Journal of Experimental Biology and Agricultural Sciences, December - 2022; Volume – 10(6) page 1331 – 1343

Journal of Experimental Biology and Agricultural Sciences

http://www.jebas.org

ISSN No. 2320 – 8694

Functional, and phylogenetic analysis of maleylacetate reductase of *Pseudomonas* sp strain PNPG3: An in-silico approach

Sk Aftabul Alam[1], Pradipta Saha[1*]

Department of Microbiology, The University of Burdwan, Golapbag, Burdwan-713104, WB, India

Received – July 25, 2022; Revision – October 13, 2022; Accepted – November 28, 2022
Available Online – December 31, 2022

DOI: http://dx.doi.org/10.18006/2022.10(6).1331.1343

KEYWORDS

Maleylacetate reductase

Pseudomonas

PNP biodegradation

Homology modeling

STRING database

ABSTRACT

Shrinking freshwater ecosystems are under tremendous pollution threat due to anthropocentric activities. Para nitrophenol (PNP), a well-documented priority pollutant extensively used in dyes, petrochemical, pharmaceutical, explosives, pesticides, leather industries, and agrochemicals, is responsible for contaminating aquatic ecosystems globally. It is highly toxic and has carcinogenic and mutagenic effects on living organisms like humans and several animal models. Bioremediation approaches mainly involving bacteria are considered the best, most eco-friendly, cost-effective, green, and clean method for effective removal PNP from its contaminated sites. This manuscript highlights the structural and functional analysis of a lower pathway enzyme involved in PNP degradation, maleylacetate reductase (MR), from *Pseudomonas* sp strain PNPG3, which was recently isolated from a freshwater ecosystem. This enzyme plays a role in converting maleylacetate to 3-oxoadipate. Despite its crucial functional role, no model is available for this protein in the protein database (PDB). Therefore, attempts were made for the computational investigation of physicochemical, functional, and structural properties, including secondary, and tertiary structure prediction, model quality analysis, and phylogenetic assessment using several standard bioinformatics tools. This enzyme has a molecular weight of about ~37.6 kDa, is acidic and thermostable, belonging to a member of iron-containing alcohol dehydrogenase. Moreover, this study will benefit the scientific community in deciphering the prediction of the function of similar proteins of interest.

* Corresponding author

E-mail: psaha@microbio.buruniv.ac.in (Pradipta Saha)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI] (http://www.horizonpublisherindia.in/).

1 Introduction

P-Nitrophenol (PNP) considered a priority environmental pollutant, is one of the most abundant chemical pollutants, which has been thoroughly and extensively used in the manufacturing of dyes, explosives, plastics, medicines, and a diverse range of chemical products (Bhushan et al. 2000; Kuang et al. 2020). Besides, PNP is reported to be released into the environment from parathion-based pesticides and several other agrochemicals resulting in possible high levels of PNP contaminations in several ecosystems. It is highly toxic and is reported to have carcinogenic and mutagenic effects on several animal models, causing dysfunction of the kidney, liver, and other crucial physiological life processes of animals (Kuang et al. 2020).

Since genomic DNA is the central repository of all the information for an organism, the study of genome sequences of PNP degrading bacterium (especially in this omics era) may provide a blueprint of the full potential of these bacteria (Koonin et al. 1996). To execute these, rational approaches involving wet laboratory-based experimental and computational work are necessary (Munn et al. 2017). It is now possible to hypothesize and speculate the possible potential of a bacterium with the current tools and methods available in public servers and databases (Booth et al. 2013).

Enzymes perform a significant role in the biodegradation of toxic nitroaromatic compounds. Maleylacetate reductase (MR) is one of the essential enzymes in aerobic bacteria for the degradation of PNP and other chloroaromatic compounds (Vikram et al. 2013). This enzyme catalyzed NADPH or NADP-dependent reduction, converting the carbon-carbon double bond (maleylacetate) to 3-oxoadipate (Vikram et al. 2013). Currently, for the vital enzyme MR, no protein model is available at PDB.

Considering the significance and application of MR, particularly its importance in the biodegradation of toxic compounds like PNP, the current study was undertaken for its computational analysis. In these contexts, we carry out in-silico bioinformatic analyses of MR (selected by genome mining) from the genome sequence of a bacterium *Pseudomonas* sp strain PNPG3. Attempts were also undertaken to predict the secondary structure, modelling the tertiary structure, and analyses of MR of the strain using various standard tools to intimate the scientific community about the structure-function details of this particular vital protein.

2 Materials and Methods

2.1 Sequence retrieval and phylogenetic analysis

The sequence of MR was retrieved from NCBI manually in FASTA format [(locus tag - MJ643_19190; protein id-MCK2122717.1; GenBank accession no of node 31 - JALLKV010000031; Whole Genome accession no- JALLKV010000000) or RAST accession no. - fig|286.4486.peg.3636]. This genome sequence was retrieved from a PNP degrading aquatic bacterium *Pseudomonas* sp strain PNPG3, isolated from the Ganges river, Chinsura, Hooghly district, West Bengal, India (Biosample Id SAMN26116625). This sequence was used later for computational investigation, including molecular modeling.

NCBI-BLAST_P search was performed to find similar protein sequences. FASTA formatted sequences were retrieved and the phylogenetic tree was constructed using MEGA11 (Tamura et al. 2021) (based on the Maximum Likelihood tree method). The evolutionary distance was computed by the pairwise distance method. The enzyme P-nitrophenol-4-monooxygenase (6AIN_A) of *P. putida* strain DLL-E4 was used as an outgroup.

2.2 Protein localization

PSORTb forecasted localization of the protein– 3.0 (Yu et al. 2010) and CELLO (Yu et al. 2004) (http://cello.life.nctu.edu.tw), CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/) (Yu et al. 2014). PSORTb provides the localization of bacterial protein most accurately. CELLO is user-friendly and simple, and no other algorithm is required for this tool. This tool provides better performance than the PSORT-B tool (Yu et al. 2004). The signal peptide sequence was determined using Signal P - 6.0 Server (Teufel et al. 2022). Information about the transmembrane protein was determined using TMHMM v 2.0 (Krogh et al. 2001).

2.3 Primary sequence analysis

Primary sequence analysis, including the physicochemical characteristics of the protein of interest, was predicted by the Expasy ProtParam Tool (Gasteiger et al. 2005). A variety of parameters like the composition of amino acid, isoelectric point (pI), molecular weight (MW), aliphatic index (AI) (aliphatic side chains occupying a relative volume of protein), and Grand average of hydropathicity (GRAVY) data of the protein were obtained from this server.

2.4 Secondary structure prediction

Prediction of the secondary structure of the protein was done by Netsurf-P (Klausen et al. 2019). Here solvent accessibility, the number of helices, strands, coils, and structure disorder can be determined. Another tool was used for this purpose- Chou and Fasman Secondary Structure Prediction Server (CFSSP) (Kumar 2013). Secondary structural elements, like the beta sheet, alpha helix, and turns from the amino acid sequence could be determined in a lined sequential view by this server (Kumar 2013).

2.5 Structural modelling and evaluation

The same protein of strain PNPG3 was nominated to forecast the tertiary structure. The I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang et al. 2015) was used for modeling the tertiary structure of MR of strain PNPG3. The modules like ERRAT, PROCHECK, and Verify 3D under the SAVES Server (http://services.mbi.ucla.edu/SAVES/) were used to evaluate the quality of the prepared structural model of the protein. The .pdb file obtained from I-TASSER was used to obtain the Ramachandran plot from the Structural Analysis and Verification Server (SAVES) (http://services.mbi.ucla.edu/SAVES/ Ramachandran/). Ramachandran plot helps to invent the energetically favourable regions which were connected to backbone dihedral angles (ψ and φ) of amino acid residues of the interested protein. The tertiary structure of the predicted model was visualized by UCSF chimera (http://www.cgl.ucsf.edu/chimera) (Pettersen et al. 2004).

