






# In Silico Identification and Characterization of a Hypothetical Protein From *Rhodobacter capsulatus* Revealing S-Adenosylmethionine-Dependent Methyltransferase Activity

Bioinformatics and Biology Insights  
Volume 16: 1–16  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11779322221094236  


Spencer Mark Mondol<sup>1</sup> , Depro Das<sup>2</sup> , Durdana Mahin Priom<sup>1</sup> ,  
M Shaminur Rahman<sup>3\*</sup>, M Rafiul Islam<sup>1</sup> and Md Mizanur Rahaman<sup>1</sup> 

<sup>1</sup>Department of Microbiology, University of Dhaka, Dhaka, Bangladesh. <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh. <sup>3</sup>Department of Microbiology, Jashore University of Science and Technology, Jashore, Bangladesh.

**ABSTRACT:** *Rhodobacter capsulatus* is a purple non-sulfur bacteria widely used as a model organism to study bacterial photosynthesis. It exhibits its extensive metabolic activities and demonstrates other distinctive characteristics such as pleomorphism and nitrogen-fixing capability. It can act as a gene transfer agent (GTA). The commercial importance relies on producing polyester polyhydroxyalkanoate (PHA), extracellular nucleic acids, and commercially critical single-cell proteins. These diverse features make the organism an exciting and environmentally and industrially important one to study. This study was aimed to characterize, model, and annotate the function of a hypothetical protein (Accession no. CAA71016.1) of *R. capsulatus* through computational analysis. The *urf7* gene encodes the protein. The tertiary structure was predicted through MODELLER and energy minimization and refinement by YASARA Energy Minimization Server and GalaxyRefine tools. Analysis of sequence similarity, evolutionary relationship, and exploration of domain, family, and superfamily inferred that the protein has S-adenosylmethionine (SAM)-dependent methyltransferase activity. This was further verified by active site prediction by CASTp server and molecular docking analysis through Autodock Vina tool and PatchDock server of the predicted tertiary structure of the protein with its ligands (SAM and SAH). Normally, as a part of the gene product of photosynthetic gene cluster (PGC), the established roles of SAM-dependent methyltransferases are bacteriochlorophyll and carotenoid biosynthesis. But the STRING database unveiled its association with NADH-ubiquinone oxidoreductase (Complex I). The assembly and regulation of this Complex I is mediated by the gene products of the *nuo* operon. As a part of this operon, the *urf7* gene encodes SAM-dependent methyltransferase. As a consequence of these findings, it is reasonable to propose that the hypothetical protein of interest in this study is a SAM-dependent methyltransferase associated with bacterial NADH-ubiquinone oxidoreductase assembly. Due to conservation of Complex I from prokaryotes to eukaryotes, *R. capsulatus* can be a model organism of study to understand the common disorders which are linked to the dysfunctions of complex I.

**KEYWORDS:** Purple bacteria, methyltransferase, hypothetical protein, homology modeling, molecular docking

**RECEIVED:** January 7, 2022. **ACCEPTED:** March 25, 2022.

**TYPE:** Original Research

**FUNDING:** The author(s) received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Md Mizanur Rahaman, Department of Microbiology, University of Dhaka, Dhaka 1000, Bangladesh. Email: razu002@du.ac.bd

## Introduction

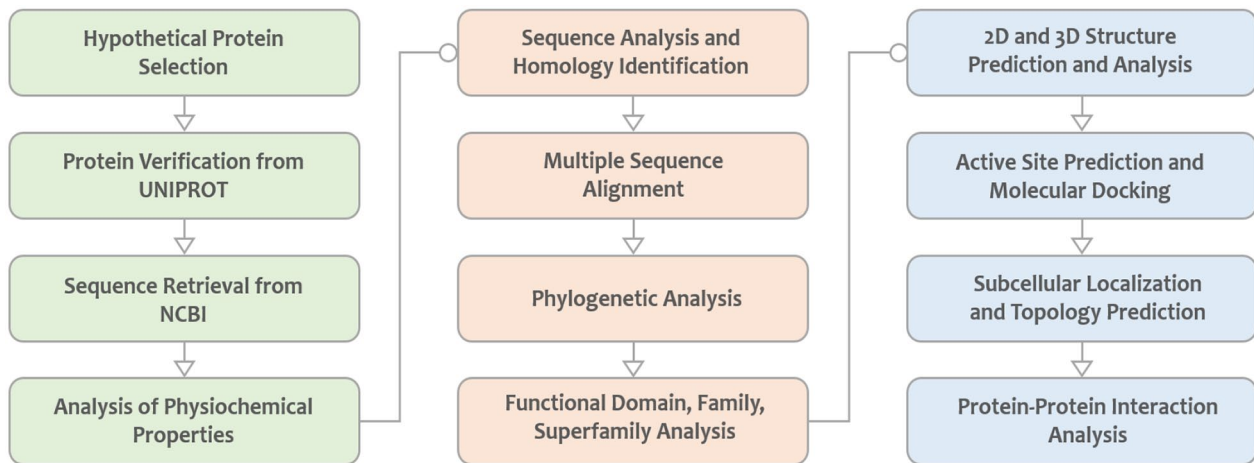
Evolution of Earth's biosphere has largely limited the primitive role of anoxygenic phototrophs, which once performed the fixation of entire global carbon, and brought about their spatial distribution.<sup>1,2</sup> The curiosities were revealed by the efforts of Erwin von Esmarch in 1887 and Hans Molisch in 1907. They first demonstrated the presence of anoxyphototrophs including *Rhodobacter capsulatus*, previously known as *Rhodospseudomonas capsulata*.<sup>3</sup> It is a gram-negative, photosynthetic, purple non-sulfur bacterium (PNSB). The individual cells are spherical, ovoid, filamentous, or rod-shaped. However, the organism exhibits comprehensive morphological properties and distinguishing features such as “zigzag” or straight chain arrangement and both flagellum-dependent and flagellum-independent motility. At present, different ecosystems around the world harbor this prokaryote, most commonly in freshwater.<sup>4-6</sup>

\*M Shaminur Rahman is now affiliated to Department of Microbiology, University of Dhaka, Dhaka, Bangladesh.

The completely sequenced genome of *R. capsulatus* contains a 3.74-Mb chromosome and a 133-kb plasmid with a median GC% of 66.6. According to the reported data, 84.1% of the open reading frames (ORFs) within the genome encode proteins which have defined functional roles, whereas 16.6% ORFs putatively code for hypothetical proteins (HPs).<sup>7</sup> By definition an HP is a predicted product expressed from an ORF whose translation has not been shown and functional relevance yet remains uncharacterized.<sup>8</sup> Even though X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the most authenticated methods to resolve the structures of biological macromolecules, attempts have been made for direct characterization from sequence information due to rapidly growing laboratory datasets and accessible computational methods. Nowadays, plenty of bioinformatic tools are available in the public domain, which have made it possible to elucidate the structural details and functional roles of HPs.<sup>9,10</sup> In this study, an effort has been made to characterize a hypothetical protein (CAA71016.1) from *R. capsulatus*, propose a 3-dimensional



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



**Figure 1.** Flowchart of methodology. NCBI indicates National Center for Biotechnology Information.

(3D) structure, and annotate its functional role as S-adenosylmethionine (AdoMet or SAM)-dependent methyltransferase (MTase) through in silico proteomics approaches.

Class I MTase is a major structural family of methyl transferring enzymes which use SAM as a cofactor and act on diverse substrates, particularly free amino acids, proteins, nucleic acids, and small bioorganic compounds.<sup>11</sup> In common with many other organisms, *R. capsulatus* has harnessed this enzymatic principle in different biochemical pathways. Notable examples include the *bcbM* gene product S-Adenosyl-L-methionine: Mg-protoporphyrin IX O-methyltransferase (MPMT), *crfF* gene product hydroxyneurosporene-O-methyltransferase, and numerous cobalamin methyltransferases which respectively catalyze different steps in bacteriochlorophyll, carotenoid, Vitamin B<sub>12</sub> (cobalamin), and siroheme biosynthesis.<sup>12–14</sup> These molecules execute different physicochemical roles that help the organism to sustain on different environmental conditions and show versatile metabolic behavior. Regarding this notion, *R. capsulatus* is competent of phototrophic anaerobic respiration, chemotrophic aerobic photosynthesis, fermentative growth, and nitrogen fixation.<sup>6,15</sup> Knowledge of these intrinsic microbial properties has led to innovation in biomonitoring and bioremediation for wastewater treatment,<sup>16</sup> developing photo bioelectrochemical cells (PBCs)<sup>17</sup> and biological hydrogen production system<sup>18</sup> as an alternative source of clean energy. In addition, *R. capsulatus* serves as a host for the production of biopolyester polyhydroxyalkanoate (PHA), extracellular nucleic acids (DNA and RNA),<sup>19</sup> cycloartenol, lupeol,<sup>20</sup> and commercially important single-cell proteins.<sup>21</sup> Despite these overwhelming significance, a large amount of HPs of *R. capsulatus* remain uncharacterized.

As previously demonstrated, bioinformatic analysis can be a feasible approach to build de novo protein models, predict new functions as well as biochemical properties, and enrich the proteome. It reduces time and labor for an indispensable wet laboratory analysis.<sup>22</sup> Considering the environmental and socioeconomic landscapes of *R. capsulatus*, in silico characterization of HPs can guide us to profoundly understand its behavior and develop new strategies for its

application, which may unlock a gateway for a sustainable future.

## Materials and Methods

The workflow of this study is presented in Figure 1.

