Universal Journal of Catalysis Science
https://ojs.wiserpub.com/index.php/UJCS/

UNIVERSAL WISER PUBLISHER

Research Article

In Silico Structural and Functional Annotations of Hypothetical Protein All4871 as a 1-Acyl-sn-Glycerol-3-Phosphate Acyltransferase from *Nostoc* sp. PCC 7120

Manish Singh Kaushik[1*], Prashant Singh[1], Vishal Singh[2], Amaresh Kumar Sahoo[2], Kamlesh Kumar Soni[3]

[1]Laboratory of Cyanobacterial Systematics and Stress Biology, Department of Botany, Institute of Science, Banaras Hindu University, Varanasi, India
[2]Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Allahabad, India
[3]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan
E-mail: manish13587@gmail.com

Received: 17 October 2022; **Revised:** 1 December 2022; **Accepted:** 12 December 2022

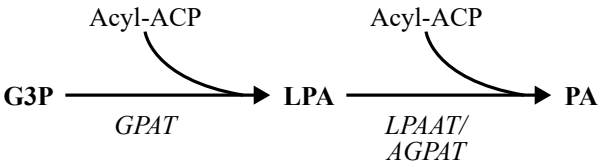
Abstract: In this study, the *in silico* structural and functional characterization of a hypothetical protein All4871 from *Nostoc* sp. PCC 7120 was performed. The 287 amino acid residue protein sequence revealed considerable homology with 1-acyl-sn-glycerol-3-phosphate acyltransferase (AGPAT) from different organisms. Multiple sequence alignment of All4871 with the AGPAT isomers showed the presence of conserved motifs (motif I-IV) crucial for acyltransferase activity. The All4871 protein depicted the characteristic topology of membrane-bound AGPATs. Both N- and C-terminus of the All4871 protein face the inner side of the membrane. The predicted tertiary structure of All4871 has an N-terminal helix and αβ catalytic core, which are characteristic features of AGPATs. Amongst the CASTp predicted binding pockets, pocket 1 acts as a groove-spanning active site crucial for acyltransferase catalysis, lysophosphatidic acid (LPA), and acyl coA (MCL) binding site. Docking studies further supported pocket 1 as the binding site for both LPA and MCL, which act as substrates for acyltransferase catalysis. So, the overall analyses have pointed toward the possibility that the hypothetical protein All4871 from *Nostoc* sp. PCC7120 may act as a 1-acyl-sn-glycerol-3-phosphate acyltransferase.

***Keywords*:** hypothetical protein All4871, AGPAT, membrane protein, LPA, MCL, Nostoc sp. PCC 7120

1. Introduction

Cyanobacteria carefully control the composition/fluidity of their membrane glycerophospholipids by using a universal two-step pathway, which regulates the positional asymmetry of fatty acids in glycerophospholipids [1-3]. In the first acylation step, the sn-glycerol-3-phosphate (G3P) is converted into 1-acyl-glycerol-3-phosphate (lysophosphatidic acid or LPA) after reacting with the acyl-ACP (fatty acyl moiety carrying acyl carrier protein), and this reaction is catalyzed by G3P acyltransferase (GPAT). The second acylation step catalyzes the conversion of 1-acyl-glycerol-3-phosphate to central intermediate phosphatidic acid (PA) with the help of enzyme 1-acyl-sn-glycerol-3-phosphate acyltransferase (LPAAT or AGPAT).

DOI: https://doi.org/10.37256/ujcs.1120232059



(1)

PA is the key intermediate in the biosynthesis of glycerophospholipids and triacylglycerols (TAG) [4]. In cyanobacteria, GPATs and AGPATs regulate the positional asymmetry of fatty acids by directing C18:0 to the sn-1 and C16:0 to the sn-2 position [5]. Unlike plants and animals, cyanobacterial AGPATs have not been well characterized. The information related to cyanobacterial GPATs and AGPATs is limited to only a few reports [3, 6, 7, 8].

The advent of advanced Next Generation Sequencing techniques has made the whole genome sequencing of newly identified organism a routine work. Although, the complete genome of a diverse group of organisms has been sequenced; however, the functional roles of as many as 30–40% of the genes have not been assigned yet [9]. Nostoc sp. PCC 7120 (hereafter denoted as Nostoc 7120) is one such model organism, whose genome was completely sequenced more than a decade ago [10, 11], however, there are still hundreds of conserved hypothetical proteins which do not have a prescribed function. Although, the central pathways regulating metabolisms in Nostoc are already known; however, these hypothetical proteins remain hurdles in understanding mechanisms regulating signaling and protein-specific responses [12]. These knowledge gaps provoked scientists to develop computational pipelines using experimental data for predicting protein structures and functions. These computational pipelines although not sufficient for establishing biological functions; however, can surely help to provide scientific openings for wet laboratory experiments.

The Nostoc 7120 contains approximately 5368 protein-encoding genes, among which 1453 (27 %) are categorized as genes code for hypothetical proteins and the assignment of each gene is listed in CyanoBase (http://www.kazusa.or.jp/cyanobase/) [10]. In the present communication, we have surveyed the CyanoBase, selected a hypothetical protein All4871, which could be a probable acyltransferase, and performed its structural and functional characterization using in VVsilico approach.

2. Materials and methods

2.1 *Sequence retrieval, domain analysis, and homology identification*

The FASTA sequence of All4871 encoding protein sequence was retrieved from the CyanoBase (http://www.kazusa.or.jp/cyanobase/) and subjected to conserved domain search using Conserved Domain Architecture Retrieval Tool (CDART; https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) [13]. The BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was then employed to perform protein homology against the non-redundant and UniProt/SwissProt database. The top 15 hits were selected and the sequence conservation was determined by multiple sequence alignment and phylogenetic analysis using the BioEdit tool [14] and MEGA6 [15], respectively.

2.2 *Protein-protein interaction analysis and topology prediction*

The probable interacting proteins for the hypothetical protein All4871 were predicted using String version 11 ([16]; https://string-db.org) and visualized in Cytoscape 3.7.2 [17]. The STRING database integrates the available information from the classification systems based on Gene Ontology and KEGG, high-throughput text-mining as well as on hierarchical clustering of the association network to generate a comprehensive network. The interaction scores represent confidence (on a scale of zero to one) in the association. Bologna Unified Subcellular Component Annotator (BUSCA) web server (http://busca.biocomp.unibo.it) was used for predicting the protein subcellular location of the hypothetical protein [18]. The topology of the hypothetical protein was predicted by using MemBrain 3.1 online web server ([19]; http://www.csbio.sjtu.edu.cn/bioinf/MemBrain/). MemBrain as a tool predicts transmembrane helices, residue–residue contacts, and the relatively accessible surface area of α-helical membrane proteins.

2.3 *Structure prediction, 3D modeling, and quality assessment*

The physicochemical properties of the hypothetical protein All4871 were determined using the Protparam tool ([20]; http://web.expasy.org/protparam/). The secondary structure of the All4871 protein was predicted by the UCLA PSIPRED server ([21] ; http://bioinf.cs.ucl.ac.uk/psipred/). The three-dimensional structure was predicted by the Phyre2 modeling tool [22] and further visualization was done using Discovery Studio version 16. The quality of the predicted structure was checked by PROCHECK [23] and verify3D of the UCLA-DOE LABS-SAVES version 6.0 ([24]; https://saves.mbi.ucla.edu/). The overall quality of the 3D model was also assessed using the ProSA-web server. The obtained Z-score determined the potential errors in the predicted 3D model and ensured that the predicted structure is within the range of scores typically found for native proteins of similar size[25]. The hypothetical protein 3D model was superimposed with the known 3D structure of the template using SuperPose version 1.0 (http://superpose.wishartlab.com/) to assess the structural similarity of All4871 with the selected template. The final validated model was then deposited to PMDB (Protein Model Data Base) [26].

2.4 *Binding pocket predictions and docking analysis*

The binding pockets in the hypothetical protein All4871 were identified using the CASTp server ([27]; http://sts.bioe.uic.edu/castp/index.html?2pk9). The top 4 pockets were selected and studied through UCSF Chimera version 1.11.2 [28]. Docking study was performed by Patchdoc server ([29]; http://www.molegro.com/products.php) and docked model was further refined using Firedock ([30]; http://bioinfo3d.cs.tau.ac.il/FireDock/). The structures of ligands [lysophosphatidic acid (LPA) and acyl-CoA (MCL)] were retrieved as SDF files from the PubChem site (http://www.ncbi.nlm.nih.gov/pccompound) and converted to PDB file using OpenBabel version 3.1.1 [31]. The ligand PDB file along with the protein PDB file (as a receptor) was imported to the Patchdoc server for docking analysis and the output was visualized in discovery studio 16.

