Canadian Journal of Microbiology

Computational Studies on Photolyase (Phr) Proteins of Cyanobacteria

Journal:	<i>Canadian Journal of Microbiology</i>
Manuscript ID	cjm-2021-0167.R1
Manuscript Type:	Article
Date Submitted by the Author:	13-Aug-2021
Complete List of Authors:	-, Rajneesh; Banaras Hindu University Faculty of Science Mondal, Soumila; Banaras Hindu University Faculty of Science Pathak, Jainendra; Pt Jawaharlal Nehru College (Affiliated to Bundelkhand University Jhansi), Department of Botany Singh, Prashant R.; Banaras Hindu University Faculty of Science Singh, Shailendra P.; Banaras Hindu University Faculty of Science Sinha, Rajeshwar; Banaras Hindu University Faculty of Science,
Keyword:	cyanobacteria, cryptochrome, cyclobutane pyrimidine dimers, photolyases, photoreactivation
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

For publication in: *Canadian Journal of Microbiology*

Computational Studies on Photolyase (Phr) Proteins of Cyanobacteria

Rajneesh, Soumila Mondal, Jainendra Pathak, Prashant R. Singh, Shailendra P. Singh, and Rajeshwar P. Sinha*

Rajneesh, Prashant R. Singh, and Rajeshwar P. Sinha* : Laboratory of Photobiology and Molecular Microbiology, Centre of Advanced Study in Botany, Institute of Science, Banaras Hindu University, Varanasi-221005, India

Soumila Mondal, and Shailendra P. Singh: Centre of Advanced Study in Botany, Department of Botany, Institute of Science, Banaras Hindu University, Varanasi 221005, India

Jainendra Pathak: Department of Botany, Pt. Jawaharlal Nehru College (Affiliated to Bundelkhand University, Jhansi), Banda-210001, India

***Corresponding author:**

Prof. Dr. Rajeshwar P. Sinha

E-mail: rpsinhabhu@gmail.com

Tel.: +91 542 2307147; **Fax:** +91 542 2366402

Abstract

Photolyases (Phrs) are enzymes that utilize blue/ultraviolet (UV-A) region of light for repairing UV-induced cyclopyrimidine dimer. We have studied Phr groups by bioinformatic analyses as well as active-site and structural modeling. The analysis of 238 amino acid sequences from 85 completely sequenced cyanobacterial genomes revealed five classes of Phrs, i.e., CPD Gr I, 6-4 Phrs/cryptochrome, Cry-DASH, Fe-S bacteria Phrs, and a group having fewer number of amino acids (276-385) in length. Distribution of Phr groups in cyanobacteria belonging to the order Synechococcales was found to be influenced by the habitats of the organisms. Class V Phrs were exclusively present in cyanobacteria. Unique motif and binding sites were reported in Group II and III. Fe-S protein binding site was only present in Group V. Active site residues and putative CPD/6-4pp binding residues are charged amino acids which were present on the surface of the proteins. Majority of hydrophilic amino acid residues were present on surface of Phrs. Sequence analysis confirmed the diverse nature of Phrs, though, sequence diversity does not affect their overall 3D structure. Protein-ligand interaction analysis identified novel CPD/6-4PP binding sites on Phrs. This structural information of Phrs can be used for the preparation of efficient Phr based formulations.

Keywords: cyanobacteria, cryptochrome, cyclobutane pyrimidine dimers, photolyases, photoreactivation, (6-4) photoproducts, ultraviolet radiation.

