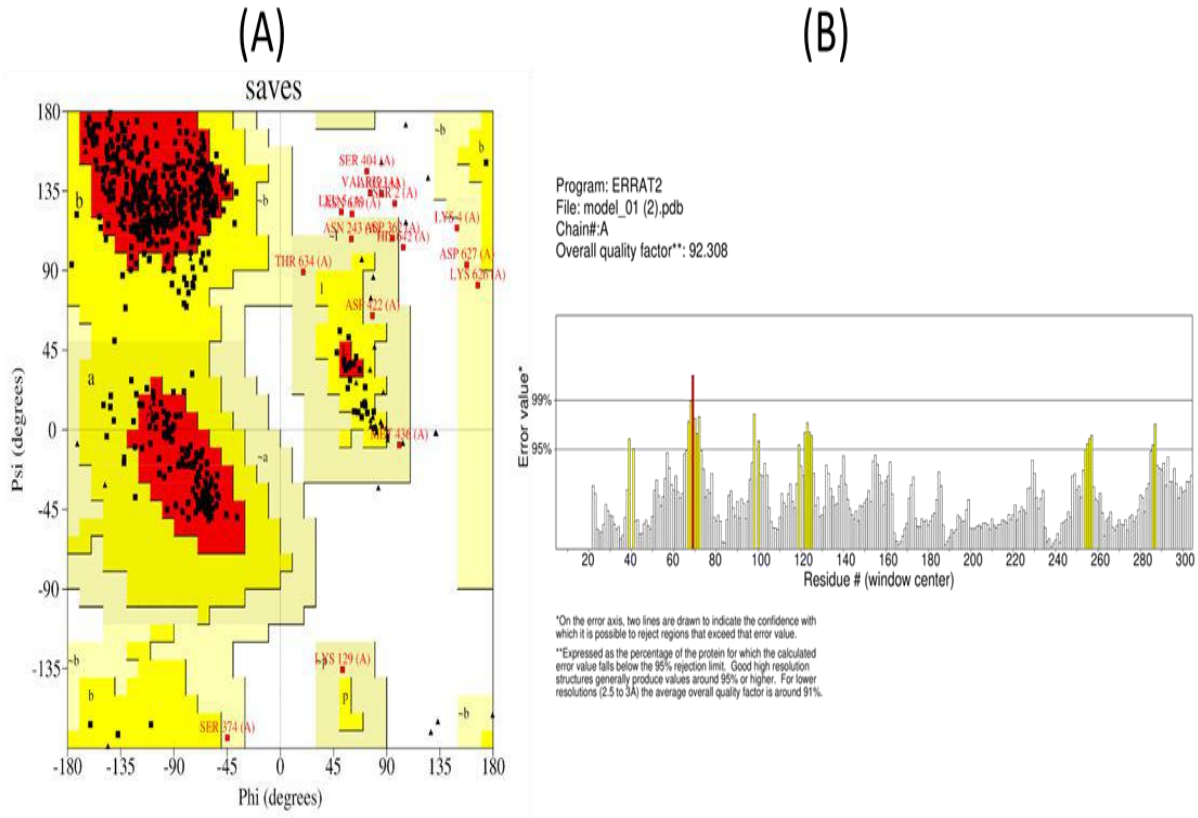

Fig. 6. (A) The Ramachandran plot of the predicted model through the Swiss Model, validated using the PROCHECK server, illustrates the tertiary structure of the hypothetical protein (HP). The beta-sheet region is represented in the first quadrant, while the right-handed and left-handed alpha-helix regions are depicted in the second and third quadrants, respectively. The plot's color-coded regions indicate residue distribution: red represents the most favored regions, yellow denotes additional allowed regions, gray indicates generously allowed regions, and white corresponds to disallowed regions. (B) The ERRAT server, which supplied the model's quality factor, was used to evaluate the target model's overall quality.