2.6 Prediction of ligand binding sites and active sites

After validation of the tertiary structure of MR, protein-ligand binding interactions were analyzed using the COFACTOR module (https://zhanglab.ccmb.med.umich.edu/COFACTOR/) (Roy et al. 2012) integrated into the I-TASSER server for protein structure prediction and structure-based function annotation (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Yang et al. 2015). Important amino acids which were involved in ligand binding sites and active sites were also predicted.

2.7 Functional analysis

The STRING server (Szklarczyk et al. 2019) (http://string.db.org) was used to predict functionally interacting proteins. MR of *P. fluorescens* F113 was selected as the query sequence for this interactome analysis. Functional motifs of the protein sequence were identified by the tool known as MOTIF search (http://www.genome.jp/tools/motif/). Highly conserved regions among the MR were investigated by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

3 Results and Discussion

3.1 Sequence retrieval and Phylogenetic analysis

Homologous amino acid sequences of MR from other strains of *Pseudomonas* were retrieved from NCBI. To compare evolutionary relationships among the MR of strain PNPG3 with the close relatives, a phylogenetic tree (based on the maximum likelihood method) was constructed using MEGA11. From this analysis, it was evident that the MR of strain PNPG3 was placed within similar sequences from various strains representing the genus *Pseudomonas*. It also formed a cluster together with *Pseudomonas* sp. strain PDS-7 (AWB99095.1). This cluster was closely related to MR of strain NyZ402 (ACZ51383.1) and the latter formed a well-defined lineage. Also, the remaining MR sequences from the representative 8 strains were well separated (Figure 1), indicating a robust probable correlation among these taxa based on their protein sequences.

3.2 Localization of Protein

Both PSORTb and CELLO data revealed that MR is a cytoplasmic protein. The cytoplasmic score for PSORTb and CELLO is 8.46 and 3.362, respectively. CELLO2GO tool also supported the result of the cytoplasmic nature of this protein (Figure 2). According to Grasso et al. (2021), the subcellular localization of protein determines the environment where it operates, and it could be strongly correlated with the determination of said protein function. The function of a protein could be correlated to the localization of

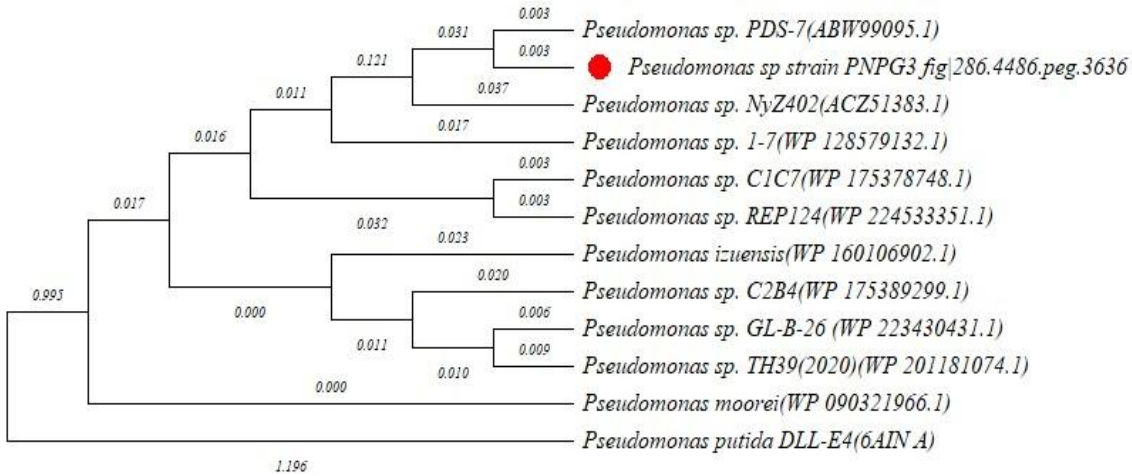


Figure 1 Phylogenetic analysis (using maximum likelihood method) of amino acid sequences of maleylacetate reductase (MR) of different strains of *Pseudomonas* sp. Values at the internodes indicate branch lengths

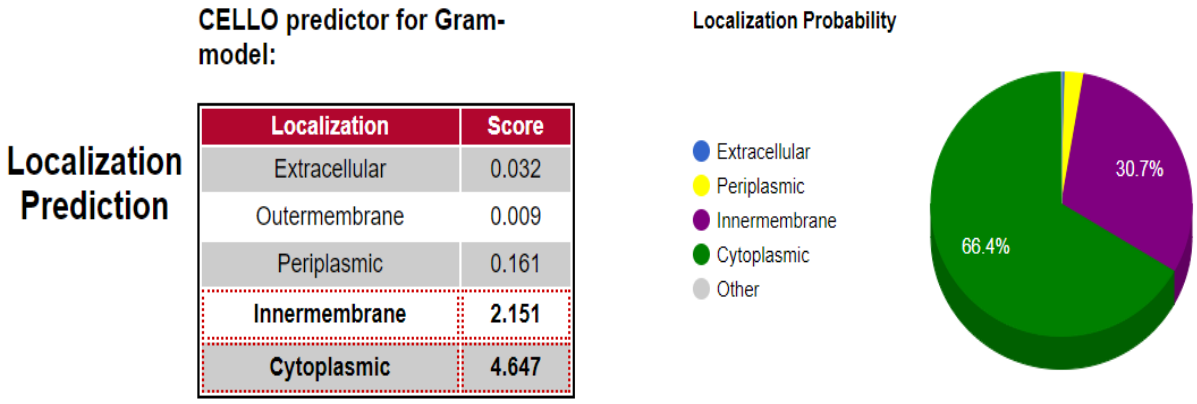


Figure 2 Localization prediction for the protein MR by CELLO2GO

the protein. SignalP - 6.0 server indicated that this protein has no signal peptide sequence. TMHMM-2.0 server indicated that no transmembrane helices are present within this protein. These results predicted that the MR protein is functionally active in the cytoplasmic region of the cell.

3.3 Primary sequence analysis of protein

Physicochemical characterization plays an important role in knowing the protein's nature (Pramanik et al. 2017). The Expasy ProtParam tool indicated the proportion of alanine (15.5%) to be higher than the other amino acids, indicating the stable nature of the protein (Figure 3). The protein consisted of 355 amino acid residues having a molecular weight of ~37.6 kDa. Molecular weight is one of the crucial criteria for the functional characterization of proteins (Prabhu et al. 2020). Theoretical pI is 5.81, indicating the acidic nature of this protein. For purification and isoelectric focusing in the development of buffer, pI prediction plays an important role (Prabhu et al. 2020). The aliphatic index of the protein was very high (95.83), indicating the protein is thermostable, indicating this protein can be used for industrial or research applications. The general concept is that less thermostable enzymes generally have limited applications in industry, that's why researchers are actively engaged in searching for highly thermostable proteins for industrial applications (Delgove et al. 2018). The grand average of hydropathicity (GRAVY) was low (0.063), indicating this protein has better interaction with water. Similar types of results were reported in the case of *Rhodococcus rhodochrous* while working with alkane-1-monooxygenase (Pal and Sengupta 2022).

