engineering proceedings

MDPI

Proceeding Paper

Structure-Based Functional Annotation of an Uncharacterized Conserved Protein of *Acinetobacter baumannii*: An *In-Silico* Approach †

Mamun Al Asad [1,2,‡], Surya Afrin Shorna [2,3,‡], Abu Saim Mohammad Saikat [2,4,*,‡] and Ekhlas Uddin [2,5]

[1] Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia 7003, Bangladesh; ronniebge22@gmail.com
[2] Department of Computational Biology and Bioinformatics, Advanced Bioscience Center for Collaborative Research (ABSCCR), Rajshahi 6250, Bangladesh; suryaafrin0@gmail.com (S.A.S.); ekhlasbt03@gmail.com (E.U.)
[3] Department of Biotechnology, Bangladesh Agriculture University, Mymensingh 2202, Bangladesh
[4] Department of Biochemistry and Molecular Biology, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj 8100, Bangladesh
[5] Department of Biochemistry and Molecular Biology, Gono Bishwabidyalay, Dhaka 1344, Bangladesh
* Correspondence: asmsaikat.bmb@gmail.com
† Presented at the 2nd International Electronic Conference on Processes: Process Engineering—Current State and Future Trends (ECP 2023), 17–31 May 2023; Available online: https://ecp2023.sciforum.net/.
‡ These authors contributed equally to this work.

Abstract: *Acinetobacter baumannii* (*A. baumannii*) is an example of an opportunistic pathogen that is generally harmless to healthy individuals but can cause serious infections such as ventilator-associated pneumonia, wound infections, and bacteremia in critically ill hospital patients. *A. baumannii* produces many proteins within its genome. By analyzing its structural and functional interpretation, bioinformatics techniques can make it easier to understand this organism. The protein is still unclear, though. As a result, this study developed an in-silico method for functional and structural characterization of the uncharacterized protein (accession ID: SSI32830.1). These provide many characteristics in silico viewpoints, such as the protein's physiochemical qualities, sub-cellular localization, three-dimensional structure, and protein-protein interactions. Protein-protein interactions are explained using the STRING software. The projected tertiary structure evaluation was conducted using the Swiss Model. The best materials are chosen utilizing structural analyses based on Ramachandran plot analysis. This research sought to understand the function of *A. baumannii*. Therefore, this investigation will increase our understanding of pathophysiology and allow us to target the protein complex specifically.

Keywords: *Acinetobacter baumannii*; uncharacterized protein; protein modeling; protein-protein interactions; computational research

Citation: Al Asad, M.; Shorna, S.A.; Saikat, A.S.M.; Uddin, E. Structure-Based Functional Annotation of an Uncharacterized Conserved Protein of *Acinetobacter baumannii*: An *in-silico* Approach. *Eng. Proc.* **2023**, *5*, x. https://doi.org/10.3390/xxxxx
Published: 17 May 2023

1. Introduction

A. baumannii is a Gram-negative pathogenic bacterium, and it is a nosocomial pathogen that causes a variety of infections in the skin, bloodstream, and urinary system, as well as high mortality and morbidity rates in humans [1,2]. Immunocompromised people are more likely to contract this bacterium, particularly if they have had a lengthy hospital stay [3]. Although this gram-negative bacterium was once regarded as a low-grade pathogen, its high resistance to most clinically available antibiotics has made it a significant public health issue in recent years [4]. Many methods have been discovered recently to manage infections with this bacterium, including vaccination, monoclonal antibodies, and phage therapy [4,5]. The extracellular proteins of *A. baumannii* have

received less attention as possible adjuvant prospects. According to the findings of various bioinformatics databases and studies, these proteins play a significant part in the pathogenesis of *A. baumannii* [6].

2. Materials and Methods

2.1. Protein Selection and Sequence Retrieval

The NCBI database (http://www.ncbi.nlm.nih.gov (accessed on)) was accessed to obtain the FASTA format of the amino acid (aa) sequence for the Uncharacterized conserved protein (*A. baumannii*).

2.2. Characterization of the Selected Protein's Physicochemical Properties

The ProtParam tool of ExPASy [7] was used to record the physicochemical characterization of proteins, including the amino acid sequence composition, instability index, aliphatic index, and GRAVY. Using SMS (v.2.0) [8], the theoretical isoelectric point (pI) and GRAVITY of the SSI32830.1 protein were also determined.