Verify3D and ERRAT were employed to verify 3D structures. The SWISS model's strong environmental profile was validated by the Verify3D analysis, which revealed that 92.73% of residues had an average 3D-1D score ≥ 0.2. The ERRAT analysis produced an overall quality factor of 92.308, signifying a reliable model (Fig. 6B).

The YASARA energy minimization server was used to fine-tune the protein. The energy of the structure prior to minimization was -238,197.1 kJ/mol, which improved significantly to -327,989.1 kJ/mol after three rounds of steepest descent energy minimization. Finally, the ProSA-web tool was employed to evaluate the model, yielding a Z-score of -6.76, consistent with a high-quality structural prediction.

Active site of the hypothetical protein

Identifying and characterizing active site residues are crucial steps in designing drugs or inhibitors. The CASTp server was used to analyze the active site of the modeled protein structure and identify the active-site amino acid residues. With a volume of 2779.748 and an area of 906.501, one of the largest pockets contained the main active sites. The active residues of the model protein as predicted by CASTp are shown in Fig. 7.

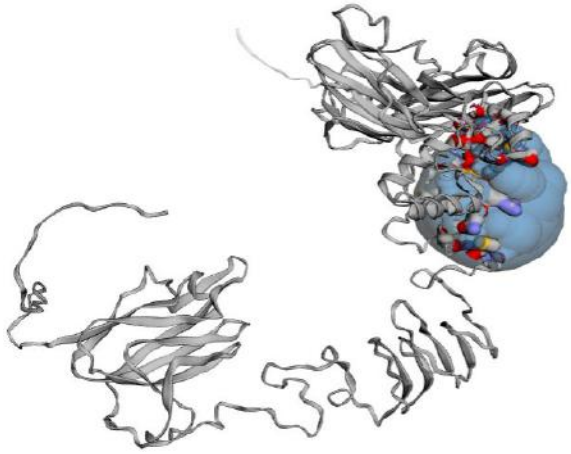

Fig. 7: The active site of KDM58575.1 was located using the CASTp server. The regions containing 906.501 and 2779.748 amino acid volumes had some of the largest active sites.

Protein-protein interaction

To attain a comprehensive understanding of biological phenomena, it is essential to consider proteins and their interrelationships with other proteins. However, current knowledge regarding protein-protein interactions is inadequate. The aim of the STRING database is to compile, assess, and integrate all publicly accessible sources of information on protein-protein interactions, while including computational predictions to improve its comprehensiveness. The purpose of this initiative is to construct a comprehensive and unbiased global network, encompassing both main (physical) and secondary (functional) connections (Marsh and Teichmann, 2014; Sowmya and Ranganathan, 2013; M. Wang et al., 2015). The protein-protein (pr-pr) interaction was identified using the STRING software (v.11.5). The program involving strings showcased the performance of various functional colleagues, as indicated by their respective scores: ARG18227.1 (0.489), ARG18228.1 (0.484), ARG18229.1 (0.506), ARG18230.1 (0.980), ARG18232.1 (0.749), ARG18233.1 (0.579), ARG18234.1 (0.622), ARG18235.1 (0.697), ARG18892.1 (0.916), ARG18900.1 (0.542) (Figure 8). Gene Ontology analysis indicates that this hypothetical protein plays a role in cellular copper ion homeostasis and copper ion binding (Table 7). Additionally, KEGG pathway analysis reveals its involvement in the bacterial two-component system of *Acinetobacter nosocomialis* (Table 7). These two-component signal transduction systems allow bacteria to detect, respond to, and adapt to environmental or intracellular changes. Each system typically comprises a sensor histidine kinase (HK) and a corresponding response regulator (RR).