3.4 Secondary structure prediction

Netsurf-P server prediction revealed that there are 14 helical structure and 9 strands within this protein (Figure 4). Prediction of secondary structure by CFSSP showed that the protein has alpha-helix 73.2%, sheets 33.5%, and turns 12.7%. From this secondary

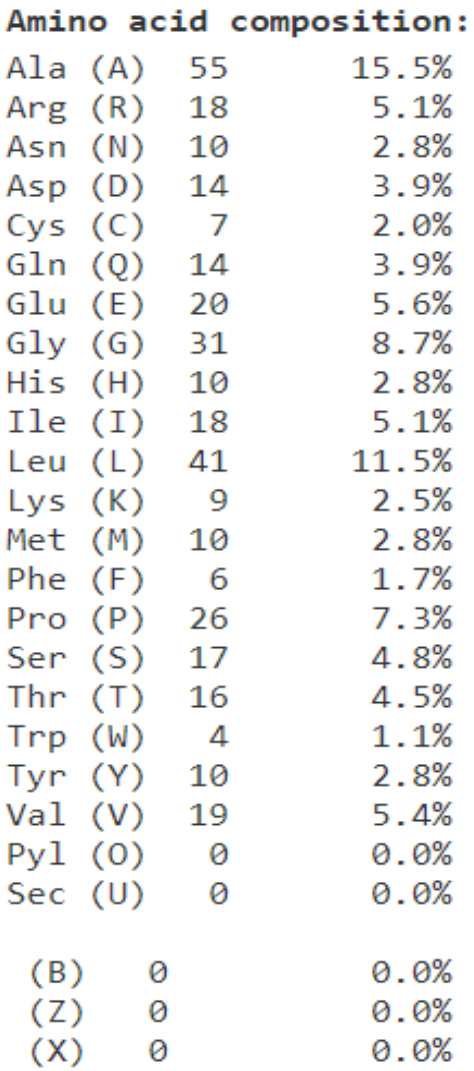


Figure 3 Amino acid composition of the sequence of MR determined by Expasy protparam tool

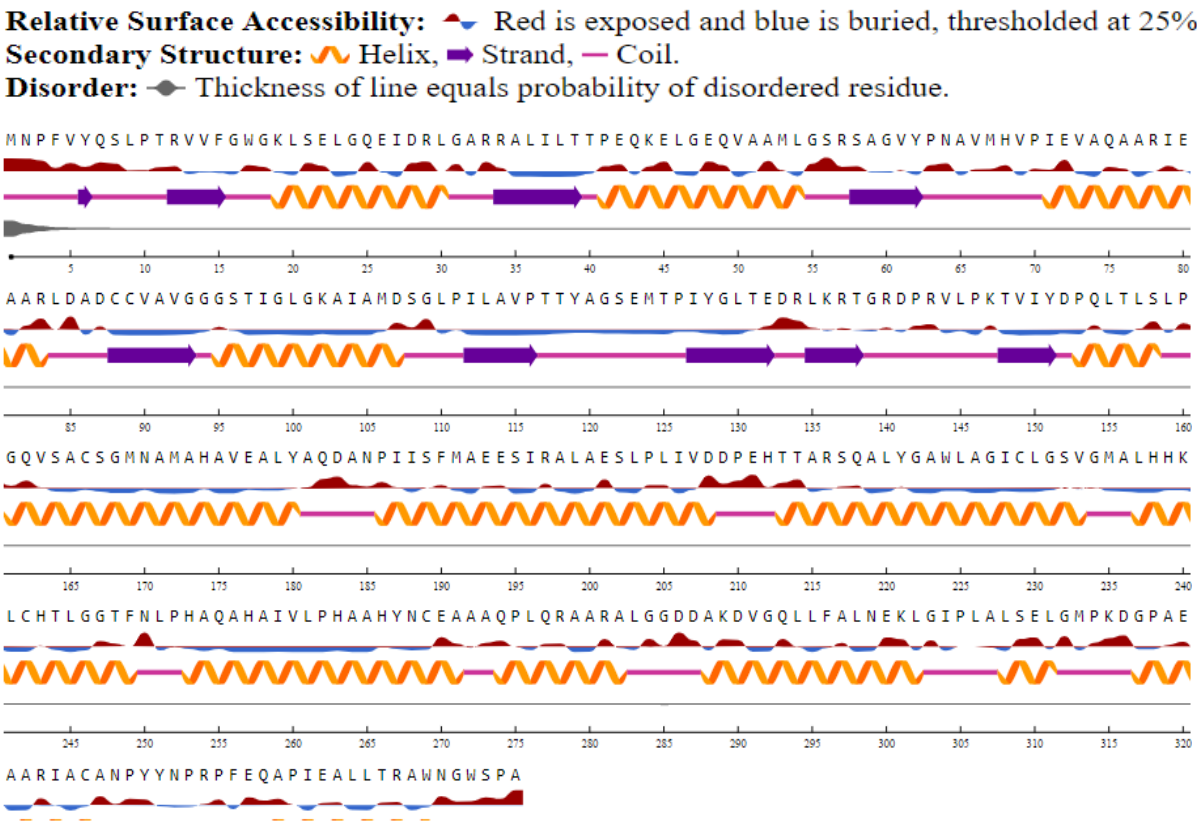


Figure 4 Secondary structure of the enzyme was determined by Netsurf-P

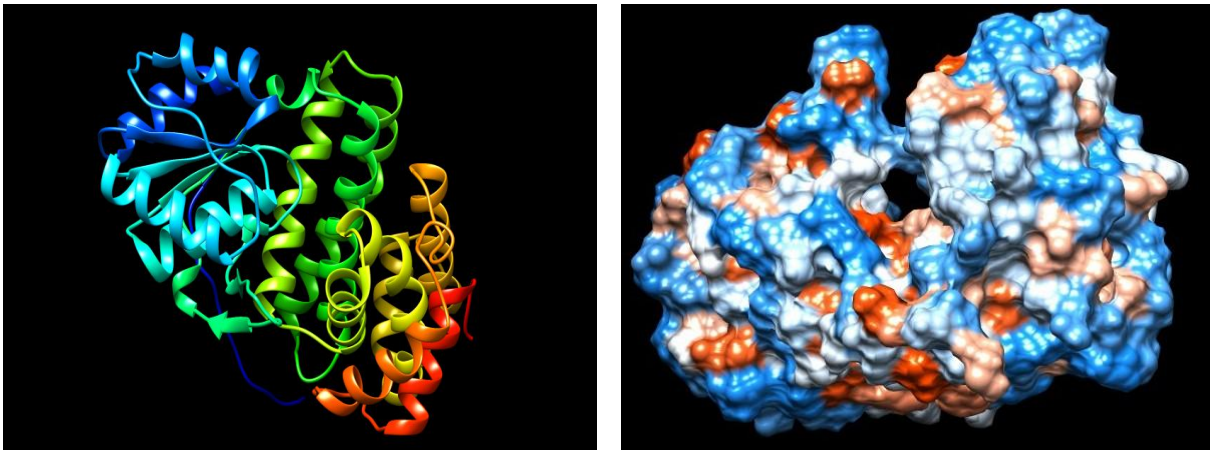


Figure 5a Predicted ribbon structure of MR visualized by UCSF chimera; Figure 5b Predicted hydrophobicity surface view of MR, showing major and minor grooves

structure prediction, it was evident that the proportion of alpha helix was much higher than the other type of protein conformations (turn and sheet). Protein-protein interactions are often mediated by an alpha helix. The higher content of alpha-helices indicated that the protein is thermostable because such secondary structures are participated to resist the high temperature in thermophilic organisms (Kumar et al. 2000; Yakimov et al. 2016).