2.3. Identification of the Subcellular Location

By using the PSORTb (v.3.0.2) [9], BUSC [10], Gneg-mPLoc [11], HMMTOP v.2.0 [12] and TMHMM server v.2.0 [13], the protein's subcellular localization was identified.

2.4. The Selected Protein's Functional Annotation

Using the CD-search tool [14] on the NCBI platform, the conserved domain in the protein SSI32830.1 was predicted. The ScanProsite tool of the ExPASy software [15] was also utilized to identify protein motifs. The evolutionary relationships of the protein SSI32830.1 were detected by the SuperFamily program [16].

2.5. Protein-Protein Interaction

The STRING v.11.0 server [17] was used to identify the likely protein-protein (pr-pr) interactions.

2.6. Identification and Verification of the Chosen Protein's Predicted Secondary Structure

Secondary structural elements were predicted using the SOPMA program. The secondary structure was identified using SPIPRED (v. 4.0) tool [18].

2.7. Prediction of the Three-Dimensional Structure and Validation of the Chosen Protein

The SWISS-MODEL was used to predict the three-dimensional structure of the chosen protein, and the PDB format was saved. The modeled 3D structure of the protein was also structurally confirmed using the PROCHECK algorithm of the SAVES program (v.6.0) [19]. The Z-score of the modeled Structure was also determined for the structural assessment using the ProSA-web tool [20].

2.8. Active Site Determination

The CASTp v.3.0 server [21] was used to predict the active sites of the modeled protein (*A. baumannii*).

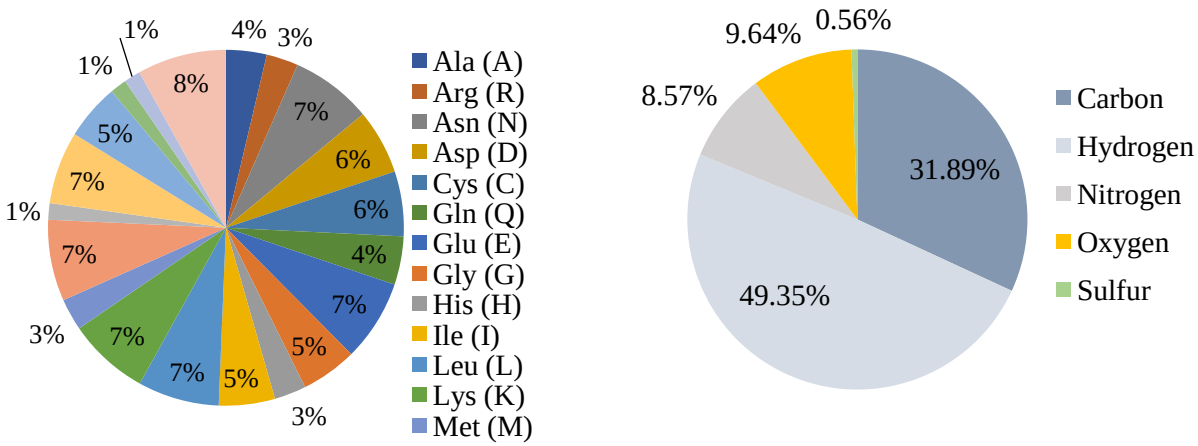
3. Results and Discussion

3.1. Protein Sequence Retrieval

The amino acid (aa) sequence of the uncharacterized conserved protein (*A. baumannii*) (accession ID: SSI32830.1) was found in the NCBI database [22]. The protein tertiary structure was modeled using a 136 amino acid-long protein sequence.

3.2. Physicochemical Properties

In order to measure physicochemical parameters, the amino acid sequence of SSI32830.1, which is present in *A baumannii*, was retrieved in FASTA format and used as a query sequence (Figure 1). Because the protein's instability index is 30.59 (below 40.00), it is stable [23]. The protein has a pH of 5.52, which corresponds to its theoretical isoelectric point (pI) and shows that the molecular weight (15,569.78 Da), aliphatic index (75.88), instability index (27.54), and GRAVY (−0.228).