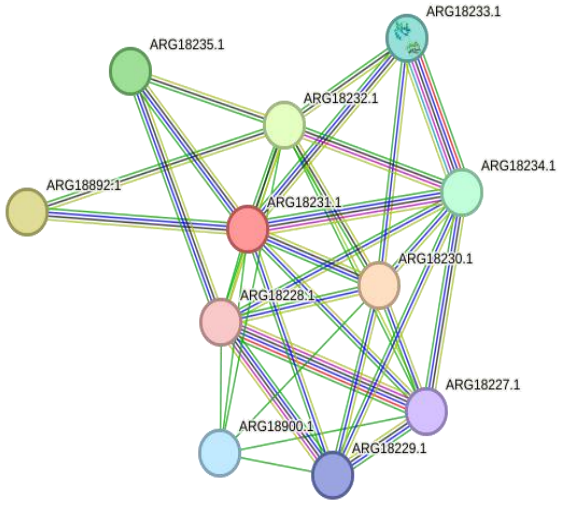

Fig. 8. Proteins with known or predicted 3D structures are represented as filled nodes, while unknown proteins are shown as empty nodes in a string network analysis of the hypothetical protein ARG18231.1.

Table 7. Functional Enrichment Analysis of the Hypothetical Protein.

Category of the function	Term ID	Description	Strength	Signal	False Discovery Rate(FDR)
GO Process	GO:0006878	Cellular copper ion homeostasis	2.54	1.77	0.00054
GO Function	GO:0005507	Copper ion binding	2.03	2.05	7.22E-05
STRING clusters	CL:5237	Mixed, incl. Copper binding periplasmic protein CusF, and Cellular copper ion homeostasis	2.24	5.01	1.46E-11
STRING clusters	CL:5164	Mixed, incl. Efflux transmembrane transporter activity, and HlyD family secretion protein	1.42	2.37	8.87E-10
STRING clusters	CL:5239	Copper binding periplasmic protein CusF, and AcrB/AcrD/AcrF family	2.44	3.2	6.94E-07
KEGG	ano02020	Two-component system	1.29	0.94	0.0046

Table 8. The summary provides information about the range and precision of interactions involving proteins.

Ligand	Binding Affinity	rmsd/ub	rmsd/lb	Interacting Molecules
2-acetamido-2-deoxy-beta-D-glucopyranose	-5.3	0	0	ARG255, GLU154, ALA15
beta-D-Mannopyranose	-5.7	0	0	GLN559, LEU556, PHE55 ASP590, ASN571

Molecular Docking Analysis

A structure-based approach to drug design, the molecular docking method predicts the mechanisms by which ligands and receptors bind by simulating molecular interactions. Docking analyses were used to investigate the interactions between a host protein and potential inhibitors using two ligands: NAG (2-acetamido-2-deoxy-beta-D-glucopyranose) and beta-D-mannopyranose. NAG is frequently used in computational docking studies as a representative sugar moiety due to its prevalence in bacterial cell walls and in host–pathogen interactions. We chose β-D-mannopyranose because sugars containing mannose are often used in studies of how microbes recognise and bind to substrates. We used Autodock Vina for docking studies and a targeted docking method to test substrate specificity. After the docking process, Discovery Studio was used to examine the interactions in more detail.

This study showed several interactions between the ligands and the protein. Beta-D-mannopyranose showed a significant propensity for binding affinity for the protein, with a docking score of -5.7 kcal/mol. NAG had a somewhat lower binding affinity, as shown by a docking score of -5.3 kcal/mol (Table 8 and Fig. 9). The compounds with the lowest docking scores were considered to have significant affinities,