Introduction

Photoreactivation is one of the unique repair process found in cyanobacteria (Rajneesh et al. 2017, 2018) which incorporates photolyase (Phr) enzyme utilizing blue light for reverting back cyclobutane pyrimidine dimers (CPD) generated by ultraviolet radiation (UVR) into the normal state. Phrs are very critical enzymes for the survival of photoautotrophic organisms, including cyanobacteria. UV-induced damage was probably even more severe before the formation of ozone shield in photosynthetic organisms like cyanobacteria (Hitomi et al. 2002). UVR is a well-known genotoxic agent that interferes with correct transmission and expression of genetic information (Rajneesh et al. 2019; Terai et al. 2020). DNA is a major target for UV-B radiation (Sinha and Häder 2002). The main lesions induced by UV-B radiation are CPD and (6-4) photoproducts, causes mutagenesis, which if not repaired in organisms lead to cell death/programmed cell death (PCD) (Batista et al. 2009; Rastogi et al. 2011; Kavakli et al. 2017; Rajneesh et al. 2018; Pathak et al. 2019). Phrs are monomeric proteins which are encoded by GC rich codons (Rajneesh et al. 2017), their size ranges from 50-61 kDa depending upon the species. Phrs are present in all three domains of life and are generally accepted as homologous of the cryptochrome/Phr family (CPF) (Rajneesh et al. 2017). Contrary to this, it was believed that (6-4) Phrs are only found in eukaryotes but recently, they have also been traced in prokaryotes but with a different chromophore (Muller and Carell 2009; Kavakli et al. 2017). On the basis of substrate, Phrs are named either as CPD Phrs or as (6-4) Phrs (Sancar 2003). CPD Phrs have been reported in several organisms including Achaea, bacteria, insects, vertebrates and plants (Kato et al. 1994; Sancar 1994; Yasui et al. 1994). With the formation of the ozone shield around the Earth's atmosphere, Phrs became less important and alternative DNA repair systems were developed in the cells such as recombination, mismatch repair and excision repair (Kato et al.

1994; Hitomi et al. 2002). In eukaryotes, it has been found that the Phr genes have been subjected to duplication and subsequent functional divergence, giving rise to the cryptochrome/blue light photoreceptors (Kanai et al. 1997; Cashmore et al. 1999; Todo 1999). All cryptochromes share sequence similarity at their N-terminal PHR domains with DNA photolyases and act as photoreceptors. They are devoid of repair function but are engaged in the photoreception and signal transduction of blue light in a broad range of organisms. Cryptochromes were known to regulate circadian rhythm, development, and growth in many organisms (Wang et al. 2015). The structure of cryptochrome revealed by Brudler et al. (2003) from *Synechocystis* sp. PCC 6803 has helped in the establishment of a new class of cryptochrome, cryptochrome DASH which offered a greater insight into cryptochrome/Phr evolutionary and structure-function relationship, and an initial framework for differentiating cryptochrome and Phr functions at the structural level. Seven major classes of CPF members have been reported, which constitute three classes of CPD Phrs (class I to III, CRY-DASH proteins), (6-4) Phrs along with animal cryptochromes, plant cryptochromes and a new novel group, designated as FeS-BCPs that contains bacterial (6-4) Phrs and photoreceptors having an iron-sulfur cluster (Kavakli et al. 2017).

The outcome of studies on cyanobacterial Phrs gives notable knowledge about the catalytic mechanism, evolution of Phrs, structure, and cryptochromes. The crystal structure of *Anacystis nidulans* (*Synechococcus elongatus*) Phrs along with DNA duplex containing CPD analog was the major breakthrough in the understanding of the mode of lesion recognition and binding by Phrs (Mees et al. 2004). Since, the Phrs were uncovered many decades ago, knowledge of the basic mechanism of repair done by Phrs using light is well established (Sancar 2003). Two cofactors play pivotal role in Phr enzymatic activity. Flavin adenine dinucleotide (FAD) is the

first one, important for catalytic activity, second act as antennae for capture of light to accelerate repair process and varies in different species. On the other hand, the (6-4) Phr repair mechanism is less understood owing in part to the reaction's complexity. No agreement on the repair mechanism has been reached in the literature (Wang et al. 2015). Crystal structure of (6-4) Phr from any cyanobacteria has not been determined until now. The true cryptochrome are absent in cyanobacteria.