2.5 *MD Simulation of membrane-protein complexation*

MD simulations for membrane (POPC)-protein (All4871) association were performed using GPU-accelerated Desmond package of Schrödinger [32] running on an HP Proliant DL380 Gen9 high-performance Linux cluster computer. The protein 3D model used for MD simulation was prepared by the protein preparation wizard in Maestro and ionized at biological pH (7.4 ± 2) applied by the PROPKA method. Further, the protein 3D model was refined to minimize the steric clashes. The energy minimization of the protein model was done with convergence criteria of 0.3 Å RMSD by employing the OPLS3e force field [33]. The TIP3P water model was used to create the hydration model. To neutralize the system, a simulation was performed with 0.15 M NaCl (counter ion) concentration. The cut-off radius of 9 Å, 2.0 fs, 300 K, and 1 bar were set for van der Waals, time step, initial temperature, and pressure of the system, respectively. The sampling interval was set to 100 ps and further, MD simulation was performed for 80 ns under the NPT ensemble. The stability of MD simulations was established by monitoring the RMSD of the protein atom positions throughout the simulation period.

3. Results

3.1 *Sequence analysis of hypothetical protein All4871*

The CyanoBase database was surveyed for studying the collection of hypothetical proteins from *Nostoc* 7120 genome (data source name: GCA_000009705.1). Out of several hypothetical proteins, we have selected All4871 for further studies. The FASTA sequence of hypothetical protein All4871 (287 amino acid) was retrieved from CyanoBase and used for the conserved domain prediction by using the CDART tool. The domain analysis revealed the presence of the PlsC domain (PHE4-SER256), which belongs to the LPLAT/acyltransferase superfamily (E-value 1.88e-43) (Figure A1). The protein-protein interaction analysis of the target hypothetical protein using the String tool also showed the association of All4871 protein with acyltransferases and other proteins involved in lipid metabolism (Figure A2). The hypothetical protein sequence was subjected to blastp analysis against non-redundant and UniProt/SwissProt

database, which showed its homology with 1-acyl-sn-glycerol-3-phosphate acyltransferases (AGPATs) from different cyanobacterial species (Tables A1 & A2). Further, the top 15 homologs obtained through blastp analysis against a non-redundant protein database were used to deduce the phylogeny, which showed the clustering of All4871 with different cyanobacterial AGPATs (Figure 1).

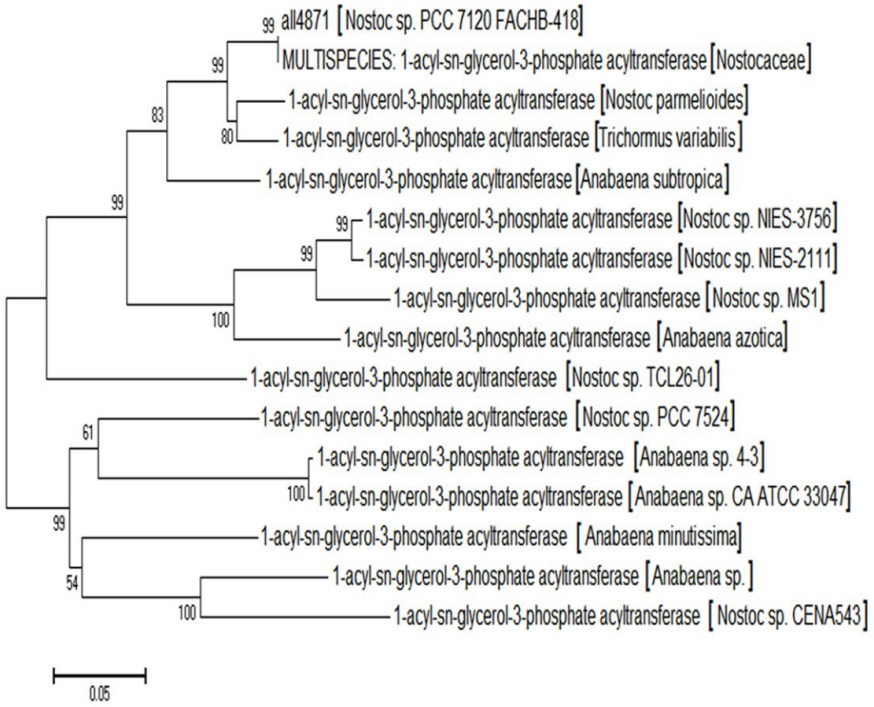


Figure 1. Phylogenetic analysis of target hypothetical protein All4871 with different cyanobacterial 1-acyl-sn-glycerol-3-phosphate (AGPAT) homologs

We have compared the sequence of All4871 with different AGPAT isoforms from *Thermotoga maritime, Escherichia coli*, *Homo sapiens*, *and Mus musculus*, which showed conservancy of the catalytically important motifs in All4871 (Figure 2). The invariant histidine and aspartate residues in motif I (HX_4D) are crucial for acyltransferase catalysis and were found highly conserved. The motif III ($XPEGX_2$), another characteristic motif of AGPATs, is also found conserved in All4871. However, the invariant arginine residue presents in motif III (PEGTR) of acyltransferase from *E. coli*, *H. sapiens*, *and M. musculus*, respectively, was found replaced by isoleucine residue in motif III (PEGNI) of All4871 (Figure 2). Motif II and IV are considered important for the acyltransferase activity but were not found well conserved in hypothetical protein All4871 (Figure 2). Further, the multiple sequence alignment of the All4871 sequence with the top 15 homologs obtained after the BLASTp analysis also showed the presence of conserved HX_4D and $XPEGX_2$ motifs in all the sequences (Figure A3). The canonical motif II and motif IV were completely replaced by other motifs in All4871, which were however found highly conserved among the cyanobacterial homologs (Figure A3). We have also predicted the physicochemical properties of the hypothetical protein All4871, which revealed the instability index, aliphatic index, and GRAVY as 53.95, 88.26, and -0.238, respectively (Table A3).

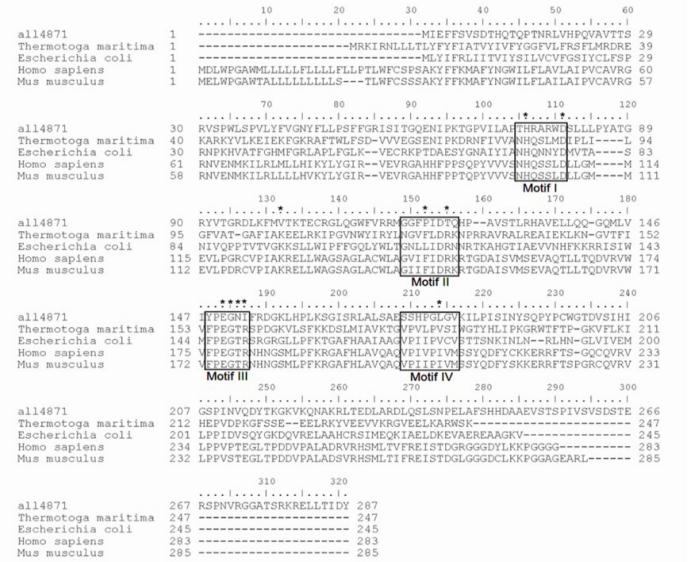


Figure 2. Multiple sequence analysis of All4871 protein sequence with the 1-acyl-sn-glycerol-3-phosphate (AGPAT) isoforms from *Thermotoga maritime, Escherichia coli, Homo sapiens, and Mus musculus*. The boxes represent the conserved motifs (I-IV) important for the acyltransferase activity. Asterisk (*) represents the amino acid residues that have important roles in acyltransferase catalysis [34]

3.2 *Sub-cellular location and topology prediction of the hypothetical protein All4871*

The BUSCA output revealed that the hypothetical protein All4871 localizes on the plasma membrane (Table A4). The topology of the target hypothetical protein was predicted through MemBrain 3.1 online web server. The MemBrain output also revealed that the hypothetical protein All4871 is a membrane protein whose N- and C-terminal both faces the inner side of the membrane (Figure 3). The hydrophobic amino acids from L39-L48 (TMH1), L82-A87 (TMH2), and S167-S172 (TMH3) build the transmembrane helices (TMH) which anchors the membrane (Figure 3). The conserved motif I lie in between TMH1 and TMH2, conserved motifs II and III are located between TMH2 and TMH3 and motif IV preceded TMH3 (Figure 3). The motif I and IV were found to be located near the membrane.