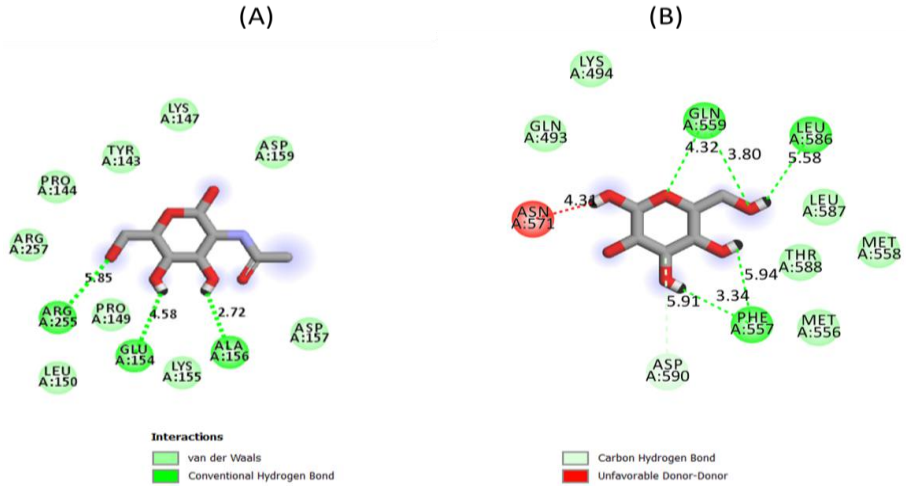

Fig. 9. The target protein's (HP) and putative ligands' intermolecular interactions are depicted as follows: (A) The target protein interacts with the ligand NAG (2-acetamido-2-deoxy-beta-D-glucopyranose); (B) the target protein interacts with the ligand beta-D-Mannopyranose.

suggesting they could be good inhibitors. This research provides valuable insights into where ligands are located and how they fit within the protein's binding site. This information is important for the design and improvement of therapeutic drugs targeting this protein.

Docking results indicated that 3-4 amino acid residues of the protein form hydrogen bonds with the bound complexes of NAG and beta-D-mannopyranose. Specifically, these interactions pertain to the residues ARG255, GLU154, and ALA156. Furthermore, complexes containing beta-D-mannopyranose ligands engage with residues GLN559, LEU556, and PHE557 (Fig. 9). These interactions, especially the hydrogen bonds within the active site residues, offer crucial insights into the mechanisms of residue-ligand binding, conformational architecture, and function, clarifying significant aspects of the protein's bilogical structure and function.

Conclusion

This study focused on the characterization of a hypothetical protein from *Acinetobacter nosocomialis*. The findings revealed that it is a multicopper oxidase involved in the copper resistance system. Through in silico analysis, the physicochemical characteristics, subcellular localisation, and structural and functional features of the protein were thoroughly characterised. The multicopper oxidase, which detoxifies, transports, and binds metal ions, indicates its involvement in bacterial survival. The SWISS-modeled protein's 3D structure confirmed its functional domain by identifying copper oxidase. The findings also revealed protein-protein interaction partners for essential bacterial pathways and KEGG pathways with respect to the bacterial two-component system. In a molecular docking study, high-binding ligands were identified as having therapeutic target potential. This examined protein may enable the pathogen *Acinetobacter nosocomialis* to be less resistant to copper toxicity, making it a possible diagnostic target for multidrug-esistant infections. Finally, these findings suggest that experimental validation and drug development are needed to address emerging diseases caused by *Acinetobacter nosocomialis.*

Acknowledgment

The authors acknowledge the Bangladesh Council of Scientific and Industrial Research for funding the study.

Authors contribution

Anika Tasnim: Conceptualization, formal analysis, methodology, review and editing; Nizam Uddin: Data curation, formal analysis, methodology.

Conflict of interest

We hereby declare that we have no conflict of interest regarding this paper.

References

Ainsworth S, Ketter PM, Yu JJ, Grimm RC, May HC, Cap AP, Chambers JP, Guentzel MN and Arulanandam BP. Vaccination with a live attenuated Acinetobacter baumannii deficient in thioredoxin provides protection against systemic Acinetobacter infection. *Vaccine*, 2017; 35(26): 3387-3394.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*1997; 25(17): 3389-3402.

Anton Y, Peleg HS and David LP. Acinetobacter baumannii: Emergence of a Successful Pathogen Clinical Microbiology Reviews, 2008; 21(3): 538-582.

Ashrafi H, Siraji MI, Showva NN, Hossain MM, Hossan T, Hasan MA, Shohael AM and Shawan MMAK. Structure to function analysis with antigenic characterization of a hypothetical protein, HPAG1_0576 from Helicobacter pylori HPAG1. *Bioinform.* 2019; 15(7): 456-466.

