sciendo

ADVANCED RESEARCH IN LIFE SCIENCES
8, 2024, 39 – 44

www.degruyter.com/view/j/arls

DOI: 10.2478/arls-2024-0005
Research Article

AlphaFold Modeling and Computational Analysis of a PHA Synthase from *Actinophytocola algeriensis*

Mahfoud Bakli [*1,2], Raul Paşcalău [3], Laura Șmuleac [3]

[1] *Département de Biologie, Faculté des Sciences de la Nature et de la Vie et Sciences de la Terre, Université de Ghardaïa, Ghardaïa, Algeria.*
[2] *Laboratoire de Valorisation et Conservation des Écosystèmes Arides (LVCEA), Faculté des Sciences de la Nature et de la Vie et Sciences de la Terre, Université de Ghardaïa, Ghardaïa, Algeria.*
[3] *University of Life Sciences "King Michael I of Romania", 119 Calea Aradului, 300645, Timişoara, Romania.*

Received May, 2024; Revised July, 2024; Accepted July, 2024

Abstract
Environmental challenges related to plastic waste underscore the urgent need for innovative solutions. Polyhydroxyalkanoates (PHA) have emerged as sustainable alternatives to conventional plastics, particularly in packaging, due to their biodegradability and biocompatibility. However, the cost of PHA production and certain physical limitations compared to synthetic polymers remain significant barriers to widespread adoption. Within this context, *Actinophytocola algeriensis*, an Actinobacteria species isolated from the Sahara desert in Algeria, holds promise for its biotechnological potential and bioactive molecules. Despite this, our understanding of its enzyme profile, notably the PHA synthase (EC 2.3.1.304), the key enzyme in PHA biosynthesis, remains limited. In this study, the 3D structure of PHA synthase was modeled utilizing the artificial intelligence program AlphaFold, followed by the structural refinement and validation. In addition, physicochemical properties and functional characterization were conducted using various bioinformatics tools. This research signifies a substantial advancement in comprehending the molecular mechanisms underlying PHA biosynthesis in *A. algeriensis*, thereby fostering the development of innovative biotechnological applications for sustainable biopolymer production.

Keywords: Polyhydroxyalkanoates, *Actinophytocola algeriensis*, PHA synthase, bioinformatics tools.

Introduction

Polyhydroxyalkanoates (PHA) are aliphatic polyesters synthesized and accumulated by various microorganisms under stress conditions, marked by an excess of carbon substrate and a scarcity of other essential nutrients like nitrogen, sulfur, phosphorus, or oxygen [1]. These biopolymers show great promise with diverse applications in biotechnology, particularly in producing environmentally friendly plastics [2]. Their biodegradability, biocompatibility, and adaptability have garnered substantial interest, serving not only as an eco-conscious alternative to conventional petroleum-based plastics but also for specialized uses, including agricultural and biomedical devices [3]. The key enzyme for PHA biosynthesis is PHA synthase (Polyhydroxyalkanoate synthase, EC 2.3.1.304), or PhaC, catalyzing the polymerization of hydroxyalkanoates (HAs) from HA-CoA and releasing CoA (coenzyme A) during the reaction [4]. These PhaCs can be classified into four classes, depending on their amino acid primary structures, substrate specificities, and subunit compositions [5]. With the exception of Class II synthases, which have a tendency to utilize medium chain length (MCL) monomers, Classes I, III, and IV synthases show a preference for short

* Corresponding author: **M. Bakli, Email:** mahfoud.bakli@gmail.com

chain length (SCL) monomers. Furthermore, Newly identified PhaCs fall outside the established four classes [6]. Over the past two decades, the actinobacteria phylum, including genera like *Nocardia*, *Rhodococcus*, *Leifsonia, Microbacterium, Paenarthrobacter*, *Arthrobacter, Kineosphaera*, and *Streptomyces* have been identified as producers of PHA and PHA synthase [7]. However, no studies have been conducted on this topic regarding the genus *Actinophytocola. Actinophytocola algeriensis*, isolated from Algeria's Sahara desert soil [8], exhibited biotechnological promise and contained bioactive molecules such as PHA synthase. Thus, more investigation is necessary to bridge this knowledge gap and uncover the potential uses of this enzyme. Nonetheless, the experimental process of expressing and purifying proteins can present challenges, demanding considerable time and effort [9]. Therefore, The approach of in silico methods offers a viable solution to these challenges, allowing researchers to gain insights into the functional and structural properties of proteins [10]. In this study, PHA synthase from *A. algeriensis* was subjected to in silico analysis, revealing its physicochemical, functional, and secondary structural properties, and undergoing protein homology modeling.