### Sequence retrieval

Hypothetical proteins (HPs) of *R. capsulatus* were searched in the NCBI Protein database (<https://www.ncbi.nlm.nih.gov/protein/>) using the keyword “Hypothetical proteins (*Rhodobacter capsulatus*).” From the resultant hits, a HP (Accession no. CAA71016.1, GI|2182083|) was randomly selected for the study and its sequence was retrieved in FASTA format for further analysis. In addition, a sequence-based peptide search was also performed in the UniProt database (<https://www.uniprot.org/peptidesearch/>) to inspect whether the protein is redundant.<sup>23</sup>

### Analysis of physicochemical properties

The physicochemical properties of the selected HP were studied using the ProtParam tool (<https://web.expasy.org/protparam/>) on the ExPASy server. This online tool executes theoretical measurements such as molecular weight, amino acid composition, total number of positive and negative residues, theoretical pI, instability index (II), aliphatic index (AI), extinction coefficient, and grand average of hydropathicity (GRAVY) value.<sup>24</sup>

### Sequence analysis and homology identification

Looking for the structural homologs and sequence similarity in different genomics and proteomics-based databases is the most basic step for the function prediction of a hypothetical or an uncharacterized protein.<sup>25</sup> The most frequently used tool for studying sequence similarity is the Basic Local Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). In relation to the previous statement, a similarity search for proteins was performed using NCBI’s BLASTp algorithm<sup>26</sup>

against a non-redundant database to make the preliminary prediction about the function of the query protein.

#### *Functional domain and family/superfamily prediction*

HPs can be classified into families and superfamilies based on their sequence feature, domain, or motif architecture and functional similarities through automated and manual curation. For this reason, different databases use different algorithms to make a prediction from an unknown protein sequence.<sup>27</sup> Thereby, for classification and precise functional annotation, we have used multiple sequence alignment (MSA)-based servers such as Pfam,<sup>28</sup> SUPERFAMILY,<sup>29</sup> and Conserved Domain Database (CDD)<sup>30</sup>; domain profile-based Conserved Domain Architecture Retrieval Tool (CDART)<sup>31</sup>; and an integrative database InterProScan.<sup>32</sup> In each case, default parameters were considered.

#### *Multiple sequence alignment and phylogenetic analysis*

At first, several protein sequences having annotated similar functionality were retrieved from the NCBI protein database. Molecular Evolutionary Genetics Analysis X (MEGA X) software<sup>33</sup> was used to carry out the MSA and phylogenetic analysis between the targeted HP and fetched dataset. The progressive ClustalW algorithm<sup>34</sup> was applied for the MSA analysis. Furthermore, a phylogenetic tree was also constructed using the similar sequence alignment to show the evolutionary distance among the related proteins. For this purpose, we have considered the default parameters (WAG model) with 500 bootstrap replications. Statistically, the WAG model is based on maximum-likelihood (ML) methods. It incorporated the best attributes of previously proposed matrices and provided an optimal result, hence was our preferable choice.<sup>35</sup>

#### *Structure prediction*

PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) of UCL Department of Computer Science was used to predict the secondary (2D) structure of the targeted HP. It uses 2 feed-forward neural networks and PSI-BLAST algorithm for analysis.<sup>36</sup> The tertiary (3D) structure was designed using MODELLER<sup>37</sup> through the HHPred<sup>38</sup> tool of the Max Planck Institute for Developmental Biology.

#### *Structure refinement and energy minimization*

YASARA Energy Minimization Server<sup>39</sup> was used to attain a minimum energy arrangement of the constructed 3D structure of the HP. Subsequently, the minimized 3D structure was further optimized using GalaxyRefine.<sup>40</sup> After analyzing all the potential structures generated by GalaxyRefine, arguably the one having the best quality and performance was selected.

#### *Model quality assessment*

Evaluation of the energy minimized and refined 3D structure was done by PROCHECK,<sup>41</sup> ERRAT,<sup>42</sup> and Verify3D<sup>43</sup> modules of the SAVES server (<https://saves.mbi.ucla.edu/>). The ExpASy server (<https://www.expasy.org/>) of the Swiss Institute of Bioinformatics (SIB) incorporates different bioinformatic tools. Between these resources, the SWISS-MODEL Structure Assessment tool and QMEAN tool were collaboratively used to estimate the QMEAN Z-score and global quality of the model. In the QMEAN server, both QMEAN<sup>44</sup> and QMEANDisCo<sup>45</sup> scoring functions were considered. To further consolidate the global quality score, the result generated by the ModFOLD server<sup>46</sup> was taken into account.

#### *Active site prediction*

The active site of the protein was identified by Computed Atlas of Surface Topography of proteins (CASTp) (<http://sts.bioe.uic.edu/castp/index.html>). The web server interlinks protein's structural and sequence information using the Protein Data Bank (PDB), UniProt, and SIFTS database for timely residue-level annotations.<sup>47</sup> This tool also predicted the active residues which were further validated by analyzing the protein-ligand interactions of the docked complex.

#### *Subcellular localization and function prediction*

A protein's optimum performance depends on the regional environment which dictates its interaction patterns and biological networks. Therefore, predicting the subcellular localization is one of the important steps in specifying the cellular function of a hypothetical or uncharacterized protein.<sup>48</sup> Prediction of the gene ontology (GO)<sup>49</sup> and protein topology<sup>50</sup> display more extensive framework of its molecular function, biological process, and location. Tools used for these objectives were CELLO2GO,<sup>49</sup> CELLO v.2.5,<sup>51</sup> PSORTb,<sup>52</sup> PSLpred,<sup>53</sup> SOSUIGramN,<sup>54</sup> Gneg-PLoc,<sup>55</sup> BUSCA,<sup>56</sup> PRED-TMBB,<sup>57</sup> TMHMM,<sup>50</sup> and HMMTOP<sup>58</sup> tools. The ProFunc<sup>59</sup> and PredictProtein<sup>60</sup> servers were used to validate the function of the hypothetical protein predicted by the CELLO2GO tool.

#### *Docking analysis*

Molecular docking is performed to study and predict intermolecular interactions between ligands and macromolecules, using open-source software and web servers.<sup>61</sup> To further validate the probable function of our HP of interest, separate docking analyses were performed between the HP and 2 different ligand molecules, S-adenosylmethionine (SAM) and S-adenosylhomocysteine (SAH). Ligand structures were fetched from PDB (<https://www.rcsb.org/>).<sup>62</sup> Afterward, the hypothetical protein-ligand docking was performed using AutoDock Vina through PyRx<sup>63</sup> and PatchDock server.<sup>64</sup>

**Table 1.** Retrieval of the hypothetical protein from the NCBI database.

PROTEIN INDIVIDUALITIES	HYPOTHETICAL PROTEIN INFORMATION
Locus	CAA71016
Definition	Hypothetical protein [ <i>Rhodobacter capsulatus</i> ]
Accession	CAA71016
Version	CAA71016.1
GI	2182083
Amino acid	257
Gene	<i>urf7</i>
Organism	<i>Rhodobacter capsulatus</i>
Fasta sequence	>CAA71016.1 hypothetical protein [ <i>Rhodobacter capsulatus</i> ] MTTEAKKSAWKFRFEGEDVAADIRTKYGAG GDLVDIYAAANGREVHKWHHYLPIYERYFEKFRGKPVRLMIEIGTW RGGSLAMWRDYFGPEAVIFGIDINPRCKDYDGEAAQVRIGSQA DPKFLAEVIAEMGGVDIILDDGSHVMKHVRASLRMLFPQLAEGGVYMIEDMHTAY WKKFGGMDTSDNIFNFRKLIDDMHRWYHGGKRRVPLFGPMISGI HVHDSIIVLEKGPVHPPVASIRGGRTAETPAETDASVR

Abbreviation: NCBI, National Center for Biotechnology Information.

### Protein-protein interaction analysis

Protein network databases aim to integrate possible protein-protein interactions (PPIs) and present them under a network topology, from which a conclusion about shared functional features of a HP can be drawn. The STRING database evaluates both functional and physical associations. It currently features 24.6 million proteins<sup>65</sup> and aims to cover 14 000 organisms by the year 2021.<sup>65</sup> It was used in our analysis because of its larger coverage. The results obtained from STRING database were further validated by protein-protein docking analysis through HADDOCK v2.4,<sup>66</sup> HDOCK,<sup>67</sup> ClusPro 2.0,<sup>68</sup> and AutoDock Vina.<sup>63</sup> The tertiary structures of NuoF, NuoG, NuoI, NuoJ, and NuoH were obtained using SWISS-MODEL server<sup>69</sup> before docking analysis. Multiple docking tools were used to obtain high confidence about the findings.

## Results and Discussion

### Sequence retrieval

The HP (Accession no. CAA71016.1, GI|2182083|) of *R. capsulatus* fetched from the NCBI database contains 257 amino acids. The retrieved sequence was further searched in UniProt which is a comprehensive, high-quality, and freely accessible resource of protein sequence along with functional information. The database entries showcased the protein to be non-redundant which might have a significant role. Further information collected from the NCBI database is listed in Table 1.

### Physicochemical properties of the protein

Both physical and chemical properties of the HP can be estimated by analyzing the analogous properties

of individual amino acids or the N-terminal residue of the protein. From the results obtained from the ProtParam tool, the HP was found to have a molecular weight of 28 971.14 Da. The theoretical pI value of a molecule is the pH at which that particular molecule carries no net electrical charge and it is also feasible to comprehend the protein charge stability. The calculated theoretical pI value of 6.84 indicated the protein to be negatively charged and considered as an acidic protein. The II is a measurement of primary structure-dependent protein stability under in vitro conditions. It is expected that an II value less than 40 (<40) would predict the protein to be stable and a value greater than 40 (>40) would predict the protein to be unstable. The II value of the HP is computed to be 36.35, which classified the protein to be stable.<sup>24</sup> A protein's AI is known as the relative volume occupied by aliphatic side chains (alanine [Ala], valine [Val], isoleucine [Ile], and leucine [Leu]). It signifies the maintenance of a thermostable structure. The computed AI value of the HP was 75.91, which indicated that the protein is stable over a wide temperature range.<sup>70</sup> For a peptide or protein, the GRAVY score is defined as the total of the hydrophathy values divided by the number of residues in the query sequence, where all the amino acids are taken into consideration. It was computed to be -0.335. The extinction coefficient is an expression of a proportionality constant in the Beer-Lambert law. It estimates the amount of light that is absorbed by proteins at a particular wavelength.<sup>71</sup> It was calculated to be 47 900 for our query protein. The high extinction coefficient indicated the presence of a high amount of tyrosine, tryptophan, and cysteine.<sup>24</sup> Besides, all the physicochemical properties of our HP are listed in Table 2. These properties will be useful for experimental handling of the protein.