Bhasin M, Garg A and Raghava GPS. PSLpred: Prediction of subcellular localization of bacterial proteins. *Bioinform.* 2005; 21(10): 2522-2524.

Biegert A, Mayer C, Remmert M, Söding J and Lupas A. The MPI Toolkit for protein sequence analysis. *Nucleic Acids Res.* 2006; 34: W335-W339.

Bitencourt-Ferreira G and de Azevedo WF. Homology modeling of protein targets with MODELLER. *Methods Mol. Biol.* 2019; 2053: 231-249.

Bowie J, Luethy R and Eisenberg D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science*, 1991; 253: 164-170.

Buchan D and Jones D. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.*2019; 47: W402-W407.

Colovos C and Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1993; 2(9): 1511-1519.

Combet C, Blanchet C, Geourjon C and Deleage G. NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* 2000; 25(3): 147-150.

Cornejo-Juárez P, Vilar-Compte D, Pérez-Jiménez C, Ñamendys-Silva SA, Sandoval-Hernández Sand Volkow-Fernández P. The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *Int. J. Infect. Dis.* 2015; 31: 31-34.

Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*2004; 32(5): 1792-1797.

Eisenberg D, Lüthy R and Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 1997; 277: 396-404.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate Jand Punta M. Pfam: The protein families database. *Nucleic Acids Res.* 2014; 42(D1): D222-D230.

Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas A and Alva V. Protein sequence analysis using the MPI bioinformatics toolkit. *Curr.Protoc. Bioinform.* 2020;72(1):e108.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD and Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003; 31(13): 3784-3788.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel RD and Bairoch AM. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*, 2005; pp. 571-607.

Hegyi H and Gerstein M. The relationship between protein structure and function: A comprehensive survey with application to the yeast genome. *J. Mol. Biol.*1999; 288: 147-164.

Hodge EA, Benhaim MA and Lee KK. Bridging protein structure, dynamics, and function using

hydrogen/deuterium-exchange mass spectrometry. *Protein Sci.* 2020; 29(4): 843-855.

Hu J, Han J, Li H, Zhang X, Liu LL, Chen F and Zeng B. Human embryonic kidney 293 cells: A vehicle for biopharmaceutical manufacturing, structural biology, and electrophysiology. *Cells Tissues Organs*, 2018; 205(1): 1-8.

Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*1999; 292(2): 195-202.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R and Hunter S. InterProScan 5: Genome-scale protein function classification. *Bioinform*. 2014; 30(9): 1236-1240.

Kanehisa M, Goto S, Kawashima S and Nakaya A. The KEGG databases at genomenet. *Nucleic Acids Res.* 2002; 30: 42-46.

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson, J, Tyka M, Baker D and Karplus K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct. Bioinf.* 2009; 77(S9): 114-122.

Krogh A, Larsson B, von Heijne Gand Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.*2001; 305(3): 567-580.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 1993; 26(2): 283-291.

Liochev S and Fridovich I. The Haber-Weiss cycle—70 years later: An alternative view. Redox Report : *Communications in Free Radical Research*, 2002; 7: 55-57.

Macomber L and Imlay J. The iron-sulfur clusters of dehydratases are primary targets of copper toxicity. *Proc. Natl. Acad. Sci. USA*. 2009; 106: 8344-8349.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV and Bryant SH. CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 2011; 39(suppl_1): D225-D229.

Marsh J, Teichmann S. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* 2014; 84: 551-575.

Mazumder L, Hasan M, Abu Rus'd Aand Islam M. In-silico characterization and structure-based functional annotation of a hypothetical protein from Campylobacter jejuni involved in propionate catabolism. *Genomics Inform.* 2021;19: e43.

McConnell MJ and Pachón J. Active and passive immunization against Acinetobacter baumannii using an inactivated whole cell vaccine. *Vaccine*, 2010; 29(1): 1-5.

Metan G, Sariguzel F and Sumerkan B. Factors influencing survival in patients with multi-drug-resistant Acinetobacter bacteraemia. *Euro.J. Int. Medi.*,2009; 20(5): 540-544.