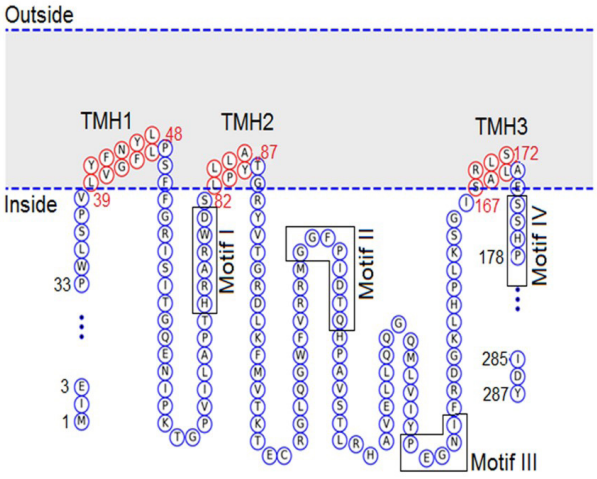


Figure 3. Schematic representation of the predicted topology of All4871 located with both N- and C- terminals on the inner side of the membrane. Amino acid residues forming transmembrane helices (TMH 1-3) and loops are presented in red and blue color, respectively. Amino acids from Met1-Val38, Pro49-Ser81, Thr88-Ile166, and Ala173-Try287 form predicted loops, while L39-L48, L82-A87, and S167-S172 form TMH spanning the membrane

3.3 *All4871 tertiary structure prediction and validation*

The secondary structure prediction of the All4871 protein revealed that the protein has 40 % α-helix, 17 % β-strand, and 6 % TMH (Figure A4). The tertiary structure of the target hypothetical protein was predicted by using the Phyre2 modeling tool against the template 5kymA (a TmPlsC protein from *Thermotoga maritima*) with 24 % identity and 67 % coverage (Figure 4a). Despite the low identity, it was found that motif I (HX_4D) and motif III ($XPEGX_2$) of hypothetical protein All4871showed high similarity with that of template protein (Figure 2). The motif III (PEGTR) of template protein has a conserved arginine which has been replaced by isoleucine residue in motif III (PEGNI) of All4871 (Figure 2). The predicted tertiary structure (PMDB id: PM0084220) contains eight α-helix and seven stranded antiparallel β-sheets, which is the characteristic feature of AGPATs. The tertiary structure of the hypothetical protein was superimposed on the template protein (Figure 4b). The small RMSD value (1.55 Å) revealed the reliability of the tertiary structure. However, it was observed that the tertiary structure of the template protein has one extra N-terminal α-helix (Figure 4b). We have further validated the predicted tertiary structure of All4871 using the PROCHECK program. The PROCHECK output revealed that the predicted tertiary structure is of good quality as most of the amino acid residues fall within the limits of favored (85.6 %), additionally allowed (12.6 %), and generously allowed region of Ramachandran plot (Figure A5). The tertiary structure of All4871 was also verified by verify3D which showed that 85.31% of the amino acid residues have scored >=0.2 in the 3D/1D profile and indicated that the environmental profile of the model is good (Figure A6). The overall model quality was estimated by the ProSA-web tool. The z-score value for the tertiary structure of All4871 and the template protein was estimated at -5.29 and -6.44, respectively, which suggested a close similarity between the template and the predicted structures and reinforced a good quality of the modeled protein (Figure A7).

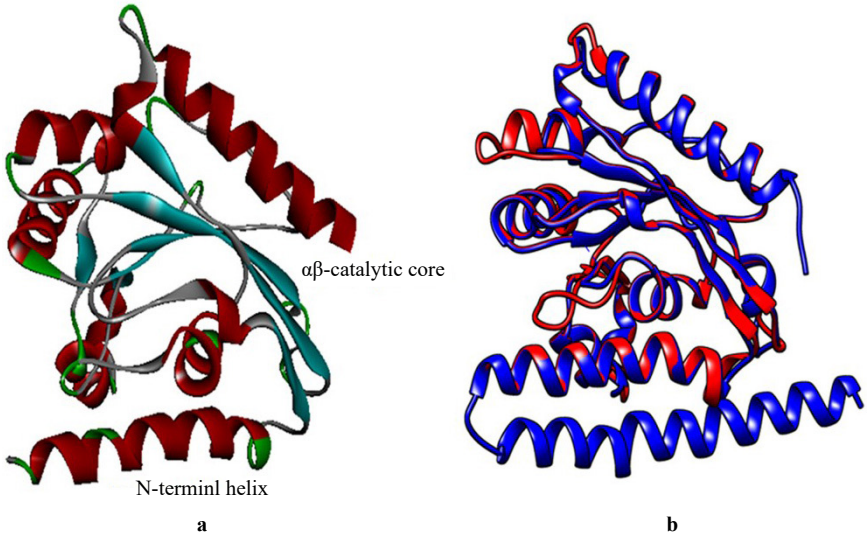


Figure 4. (a) Predicted Tertiary structure of the hypothetical protein All4871 showing eight α-helix and seven stranded antiparallel β-sheets. (b) The superimposed tertiary structure of All4871 protein (red color) with the template protein 5kymA (blue color)

3.4 *Prediction of binding pockets and docking analysis of the hypothetical protein All4871*

The binding pockets for the hypothetical protein All4871 were predicted and shown in Figure 5. We have selected the top 4 binding pockets (pockets 1-4) on the All4871 protein predicted by the CASTp server. The amino acid residues involved in the formation of binding pockets are presented in Table 1.

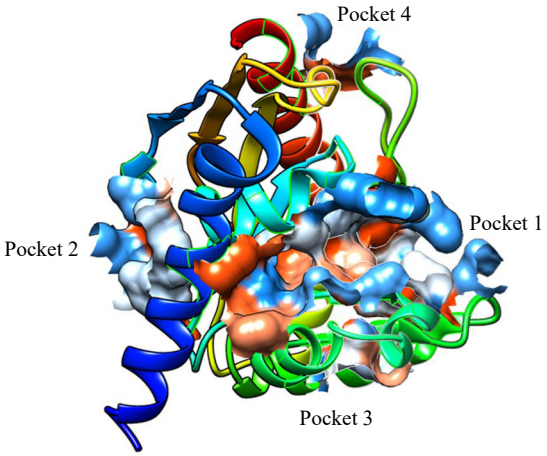


Figure 5. Predicted active sites (pocket 1-4) located on the tertiary structure of the hypothetical protein All4871

Table 1. The amino acids involved in forming active sites of hypothetical protein All4871 predicted via CASTp server

Pocket	Amino Acids
1	LEU39, VAL42, HIS75, ALA77, ARG78, ASP80, SER81, LEU82, PRO85, ARG90, LEU97, PHE99, MET100, VAL101, THR102, GLU105, GLN110, PHE113, VAL114, MET117, ILE122, GLN125, ALA128, VAL129, THR131, LEU132, VAL146, ILE147, TYR148, GLY151, ASN152, ILE 153, LYS163, SER164, GLY165, ILE166, ARG168, LEU169
2	TYR40, GLN60, ILE63, PRO64, LYS65, THR66, GLY67, ALA87, THR88, GLY89, ARG90, TYR91, VAL92, MET144
3	VAL101, GLU105, CYS106, GLY111, VAL114, ARG115, GLY119, PRO121
4	LYS158, LEU159, ALA232, LEU235, GLN236

The docking analysis of hypothetical protein All4871 was performed with the two different substrates of acyltransferase i.e., LPA and MCL. The details of the docking analysis are presented in Table A5 and Figure 6.