Materials and methods

Sequence retrieval

The protein sequence of PHA synthase from *Actinophytocola algeriensis* was obtained from UniProt (accession no. A0A7W7VBE2) in FASTA format and used for computational analyses.

Prediction of physicochemical properties

The ExPasy ProtParam tool (http://expasy.org/tools/protparam.html) computed physiochemical parameters for the PHA synthases, including molecular weight, pI, amino acid composition, extinction coefficient, instability index, aliphatic index, and GRAVY [11].

Secondary structure characterization

The SOPMA tool (self-optimized prediction method with alignment), which is available on the online NPS@ server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), was applied to predict the secondary structure of the PHA synthase protein [12].

AlphaFold modeling and evaluation of the tertiary structure

The tertiary structure of the target query protein sequence was determined using the AlphaFold protein structure database (https://alphafold.ebi.ac.uk/) [13]. AlphaFold employs a groundbreaking machine learning technique that utilizes multi-sequence alignments to build a deep learning algorithm, incorporating both physical and biological principles of protein structure [13]. The generated strutural model was subjected to energy minimization using the ModRefiner server, available at http://zhanglab.ccmb.med.umich.edu/ModRefiner/ [14]. Indeed, the critical phase of homology modeling involves assessing the model's quality, ensuring the reliability of the predicted protein structure. In this study, the generated PHA synthase refined model underwent validation and verification by Ramachandran plot analysis using PROCHECK tool on SAVES v6.0 server (https://saves.mbi.ucla.edu/) [15]. The PyMOL Molecular Graphics System (Schrödinger LLC, version 2.3) was used to generate and optimize image of the 3D structure model [16].

Functional analyses

The SOSUI server (http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html) [17] identified the protein's transmembrane tendency in the target protein by distinguishing between membrane and soluble proteins from the amino acid sequence. For examining the functional interaction between the PHA synthase of *A. algeriensis* and its ligand, we utilized the COFACTOR online meta-server at Zhang's Lab (https://zhanggroup.org/COFACTOR2/) [18] and subsequently visualized the results using PyMOL software. Furthermore, the PHA synthase protein sequence was analyzed using the NCBI Conserved Domain Database (CDD) search service (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) to identify conserved functional domains [19].

Results and discussion

Physicochemical properties prediction

The amino acid sequence of PHA synthase from *Actinophytocola algeriensis*, obtained from UNIPROT (ID: A0A7W7VBE2), consists of 350 amino acids. The physicochemical properties (Table 1) were calculated using Expasy's ProtParam. The PHA synthase protein showed a molecular weight measuring 38.31 kDa. The protein sequence contains an equal number of positively and negatively charged residues. The protein exhibited a high extinction coefficient value, indicating elevated levels of Trp and Tyr, enhancing quantitative analysis of interactions in solution. The computed isoelectric point was 7.10, reflecting the neutral nature of the protein. An instability index (II) of 35.35, below the threshold of 40, signifies its stability. In addition, the PHA synthase protein has an aliphatic index (AI) of 100.86, indicating a higher proportion of aliphatic side chains (such as alanine, valine, isoleucine, and leucine) which enhances structural stability.

This increase in the aliphatic index contributes to improved enzyme thermostability [20]. The grand average of hydropathy (GRAVY) value for a peptide or protein is computed by summing the hydropathy values of all amino acids and dividing by the number of residues in the sequence [21]. A calculated GRAVY value of the protein is 0.071. A positive GRAVY score indicated that the PHA synthase protein was typically hydrophobic. In summary, the neutral and thermostable properties of the target PHA synthase enzyme from *A. algeriensis* may offer significant potential for application in the biodegradable plastics industry [22].

Table1. Parameters computed for the PHA synthase from *A. algeriensis* using Expasy's Protparam

Property	Value
Number of amino acids (AA)	350
Molecular weight (Da)	38319.24
Theoretical pI	7.10
Total number of negatively charged residues (Asp + Glu)	39
Total number of positively charged residues (Arg + Lys)	39
Extinction coefficient (EC)	35535
Instability index (II)	35.35 (Stable)
Aliphatic index (AI)	100.86
Grand average of hydropathicity (GRAVY)	0.071

Secondary structure characterization

The secondary structure predication of the *A. algeriensis* PHA sythase protein was predicted by the online tool SOPMA showing that the protein had a significant percentage of alpha helices and random coils followed by moderate content of extended strands and beta turns (Table 2).