(**a**) (**b**)

Figure 1. Physicochemical parameters of the selected protein. (**a**) The protein contains Ala (5, 3.70%), Arg (4, 2.90%), Asn (10, 7.40%), Asp (8, 5.90%), Cys (8, 5.90%), Gln (6, 4.40%), Glu (10, 7.40%), Gly (7, 5.10%), His (4, 2.90%), Ile (7, 5.10%), Leu (10, 7.40%), Lys (10, 7.40%), Met (4, 2.90%), Phe (10, 7.40%), Pro (2, 1.50%), Ser (9, 6.60%), Thr (7, 5.10%), Trp (2, 1.50%), Tyr (2, 1.50%), and Val (11, 8.10%). (**b**) The atomic composition of the protein as of carbon (685, 31.89%), hydrogen (1060, 49.35%), nitrogen (184, 8.57%), oxygen (207, 9.64%), and sulfur (12, 0.56%).

3.3. Subcellular Location Determination

The protein's subcellular localization was evaluated using the PSORTb (v. 3.0.2) and BUSCA tools (Accession No. SSI32830) (Table 1).

Table 1. Subcellular localization assessment.

Analysis Tool/Server	Location of Protein
PSORTb (v.3.0.2)	Cytoplasm
BUSCA	Cytoplasm
Gneg-mPLoc	Cell inner membrane
HMMTOP (v.2.0)	No transmembrane helices present
TMHMM (v.2.0)	No transmembrane helices present

3.4. Functional Annotation of the Selected Protein

The domain that appears in identical protein sequences is identified by the NCBI CDD tool. When comparing a test sequence to position-specific rating datasets derived from conserved domain (CD) alignments in the CD protein cluster, CD-Search utilizes RPS-BLAST (8, 29). The CD search engine discovered a conserved domain in the protein SSI32830.1 (30–32). The CD search engine identified a conserved domain in the protein SSI32830.1 as specific domain hits (accession no. COG3791).

3.5. Protein-Protein Interaction

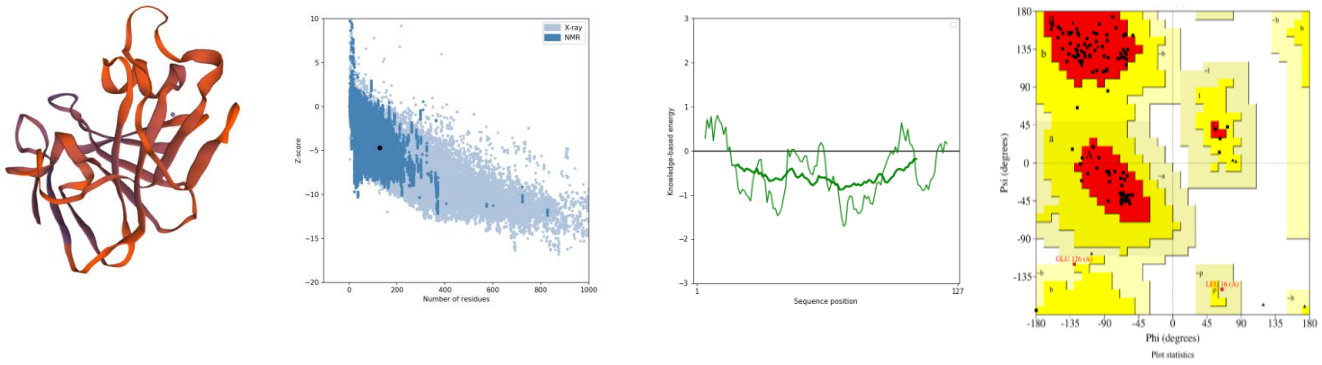
Protein-protein interactions are primarily concerned with understanding how biological systems function. Such connections provide an insightful platform for annotating functional, structural, and evolutionary properties of proteins and filtering, evaluating, and validating functional genomics data. The tool can map the relationships between various species and predict subsequent studies [24,25]. The empty nodes in the picture represent proteins with an unknown three-dimensional structure.

3.6. Identification and Validation of the Predicted Secondary Structure of the Selected Protein

The uncharacterized conserved protein (*A. baumannii*) was predicted using secondary structural elements using the default SOPMA program parameters (window width of 17, number of states of 4, and similarity threshold of 8). The Sequence Plot structure and the PSIPRED Cartoon were predicted using the SPIPRED v.4.0 program.

3.7. The Three-Dimensional Protein Structure Anticipation and Assessment

Ramachandran plot analysis and Interactive workplace analysis of the Swiss-Model server were used to compare the modeled tertiary structures to assess the structures' quality [8,35]. The best template was chosen as a CENP-V/GFA domain-containing protein, and ultimately the tertiary modeled protein structure (Figure 2) was stored in PDB format.