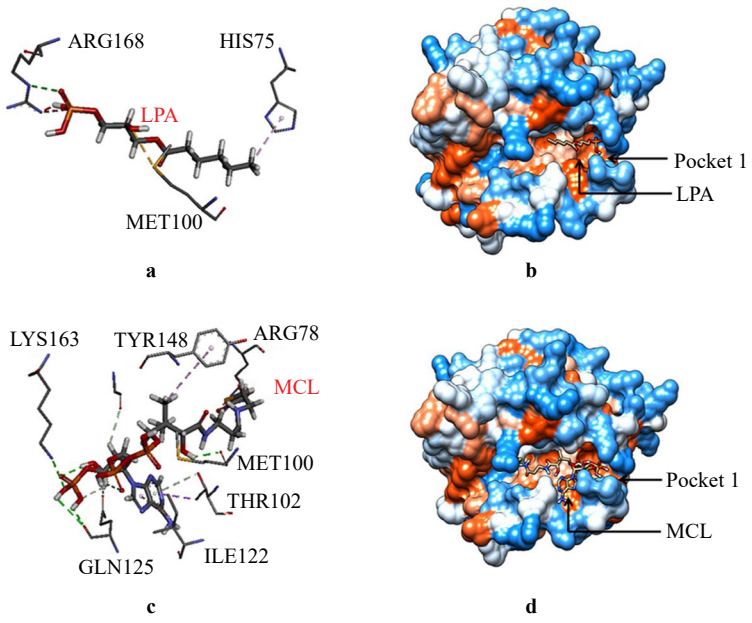


Figure 6. The docked complexes of hypothetical protein All4871. Amino acid residues interacting with LPA (a) and MCL (c). The hydrophobic surface representation of the hypothetical protein All4871 shows pockets/grooves for binding LPA (b) and MLC (d). LPA and MLC bind to CASTp predicted pocket 1

The LPA interacts with the amino acid residues His75, Met100, and Arg168 (Figure 6a). The LPA interacting residues His75, Met100, and Arg168 are involved in the formation binding pocket 1 predicted by CASTp analysis (Figure 6b & Table 1). Furthermore, MCL interacts with amino acid residues Arg78, Met100, Thr102, Ile 122, Gln125, Tyr148, Gly151, and Lys163 (Figure 6c), which are also the part of binding groove identified as binding pocket 1 by CASTp analysis (Figure 6d & Table 1). The overall docking analysis revealed that the CASTp predicted binding pocket 1 is the location spanning the active site important for acyltransferase activity, LPA binding groove, and acyl chain binding groove (Figure 7a). We have also shown the active site locale consists of the amino acid residues His75, Asp80, Tyr148 and Ile153, which were the part of two of the characteristic conserved motifs i.e., HX4D and XPEGX2, important for the acyltransferase activity (Figure 7b).

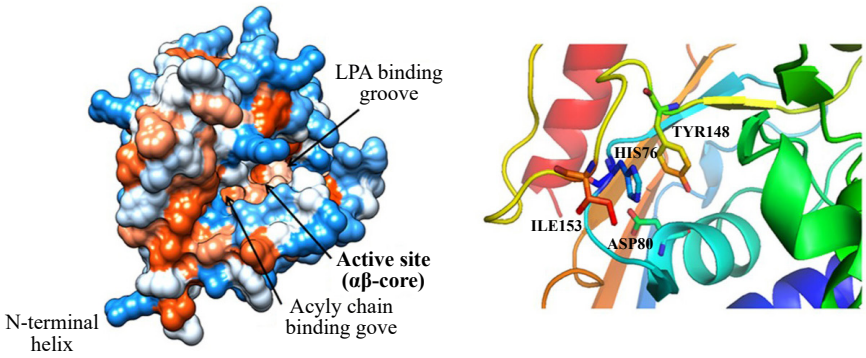


Figure 7. (a) The 3D model of All4871 representing the active site, LPA binding groove, acyl chain binding groove, and N-terminal helix. (b) The zoomed view of the active site locale of the hypothetical protein All 4871

3.5 *Analysis of membrane-protein association using MD simulation*

To get an insight into the membrane-bound nature of the All4871 protein, we performed MD simulations on the POPC membrane and predicted the tertiary structure of All4871. The simulation was run for 80 ns following the equilibration of the MD simulation. The RMSD values of side chains, Cα atoms, and heavy atoms were estimated for the All4871-POPC complex, which showed less variation and suggested that the protein model was stable while interacting with the POPC lipid bilayer (Figure A8). Figure 8a is showing the MD simulation-SU trajectory showed that the amino acids from the red helix and red loop or N-terminal helix motif (SER29-GLY53, dark blue color), yellow helix/loop, and β sheet (PRO85- GLY94, cyan color), neon green helix (CYS106, LEU109, GLY111-MET117; light green) blue loop (CYS197, TRP198, AND GLY199, orange color) are involved in the interaction between the protein and lipid bilayer (Figure 8b & Table 2). The MD simulation results fully supported the membrane-bound nature of the All4871 protein, where the N-terminal helix motif along with other domains from the αβ core helps in stabilizing the preferred orientation of All4871 in the lipid bilayer. The MD simulation data is also in coherence with the TMH prediction data.

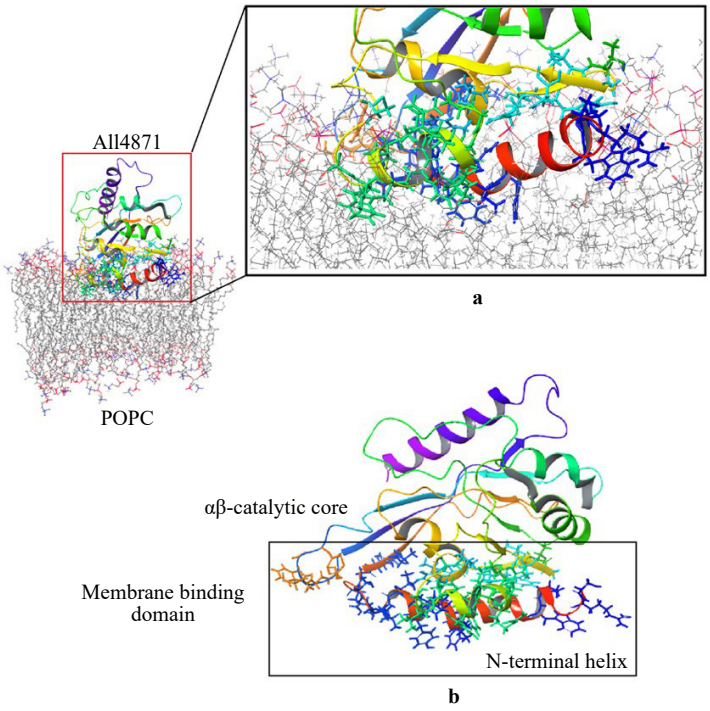


Figure 8. MD-derived complex of All4871 bound in the POPC lipid bilayer. (a) The snapshot image of the All4871-POPC complex was acquired at 80 ns following the equilibration of the MD simulation. (b) Image showing the domains and amino acids of All4871 involved in interaction with the lipid bilayer (POPC). The details of amino acids are presented in Table 2

Table 2. The MD simulation-derived domains and the amino acids of All4871 were involved in the interaction with the lipid bilayer

Interacting domain	Interacting amino acids
Red spiral and red loop (N-terminal helix)	SER29, ARG30, PRO 33, PHE41, VAL42, ASN44, PHE46, LEU48, PRO49, SER50, PHE52, GLY53,
Yellow spiral / loop and β Sheet	PRO85, TYR86, ALA87, GLY89, ARG90, TYR91, VAL92, THR93, GLY94
Neon Green	CYS106, LEU109, GLY111, TRP112, PHE113, VAL114, ARG115, ARG116, MET117
Blue loop	CYS197, TRP198 AND GLY199

4. Discussion

As the initial approach of investigation, the domain investigation, phylogenetic construction, and protein-protein interaction analysis revealed that the hypothetical protein All4871 may belong to the acyltransferase family. The homology studies of the All4871 protein showed the presence of conserved motifs (motif I-IV), which are considered the characteristic feature of the members of the acyltransferase family [3, 4]. In most of the AGPATs, motif I (HX_4D) is the most conserved region and a critical component of the acyltransferase catalytic machinery [34]. In addition to acyltransferase catalysis, motif I along with motif IV forms the binding site of acyl-CoA [4]. In motif I, the invariant histidine (His75 in All4871) with a lone pair electron act like a base to attract proton from 2-hydroxyl group present at sn-1 position of LPA, which leads to a nucleophilic attack on the thioester bond in the acyl-CoA [34]. On the other hand, conserved aspartate (Asp80 in All4871) in motif I is involved in maintaining the lone pair electrons on the Nε2 nitrogen of invariant histidine, which is required to facilitate the nucleophilic attack on the thioester substrate [34]. Mutations in either of the invariant histidine or aspartate in motif I resulted in a complete loss of acyltransferase activities [4]. In AGPATs, Motif II and III have a role in LPA binding [4]. In human AGPAT1, R149K substitution in motif II and T180S substitution in motif III resulted in an increase in the Km value for LPA; hence, reduced the affinity for LPA [4]. In AGPATs, threonine and arginine residues in motif III are essential for the acyltransferase activity [4], however, in

All4871, these crucial amino acids have been replaced by aspargine (Asn152) and isoleucine (Ile153), respectively. The arginine residue in motif III has a role in binding the 3′-phosphate of LPA, mutating which leads to complete loss of acyltransferase activity. Although, arginine in motif III is essential for acyltransferase activity; however, its replacement with isoleucine is a common feature in almost all cyanobacterial AGPATs (as observed in Figure A3). So, understanding the impact of R153I substitution in motif III of All4871 protein on LPA binding and acyltransferase activity needs a thorough investigation. Motif II and motif IV were completely replaced by other motifs in All4871. However, motif II and motif IV observed in hypothetical protein All4871 are highly conserved among the cyanobacterial homologs, suggesting their common roles in the cyanobacterial acyltransferases. In AGPATs, motif II and motif IV are involved in the binding of the phosphopantetheine arm of the acyl donor and stabilization of motif 1, respectively [34].