**Table 2.** Physicochemical parameters of the hypothetical protein (CAA71016.1).

PROTPARAM PARAMETERS	VALUES
Number of amino acids	257
Molecular weight	28971.14
Theoretical pI	6.84
Total number of negatively charged residues (Asp + Glu)	35
Total number of positively charged residues (Arg + Lys)	34
Atomic composition	Carbon C: 1303 Hydrogen H: 1998 Nitrogen N: 364 Oxygen O: 364 Sulfur S: 12
Formula	$C_{1303}H_{1998}N_{364}O_{364}S_{12}$
Total number of atoms	4041
Estimated half-life	30 hours (mammalian reticulocytes, in vitro) >20 hours (yeast, in vivo) >10 hours ( <i>Escherichia coli</i> , in vivo).
Instability index (II)	36.35 (Stable)
Aliphatic index	75.91
Grand average of hydropathicity (GRAVY)	-0.335
Extinction coefficients(M <sup>-1</sup> cm <sup>-1</sup> )	47900 Abs 0.1% (= 1 g/L) 1.653, assuming all pairs of Cys residues form cystines 47900 Abs 0.1% (= 1 g/L) 1.653, assuming all Cys residues are reduced

### Sequence similarity, alignment, and phylogenetic tree

The BLASTp results of the HP against non-redundant databases showed significant homology with other methyltransferase proteins, precisely with class I SAM-dependent methyltransferase from different species. The fetched methyltransferase proteins from BLASTp results for MSA are listed in Table 3. The MSA depicted the sequence similarity in between the targeted hypothetical protein and other methyltransferase proteins (Figure 2). Phylogenetic analysis was carried out for further confirmation of homology identification and to find out the evolutionary distance among our target protein and aligned methyltransferase proteins. The phylogenetic tree was constructed based on the alignment and BLASTp result, which showed similar concept about the HP (Figure 3).

### Domain, family, and superfamily prediction

The results obtained from NCBI Conserved Domain (CD) Search, CDART, Pfam, SUPERFAMILY, and InterProScan revealed that the HP sequence was found to have methyltransferase domain. The protein belongs to the Methyltransf\_24 family and the S-adenosyl-L-methionine-dependent methyltransferases superfamily. The Pfam server identified a conserved methyltransferase domain from 70 to 168 amino acid residues with an *e*-value of 9.9e-09. Furthermore, the presence of a methyltransferase domain in the targeted protein was evidently predicted from NCBI CD-search tool. It ranged from 70 to 168 amino acid residues with an *e*-value of 2.79e-12. The results obtained from the previously mentioned tools are summarized in Table 4. It indicated the inference of the HP having a methyltransferase activity.

**Table 3.** Data from BLASTp result against nonredundant protein sequences.

ACCESSION	ORGANISM	PROTEIN NAME	PERCENT IDENTITY	e-VALUE
WP_110803842.1	<i>Rhodobacter viridis</i>	Class I SAM-dependent methyltransferase	76.68	9e-142
WP_146344766.1	<i>Phaeobacter marinintestinus</i>	Class I SAM-dependent methyltransferase	63.04	2e-102
WP_113287895.1	<i>Rhodosalinus sp. E84</i>	Class I SAM-dependent methyltransferase	62.45	7e-101
WP_025045079.1	<i>Sulfitobacter geojensis</i>	Class I SAM-dependent methyltransferase	61.47	1e-99
WP_025053297.1	<i>Sulfitobacter noctilucae</i>	Class I SAM-dependent methyltransferase	61.04	5e-97
WP_057816543.1	<i>Roseovarius indicus</i>	Class I SAM-dependent methyltransferase	59.83	7e-97
WP_185797543.1	<i>Gemmobacter straminiformis</i>	Class I SAM-dependent methyltransferase	58.15	2e-95
WP_102108179.1	<i>Kandeliimicrobium roseum</i>	Class I SAM-dependent methyltransferase	61.04	4e-95
WP_162205095.1	<i>Microcystis aeruginosa</i>	Class I SAM-dependent methyltransferase	55.60	3e-92

Abbreviations: BLAST, Basic Local Alignment Search Tool; SAM, S-adenosylmethionine.

### Secondary and tertiary structure analysis

The secondary structure (2D) of the HP was predicted by PSIPRED server (Figure 4) with a good confidence of prediction. The tertiary structure (3D) was predicted by MODELLER using multiple templates having a probability greater than 99% (Figure 5). It was further energy minimized by YASARA energy minimization server. The energy calculated before energy minimization was  $-299\,229.5$  kJ/mol. After 2 rounds of energy minimization, it was changed to  $-128\,802.2$  kJ/mol. The score also improved from  $-3.37$  to  $-0.65$  after energy minimization. This indicated that the predicted 3D model became more stable after energy minimization compared to the initial one. This structure was further refined using the GalaxyRefine server and then the quality assessment of the model was carried out.

Ramachandran plot analysis (Figure 6A) results revealed that the most favored region, additional allowed region, generously allowed region, and disallowed region covered 93.8%, 5.3%, 0.0%, and 1.0% of residues, respectively. These results showed that majority of the amino acids follow a phi-psi distribution that is consistent with a right-handed  $\alpha$ -helix. Hence, the protein adopts a flexible and stable structure.<sup>72</sup> The structure passed in the validation analysis by Verify3D and the graph (Figure 6D) showed that 89.20% of the residues have 3D-1D score  $\geq 0.2$  on average. The overall quality factor predicted by the ERRAT server was 97.826, which indicated the model to be a good-quality structure as high-resolution structures produce values around 95% or higher on ERRAT. The graph (Figure 6C) generated on ERRAT showed that no residue crossed the 99% rejection limit which is also an indication of good-quality and high-resolution structure. The results obtained from the ModFOLD server showed that

the structure have a  $P$ -value of  $8.322\text{E-}4$  and a global model quality score of 0.6722. The  $P$ -value indicates the confidence of the prediction of the model to be in CERT category. It designates the structure to be valid and indicates a very high confidence of prediction. The  $P$ -value less than .001 denotes that the model has less than a 1/1000 chance of being incorrect. The QMEAN4 value predicted by the QMEAN server was  $-0.57$  and the value was transformed into a  $Z$ -score. It is depicted in the estimated absolute model quality graph (Figure 6B) where our protein model was in the dark region. It has a  $|Z\text{-score}| < 1$  which infers the model scores to be expected from an experimentally determined structure of similar size. The global score of the protein structure was calculated to be  $0.63 \pm 0.05$  which validated the global score predicted by the ModFOLD server.

### Active site detection and docking analysis

The active site of the protein predicted by the CASTp server found that 25 amino acids are involved in the potent active site. The predicted active site of the protein with their amino acid residues is depicted in Figure 7. Further docking analysis between the HP and the ligands (SAM and SAH) was carried out considering the amino acids involved in the active site predicted by CASTp server. S-adenosylmethionine is an exigent molecule and the principle biological methyl donor, found in almost all living organisms. S-adenosylmethionine-dependent methyltransferase enzymes use SAM as methyl donor.<sup>11</sup> After donating the methyl group, SAM converts into SAH which acts as a potent competitive inhibitor of methyltransferase depending on the available concentration of SAM and SAH molecules in physiological condition.<sup>73</sup> The docking analyses were carried out through Autodock vina on the