(**a**) (**b**) (**c**) (**d**)

Figure 2. Tertiary structural prediction (**a**) The predicted three-dimensional structure, (**b**) The overall model quality by Z-score (−7.26), and (**c**) The local model quality assessment. (**d**) Ramachandran plot analysis of uncharacterized conserved protein of *A. baumannii* protein structure predicted by UCLA SAVES V.6.0.

3.8. Active Site Determination

The modeled protein's 21 predicted active sites using the CASTp v.3.0 algorithm (Figure 7). A database server called CASTp can locate regions on proteins, outline those regions, determine their dimensions, and compute the areas of those regions. The top active sites were found between the area of 126.241 and 101.535 of the modeled protein.

4. Conclusions

The aggressive bacterium *A. baumannii* is often found in healthcare settings where it causes illnesses. The subsequent upsurge in prevalence attributable to infected frontline soldiers arriving from battle zones—and the substantial uptick in the incidence of multidrug-resistant (MDR) strains have greatly enhanced the prominence of this new opportunistic infection. Infections caused by *A. baumannii* have been linked to a wide variety of organ systems, and they have been shown to have a wide range of clinical manifestations and consequences. Although this developing pathogen has been studied

extensively, a minor is recognized about its natural pathogenic potency or pathogenicity assortment. This research aims to characterize the conserved protein of the bacteria that is essential for its survival. Secondary and tertiary features of the chosen protein revealed protein structure-based correlations and, therefore, more understandable traits. Because of this, the selected protein may be used as a target for developing protein-based drugs and vaccines against the protein, reducing the prevalence of bacterial infections.

Author Contributions: Conceptualization, A.S.M.S.; methodology, A.S.M.S. and M.A.A.; software, M.A.A., S.A.S. and A.S.M.S.; validation, A.S.M.S. and E.U.; formal analysis, M.A.A., S.A.S. and A.S.M.S.; investigation, A.S.M.S.; resources, M.A.A., S.A.S., A.S.M.S. and E.U.; data curation, M.A.A., S.A.S. and A.S.M.S.; writing—M.A.A., S.A.S. and A.S.M.S.; writing—review and editing, A.S.M.S; visualization, M.A.A. and A.S.M.S.; supervision, A.S.M.S. and E.U.; project administration, A.S.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Upon request to corresponding author.

Acknowledgments: The authors want to acknowledge the Department of Computational Biology and Bioinformatics, Advanced Bioscience Center for Collaborative Research (ABSCCR), Rajshahi 6250, Bangladesh, for supporting this study.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Peleg, A.Y.; Seifert, H.; Paterson, D.L. *Acinetobacter baumannii*: Emergence of a Successful Pathogen. *Clin. Microbiol. Rev.* **2008**, *21*, 538–582. https://doi.org/10.1128/cmr.00058-07.
2. Saikat, A.S.M.; Paul, A.K.; Dey, D.; Das, R.C.; Das, M.C. In-Silico Approaches for Molecular Characterization and Structure-Based Functional Annotation of the Matrix Protein from *Nipah henipavirus*. *Chem. Proc.* **2022**, *12*, 21. https://doi.org/10.3390/ecsoc-26-13522.
3. Abhari, S.S.; Badmasti, F.; Modiri, L.; Aslani, M.M.; Asmar, M. Circulation of imipenem-resistant Acinetobacter baumannii ST10, ST2 and ST3 in a university teaching hospital from Tehran, Iran. *J. Med. Microbiol.* **2019**, *68*, 860–865. https://doi.org/10.1099/jmm.0.000987.
4. Garcia-Quintanilla, M.; Pulido, M.R.; McConnell, M.J. First steps towards a vaccine against Acinetobacter baumannii. *Curr. Pharm. Biotechnol.* **2014**, *14*, 897–902. https://doi.org/10.2174/1389201014666131226123511.
5. Khan, R.A.; Hossain, R.; Siyadatpanah, A.; Al-Khafaji, K.; Khalipha, A.B.R.; Dey, D.; Asha, U.H.; Biswas, P.; Saikat, A.S.M.; Chenari, H.A.; et al. Diterpenes/Diterpenoids and Their Derivatives as Potential Bioactive Leads against Dengue Virus: A Computational and Network Pharmacology Study. *Molecules* **2021**, *26*, 6821. https://doi.org/10.3390/molecules26226821.
6. Li, H.; Tan, H.; Hu, Y.; Pan, P.; Su, X.; Hu, C. Small protein A and phospholipase D immunization serves a protective role in a mouse pneumonia model of Acinetobacter baumannii infection. *Mol. Med. Rep.* **2017**, *16*, 1071–1078. https://doi.org/10.3892/mmr.2017.6688.
7. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. https://doi.org/10.1093/nar/gkg563.
8. Ravella, D.; Kumar, M.U.; Sherlin, D.; Shankar, M.; Vaishnavi, M.K.; Sekar, K. SMS 2.0: An Updated Database to Study the Structural Plasticity of Short Peptide Fragments in Non-redundant Proteins. *Genom. Proteom. Bioinform.* **2012**, *10*, 44–50. https://doi.org/10.1016/s1672-0229(11)60032-6.
9. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. https://doi.org/10.1093/bioinformatics/btq249.
10. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Profiti, G.; Casadio, R. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* **2018**, *46*, W459–W466. https://doi.org/10.1093/nar/gky320.
11. Shen, H.-B.; Chou, K.-C. Gneg-mPLoc: A top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. *J. Theor. Biol.* **2010**, *264*, 326–333. https://doi.org/10.1016/j.jtbi.2010.01.018.
12. Tusnady, G.E.; Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **2001**, *17*, 849–850. https://doi.org/10.1093/bioinformatics/17.9.849.
13. Yang, M.; Derbyshire, M.K.; Yamashita, R.A.; Marchler-Bauer, A. NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. *Curr. Protoc. Bioinform.* **2019**, *69*, e90. https://doi.org/10.1002/cpbi.90.