The topology analysis of hypothetical protein All4871 showed the presence of transmembrane helices (TMH 1-3), which pointed towards the fact that this is probably a membrane protein. All4871 has exposed all its catalytic motifs towards the inner side of the membrane which is the characteristic topology of many acyltransferases [35]. The All4871 topology revealed that motif I and IV were arranged near the inner side of the membrane, while motifs II and III were found comparatively away from the inner side of the membrane. It has been reported that the binding site of acyl-CoA (acyltransferase substrate) is located near the membrane surface [4]. The above fact supported the location of motifs I and IV near the inner side of the membrane, as these two motifs together form the predicted binding site of the acyl-CoA in All4871. Moreover, the location of motifs II and III away from the inner side of the membrane might be involved in the binding polar LPA or molecules lacking hydrophobic moieties like glycerol-3-phosphate (G3P) [4].

Furthermore, the predicted tertiary structure of All4871 is shown to have an N-terminal helix and characteristic αβ-core of AGPATs [34, 36]. In structurally conserved acyltransferases, the presence of hydrophobic and aromatic amino acids in the N-terminal helix helps in anchoring the αβ catalytic core to the membrane bilayer. Our MD simulation data also supported the involvement N-terminal helix along with other domains from the catalytic αβ core in stabilizing the orientation of All4871 in the lipid bilayer. The N-terminal helix motif of TmPlsC helps in bringing the catalytic domain into proximity of the membrane bilayer, where the kink in helix α1 allows it to enter the membrane, while the hydrophobic and basic residues interact with the surface of the membrane [34]. The structural comparison of the 3D model of All4871 with the template 5kymA (a TmPlsC protein from *Thermotoga maritima*) showed a similarity between the two structures. However, the template protein contains an extra N-terminal antiparallel α-helix which renders a comparatively tight association with the membrane. Robertson et al. [34] reported that the truncation of the N-terminal helix (α1 helix or α1 and α2 helix both) resulted in the removal of TmPlsC from the membrane. The N-terminal helix is also responsible for exposing the αβ catalytic core to the cytosolic face; therefore, truncated protein (without the N-terminal helix) was catalytically inactive [34]. Based on CASTp analysis and docking analysis, the predicted binding pocket 1 was considered as the pocket spanning the location of the active site (αβ catalytic core), LPA binding groove, and acyl chain binding groove. The hypothetical protein All4871 showed the predicted binding sites for the AGPAT substrates i.e., LPA and MCL. Interestingly, in the All4871-LPA complex, the 3′-phosphate of LPA was bound to the Arg168 residue. Therefore, it might be possible that in cyanobacterial AGPATs, Arg168 has taken over the role of conserved arginine of AGPAT motif III which has been substituted by isoleucine (R153I) in All4871. However, a thorough investigation is needed to establish the above assumption.

5. Conclusion

The overall analysis suggested All4871 protein might act as a 1-acyl-sn-glycerol-3-phosphate acyltransferase (AGPAT). Although the hypothetical protein was predicted as an AGPAT, a thorough experimental validation is needed to confirm the biological function. This study is still very helpful in providing an opening for studying and understanding cyanobacterial AGPATs.

Conflict of interest

The authors declare no conflict of interest, financial or otherwise.

Acknowledgement

MSK is thankful to the Council of Scientific and Industrial Research (File No: 09/0013(11470)/2021-EMR-I) for providing the fellowship.

Appendix

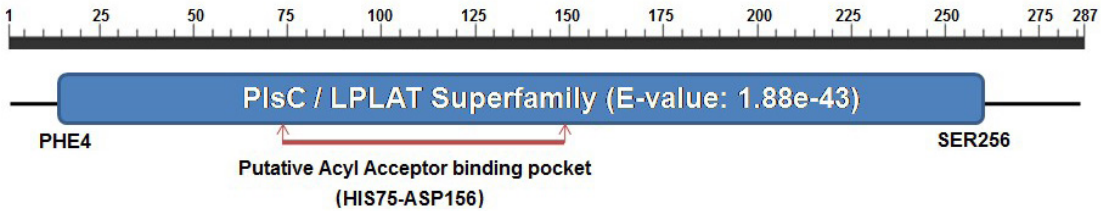


Figure A1. Schematic representation of domain organization of the hypothetical protein All4871 from Anabaena sp. PCC 7120

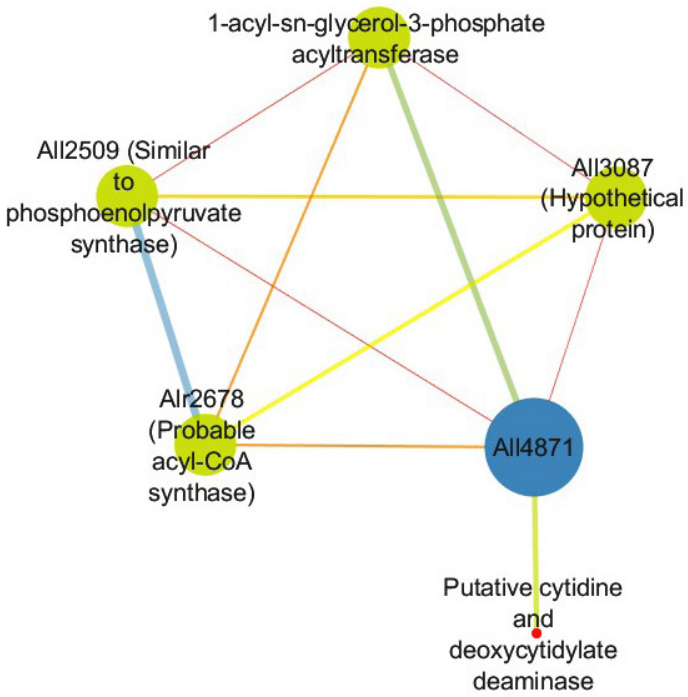


Figure A2. Protein-protein interaction network for the hypothetical protein All4871. Protein-protein interaction analysis was performed by using String version 11 (https://string-db.org/) and further analyzed by Cytoscape 3.4.1 using the betweenness centrality algorithm. The color range represents betweenness values