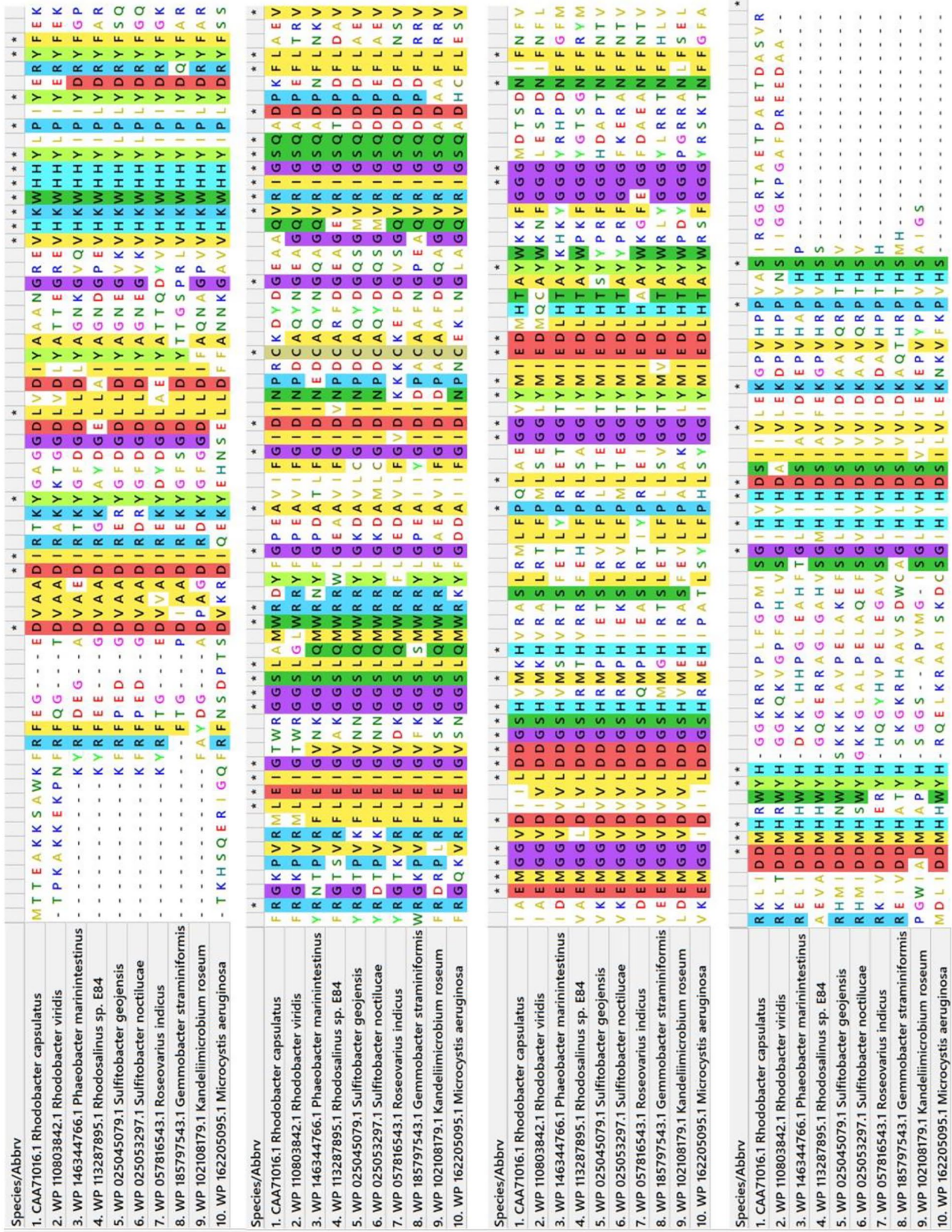
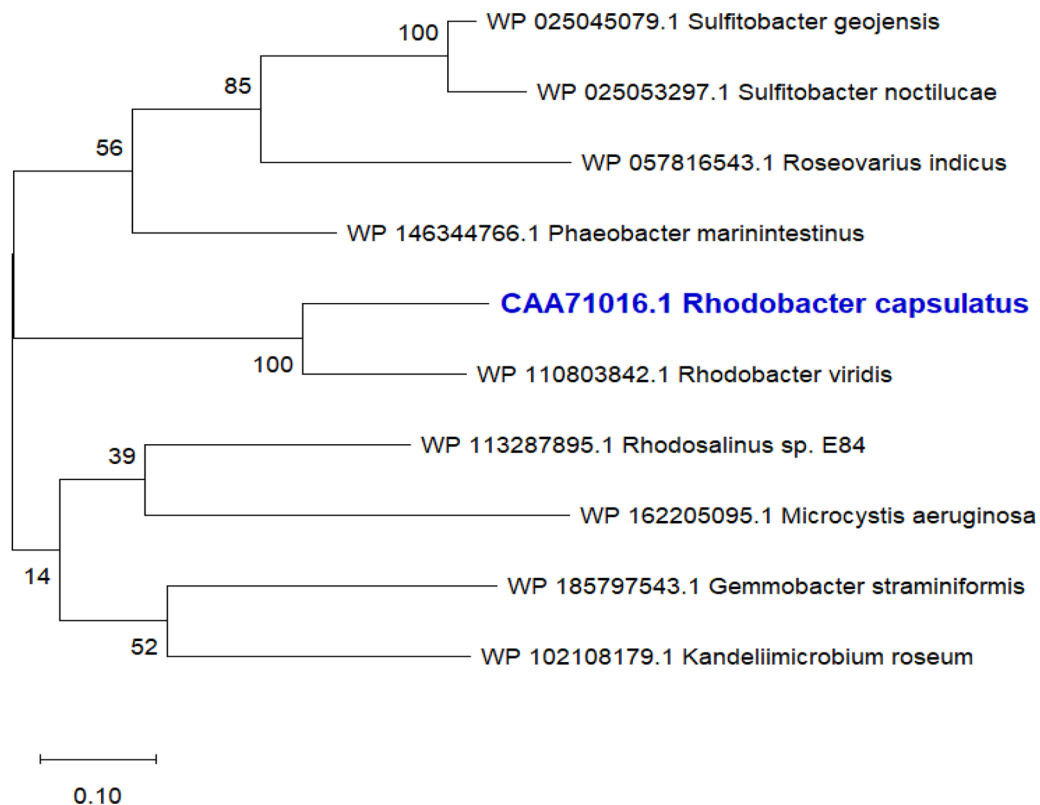


Figure 2. MSA among different methyltransferase proteins and targeted hypothetical protein using ClustalW algorithm by MEGA X software. MEGA X indicates Molecular Evolutionary Genetics Analysis X; MSA, multiple sequence alignment.



**Figure 3.** Evolutionary analysis of different methyltransferase proteins with the target protein (CAA71016.1 *Rhodobacter capsulatus*). The phylogenetic tree follows WAG replacement matrices which is based on maximum-likelihood (ML) methods. The branch lengths reflect the degree of divergence of each sequence.

**Table 4.** Protein domain, family, and superfamily analysis.

TOOLS	RESULTS
NCBI Conserved Domain Search	Domain: Methyltransferase
	Family: Methyltransf_24
	Superfamily: Class I S-adenosyl-L-methionine-dependent methyltransferases (SAM or AdoMet-MTase)
Pfam	Domain: Methyltransferase
	Family: Methyltransf_24
Superfamily	Superfamily: S-adenosyl-L-methionine-dependent methyltransferases (SAM or AdoMet-MTase)
InterProScan	Superfamily: S-adenosyl-L-methionine-dependent methyltransferases (SAM or AdoMet-MTase)
CDART (Conserved Domain Architecture Retrieval Tool)	Superfamily: S-adenosyl-L-methionine-dependent methyltransferases (SAM or AdoMet-MTase)

Abbreviations: NCBI, National Center for Biotechnology Information; SAM, S-adenosylmethionine.

PyRx server. The binding affinity (kcal/mol) of SAM and SAH with the target protein was  $-7.1$  and  $-6.7$  kcal/mol, respectively (Table 5). It indicated a strong interaction of the ligands with the target protein. The interacting residues and the interactions of the ligands with the target protein are depicted in Figure 8. The molecular docking analysis was also

carried out using the PatchDock server through interaction refinement with FireDock server. It also showed promising results (Table 5) indicating that the ligands bind efficiently with the target protein.

Another set of docking analyses was performed without marking the active site amino acids, targeting the whole





Figure 4. Secondary structure analysis by using PSIPRED server.