Phrs have immense ability of photorepair, hence, constitute an important ingredient in topical creams (Ramírez et al. 2021). In medical field, Phrs are used for reversing skin damage, and in preventing diseases such as actinic keratosis, cancer and premature photoaging (Ramírez et al. 2021). Hence, research should focus on studying new Phrs from organisms such as cyanobacteria inhabiting extreme environments, where these high irradiances of UVR plays a crucial role for the Phrs to have optimum activity in the process of photorepair. Earlier studies showed that members of the CRY/Phr superfamily were distributed in the genomes of organisms across all kingdoms but vary greatly among different taxa. Present investigation was aimed to study the distribution of different types of Phrs proteins in cyanobacteria and various unique characteristics of these cyanobacterial Phrs proteins. Comprehensive search of different types of Phrs proteins sequences existing in cyanobacteria have been performed by selecting 85 fully sequenced cyanobacterial genomes of various habitats by using different bioinformatics tool and techniques. This work is first of its kind in cyanobacteria, extending the knowledge of detailed phylogenetic and structural analyses of cyanobacterial Phrs.

Materials & Methods

Retrieval of cyanobacterial Phrs sequence from the biological database

The Phr protein sequences were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/genome/>) by using blastP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The blastP was used for sequences search on the default setting to get proper results. In this study 85 fully sequenced cyanobacterial genomes were selected (<http://www.kazusa.or.jp/cyano/>). Partially annotated, redundant, and incomplete sequences were discarded. The NCBI conserved domain server was used for finding domain present in Phrs protein (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Marchler-Bauer et al. 2010). The physicochemical parameters such as the molecular weight, isoelectric point (pI), extinction coefficient, aliphatic index, amino acid property, instability index, and grand average hydropathy (GRAVY) were calculated using the Prot-Param tool of the ExPASy proteomics server (<http://web.expasy.org/protparam/>) (Thompson et al. 1994).

Bioinformatics and phylogenetic analysis

For phylogenetic analyses, Phr protein sequences were aligned using the CLUSTAL W algorithm of MEGA7 (Kumar et al. 2016). The phylogenetic analysis was performed with IQ-TREE (Nguyen et al. 2015) through an online web server (<http://iqtree.cibiv.univie.ac.at>) (Trifinopoulos et al. 2016). The best-fit evolutionary model (WAG+F+G4) was identified with Model Finder based on Bayesian Information Criterion (BIC) (Kalyaanamoorthy et al. 2017). Boot-strap support (BS) values were estimated by means of an ultrafast bootstrap (UFBoot2) approach (Hoang et al. 2018). The WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) diagram of Phr was built with the help of MSA files (Crooks et al. 2004). Hidden Markov Model (HMM) (Eddy 1996) and the logo was built using Skyline (<http://skylign.org/>) where the letter's height represents the probability for the occurrence of a particular amino acid at a given position

in Phr sequence (Wheeler et al. 2014). Gapped columns in MSA were automatically eliminated from the Skylign logo at the time of the HMM profile building.

Theoretical modelling and structural analyses

Phrs theoretical model structures were generated for structural analysis. Homology [Swiss-Model (<https://swissmodel.expasy.org/>) (Guex and Peitsch 1997)], Modeller version 9.18 (Šali and Blundell 1993) and threading algorithm-based [RaptorX (<http://raptorx.uchicago.edu/StructurePrediction/predict/>) server (Kallberg et al. 2012)] approach was used for modeling of structures. The blast was used for the searching template in PDB by using Phr protein sequences. The crystallographic structures having a high structural identity, better resolution, and top query cover were used as a template. Theprechecks (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) (Laskowski et al. 1993, 2001) and rampage server (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell et al. 2003) was used for validation of modelled structures. Chimera (UCSF 1.11.2 software) was used for visualization and structural analysis of hydrophobic surface, electrostatic charges and residue conservation of modelled structures.