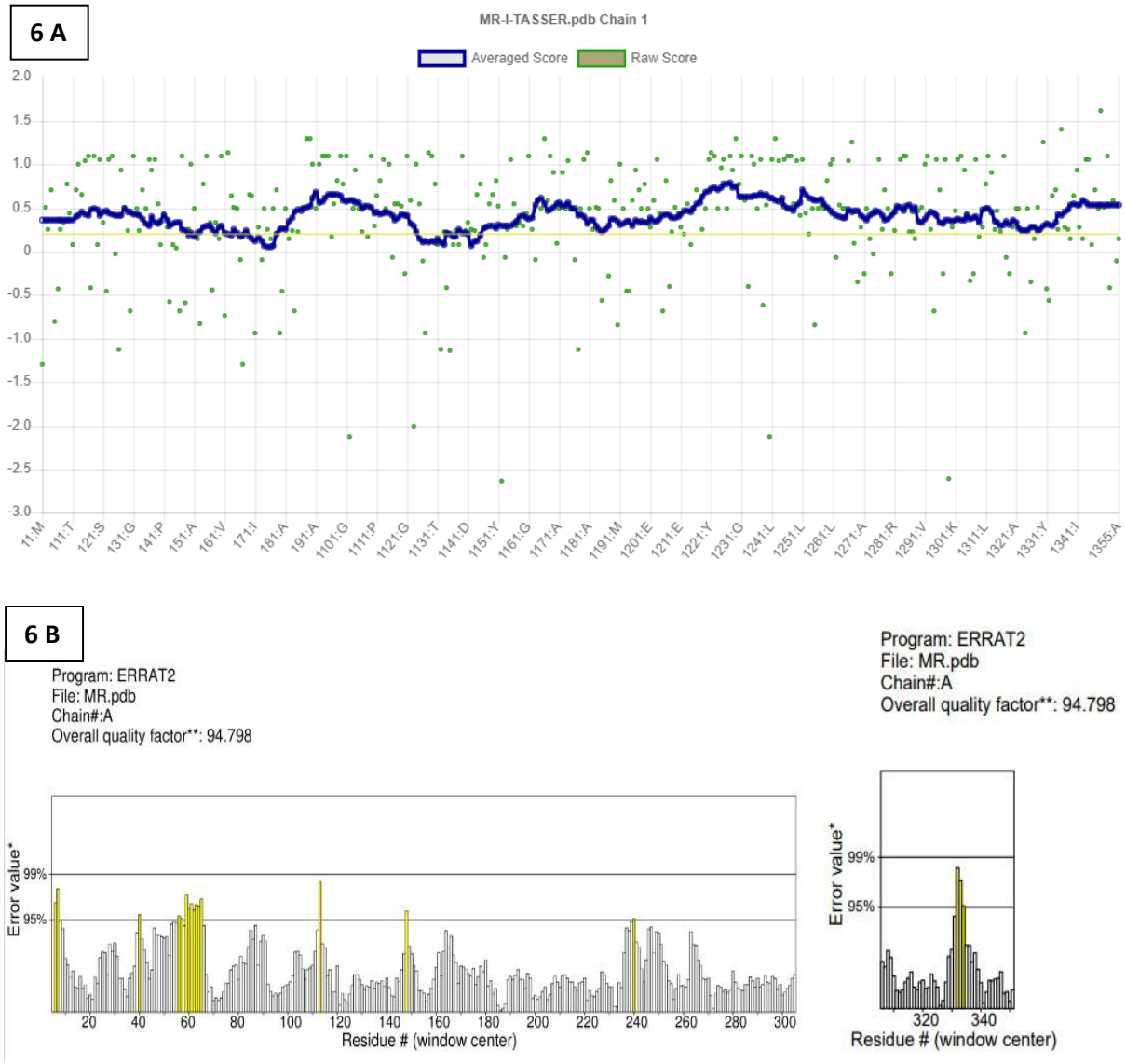
3.5 Structure analysis and model quality assessment

The I-TASSER server predicted the high-quality tertiary structure of a protein. The server generates only one model for which the C-scores is 0.81. The positive C-score of MR protein secures high confidence for the generated model. The Estimated TM-score of 0.82±0.08 and estimated RMSD of 4.9±3.2Å show a positive

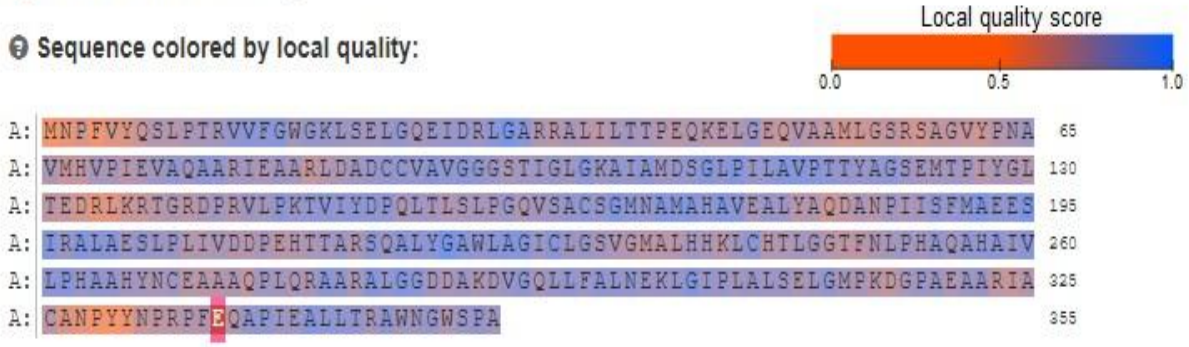
correlation with its native protein structures. The chosen template proteins like 3W5S, 6SCI, 3JZD, 1VLJ, 5BR4, 7FJG, and 6JKO by the I-TASSER server are different by origin but functionally similar i.e., oxidoreductase. From these parametric values of TM-align, the structural model of MR protein is certainly similar to the following proteins like 6SCI, 3JZD, 3BFJ, 3OWO, 4FR2, 7DAG,3ZDR,3HL0, and 2BI4 that have Oxidoreductase function. Thus, based on structural similarity with other proteins having oxidoreductase function, it can be predicted that in biological system MR protein may also have oxidoreductase properties. The helix, sheet, and turns of the predicted model was visualized by UCSF chimera, shown in Figure 5a, whereas the hydrophobicity surface was visualized in Figure 5b. Blue patches were indicated by hydrophilic area and red patches were hydrophobic in nature.

The confidence score (C-score) of the predicted model was 0.81. C-score generally ranges from [-5,2], where the higher values indicated the model was high confidence and vice-versa. The estimated TM and RMSD was reported 0.81±0.8 and 4.9±3.2Å respectively. All these values indicated that the model was highly significant. The model was selected for further analysis.