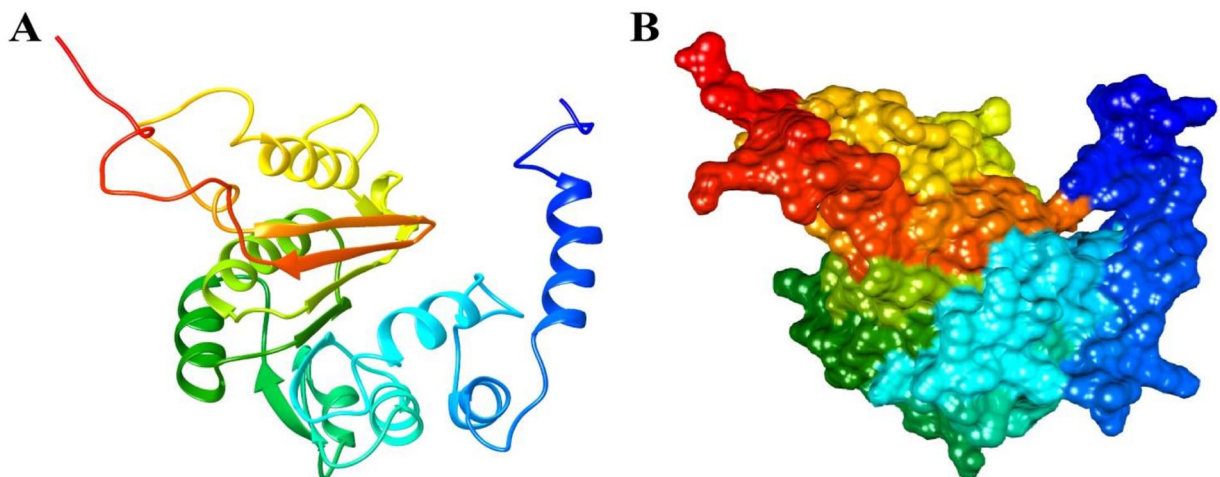
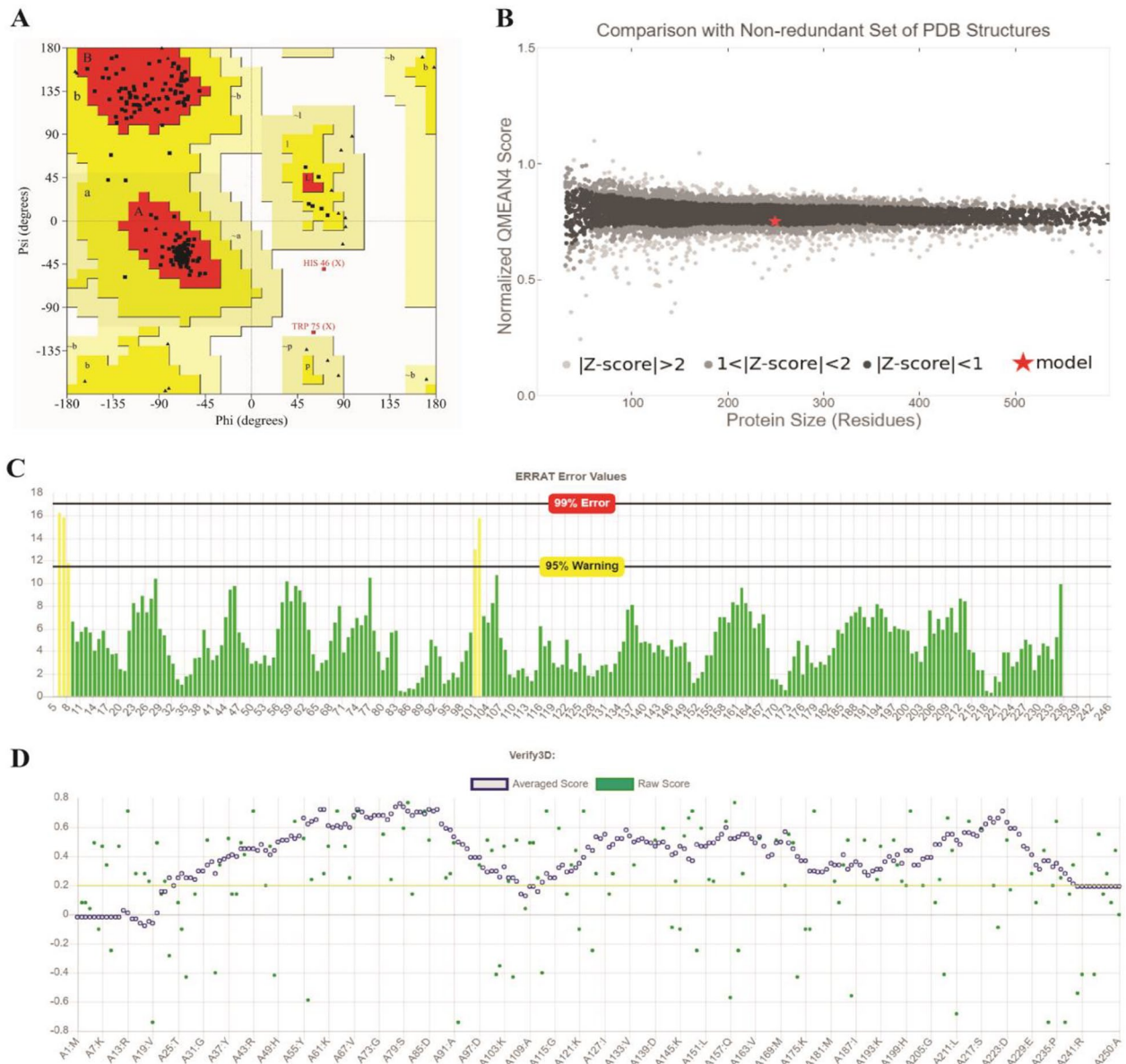
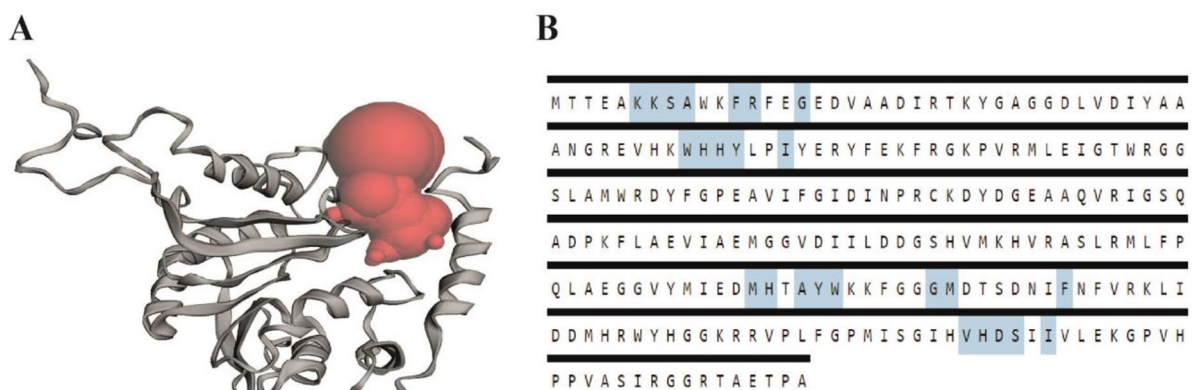


Figure 5. Illustration of predicted 3-dimensional structure of the hypothetical protein: (A) ribbon diagram and (B) surface diagram.



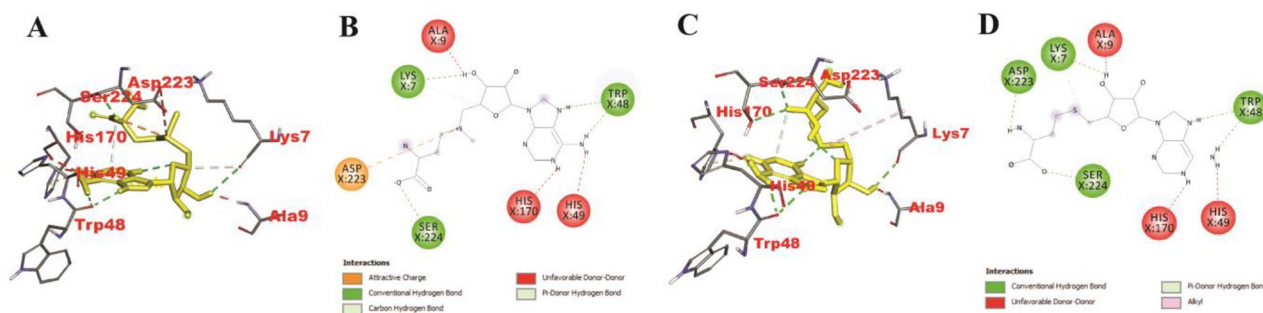
**Figure 6.** Quality assessment of the predicted tertiary structure. (A) Ramachandran plot of modeled structure validated by PROCHECK program. (B) Graphical presentation of estimation of absolute quality of model with QMEAN. (C) Graphical representation of ERRAT value estimated overall quality factor of 97.826. (D) Graphical representation of the averaged 3D-1D scores of the amino acid residues of the tertiary structure determined by VERIFY3D server. PDB indicates Protein Data Bank.



**Figure 7.** Active site of the hypothetical protein. (A) The sphere indicates the active site/pocket of the protein. (B) The marked amino acid residues construct the active site of the protein.

**Table 5.** Docking study of the ligands to the target protein.

DOCKING ANALYSIS BY AUTODOCK VINA THROUGH PYRX SERVER				
CATEGORY	LIGAND	BINDING AFFINITY (kcal/mol)	RMSD	INTERACTING RESIDUES
Selecting the active sites	S-adenosyl methionine (SAM)	-7.1	0.0	Lys7, Ala9, Trp48, His49, His170, Asp223, Ser224
	S-adenosyl homocysteine (SAH)	-6.7	0.0	Lys7, Ala9, Trp48, His49, His170, Asp223, Ser224
Without selecting the active sites(Blind dock)	S-adenosyl methionine (SAM)	-6.6	0.0	Lys8, Trp48, His49, His170, Asp223, Ser224
	S-adenosyl homocysteine (SAH)	-6.4	0.0	Lys7, Ser8, Ala9, Arg13, His49, Ser224
DOCKING ANALYSIS THROUGH PATCHDOCK-FIREDOCK SERVER				
LIGAND(S)	RANK	GLOBAL ENERGY	ATTRACTIVE VDW	REPULSIVE VDW
S-adenosyl methionine (SAM)	01	-44.22	-17.94	5.65
	02	-42.71	-15.96	1.49
S-adenosyl homocysteine (SAH)	01	-46.03	-20.44	5.07
	02	-45.74	-19.96	5.21



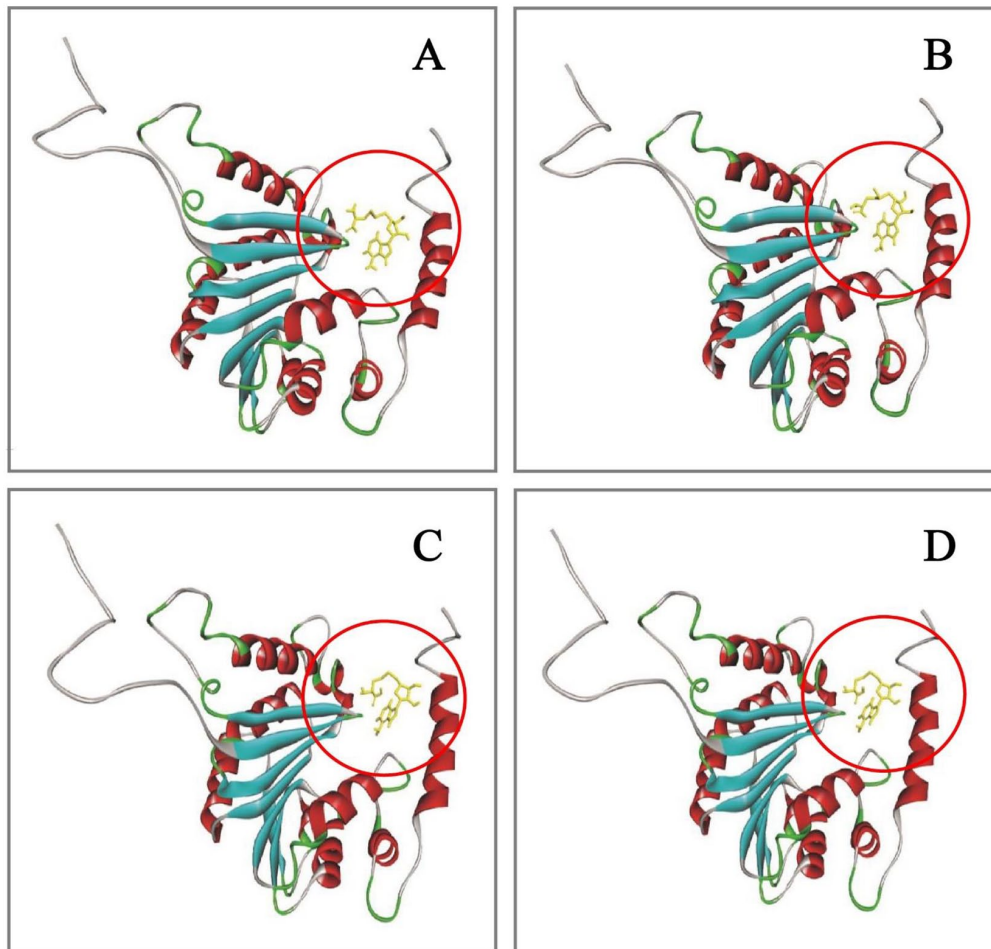
**Figure 8.** Molecular docking (targeted protein-ligand interactions). (A) 3D interaction between the targeted protein and ligand (SAM). (B) 2D interaction between the targeted protein and ligand (SAM). (C) 3D interaction between the targeted protein and ligand (SAH). (D) 2D interaction between the targeted protein and ligand (SAH). SAH indicates S-adenosylhomocysteine; SAM, S-adenosylmethionine.