AlphaFold modeling and evaluation of the tertiary structure

The modeling of the three-dimensional structure of the target protein was performed by AlphaFold program. Subsequently, the 3D built model was refined by the ModRefiner server. This refined modeled structure was visualized using PyMOL, as depicted in Figure 1.A. To validate the generated structural model, a Ramachandran plot analysis was performed using the PROCHECK tool. The distribution of Phi and Psi angles from the Ramachandran map in the refined protein model, generated by non-glycine and non-proline residues, was summarized in Table 3 and Figure 1.B. The highest favored regions in the Ramachandran plot accounted for 93.0% of the total, with no presence of disallowed regions (0%), indicating the characteristics of a high-quality model.

Table 2. Secondary structure prediction of the the PHA synthase from *A. algeriensis* by SOPMA SERVER

Property	Value
Alpha helix	43.14 %
310 helix	0.00 %
Pi helix	0.00 %
Beta bridge	0.00 %
Extended strand	14.29 %
Beta turn	5.14 %
Bend region	0.00 %
Random coil	37.43 %
Ambiguous state	0.00 %
Other states	0.00 %

Table 3. Ramachandran plot analysis of refined PHA synthase model using PROCHECK

Favored region (%)	Additional allowed region (%)	Generously allowed region (%)	Disallowed region (%)
93.0	6.4	0.7	0.0

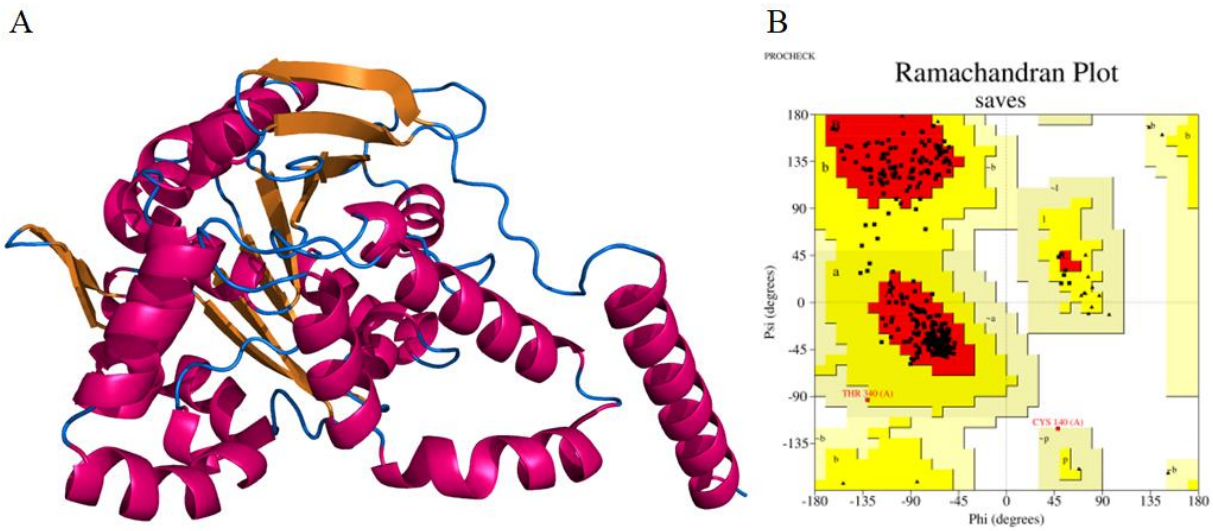

Figure 1. Tertiary structure prediction and validation for the PHA synthase from *A. algeriensis*. (A) The protein's tertiary structure was predicted using AlphaFold modeling, refined by ModRefiner, and visualized with PyMOL. The structural features, including α-helix (pink), β-sheet (orange), and loop (marine), were identified. (B) The predicted and refined structure's Ramachandran plot was validated using the PROCHECK program, classifying residues based on their conformational preferences. Regions with favored (A, B, and L), additional allowed (a, b, l, and p), generously allowed (~a, ~b, ~l, and ~p), and disallowed conformations are highlighted accordingly. Non-glycine and non-proline residues are represented by filled black squares, while glycines (excluding those at the ends of the polypeptide chains) are depicted as filled black triangles.