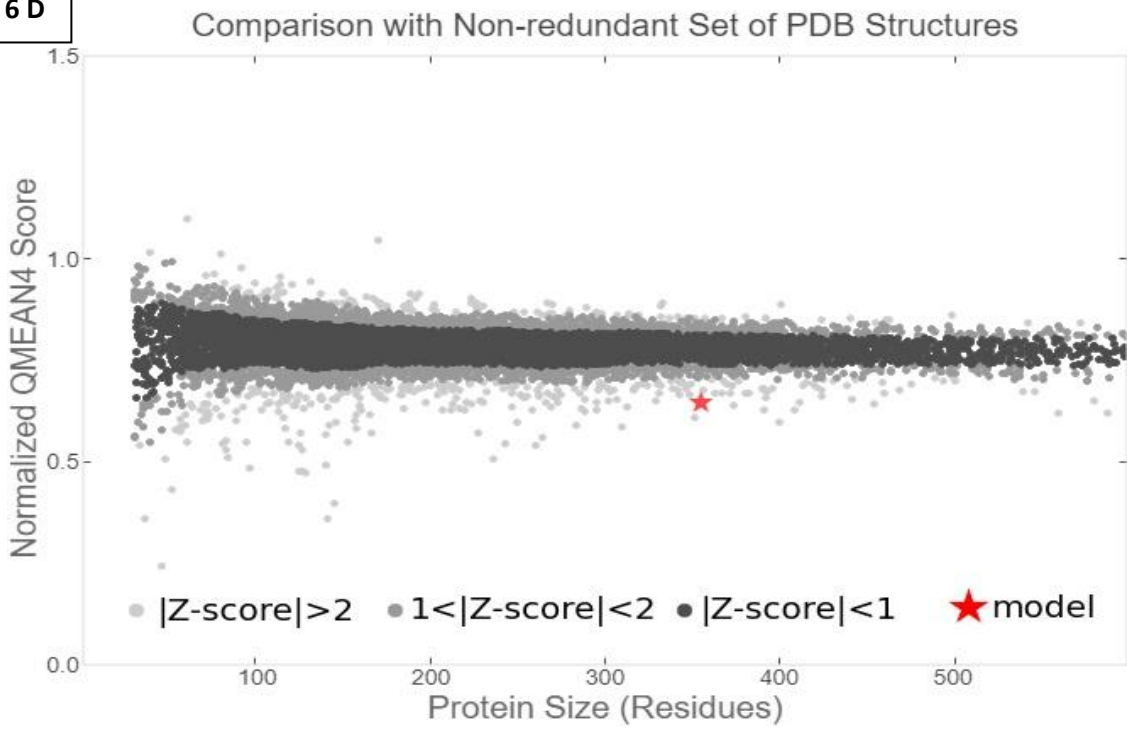
The quality of the build 3D model (.pdb) was assessed by PROCHECK, Verify 3D, ERRAT, and QMEAN Programmes. The average 3D-1D score of the protein model showed that 91.83% of residues have ≥ 0.2 scores in Verify 3D server; whereas the threshold value to attain the best fit and acceptable model is 80% (Laskowski et al. 1993) (Figure 6a). So, it could be concluded that the model was accepted and passed through Verify 3D server. The ERRAT value for this protein model was 94.7977 for both chains



6C



6 D



(Figure 6b). A higher score of ERRAT value indicates the good quality of this model. The accepted threshold value for a high-quality model is >50 (Pal et al. 2021). QMEAN is a scoring function that calculates the global as well as local per residue quality analysis of the protein of interest, based on one single model (Benkert et al. 2011). The QMEAN score close to zero denotes the good quality of the model. QMEAN4 score for this predicted protein of MR was -3.64 (Figure 6c). The QMEAN score of less than -4 is an indication of a low-quality model (Pal et al. 2021). The Z score of the query protein sequence (MR of strain PNPG3) was within an acceptable range, i.e., 1< [Z score] <2 in comparison with a non-redundant set of PDB structures (Figure 6d).

The Ramachandran plot was constructed to find the locations of amino acid residues of the protein of interest. According to the PROCHECK result, Ramachandran plot data (Figure 6e, Table 1) showed 87.2% residues were in the maximum favored regions; whereas 10.8% and 2% residues existed in the additional allowed regions and the generously allowed regions respectively. All these values were somewhat nearer to the expected value and statistically significant. Generally, a good quality model usually shows more than 90% residues in the favored regions (Berman et al. 2000; Yadav et al. 2013). A similar type of model validation was reported by Pramanik et al. (2018) and Pal et al. (2021).

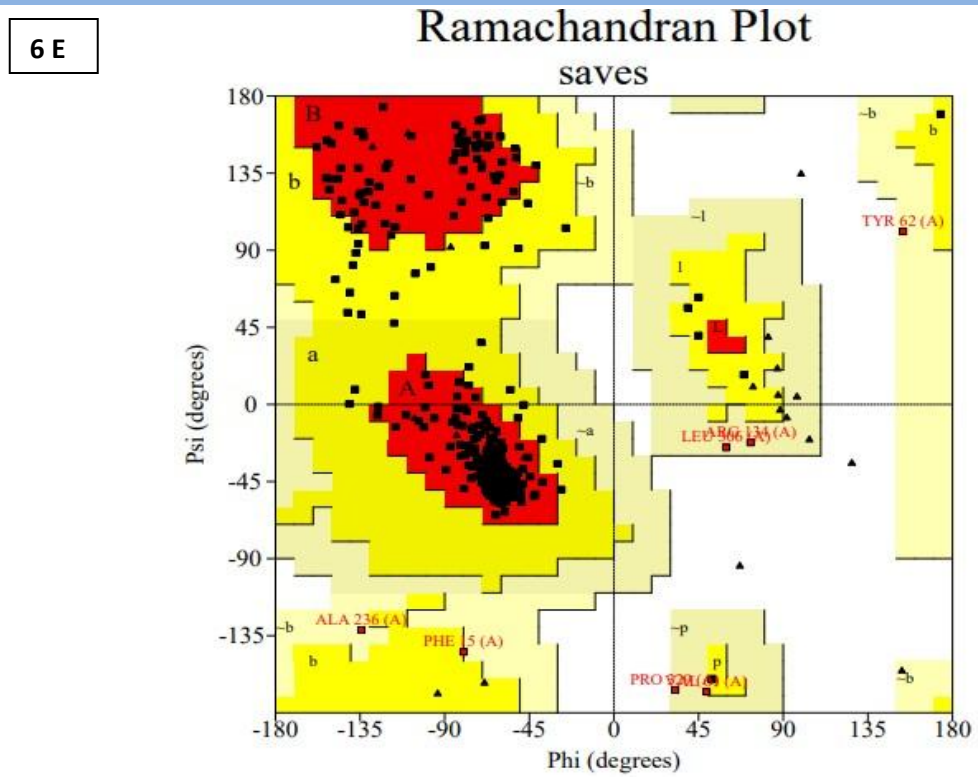


Figure 6(a-e) Evaluation of protein model of MR of *Pseudomonas* sp strain PNPG3; Figure 6a Average 3D-1D score of the protein model determined by SAVES server; Figure 6b The ERRAT value for this protein model determined by SAVES server; Figure 6c QMEAN4 score for this predicted protein of MR; Figure 6d The Z score of the predicted protein of MR; Figure 6e Ramachandran plot of the predicted protein model of MR

Table 1 Statistics of Ramachandran plot

Statistics	Number of amino acid residues	Percentage
Residues in the most favoured regions [A, B, L]	258	87.2%
Residues in the additional allowed regions [a, b, l, p]	32	10.8%
Residues in the generously allowed regions [~a, ~b, ~l, ~p]	6	2%
Residues in disallowed regions	0	0%
Number of nonglycine and nonproline residues	296	
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	31	
Number of proline residues	26	
Total number of residues	355	