protein, using the Autodock vina on the PyRx server. It helped to reinspect the active site predicted by CASTp server and find out whether the ligands actually interact within the predicted active site or some other site of the protein (Table 5). The comparative analysis of active sites through docking showed that the ligands interact firmly with the protein within the pocket inferred by CASTp server and validated the active site detection to be a preferably precise prediction. The comparative active sites of interaction are depicted in Figure 9. Overall, the results obtained from these docking analyses strongly justify

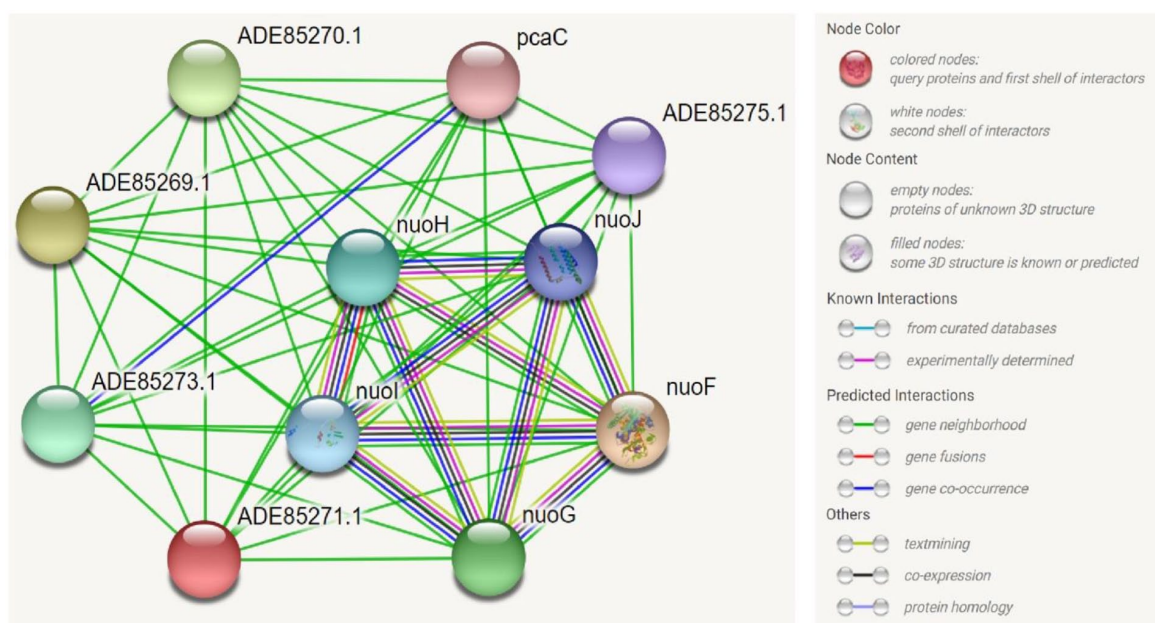
the precision of prediction of the target protein to be a SAM-dependent methyltransferase.

#### *Subcellular localization nature and functional annotation*

The subcellular localization prediction of a protein involves finding out where the protein actually resides within a cell. Subcellular localization predicted by the CELLO2GO and CELLO v2.5 server revealed that the protein is predicted to be



**Figure 9.** The comparative analysis of active site of the protein. The ligand(s) docked inside the same pocket (Circled) in all of the 4 cases indicating toward the precise active site determination by CASTp server. (A) Protein-ligand (SAM) docking analysis after marking the active residues. (B) Protein-ligand (SAM) docking analysis without marking the active residues. (C) Protein-ligand (SAH) docking analysis after marking the active residues. (D) Protein-ligand (SAH) docking analysis without marking the active residues. CASTp indicates Computed Atlas of Surface Topography of proteins; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine.



**Figure 10.** STRING network analysis of the target hypothetical protein (ADE85271.1) depicting the interactions with other proteins.

**Table 6.** Subcellular localization analysis.

S. NO.	SERVER NAME	LOCALIZATION
1	CELLO2GO	Cytoplasm
2	CELLO v.2.5	Cytoplasm
3	PSORTb v3.0.2	Cytoplasm
4	PSLpred	Cytoplasm
5	Gneg-Ploc	Cytoplasm
6	SOSUIGramN	Cytoplasm
7	BUSCA	Cytoplasm
8	PRED-TMBB	Cytoplasm

localized in the cytoplasm of the cell. The result was further validated by PSORTb, PSLpred, SOSUIGramN, Gneg-PLoc, BUSCA, and PRED-TMBB tools which also predicted the protein to be a cytoplasmic protein (Table 6). The TMHMM and HMMTOP servers predicted the absence of transmembrane helices. The absence of transmembrane helices overrules the possibility of the HP to be a transmembrane protein. Gene ontology results from CELLO2GO tool predicted the molecular function of the protein and its involvement in biological processes. The tool revealed that the major molecular function of the target protein is to impart methyltransferase activity. It also predicted that the protein is mainly involved in the biosynthetic process. Besides, the protein also has a probability to have involvement in protein complex assembly, cellular component assembly, and macromolecular complex assembly. The ProFunc and PredictProtein servers also validated the result by predicting our query protein as a methyltransferase protein.

#### Protein-protein interaction analysis

STRING is a web-based database of known and predicted PPIs that includes direct and indirect associations. Protein-protein interaction network analysis obtained from this database revealed that our HP of interest has interaction with other proteins, some having experimentally known functions and some whose functions are not yet experimentally annotated (Figure 10). Our targeted protein has a strong predicted interaction with NuoF (NADH-quinone oxidoreductase subunit F) and also has a moderate interaction with NuoH (NADH-quinone oxidoreductase subunit H), NuoI (NADH-quinone oxidoreductase subunit I), NuoG, and NuoJ. Besides, the protein has also interaction with several proteins having functions which are not yet annotated. NuoF, NuoH, NuoI, NuoG, and NuoJ are among the 14 subunits of Complex I of *R capsulatus*.<sup>74</sup> Two motifs in the NuoF subunit are likely to be involved in the binding of NADH and FMN.<sup>75</sup> NuoG subunit may ligate an extra iron-sulfur (FeS) cluster required for the assembly of Complex I.<sup>76,77</sup> NuoH subunit is one of the

most conserved subunits in Complex I. It is located in the membranous part and assists Complex I assembly.<sup>78</sup> Whereas, subunit NuoI is essential for the connection between the membranous domain and peripheral domain, in Complex I.<sup>79</sup>

As a part of the *nuo* gene cluster, *urf7* gene product encodes a SAM-dependent methyltransferase. Previously, the roles of this class of enzymes associated with bacterial mitochondrial complex I have been addressed both for prokaryotes and eukaryotes.<sup>74</sup> The results obtained from STRING database were further evaluated by protein-protein docking analysis. It revealed that NuoF has the highest binding affinity with the targeted HP (Predicted SAM-dependent Methyltransferase). The subunit NuoG showed strong binding affinity after NuoF. The other 3 subunits (NuoI, NuoJ, NuoH) showed relatively lower binding affinity than NuoF and NuoG (Table 7). The outcome of the protein-protein docking analysis aligned with the confidence score obtained from STRING database presented in Table 7.

Prior studies have noted that SAM-dependent methyltransferases are involved in regulation or subunit assembly of Complex I.<sup>74,80</sup> In some lower and higher eukaryotes, the roles of methylation-dependent regulation in mitochondrial Complex I have been suggested to be associated with conserved amino acid residues, notably with arginine. However, the roles of histidine and lysine methyltransferases also have been documented.<sup>81-83</sup> Further docking analysis by Autodock Vina through the PyRx server showed higher binding affinity of arginine than histidine and lysine with the HP (Table 7). The previously discussed protein-protein docking analysis also revealed maximally evident interaction of the HP with the arginine residues of NuoF and NuoG subunit (Figure 11). Considering all compelling evidences and significant results, it can be strongly theorized that the predicted SAM-dependent methyltransferase plays a noteworthy role in the regulation of Complex I assembly as a protein arginine methyltransferase (PRMT).