Functional analyses

Transmembrane regions were detected through the PHA synthase protein sequence analysis on the SOSUI server. Table 4 represents the transmembrane regions identified for the target protein (Table 4). COFACTOR was used to predict the consensus ligand binding residues located within the predicted active site of the PHA synthase enzyme. Accordingly, the results indicated that the enzyme interacted with the ligand (PDB CCD ID: J6Z) at the following consensus binding residues; Cys 140, Leu 141, Pro 167, Asp 169, Ile 300, and His 328, with a higher confidence score (C-scoreLB) of 0.14, suggesting a more reliable prediction (Figure 2). This observation is in line with the notion that a conserved cysteine residue functions as the enzyme's active site, a feature shared by all PHA synthases [23]. Moreover, among other conserved residues, histidine and aspartate, together with the conserved cysteine, form the enzyme's catalytic triad. This triad, like that of esterases, implies an α/β hydrolase fold in PHA synthases [24]. Additionally, protein domains are closely associated with their structure, thus predicting these domains could be beneficial in inferring protein function. The NCBI Conserved Domain Database has identified conserved domains within the target *A. algeriensis* PHA synthase protein sequence, namely: PhaC, which is a poly-beta-hydroxybutyrate synthase involved in lipid transport and metabolism (accession: COG3243), PRK07868, which is an acyl-CoA synthetase (accession: PRK07868), and Abhydrolase_1, which had and alpha/beta hydrolase fold; this catalytic domain is found in a very wide range of enzymes (accession: pfam 00561) (Figure 3).

Table 4. Transmembrane regions identified in the PHA synthase using SOSUI server

No.	N terminal	Transmembrane sequence	C terminal	Type	Length
1.	130	GGAPVHVVAWCLGGILSLLTHAD	152	Primary	23
2.	159	ASIATIAAPIDMTAIPLVAPIKP	181	Secondary	23

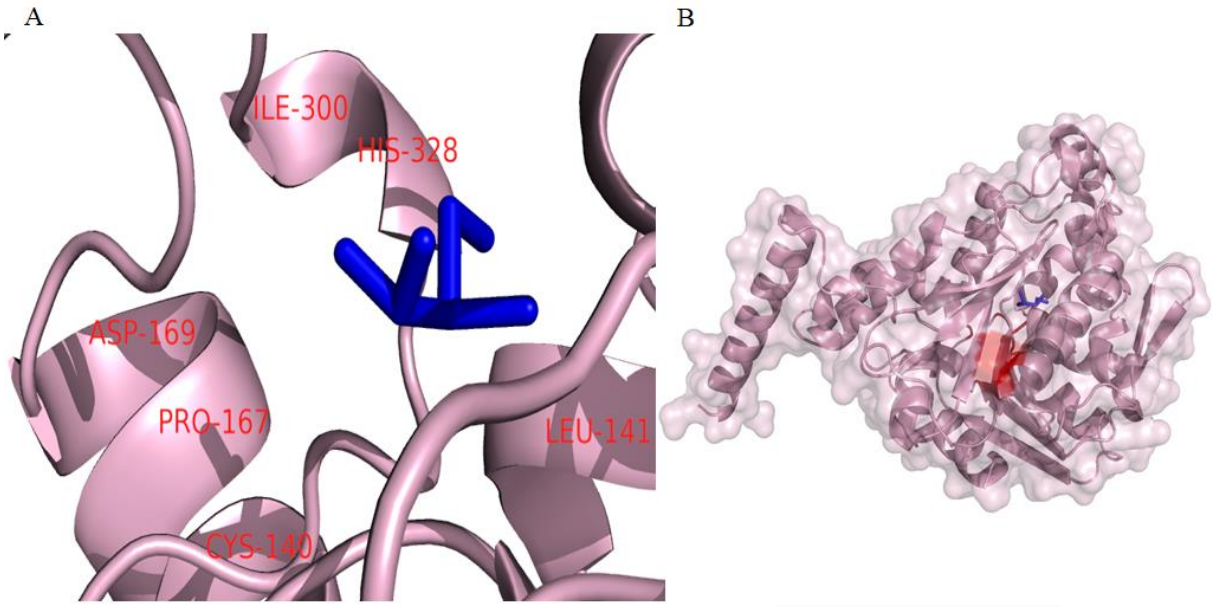

Figure 2. Ligand binding sites of the *A. algeriensis* PHA synthase predicted by COFACTOR. (A) Ligand binding sites (Cys 140, Leu 141, Pro 167, Asp 169, Ile 300, and His 328). (B) Surface view transparency of the protein with ligand. Sticks represent the ligand (blue). Ligand binding sites (red) within the protein structure (pink)