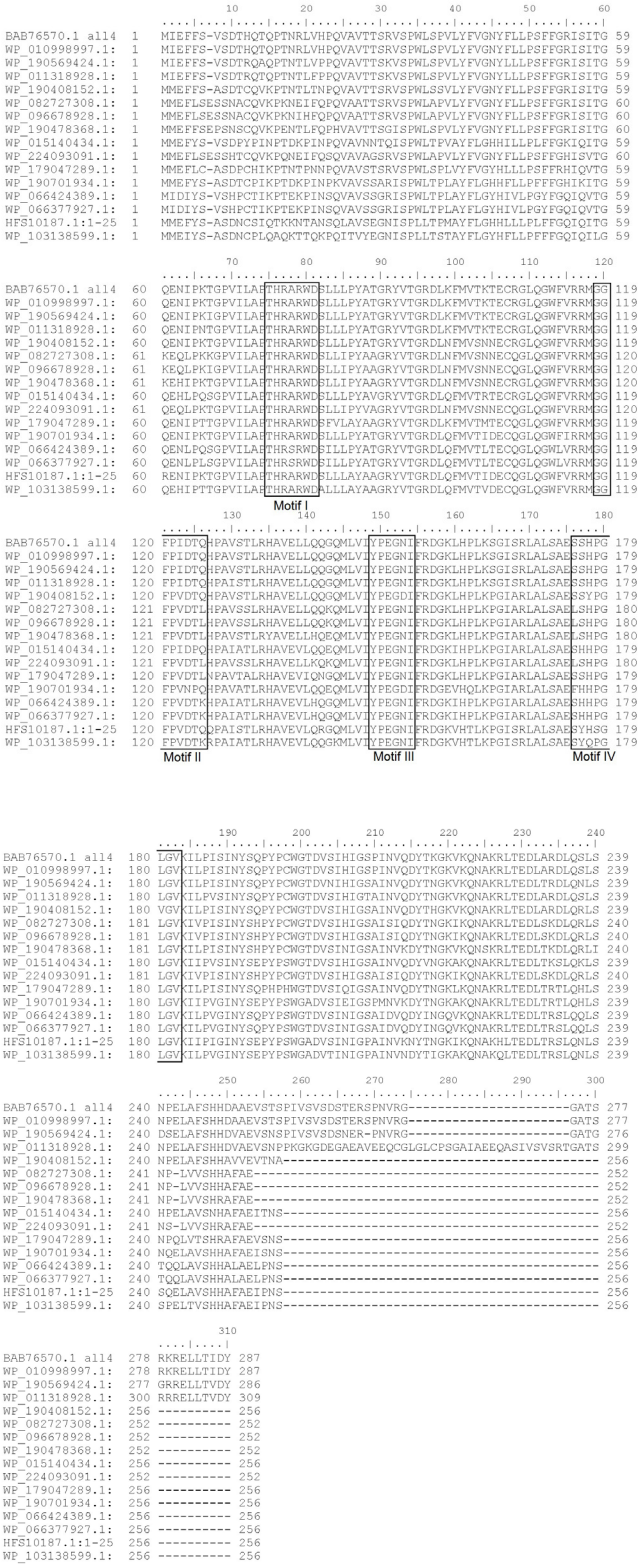


Figure A3. Multiple sequence alignment of the amino acid sequence of hypothetical protein All4871 (BAB76570.1) against the top 15 homologs from different cyanobacterial species. Motif I-IV represents the motifs crucial for the acyl transferase activity

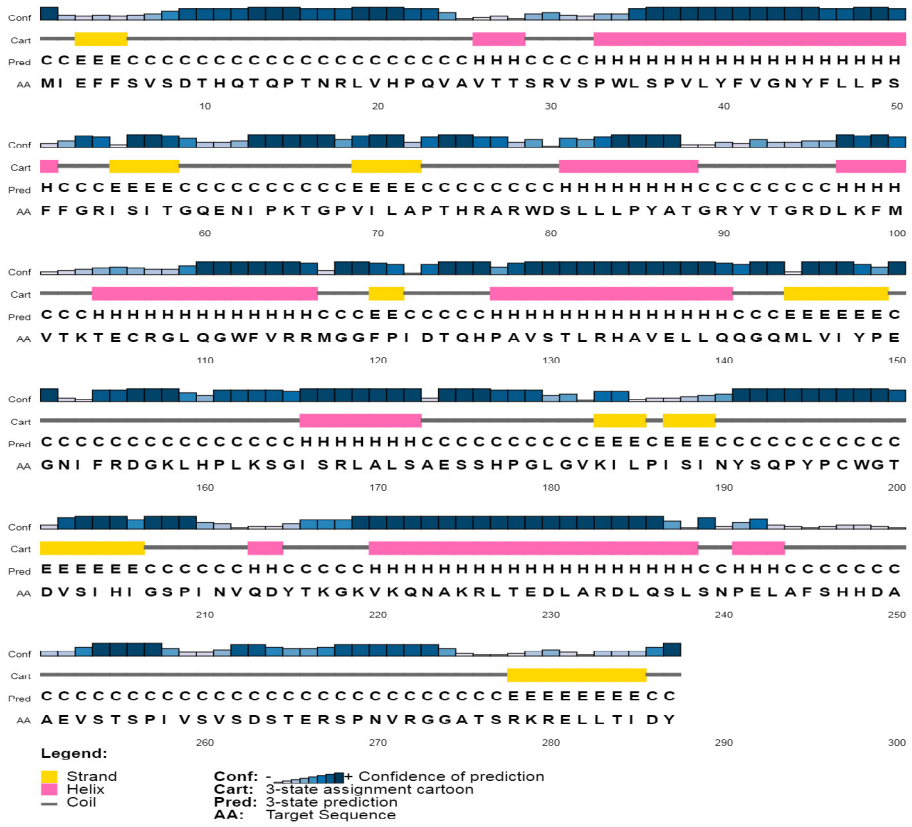


Figure A4. The secondary structure of the hypothetical protein All4871predicted by using the PSIPRED tool

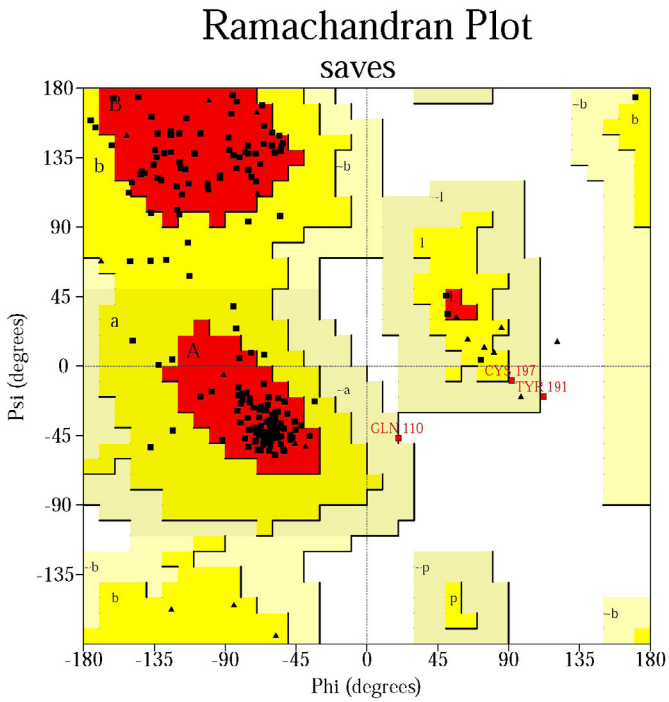


Figure A5. Ramachandran plot analysis for the hypothetical proteins All4871. The 85.6 % (149 amino acids), 12.6 % (22 amino acids), and 0.6 % (1 amino acid) falls under most favored (A, B, L), additionally allowed (a, b, l, p) and generously allowed regions (~a, ~b, ~l, ~p), respectively. Only 1.1 % (2 amino acids) was in the disallowed region

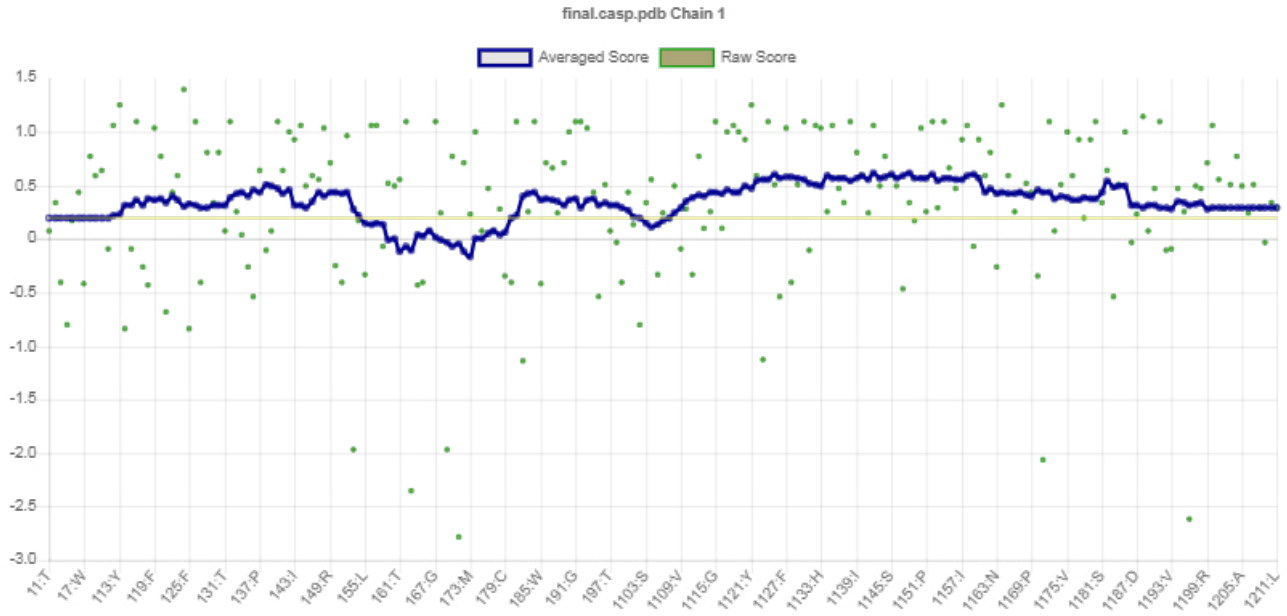


Figure A6. Verify 3D output for the hypothetical protein All4871. 85.31% of the amino acid residues have scored >=0.2 in this 3D/1D profile