#### Conclusions

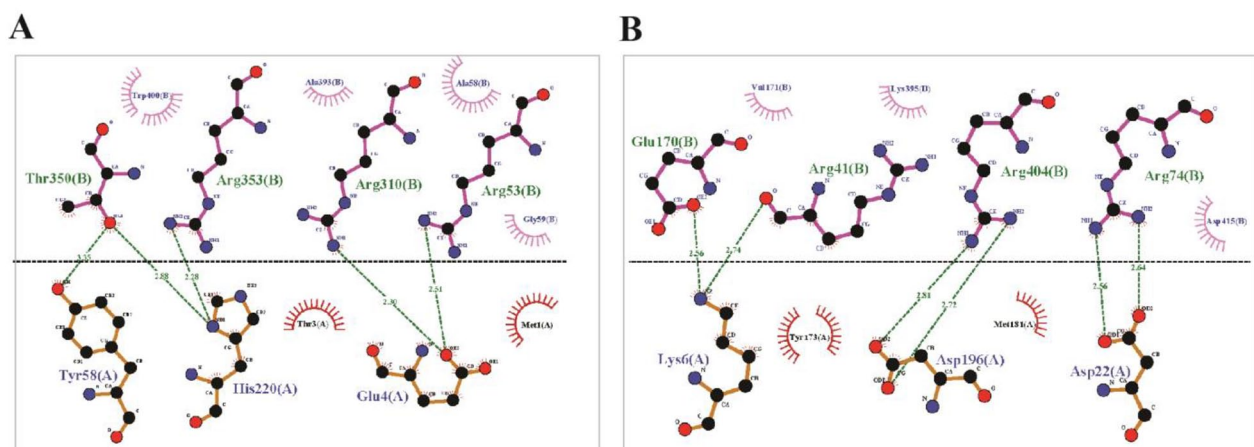
The study was designed to explore and annotate a hypothetical protein of an unknown function of *R capsulatus* through an in silico approach. Different computational tools and extensive bioinformatics workflow established its 3D structure and biological function. Our targeted hypothetical protein was predicted to be a SAM-dependent methyltransferase protein. The respective genes encoding different SAM-dependent methyltransferases are mostly responsible for catalyzing key steps in photosynthetic pigment biosynthesis. However, with the exception of this heavily studied role, the characterized protein of this study was predicted and proposed to be associated with the assembly of bacterial respiratory complex I. Throughout the process of evolution, the central subunits of complex I are conserved from prokaryotes to eukaryotes, including in humans. Deficiency in complex I is also associated with several

**Table 7.** Study of protein-protein interaction through docking analysis.

PROTEIN-PROTEIN DOCKING ANALYSIS				
BOUND PAIR	STRING INTERACTION CONFIDENCE SCORE	BINDING FREE ENERGY, $\Delta G$ (KCAL/MOL)		
		HDOCK	CLUSPRO 2.0	HADDOCK
NuoF-Methyltransferase	0.846	-14.1	-13.5	-10.3
NuoG-Methyltransferase	0.660	-12.1	-13.2	-10.8
NuoI-Methyltransferase	0.576	-10.1	-12.4	-10.0
NuoJ-Methyltransferase	0.576	-8.9	-10.3	-7.4
NuoH-Methyltransferase	0.576	-8.3	-11.1	-6.5

DOCKING ANALYSIS BY AUTODOCK VINA THROUGH PYRX				
RECEPTOR	LIGAND	BINDING AFFINITY (KCAL/MOL)		
		DOCK 1	DOCK 2	DOCK 3
Methyltransferase	Arginine	-5.3	-5.9	-5.1
	Histidine	-4.7	-5.5	-4.5
	Lysine	-4.1	-4.9	-4.0

**Figure 11.** Protein-protein interaction through docking analysis. (A) Interaction between NuoF (chain B) and targeted HP (chain A). (B) Interaction between NuoG (chain B) and targeted HP (chain A). HP indicates hypothetical protein.

human disorders. Most vigorous ones are associated with encephalomyopathy, Parkinson's disease (PD), Down syndrome, etc. Previously, *R. capsulatus* has been harnessed as a model organism to study for its commercial aspects. Due to high level of sequence conservation, establishing the structural and functional roles of unannotated protein as SAM-dependent methyltransferase can help to facilitate experimental studies and unfold new treatment strategies for critical human disorders.

### Author Contributions

SMM conceived and designed the experiment. SMM, DD, and DMP carried out the primary investigation and literature review. SMM and DD performed data validation and formal

analysis, interpreted the results, and wrote the manuscript. SMM, DD, and DMP primarily revised and edited the manuscript. MMR, MSR, and MRI reviewed the manuscript and supervised the study. All authors have read and agreed to submit the final version of the manuscript.

### ORCID iDs

Spencer Mark Mondol  <https://orcid.org/0000-0002-2240-8272>

Depro Das  <https://orcid.org/0000-0003-0969-0071>

Durdana Mahin Priom  <https://orcid.org/0000-0003-1830-4283>

Md Mizanur Rahaman  <https://orcid.org/0000-0001-6939-0500>

## REFERENCES

- Overmann J, Garcia-Pichel F. The phototrophic way of life. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, eds. *The Prokaryotes: Prokaryotic Communities and Ecophysiology*. Berlin, Germany: Springer; 2013:203-257. doi:10.1007/978-3-642-30123-0\_51.
- Cardona T. Thinking twice about the evolution of photosynthesis. *Open Biol*. 2019;9:180246. doi:10.1098/rsob.180246.
- Gest H, Blankenship RE. Time line of discoveries: anoxygenic bacterial photosynthesis. *Photosynth Res*. 2004;80:59-70. doi:10.1023/B:PRES.0000030448.24695.ec.
- Shelwell KJ, Taylor TA, Beatty JT. Photoresponsive flagellum-independent motility of the purple phototrophic bacterium *Rhodobacter capsulatus*. *J Bacteriol*. 2005;187:5040-5043. doi:10.1128/JB.187.14.5040-5043.2005.
- Shelwell KJ, Beatty JT. Coordinated, long-range, solid substrate movement of the purple photosynthetic bacterium *Rhodobacter capsulatus*. *PLoS ONE*. 2011;6:e19646. doi:10.1371/journal.pone.0019646.
- Pujalte MJ, Lucena T, Ruvira MA, Arahal DR, Macián MC. The family Rhodobacteraceae BT—the prokaryotes: Alphaproteobacteria and Betaproteobacteria. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, eds. *The Prokaryotes*. Berlin, Germany: Springer; 2014:439-512. doi:10.1007/978-3-642-30197-1\_377.
- Strnad H, Lapidus A, Paces J, et al. Complete genome sequence of the photosynthetic purple nonsulfur bacterium *Rhodobacter capsulatus* SB 1003. *J Bacteriol*. 2010;192:3545-3546. doi:10.1128/JB.00366-10.
- Ijaq J, Chandrasekharan M, Poddar R, Bethi N, Sundararajan VS. Annotation and curation of uncharacterized proteins—challenges. *Front Genet*. 2015;6:119. doi:10.3389/fgene.2015.00119.
- Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*. 2007;8:995-1005. doi:10.1038/nrm2281.
- Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. *Nat Biotechnol*. 2012;30:1072-1080. doi:10.1038/nbt.2419.
- Schubert HL, Blumenthal RM, Cheng X. 1 protein methyltransferases: their distribution among the five structural classes of AdoMet-dependent methyltransferases. In: Clarke SG, Tamanoi F, eds. *Protein Methyltransferases (The enzymes)*. Vol. 24. New York, NY: Academic Press; 2006:3-28. doi:10.1016/B1874-6047(06)80003-X.
- Takaichi S. Distribution and biosynthesis of carotenoids. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT, eds. *The Purple Phototrophic Bacteria*. Dordrecht, The Netherlands: Springer; 2009:97-117. doi:10.1007/978-1-4020-8815-5\_6.
- Warren MJ, Deery E. Vitamin B12 (cobalamin) biosynthesis in the purple bacteria. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT, eds. *The Purple Phototrophic Bacteria*. Dordrecht, The Netherlands: Springer; 2009:81-95. doi:10.1007/978-1-4020-8815-5\_5.
- Zappa S, Li K, Bauer CE. The tetrapyrrole biosynthetic pathway and its regulation in *Rhodobacter capsulatus*. *Adv Exp Med Biol*. 2010;675:229-250. doi:10.1007/978-1-4419-1528-3\_13.
- Madigan MT, Jung DO. An overview of purple bacteria: systematics, physiology, and habitats. In: Hunter CN, Daldal F, Thurnauer MC, Beatty JT, eds. *The Purple Phototrophic Bacteria*. Dordrecht, The Netherlands: Springer; 2009:1-15. doi:10.1007/978-1-4020-8815-5\_1.
- Kis M, Sipka G, Asztalos E, Rázga Z, Maróti P. Purple non-sulfur photosynthetic bacteria monitor environmental stresses. *J Photochem Photobiol B*. 2015;151:110-117. doi:10.1016/j.jphotobiol.2015.07.017.
- Grattieri M, Patterson S, Copeland J, Klunder K, Mintzer SD. Purple bacteria and 3D redox hydrogels for bioinspired photo-bioelectrocatalysis. *ChemSuschem*. 2020;13:230-237. doi:10.1002/cssc.201902116.
- Kim D-H, Kim M-S. Hydrogenases for biological hydrogen production. *Biorenew Technol*. 2011;102:8423-8431. doi:10.1016/j.biortech.2011.02.113.
- Higuchi-Takeuchi M, Numata K. Marine purple photosynthetic bacteria as sustainable microbial production hosts. *Front Bioeng Biotechnol*. 2019;7:258. doi:10.3389/fbioe.2019.00258.
- Loeschcke A, Dienst D, Wewer V, et al. The photosynthetic bacteria *Rhodobacter capsulatus* and *Synechocystis* sp. PCC 6803 as new hosts for cyclic plant triterpene biosynthesis. *PLoS ONE*. 2017;12:e0189816. doi:10.1371/journal.pone.0189816.
- Alloul A, Wuyts S, Lebeer S, Vlaeminck SE. Volatile fatty acids impacting phototrophic growth kinetics of purple bacteria: paving the way for protein production on fermented wastewater. *Water Res*. 2019;152:138-147. doi:10.1016/j.watres.2018.12.025.
- Ijaq J, Malik G, Kumar A, et al. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC Bioinformatics*. 2019;20:14. doi:10.1186/s12859-018-2554-y.
- Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2018;47:D506-D515. doi:10.1093/nar/gky1049.
- Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press; 2005:571-607. doi:10.1385/1-59259-890-0.
- Koonin EV, Galperin MY. Evolutionary concept in genetics and genomics. In: Koonin E, Galperin MY, eds. *Sequence—Evolution—Function: Computational Approaches in Comparative Genomics*. Boston, MA: Springer; 2003:25-49. doi:10.1007/978-1-4757-3783-7\_3.
- Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36:W5-W9. doi:10.1093/nar/gkn201.
- Wu CH, Huang H, Yeh L-SL, Barker WC. Protein family classification and functional annotation. *Comput Biol Chem*. 2003;27:37-47. doi:10.1016/s1476-9271(02)00098-1.
- El-Gebali S, Mistry J, Bateman A, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47:D427-D432. doi:10.1093/nar/gky995.
- Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*. 2001;313:903-919. doi:10.1006/jmbi.2001.5080.
- Marchler-Bauer A, Bo Y, Han L, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017;45:D200-D203. doi:10.1093/nar/gkw1129.
- Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. *Genome Res*. 2002;12:1619-1623. doi:10.1101/gr.278202.
- Mitchell AL, Attwood TK, Babbitt PC, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47:D351-D360. doi:10.1093/nar/gky1100.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549. doi:10.1093/molbev/msy096.
- Daugelaite J, O'Driscoll A, Sleator RD. An overview of multiple sequence alignments and cloud computing in bioinformatics. *ISRN Biomath*. 2013;2013:615630. doi:10.1155/2013/615630.
- Whelan S, Goldman N. A General empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001;18:691-699. doi:10.1093/oxfordjournals.molbev.a003851.
- Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402-W407. doi:10.1093/nar/gkz297.
- Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics*. 2006; Chapter 5:Unit-5.6. doi:10.1002/0471250953.bi0506s15.
- Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244-W248. doi:10.1093/nar/gki408.
- Krieger E, Joo K, Lee J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins*. 2009;77:114-122. doi:10.1002/prot.22570.
- Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res*. 2013;41:W384-W388. doi:10.1093/nar/gkt458.
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*. 1996;8:477-486. doi:10.1007/BF00228148.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2:1511-1519. doi:10.1002/pro.556020916.
- Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991;253:164-170. doi:10.1126/science.1853201.
- Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2010;27:343-350. doi:10.1093/bioinformatics/btq662.
- Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics*. 2019;36:1765-1771. doi:10.1093/bioinformatics/btz828.
- McGuffin LJ, Buenavista MT, Roche DB. The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res*. 2013;41:W368-W372. doi:10.1093/nar/gkt294.
- Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46:W363-W367. doi:10.1093/nar/gky473.
- Dönnies P, Höglund A. Predicting protein subcellular localization: past, present, and future. *Genomics Proteomics Bioinformatics*. 2004;2:209-215. doi:10.1016/S1672-0229(04)02027-3.
- Yu C-S, Cheng C-W, Su W-C, et al. CELLO2GO: a web server for protein subcellular localization prediction with functional gene ontology annotation. *PLoS ONE*. 2014;9:e99368. doi:10.1371/journal.pone.0099368.
- Möller S, Croning MDR, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001;17:646-653. doi:10.1093/bioinformatics/17.7.646.