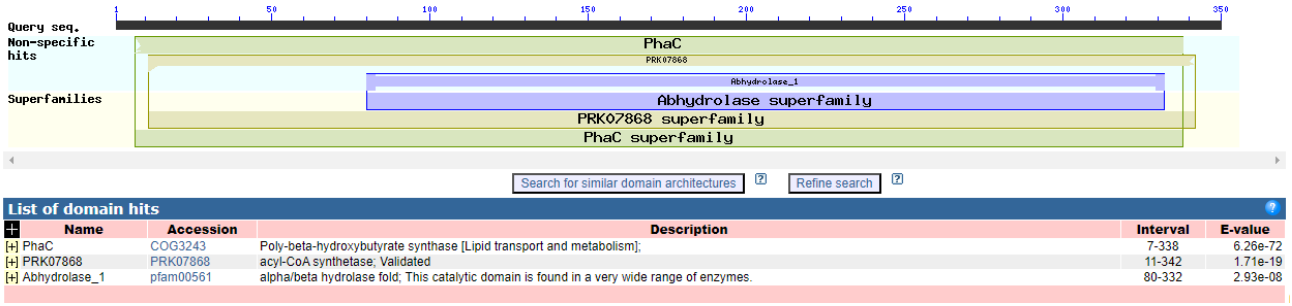

Figure 3. Analysis of *A. algeriensis* PHA Synthase Domains Using the CDD Program

Conclusions

The computational analysis of the PHA synthase from *A. algeriensis* uncovered several noteworthy characteristics; its neutrality, thermostability, hydrophobicity, and membranar localization. The protein's secondary structure prediction consists mostly of alpha helices and random coils, with additional elements of extended strands and beta turns. Structural modeling and functional analyses were conducted, suggesting that this computational approach holds promise for designing PHA synthases with industrial applications. However, additional *in vivo* studies, along with experimental investigations like enzyme kinetics and substrate specificity, are warranted to fully explore its industrial potential.

Acknowledgements

This work was supported by PRFU “Projet de Recherche et Formation Universitaire” Project Research, D01N01UN460120220001.

References

1. **Riaz, S., Rhee, K. Y., & Park, S. J.** (2021). Polyhydroxyalkanoates (PHAs): biopolymers for biofuel and biorefineries. *Polymers, 13*(2), 253. DOI: 10.3390/polym13020253

2. **Zhou, W., Bergsma, S., Colpa, D. I., Euverink, G.-J. W., & Krooneman, J.** (2023). Polyhydroxyalkanoates (PHAs) synthesis and degradation by microbes and applications towards a circular economy. *Journal of Environmental Management, 341*, 118033. DOI: 10.1016/j.jenvman.2023.118033

3. **Acharjee, S. A., Bharali, P., Gogoi, B., Sorhie, V., Walling, B., & Alemtoshi.** (2023). PHA-based

bioplastic: A potential alternative to address microplastic pollution. *Water, Air, & Soil Pollution, 234*(1), 21. DOI: 10.1007/s11270-022-06029-2

4. **Teh, A.-H., Chiam, N.-C., Furusawa, G., & Sudesh, K.** (2018). Modelling of polyhydroxyalkanoate synthase from *Aquitalea* sp. USM4 suggests a novel mechanism for polymer elongation. *International journal of biological macromolecules, 119*, 438-445. DOI: 10.1016/j.ijbiomac.2018.07.147

5. **Neoh, S. Z., Chek, M. F., Tan, H. T., Linares-Pastén, J. A., Nandakumar, A., Hakoshima, T., & Sudesh, K.** (2022). Polyhydroxyalkanoate synthase (PhaC): The key enzyme for biopolyester synthesis. *Current Research in Biotechnology, 4*, 87-101. DOI: 10.1016/j.crbiot.2022.01.002