14. De Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, W362–W365. https://doi.org/10.1093/nar/gkl124.
15. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. https://doi.org/10.1093/nar/gkw1129.
16. Saikat, A.S.M.; Uddin, E.; Ahmad, T.; Mahmud, S.; Imran, A.S.; Ahmed, S.; Alyami, S.A.; Moni, M.A. Structural and Functional Elucidation of IF-3 Protein of Chloroflexus aurantiacus Involved in Protein Biosynthesis: An In Silico Approach. *BioMed Res. Int.* **2021**, *2021*, 9050026. https://doi.org/10.1155/2021/9050026.
17. Buchan, D.W.A.; Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **2019**, *47*, W402–W407. https://doi.org/:10.1093/nar/gkz297.
18. Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291. https://doi.org/10.1107/s0021889892009944.
19. Wiederstein, M.; Sippl, M.J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. https://doi.org/10.1093/nar/gkm290.
20. Saikat, A.S.M.; Islam, R.; Mahmud, S.; Imran, A.S.; Alam, M.S.; Masud, M.H.; Uddin, E. Structural and Functional Annotation of Uncharacterized Protein NCGM946K2_146 of Mycobacterium Tuberculosis: An In-Silico Approach. *Proceedings* **2020**, *66*, 13. https://doi.org/10.3390/proceedings2020066013.
21. Sayers, E.W.; Beck, J.; E Bolton, E.; Bourexis, D.; Brister, J.R.; Canese, K.; Comeau, D.C.; Funk, K.; Kim, S.; Klimke, W.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2020**, *49*, D10–D17. https://doi.org/10.1093/nar/gkaa892.
22. Saikat, A.S.M.; Khalipha, A.B.R. Structure Prediction, Characterization, and Functional Annotation of Uncharacterized Protein BCRIVMBC126_02492 of Bacillus cereus: An In Silico Approach. *Am. J. Pure Appl. Biosci.* **2020**, *2*, 104–111. https://doi.org/10.34104/ajpab.020.01040111.
23. Peng, X.; Wang, J.; Peng, W.; Wu, F.-X.; Pan, Y. Protein–protein interactions: Detection, reliability assessment and applications. *Briefings Bioinform.* **2017**, *18*, 798–819. https://doi.org/10.1093/bib/bbw066.
24. Saikat, A.S.M.; Das, R.C.; Das, M.C. Computational Approaches for Structure-Based Molecular Characterization and Functional Annotation of the Fusion Protein of *Nipah henipavirus*. *Chem. Proc.* **2022**, *12*, 32. https://doi.org/10.3390/ecsoc-26-13530.
25. Saikat, A.S.M. An In Silico Approach for Potential Natural Compounds as Inhibitors of Protein CDK1/Cks2. *Chem. Proc.* **2021**, *8*, 5. https://doi.org/10.3390/ecsoc-25-11721.