51. Yu C-S, Lin C-J, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 2004;13:1402-1406. doi:10.1110/ps.03479604.
52. Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;26:1608-1615. doi:10.1093/bioinformatics/btq249.
53. Bhasin M, Garg A, Raghava GPS. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics.* 2005;21:2522-2524. doi:10.1093/bioinformatics/bti309.
54. Imai K, Asakawa N, Tsuji T, et al. SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in gram-negative bacteria. *Bioinformation.* 2008;2:417-421. doi:10.6026/97320630002417.
55. Shen H-B, Chou K-C. Gneg-mPLoc: a top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. *J Theor Biol.* 2010;264:326-333. doi:10.1016/j.jtbi.2010.01.018.
56. Savojardo C, Martelli PL, Fariselli P, Profti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* 2018;46:W459-W466. doi:10.1093/nar/gky320.
57. Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ. PRED-TMBB: a web server for predicting the topology of  $\beta$ -barrel outer membrane proteins. *Nucleic Acids Res.* 2004;32:W400-W404. doi:10.1093/nar/gkh417.
58. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics.* 2001;17:849-850. doi:10.1093/bioinformatics/17.9.849.
59. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* 2005;33:W89-W93. doi:10.1093/nar/gki414.
60. Yachdav G, Kloppmann E, Kajan L, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* 2014;42:W337-W343. doi:10.1093/nar/gku366.
61. Sethi A, Joshi K, Sasikala K, Alvina M. Molecular docking in modern drug discovery: principles and recent applications. In: Gaitonde V, Karmakar P, Trivedi A, eds. *Drug Discovery and Development—New Advances.* Rijeka, Croatia: IntechOpen; 2019:27-48. doi:10.5772/intechopen.85991.
62. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res.* 2000;28:235-242. doi:10.1093/nar/28.1.235.
63. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol.* 2015;1263:243-250. doi:10.1007/978-1-4939-2269-7\_19.
64. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 2005;33:W363-W367. doi:10.1093/nar/gki481.
65. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2018;47:D607-D613. doi:10.1093/nar/gky1131.
66. Honorato RV, Koukos PI, Jiménez-García B, et al. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front Mol Biosci.* 2021;8:729513. doi:10.3389/fmolb.2021.729513.
67. Yan Y, Tao H, He J, Huang S-Y. The HDock server for integrated protein-protein docking. *Nat Protoc.* 2020;15:1829-1852. doi:10.1038/s41596-020-0312-x.
68. Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. Performance and its limits in rigid body protein-protein docking. *Structure.* 2020;28:1071-1081.e3. doi:10.1016/j.str.2020.06.006.
69. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46:W296-W303. doi:10.1093/nar/gky427.
70. Enany S. Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. *J Infect Public Health.* 2014;7:296-307. doi:10.1016/j.jiph.2014.02.002.
71. Herzog B, Schultheiss A, Giesinger J. On the validity of Beer-Lambert law and its significance for sunscreens. *Photochem Photobiol.* 2018;94:384-389. doi:10.1111/php.12861.
72. Gupta S. Computational sequence analysis and structure prediction of jack bean urease. *Int J Adv Res.* 2015;3:185-191.
73. Kharbada KK, Barak AJ. 57—defects in methionine metabolism: its role in ethanol-induced liver injury. In: Preedy VR, Watson RR, eds. *Comprehensive Handbook of Alcohol Related Pathology.* Oxford, England: Academic Press; 2005:735-747. doi:10.1016/B978-012564370-2/50059-3.
74. Meijer W, Tabita F. Complex I and its involvement in redox homeostasis and carbon and nitrogen metabolism in *Rhodobacter capsulatus*. *J Bacteriol.* 2002;183:7285-7294. doi:10.1128/JB.183.24.7285-7294.2001.
75. Fearnley IM, Walker JE. Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. *Biochim Biophys Acta Bioenerg.* 1992;1140:105-134. doi:10.1016/0005-2728(92)90001-I.
76. Birrell JA, Morina K, Bridges HR, Friedrich T, Hirst J. Investigating the function of [2Fe-2S] cluster N1a, the off-pathway cluster in complex I, by manipulating its reduction potential. *Biochem J.* 2013;456:139-146. doi:10.1042/BJ20130606.
77. Herter SM, Schiltz E, Drews G. Protein and gene structure of the NADH-binding fragment of *Rhodobacter capsulatus* NADH:ubiquinone oxidoreductase. *Eur J Biochem.* 1997;246:800-808. doi:10.1111/j.1432-1033.1997.t01-1-00800.x.
78. Roth R, Hägerhäll C. Transmembrane orientation and topology of the NADH:quinone oxidoreductase putative quinone binding subunit NuoH. *Biochim Biophys Acta Bioenerg.* 2001;1504:352-362. doi:10.1016/S0005-2728(00)00265-6.
79. Chevallet M, Dupuis A, Lunardi J, Van Belzen R, Albracht SPJ, Issartel J-P. The NuoI subunit of the *Rhodobacter capsulatus* respiratory complex I (equivalent to the bovine TYKY subunit) is required for proper assembly of the membrane and peripheral domains of the enzyme. *Eur J Biochem.* 1997;250:451-458. doi:10.1111/j.1432-1033.1997.0451a.x.
80. Shahul Hameed UF, Sanislav O, Lay ST, et al. Proteobacterial origin of protein arginine methylation and regulation of complex I assembly by MidA. *Cell Rep.* 2018;24:1996-2004. doi:10.1016/j.celrep.2018.07.075.
81. Zurita Rendón O, Silva Neiva L, Sasarman F, Shoubridge EA. The arginine methyltransferase NDUF7 is essential for complex I assembly and early vertebrate embryogenesis. *Hum Mol Genet.* 2014;23:5159-5170. doi:10.1093/hmg/ddu239.
82. Mimaki M, Wang X, McKenzie M, Thorburn DR, Ryan MT. Understanding mitochondrial complex I assembly in health and disease. *Biochim Biophys Acta.* 2012;1817:851-862. doi:10.1016/j.bbabi.2011.08.010.
83. Carilla-Latorre S, Gallardo ME, Annesley SJ, et al. MidA is a putative methyltransferase that is required for mitochondrial complex I function. *J Cell Sci.* 2010;123:1674-1683. doi:10.1242/jcs.066076.