3.6 Ligand binding sites and active sites determination of the protein

Ligand binding sites of a protein provide significant information related to their functional studies. The I-TASSER server indicated that NAD, Zn, and ADP were the ligands for this protein molecule; whereas COFACTOR indicated NAD, and Cobalt (Co) were the ligand. Threonine, Glutamine, Methionine, Histidine, Glycine, Glycine, Serine, Threonine, Threonine, Alanine, Serine, Threonine, Isoleucine, Lysine, Leucine, Serine, Leucine, Proline, Valine, Serine, Asparagine, Histidine were identified as ligand binding residues (determined by I-TASSER) at 40,43, 67, 68, 94, 95, 96, 117, 118, 120, 122, 125, 127, 136, 155, 158, 159, 160,163,167, 170, 253 positions respectively for NAD (Figure 7a). Similarly, Asparagine, Histidine, Histidine, and Histidine have been identified as ligand binding residues at 170, 174, 239, and

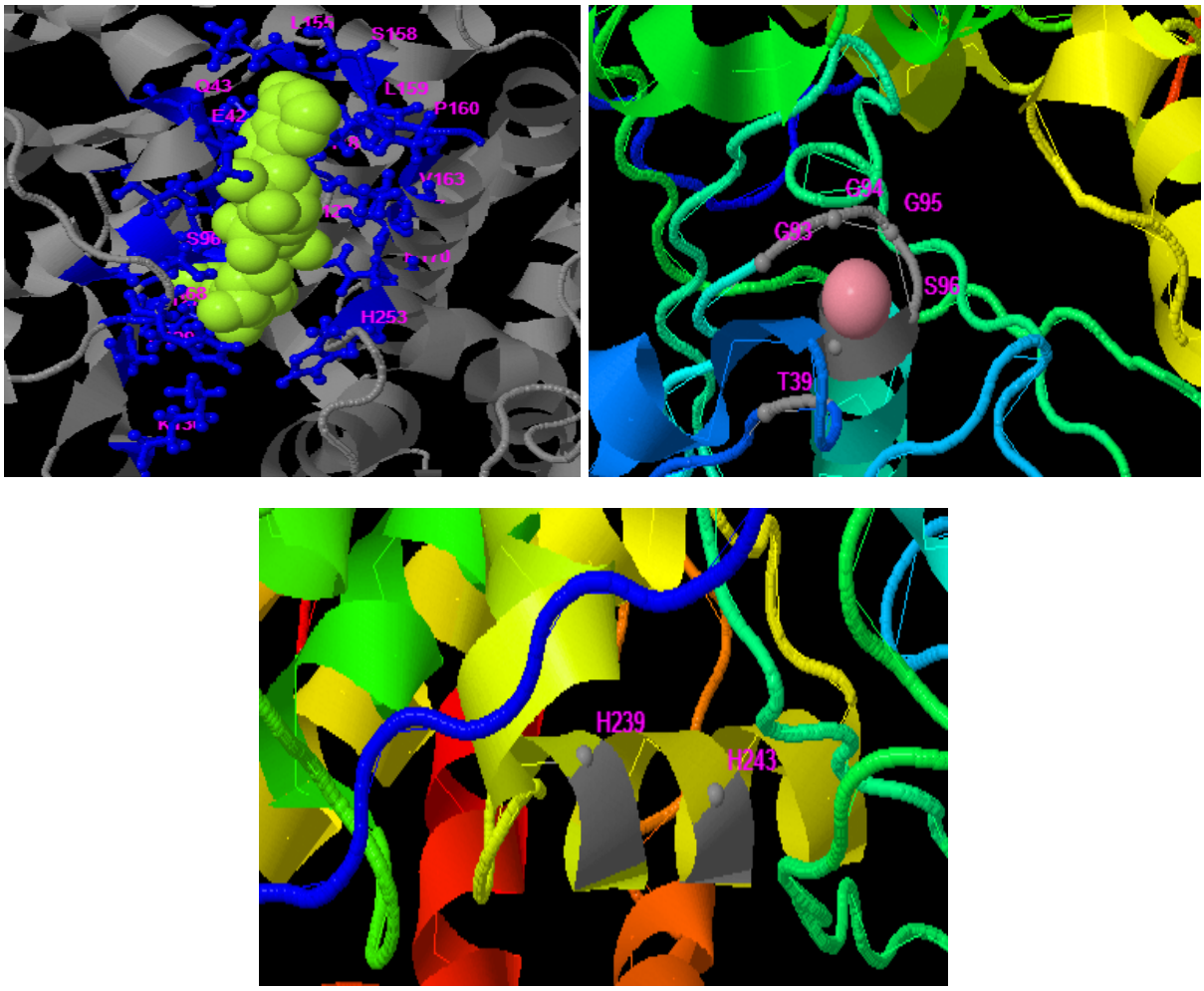
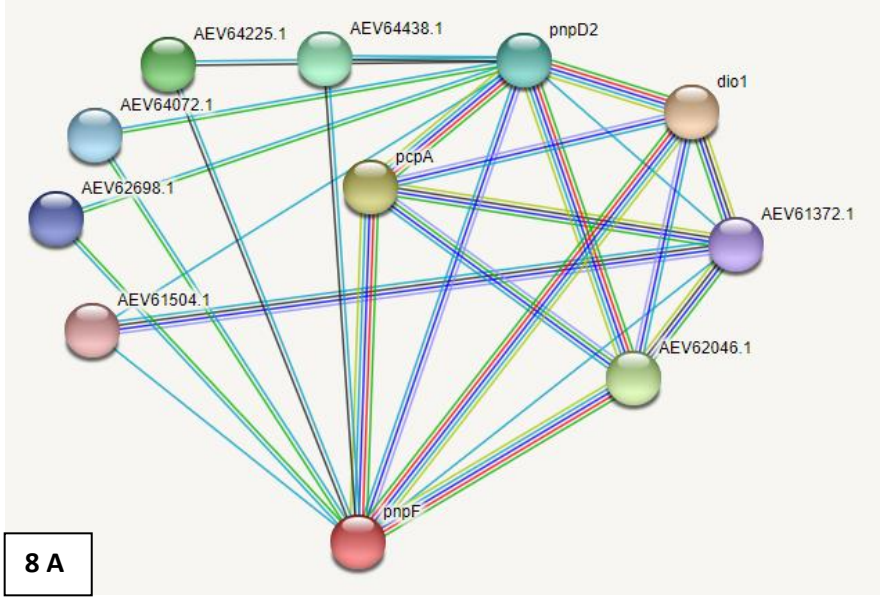


Figure 7(a-c) Ligand binding side of the protein MR; Figure 7a Ligand binding sites of NAD at the protein MR; Figure 7b Ligand binding sites of Zn at the protein MR; Figure 7c found amino acid Histidine[239] and Histidine[243]

253 respectively for Zn (Figure 7b). Threonine, Glycine, Glycine, Glycine, Serine, and Threonine were identified as ligand binding residues (determined by COFACTOR) at 39, 93, 94, 95, 96, and 97 respectively for Co. Active sites of the predicted protein were determined by the COFACTOR server. According to this server active sites of this model were found to be Histidine[239] and Histidine[243] (Figure 7c).