Möller S, Croning MD and Apweiler R. Valuation of methods for the prediction of membrane spanning regions. *Bioinform.* 2001; 17(7): 646-653.

Paul S, Saha M, Bhoumik N and Talukdar S. In silico structural and functional annotation of mycoplasma genitalium hypothetical protein MG_377. *Int. J. Biol.* 2015;19: 15-24.

Pawel S, Martin-Galiano AJ, Mikolajka A, Girschick T, Holak T and Frishman D. Protein solubility: Sequence based prediction and experimental verification. *Bioinform.* 2007;23: 2536-2542.

Rahman A, Susmi TF, Karim ME and Hossain MU. Functional annotation of an ecologically important protein from Chloroflexusaurantiacus involved in polyhydroxyalkanoates (PHA) biosynthetic pathway. *SN Appl. Sci.* 2020; 2(11): 1810.

Restrepo-Montoya D, Vizcaíno C, Niño LF, Ocampo M, Patarroyo ME and Patarroyo MA. Validating subcellular localization prediction tools with

mycobacterial proteins. *BMC Bioinform.* 2009; 10(1): 134.

Schneider G and Fechner U. Advances in the prediction of protein targeting signals. *Proteomics*, 2004; 4(6): 1571-1580.

Sigrist C, Castro E, Cerutti L, Cuche B, Hulo N, Bridge A, Bougueleret L and Xenarios I. New and continuing developments at PROSITE. *Nucleic Acids Res.*2012; 41: D344-D347.

Sowmya G and Ranganathan S. Protein-protein interactions and prediction: A Comprehensive overview. *Protein Pept. Lett.* 2013; 21(8): 779-789.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering C. von. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019; 47(D1): D607-D613.

Szklarczyk D, Gable A, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva N, Morris J, Bork P, Jensen Land von Mering C. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2018; 47(D1): D607-D613.

Tian W, Chen C, Lei X, Zhao J and Liang J. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018; 46(W1): W363-W367.

Tusnady G and Simon I. The HMMTOP transmembrane topology prediction server. *Bioinform.* 2001;17: 849-850.

Uversky VN. Protein intrinsic disorder and structure-function continuum. *Prog. Mol. Biol. Transl. Sci.* 2019; 166: 1-17.

Wang M, Herrmann C, Simonovic M, Szklarczyk Dand von Mering C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues and cell-lines. *Proteomics*, 2015;15(18): 3163-3168.

Warnes SL, Caves V and Keevil CW. Mechanism of copper surface toxicity in Escherichia coli O157:H7 and Salmonella involves immediate membrane depolarization followed by slower rate of DNA destruction which differs from that observed for Gram-positive bacteria: Rapid death of Gram-negative bacteria on dry copper surfaces. *Environ. Microbiol.* 2012; 14(7): 1730-1743.

Wiederstein M and Sippl M. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007; 35: W407-W410.

Williams CL, Neu HM, Gilbreath JJ, Michel SLJ, Zurawski DV and Merrell DS. Copper resistance of the emerging pathogen acinetobacter-baumannii. *Appl. Environ. Microbiol.* 2016; 82(20): 6174-6188.

Wilson D, Madera M, Vogel C, Chothia C and Gough J. The superfamily database in 2007: Families and functions. *Nucleic Acids Res.*2007; 35: D308-D313.

Yu CS, Chen YC, Lu CH and Hwang JK. Prediction of protein subcellular localization. *Proteins: Struct. Funct. Bioinf.* 2006; 64(3): 643-651.

Yu CS, Lin CJ and Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 2004; 13(5): 1402-1406.

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, LoR, Dao P, Sahinalp SC, Ester M, Foster LJ and Brinkman FSL. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinform.* 2010); 26(13): 1608-1615.

Zhou P. Determining protein half-lives. *Methods Mol. Biol.* 2004; 284: 67-77.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN and Alva V. A Completely reimplemented MPI bioinformatics toolkit with a new HHpredserver at its core. *J. Mol. Biol. 2018*; 430(15): 2237-2243.