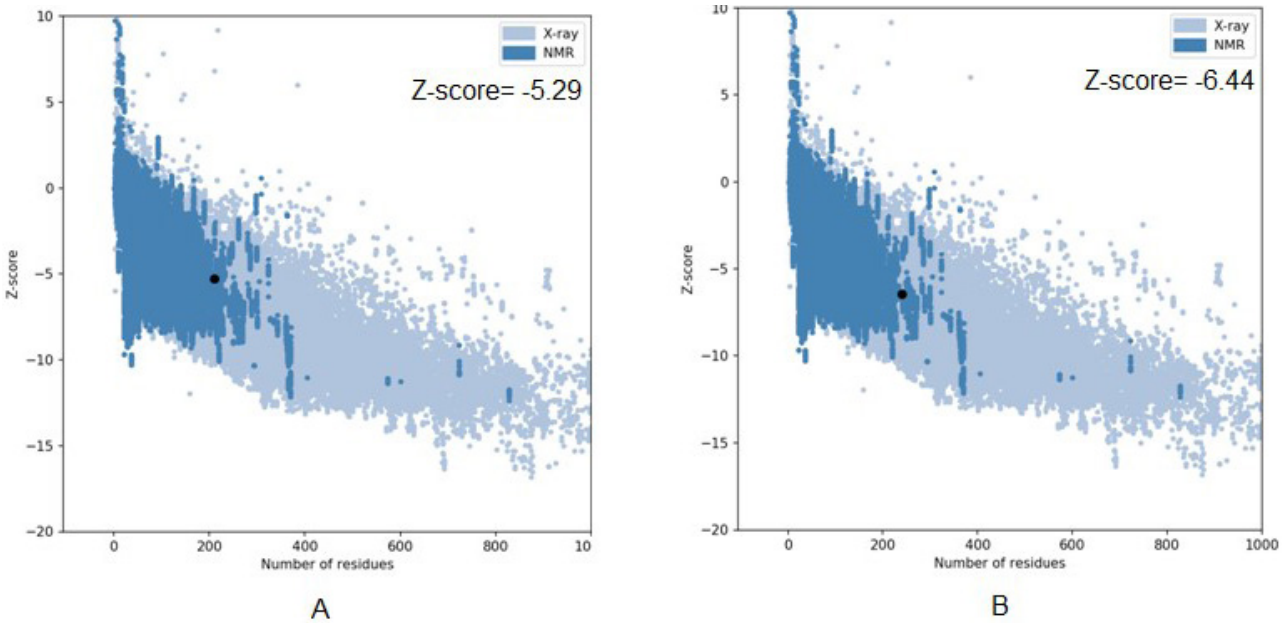


Figure A7. ProSA analysis hypothetical proteins All4871. The light and dark blue areas of the plot represent Z-scores (black dot) of the protein chains in PDB determined by X-ray crystallography and NMR spectroscopy, respectively

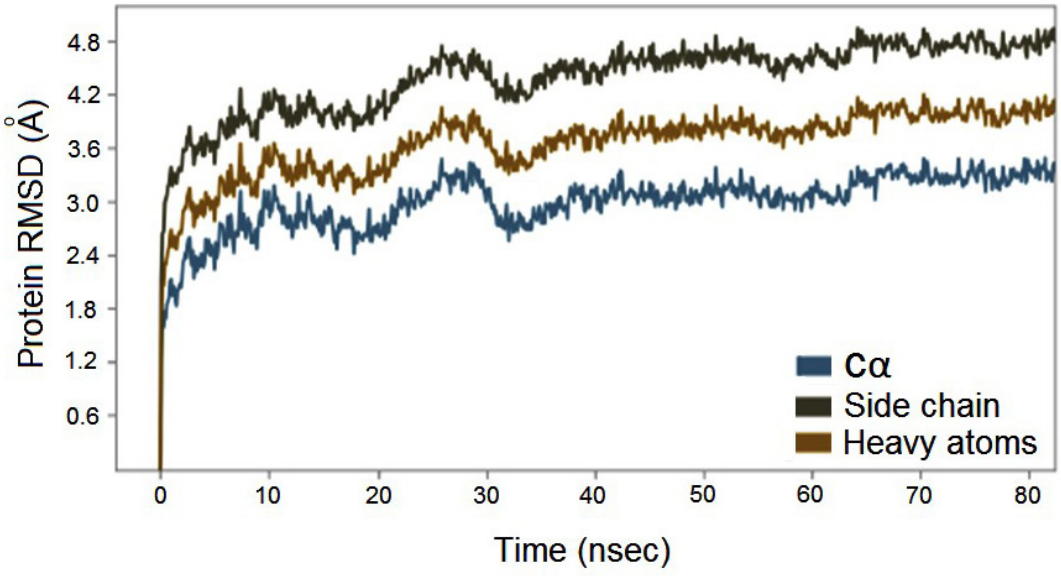


Figure A8. RMSD evolution of All4871-POPC complex

Table A1. Similar proteins were obtained from the non-redundant protein sequence database

Accession No.	Organism	Protein Name	Identity (%)	Score	E-value
WP_010998997.1	*Nostocaceae*	AGAPT	100.00	592	0.0
WP_190569424.1	*Nostoc parmelioides*	AGAPT	92.33	544	0.0
WP_011318928.1	*Trichormus variabilis*	AGAPT	83.50	518	0.0
WP_190408152.1	*Anabaena subtropica*	AGAPT	89.06	481	4e-170
WP_082727308.1	*Nostoc sp. NIES-3756*	AGAPT	80.63	431	3e-150
WP_096678928.1	*Nostoc sp. NIES-2111*	AGAPT	80.63	427	6e-149
WP_190478368.1	*Anabaena azotica*	AGAPT	81.42	424	1e-147
WP_015140434.1	*Nostoc sp. PCC 7524*	AGAPT	75.78	423	3e-147
WP_224093091.1	*Nostoc sp. MS1*	AGAPT	78.66	417	5e-145
WP_179047289.1	*Nostoc* sp. TCL26-01	AGAPT	78.12	416	3e-144
WP_190701934.1	*Anabaena* minutissima	AGAPT	74.61	414	1e-143
WP_066424389.1	*Anabaena* sp. 4-3	AGAPT	71.88	402	9e-139
WP_066377927.1	*Anabaena sp. CA = ATCC 33047*	AGAPT	71.88	400	4e-138
HFS10187.1	*Anabaena sp.*	AGAPT	71.48	389	5e-134
WP_103138599.1	*Nostoc sp. CENA543*	AGAPT	69.14	374	6e-128

AGPAT: 1-acyl-sn-glycerol-3-phosphate acyltransferase

Table A2. Similar proteins obtained from non-redundant UniProtKB/SwissProt sequence database

Accession No.	Organism	Protein Name	Identity (%)	Score	E-value
Q8DNY1.1	*Streptococcus pneumoniae*	AGAPT	26	60.5	2e-09
O07584.1	*Bacillus subtilis*	AGAPT	22.79	54.3	1e-07
Q8GXU8.1	*Arabidopsis thaliana*	AGAPT	28.67	53.9	4e-07
P32784.3	*Saccharomyces cerevisiae*	AGAPT	30.77	53.9	7e-07
Q9LLY4.2	*Brassica napus*	AGAPT	27.97	52.4	1e-06
Q9P7P0.1	*Schizosaccharomyces pombe*	AGAPT	28.00	51.6	4e-06
Q42870.1	*Limnanthes douglassi*	AGAPT	26.36	46.2	1e-04
Q42868.1	*Limnanthes alba*	AGAPT	25.58	43.1	0.001

AGPAT: 1-acyl-sn-glycerol-3-phosphate acyltransferase; GPAT: glycerol-3-phosphateO-acyltransferase; UAT: Uncharacterized acyltransferase

Table A3. Physicochemical properties of hypothetical proteins All4871

S. No.	Parameters	Values
1	Number of amino acids	287
2	Molecular weight	31891.43
3	Theoretical pI	9.46
4	Total number of atoms	4505
5	Extinction coefficients	35535
6	Instability index	53.96
7	Aliphatic index	88.26
8	GRAVY	-0.238