3.7 Functional analysis

Ten potential interacting partners of pnpF encoding protein MR were revealed by STRING analysis. The pnpF protein was MR, metal ion binding, cation transmembrane transported activity, dio1 was the closest protein with the shortest node, having catechol-1,2-dioxygenase activity, while the distantly interacting protein was recorded to be AEV61504.1 and AEV61372.1 both having 3-oxoadipate enol-lactonase activity (Figure 8a, 8b). Two functional motifs were detected by the motif finder tool (Figure 9). Both motifs were found to the member of iron-containing alcohol dehydrogenase. Pramanik et al. (2017) reported that the alkaline phosphatase of *P. aeruginosa* PA01 contains two motifs. The multiple sequence alignment among the eleven MR protein sequences revealed several stretches of conserved amino acid sequences (designated as*). Such conserve sequences are – YQSLPTRVVFGWGKLS-RLGARRALILTTPEQ-LGEQVAA-AVMHVP-EVAQAAR-EAARL-ADCC-GGGSTIGLGKAIAMDSGLPI-PTTYAAQDANPII-FMAEESIR-QALYGAW-AGICLGS-GMALHHK-CHTLGGTFNLPHAQAHIVIPHAAHYN-EAAA-RAARALGG-KLGIPLAL-GPAEAA-IACA-PYYNPRPFEQ-PIEALL (Figure 10). Pramanik et al. (2018) while working on myo-inositol hexakisphosphate phosphohydrolase of *Enterobacter*, revealed that DG-DP-LG was a conserved sequence of that enzyme. *Aspergillus niger* 3-phytase A and 3-phytase-B have conserved sequences like RHGXRXP-HD as explored by a multiple sequence alignment study by Niño-Gómez et al. (2017).



8 B

Your Input:

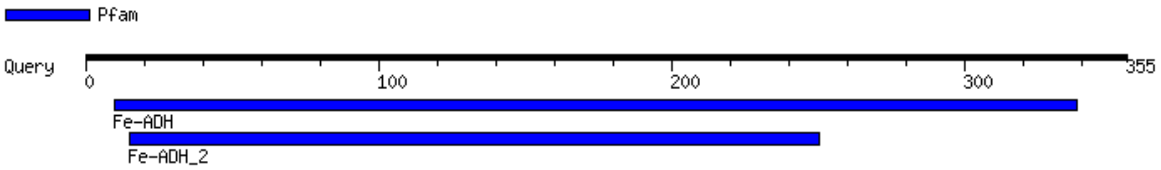
		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
pnpF	PnpF; Maleylacetate reductase; Protein involved in maleylacetate reductase activity, metal ion binding, cation transmembrane transporter activity, oxidation reduction, cation transport and transmembrane transport (355 aa)									

Predicted Functional Partners:

dio1	Dio1; Catechol 1,2-dioxygenase 1; Protein involved in catechol 1,2-dioxygenase activity, ferric iron binding, hydroxyquinol 1,...	●	●	●			●	●		0.990
pcpA	PcpA; Catechol 1,2-dioxygenase; Protein involved in oxidoreductase activity, acting on single donors with incorporation of ...	●	●	●			●	●		0.977
AEV62046.1	Hydroxyquinol 1,2-dioxygenase; Protein involved in catechol 1,2-dioxygenase activity, ferric iron binding, oxidation reductio...	●	●	●			●	●		0.975
AEV64225.1	Dienelactone hydrolase family protein; Protein involved in carboxymethylenebutenolidase activity, phenylpropanoid catabo...				●		●			0.903
AEV64438.1	Dienelactone hydrolase family; Protein involved in carboxymethylenebutenolidase activity, aromatic compound catabolic p...				●		●			0.903
pnpD2	PnpD; Iron-containing alcohol dehydrogenase; Protein involved in iron ion binding, maleylacetate reductase activity, fatty a...			●			●		●	0.902
AEV64072.1	annotation not available	●					●			0.902
AEV62698.1	annotation not available	●					●			0.901
AEV61372.1	Beta-ketoadipate enol-lactone hydrolase; Protein involved in 3-oxoadipate enol-lactonase activity, beta-ketoadipate pathwa...						●			0.900
AEV61504.1	3-oxoadipate enol-lactonase; Protein involved in transcription factor activity, 3-oxoadipate enol-lactonase activity, beta-ket...						●			0.900

Figures 8a and 8b Result of STRING based interactome study of maleylacetate reductase

Number of found motifs: 2



Pfam (2 motifs)

Pfam	Position(Independent E-value)		Description
Fe-ADH	10..338(1.4e-81)	Detail	PF00465, Iron-containing alcohol dehydrogenase
Fe-ADH_2	15..250(1.1e-16)	Detail	PF13685, Iron-containing alcohol dehydrogenase

Figure 9 The motif of maleylacetate reductase determined by themotif finder

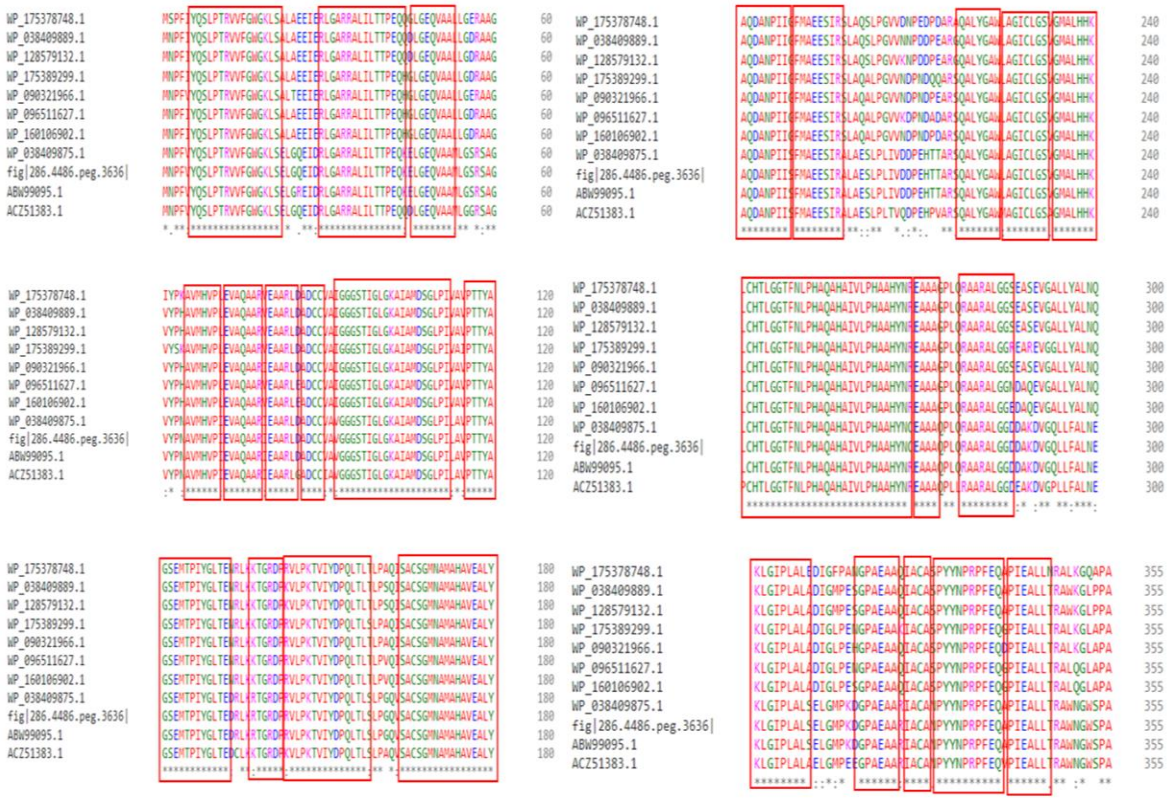


Figure 10 Multiple sequence alignment of 11 MR sequences of *Pseudomonas* spp. showing highly conserved amino acid residues. An * (asterisk) meansfully conserved amino acids, a: (colon) suggests strongly similar type of amino acids and a. (period) conveysweakly similar typesof amino acids. Hence, the hierarchy of conservation is (*) > (:) > (.). The highlighted area in the sequence indicates thehighly conserved region of the protein