6. **Tan, I. K. P., Foong, C. P., Tan, H. T., Lim, H., Zain, N.-A. A., Tan, Y. C., Hoh, C.C., Sudesh, K.** (2020). Polyhydroxyalkanoate (PHA) synthase genes and PHA-associated gene clusters in *Pseudomonas* spp. and *Janthinobacterium* spp. isolated from Antarctica. *Journal of biotechnology, 313*, 18-28. DOI: 10.1016/j.jbiotec.2020.03.006

7. **Matias, F., Bonatto, D., Padilla, G., Rodrigues, M. F. d. A., & Henriques, J. A. P.** (2009). Polyhydroxyalkanoates production by actinobacteria isolated from soil. *Canadian journal of Microbiology, 55*(7), 790-800. DOI: 10.1139/w09-029

8. **Bouznada, K., Bouras, N., Schumann, P., Spröer, C., Sabaou, N., & Klenk, H.-P.** (2016). *Actinophytocola algeriensis* sp. nov., an actinobacterium isolated from Saharan soil. *International journal of systematic and evolutionary microbiology, 66*(7), 2760-2765. DOI: 10.1099/ijsem.0.001136

9. **Bakli, M., Karim, L., Mokhtari-Soulimane, N., Merzouk, H., & Vincent, F.** (2020). Biochemical characterization of a glycosyltransferase Gtf3 from *Mycobacterium smegmatis*: a case study of improved protein solubilization. *3 Biotech, 10*(10), 1-13. DOI: 10.1007/s13205-020-02431-x

10. **Bakli, M., Bouras, N., Paşcalău, R., & Șmuleac, L.** (2023). In silico structural and functional characterization of an endoglucanase from *Actinoalloteichus hoggarensis*. *Advanced Research in Life Sciences, 7*(1), 135-141. DOI: 10.2478/arls-2023-0015

11. **Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A.** (2005). Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*, 571-607. DOI: 10.1385/1-59259-890-0:571

12. **Geourjon, C., & Deleage, G.** (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics, 11*(6), 681-684. DOI: 10.1093/bioinformatics/11.6.681

13. **Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A.** (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. DOI: 10.1038/s41586-021-03819-2

14. **Xu, D., & Zhang, Y.** (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical journal, 101*(10), 2525-2534. DOI: 10.1016/j.bpj.2011.10.024

15. **Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M.** (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography, 26*(2), 283-291. DOI: 10.1107/S0021889892009944

16. **Schrodinger, L.** (2010). The PyMOL molecular graphics system. *Version, 1*(5), 0.

17. **Hirokawa, T., Boon-Chieng, S., & Mitaku, S.** (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics (Oxford, England), 14*(4), 378-379. DOI: 10.1093/bioinformatics/14.4.378

18. **Roy, A., Yang, J., & Zhang, Y.** (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic acids research, 40*(W1), W471-W477. DOI: 10.1093/nar/gks372

19. **Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R. A.** (2023). The conserved domain database in 2023. *Nucleic acids research, 51*(D1), D384-D388. DOI: 10.1093/nar/gkac1096

20. **Ikai, A.** (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry, 88*(6), 1895-1898. DOI: 10.1093/nar/gks372

21. **Kyte, J., & Doolittle, R. F.** (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology, 157*(1), 105-132. DOI: 10.1016/0022-2836(82)90515-0

22. **Chek, M. F., Kim, S.-Y., Mori, T., Arsad, H., Samian, M. R., Sudesh, K., & Hakoshima, T.** (2017). Structure of polyhydroxyalkanoate (PHA) synthase PhaC from Chromobacterium sp. USM2, producing biodegradable plastics. *Scientific reports, 7*(1), 5312. DOI: 10.1038/s41598-017-05509-4

23. **Singh, A. K., Sharma, L., Srivastava, J. K., Mallick, N., & Ansari, M. I.** (2018). Microbially originated polyhydroxyalkanoate (PHA) biopolymers: an insight into the molecular mechanism and biogenesis of PHA granules. *Sustainable biotechnology-enzymatic resources of renewable energy*, 355-398. DOI: 10.1007/978-3-319-95480-6_14

24. **Behera, S., Priyadarshanee, M., & Das, S.** (2022). Polyhydroxyalkanoates, the bioplastics of microbial origin: Properties, biochemical synthesis, and their applications. *Chemosphere, 294*, 133723. DOI: 10.1016/j.chemosphere.2022.133723