GRAVY: Grand average of hydropathicity

Table A4. The BUSCA output showing the predicted GO-term and protein features

Protein	Accession No.	GO-id	GO-term	Score	Feature
All4871	BAB76570.1	GO:0005886	C-Plasma membrane	0.57	Transmembrane helix

GO: Gene ontology

Table A5. The patch dock output for the docking of hypothetical proteins All4871 with LPA, G3P, and MCL

Protein	Ligand	Score	Area	(ACE)	Transformation
All4871	LPA	3392	386.9	-45.01	-0.33,-0.42,-1.04,36.44,96.55, 58.64914
	MCL	7036	805	-188.04	1.90,-0.002,-0.11,38.59,153.69, 18.85

ACE: Atomic contact energy

References

[1] Saito, M., Endo, K., Kobayashi, K., Watanabe, M., Ikeuchi, M., Murakami, A., Murata, N., & Wada, H. High myristic acid content in the cyanobacterium Cyanothece sp. PCC 8801 results from substrate specificity of lysophosphatidic acid acyltransferase. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 2018; 1863(9), 939-947. 10.1016/j.bbalip.2018.05.011

[2] Sato, N., & Wada, H. Lipid biosynthesis and its regulation in cyanobacteria. In: Wada, H., & Murata, N. (eds.) *Lipids in photosynthesis. Advances in photosynthesis and respiration*. Dordrecht: Springer; 2009, p.157-177.

[3] Weier, D., Müller, C., Gaspers, C., & Frentzen, M. Characterisation of acyltransferases from Synechocystis sp. PCC6803. *Biochemical and biophysical research communications*, 2005; 334(4), 1127-1134. 10.1016/j.bbrc.2005.06.197.

[4] Yamashita, A., Nakanishi, H., Suzuki, H., Kamata, R., Tanaka, K., Waku, K., & Sugiura, T. Topology of acyltransferase motifs and substrate specificity and accessibility in 1-acyl-sn-glycero-3-phosphate acyltransferase 1. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, 2007; 1771(9), 1202-1215. 10.1016/j.bbalip.2007.07.002.

[5] Higashi, S., & Murata, N. An in vivo study of substrate specificities of acyl-lipid desaturases and acyltransferases in lipid synthesis in Synechocystis PCC6803. *Plant physiology*, 1993; 102(4), 1275-1278. 10.1104/pp.102.4.1275.

[6] Okazaki, K., Sato, N., Tsuji, N., Tsuzuki, M., & Nishida, I. The significance of C16 fatty acids in the sn-2 positions of glycerolipids in the photosynthetic growth of Synechocystis sp. PCC6803. *Plant physiology*, 2006; 141(2), 546-556. 10.1104/pp.105.075796

[7] Lem, N. W., & Stumpf, P. K. In vitro fatty acid synthesis and complex lipid metabolism in the cyanobacterium Anabaena variabilis: I. Some characteristics of fatty acid synthesis. *Plant physiology*, 1984; 74(1), 134-138. 10.1104/pp.74.1.134.

[8] Bagnato, C., Prados, M. B., Franchini, G. R., Scaglia, N., Miranda, S. E., & Beligni, M. V. Analysis of triglyceride synthesis unveils a green algal soluble diacylglycerol acyltransferase and provides clues to potential enzymatic components of the chloroplast pathway. *BMC genomics*, 2017; 18(1), 1-23.

[9] Sathyanarayanan, N., & Nagendra, H. G. Genome wide survey and molecular modeling of hypothetical proteins containing 2Fe-2S and FMN binding domains suggests Rieske Dioxygenase Activity highlighting their potential roles in bioremediation. *Bioinformation*, 2014; 10(2), 68. 10.6026/97320630010068.

[10] Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., & Kimura, T. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120 (supplement). *DNA research*, 2001; 8(5), 227-253. 10.1093/dnares/8.5.205.

[11] Herrero, A., Stavans, J., & Flores, E. The multicellular nature of filamentous heterocyst-forming cyanobacteria. *FEMS microbiology reviews*, 2016; 40(6), 831-854. 10.1093/femsre/fuw029.

[12] Galperin, M. Y., & Koonin, E. V. 'Conserved hypothetical'proteins: prioritization of targets for experimental study. *Nucleic acids research*, 2004; 32(18), 5452-5463. 10.1093/nar/gkh885.

[13] Geer, L. Y., Domrachev, M., Lipman, D. J., & Bryant, S. H. CDART: protein homology by domain architecture. *Genome research*, 2002; 12(10), 1619-1623. 10.1101/gr.278202.

[14] Hill, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*. 1999; 41, 95-98.

[15] Kumar, S., Tamura, K., & Nei, M. Molecular evolutionary genetics analysis for microcomputers. *Computer applications in the biosciences*, 1994; 10(2): 189-191. 10.1093/bioinformatics/10.2.189.

[16] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., & Bork, P. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, 2017(45), D362-D368. 10.1093/nar/gkw937

[17] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 2003; 13(11), 2498-2504. 10.1101/gr.1239303.

[18] Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., & Casadio, R. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic acids research*, 2018; 46(W1), W459-W466. 10.1093/nar/gky320.

[19] Yin, X., Yang, J., Xiao, F., Yang, Y., & Shen, H.-B. MemBrain: an easy-to-use online webserver for transmembrane protein structure prediction. *Nano-Micro Letters*, 2018; 10(1), 1-8. 10.1007/s40820-017-0156-2.

[20] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. Protein identification and analysis tools on the ExPASy server. In: Walker, J.M. (eds) *The proteomics protocols handbook. Springer protocols handbook*, New Jersey, US: Humana press; 2005, p. 571-607.

[21] Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, 1999; 292(2), 195-202. 10.1006/jmbi.1999.3091.
[22] Kelley, L., Mezulis, S., Yates, C., Wass, M., & Sternberg, M. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocol*, 2015; 10(6), 845-858. 10.1038/nprot.2015.053
[23] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography,* 1993; 26(2), 283-291. https://doi.org/10.1107/S0021889892009944.
[24] Lüthy, R., Bowie, J. U., & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature*, 1992; 356(6364), 83-85. 10.1038/356083a0.
[25] Wiederstein, M., & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 2007; 35(suppl_2), W407-W410. https://doi.org/10.1093/nar/gkm290.
[26] Castrignano, T., De Meo, P. D. O., Cozzetto, D., Talamo, I. G., & Tramontano, A. The PMDB protein model database. *Nucleic acids research*, 2006; 34(suppl_1), D306-D309. 10.1093/nar/gkj105.
[27] Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, 2018; 46(W1), W363-W367. https://doi.org/10.1093/nar/gky473.
[28] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 2004; 25(13), 1605-1612. https://doi.org/10.1002/jcc.20084.
[29] Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*, 2005; 33(suppl_2), W363-W367. 10.1093/nar/gki481.
[30] Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., & Wolfson, H. J. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic acids research*, 2008; 36(suppl_2), W229-W232. 10.1093/nar/gkn186.
[31] O'Boyle, N., & Morley, B. M. J. C. A. C. Vandermeersch T. Hutchison G R. Open Babel: An open chemical toolbox. *Cheminform*, 2011; 3(33), 1-14. 10.1186/1758-2946-3-33.
[32] *Schrödinger Release 2017-1*: Maestro, Schrödinger, LLC, New York, NY, 2017.
[33] Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R., Felts, A. K., Halgren, T. A., Mainz, D. T., & Maple, J. R. Integrated modeling program, applied chemical theory (IMPACT). *Journal of computational chemistry,* 2005; 26(16), 1752-1780. 10.1002/jcc.20292.
[34] Robertson, R. M., Yao, J., Gajewski, S., Kumar, G., Martin, E. W., Rock, C. O., & White, S. W. A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. *Nature structural & molecular biology*, 2017; 24(8), 666-671. https://doi.org/10.1038/nsmb.3436.
[35] Gonzalez-Baró, M. R., Granger, D. A., & Coleman, R. A. Mitochondrial glycerol phosphate acyltransferase contains two transmembrane domains with the active site in the N-terminal domain facing the cytosol. *Journal of Biological Chemistry*, 2001; 276(46), 43182-43188. 10.1074/jbc.M107885200.
[36] Albesa-Jové, D., Svetlíková, Z., Tersa, M., Sancho-Vaello, E., Carreras-González, A., Bonnet, P., Arrasate, P., Eguskiza, A., Angala, S. K., & Cifuente, J. O. Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nature communications*, 2016; 7(1), 1-12. 10.1038/ncomms10906.