Conclusion

Maleylacetate reductase, one of the major lower pathway enzymes found in different aerobic microbes, plays a vital role in the biodegradation and bioremediation of toxic xenobiotic compounds like PNP. However, no model for this protein is currently available in PDB. Thus, theoretical knowledge of this enzyme's structural, functional, and phylogenetic properties are significant to get further insight into biodegradation study by microorganisms. Computational-based analysis of protein structure has nowadays become a very beneficial tool for correlating the structure-function features of the protein as sometimes the crystal structure of the protein is not obtained readily. Here, attempts were made to determine and predict the structural-functional correlation of maleylacetate reductase of *Pseudomonas* sp strain PNPG3. The protein is thermostable and acidic, having a molecular weight of 37.6KDa. The number of alpha helices of this protein is much higher than the other types of secondary structures. Ligand binding sites and active sites of the enzyme was predicted. A model of this protein was generated with a good quality score. Both motifs were shown to be a member of iron-containing alcohol dehydrogenase. Highly conserved regions of the protein were determined by multiple sequence alignment. This work might be helpful to the scientific world of bioinformatics research. Researchers could get information about the protein's physicochemical properties, protein-protein interaction, structure, and conserve domain, and structural motifs essential for detoxification and cost-effective bioremediation in the aquatic ecosystem.

Acknowledgment

The authors are grateful to the Department of Microbiology, The University of Burdwan, West Bengal, India. The authors gratefully acknowledge Dr. Debdoot Gupta for useful scientific discussions. Authors thankfully acknowledge Dr. Pinaki Biswas, Scientist B at National Institute of Cholera and Enteric Disease, for critically evaluation of the manuscript. We are thankful to Suchetana Pal, JRF, Department of Microbiology, BU, for sharing knowledge related to bioinformatics work.

Ethical Approval

The manuscript does not involve any work or studies with human and or animal participants performed by any of the authors.

Conflict Declarations

The authors declare that they do not have any conflict of interest connected to the manuscript.

References

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*, 343-350.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., et al. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*, 235–242.

Bhushan, B., Chauhan, A., Samanta, S.K., & Jain, R.K. (2000). Kinetics of biodegradation of p-nitrophenol by different bacteria. *Biochemical and Biophysical Research Communications*, *274*, 626-630.

Booth, S.C., Weljie, A.M., & Turner RJ (2013). Computational tools for the secondary analysis of metabolomics experiments. *Computational and Structural Biotechnology Journal*, *4*, e201301003.

Delgove, M.A., Elford, M.T., Bernaerts, K.V., & De Wildeman, S.M. (2018). Application of a thermostable Baeyer–Villiger monooxygenase for the synthesis of branched polyester precursors. *Journal of Chemical Technology & Biotechnology*, *93*, 2131-2140.

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, pp. 571-607.

Grasso, S., van Rij, T., & van Dijl, J.M. (2021). GP4: an integrated Gram-Positive Protein Prediction Pipeline for subcellular localization mimicking bacterial sorting. *Briefings in Bioinformatics*, *22*, 302.

Klausen, M.S., Jespersen, M.C., Nielsen, H., Jensen, K.K., et al. (2019). NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins: Structure, Function, and Bioinformatics*, *87*, 520-527.

Koonin, E.V., Mushegian, A.R., & Rudd, K.E. (1996). Sequencing and analysis of bacterial genomes. *Current Biology*, *6*, 404-416.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*, 567-580.

Kuang, S., Le, Q., Hu, J., Wang, Y., et al. (2020). Effects of p-nitrophenol on enzyme activity, histology, and gene expression in *Larimichthys crocea*. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *228*, 108638.

Kumar, S., Tsai, C.J., & Nussinov, R. (2000). Factors enhancing protein thermostability. *Protein engineering*, *13*, 179-191.

Kumar, T.A. (2013). CFSSP: Chou and Fasman secondary structure prediction server. *Wide Spectrum*, *1*, 15-19.

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*, 283-291.

Munn, M., Knuth, R., Van Horne, K., Shouse, A.W., & Levias, S. (2017). How do you like your science, wet or dry? How two lab experiences influence student understanding of science concepts and perceptions of authentic scientific practice. *CBE—Life Sciences Education*, *16*, 39.

Niño-Gómez, D.C., Rivera-Hoyos, C. M., Morales-Álvarez, E.D., Reyes-Montaño, E.A., et al. (2017). "In silico" characterization of 3-phytase A and 3-phytase B from *Aspergillus niger*. *Enzyme Research*, 2017. DOI: https://doi.org/10.1155/2017/9746191.

Pal, S., Biswas, P., Ghosh, R., & Dam, S. (2021). In silico analysis and molecular identification of an anaphase-promoting complex homologue from human pathogen *Entamoeba histolytica*. *Journal of Genetic Engineering and Biotechnology*, *19*, 1-10.

Pal, S., & Sengupta, K. (2020). Computational-based insights into the phylogeny, structure, and function of *Rhodococcus* alkane-1-monooxygenase. *3 Biotech*, *10*, 1-11.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605-1612.

Prabhu, D., Rajamanikandan, S., Anusha, S., Chowdary, M.S., Veerapandiyan, M., Jeyakanthan, J. (2020). In silico functional annotation and characterization of hypothetical proteins from *Serratia marcescens* FGI94. *Biology Bulletin*, *47*, 319-331.

Pramanik, K., Ghosh, P.K., Ray, S., Sarkar, A., Mitra, S., Maiti, T.K. (2017). An in silico structural, functional and phylogenetic analysis with three-dimensional protein modeling of alkaline phosphatase enzyme of *Pseudomonas aeruginosa*. *Journal of Genetic Engineering and Biotechnology*, *15*, 527-537.

Pramanik, K., Kundu, S., Banerjee, S., Ghosh, P.K., & Maiti, T.K. (2018). Computational-based structural, functional and phylogenetic analysis of *Enterobacter* phytases. 3 Biotech, *8*, 1-12.

Roy, A., Yang, J., & Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research, 40,* 471477.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research, 47,* 607-613.

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

Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology, 40,* 1023–1025.

Vikram, S., Pandey, J., Kumar, S., & Raghava, G.P.S. (2013). Genes involved in degradation of para-nitrophenol are differentially arranged in form of non-contiguous gene clusters in *Burkholderia* sp. strain SJ98. *PLoS One, 8,* e84766.

Yadav, P.K., Singh, G., Gautam, B., Singh, S., et al. (2013). Molecular modeling, dynamics studies and virtual screening of Fructose 1, 6 biphosphate aldolase-II in community acquired-methicillin resistant *Staphylococcus aureus* (CA-MRSA). *Bioinformation, 9,* 158.

Yakimov, A.P., Afanaseva, A.S., Khodorkovskiy, M.A., & Petukhov, M.G. (2016). Design of stable α-helical peptides and thermostable proteins in biotechnology and biomedicine. *Acta Naturae, 8,* 70-81.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods, 12,* 7-8.

Yu, C.S., Cheng, C.W., Su, W.C., Chang, K.C., et al. (2014). CELLO2GO: a web server for protein subCELlularLOcalization prediction with functional gene ontology annotation. *PloS one, 9,* e99368.

Yu, C.S., Lin, C.J., & Hwang, J.K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science, 13,* 1402-1406.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics, 26,* 1608-1615.