Phr and CPD, 6-4 photoproduct interaction studies

Molecular docking protocols are widely used for predicting the binding affinities for a number of ligands. In current work, in order to get accurate results, all the docking experiments were performed with the default parameters. The time to dock one ligand was approximately 1-2 min. Intermediary steps, such as pdbqt files for protein and ligands preparation and grid box creation were completed using graphical user interface program Auto Dock Suit 4.0 (ADT). ADT assigned polar hydrogen's, united atom Kollman charges, solvation parameters, and fragmental volumes to the protein. Auto Dock saved the prepared file in PDBQT format. Auto

Grid was used for the preparation of the grid map using a grid box. The grid size was set to $80 \times 80 \times 80$ XYZ points with a grid spacing of 0.375 Å. A scoring grid is calculated from the ligand structure to minimize the computation time. Auto Dock was employed for docking using protein and ligand information along with grid box properties in the configuration file. Auto Dock iterated local search global optimizer (Baxter 1981; Blum et al. 2008). During the docking procedure, protein was considered rigid and ligand was assigned a flexible torsion tree. The confirmation with the lowest energy of binding was extracted and aligned with the receptor structure for further analysis.

Results

Different types of Phrs in cyanobacteria

Different types of the cryptochrome/Phr family referred to as GR1 to GR5 were grouped on the basis of their sequence analysis (Ozturk et al. 2008; Asimgil et al. 2012). In present study, a total of 238 Phrs sequences were retrieved from NCBI database. The phylogenetic analysis showed that the tree is clearly divided into five sub-groups (Fig. 1), where Group I, II and III showed the close relationship among them but were found to be distantly related to Group IV and V (Fig. 1). They all might share a common ancestor, which gradually get separated into groups. Out of 238 Phr sequences, 79 sequences belong to group I, 31 sequences form group II, 33 sequences form group III, 43 sequences form group IV, and 52 sequences form group V. We found that Group I Phrs (CPD Phrs) were evenly distributed in all cyanobacterial genome except in *Acaryochloris marina* MBIC 11017, where CPD Phr was showing more similarity to the eukaryotic CPD Phr (Fig. 2). The distribution of the groups II to V Phrs was found to be uneven in all the studied cyanobacterial orders. The cyanobacterial strains such as *Prochlorococcus marinus*, e.g., MIT 9313, CCMP1375 and MIT 9303 and *Atelocyanobacterium thalassa* were

found to lack Phrs (Fig. 2). Group III Phrs were completely absent in the order Nostocales. A habitat-wise distribution trend was observed in order *Synechococcales* which was not seen in any other cyanobacterial order. Two copies of group III Phrs (Cry_DASH) in *Gloeobacter violaceus* PCC 7421 and *Gloeobacter kilaueensis* JS1 and two copies of Group II Phrs (putative cryptochrome or eukaryotic type 6-4 Phrs) were present in *Nostoc punctiforme* PCC 73102 (Fig. 2). Phrs also present in plasmid found in *Nostoc punctiforme* PCC 73102.

Their detailed domain analysis showed that these five Groups of Phrs shared some unique characteristics features (Fig. 3). They all possess a light-sensing/capturing domain except Group IV and catalytic domain-joined together *via* a loop, whose length varies in each Group. In Group I DNA interacting helix turn helix motif is also present apart from the main DNA interacting site. Group II has an extra V-ATPase G (from amino acid number 462-510) at their C-terminal. Group III has arginine-succinate synthase and universal stress protein UspE sites within its structure and helix turn helix in light sensing and catalytic domain. Group V has a Fe-S binding site which was not common in Phrs (Fig. 3). All five groups of Phrs have a conserved C-terminal catalytic site, where FAD cofactor interacts, and a similar conservation pattern was also seen at the N-terminal light-capturing site where hydroxy-10-(D-ribo-2,3,4,5-tetrahydroxypentyl)-5-deazaisoalloxazine binds in cyanobacteria. Group IV Phrs are shorter in length in comparison to other group of Phrs having smaller light sensing and catalytic domain.

Different Phr groups have different physiochemical properties because of variation in their amino acid composition (Fig. 4). Group I Phrs sequences were analyzed for important residue in ligand binding and activity. All Groups of Phrs have highly conserved N and C terminal residue, still, some residue is unique in each Phr group. All Phr Group members have three conserved tryptophan residue (Fig. 4) which are important in enzyme catalysis and arginine

residue play pivotal in their activity. Group I Phrs have shorter C terminal end and methionine (M⁴¹¹) at FAD binding region is crucial for their activity (Fig. 4). Group II Phrs have slightly higher number of amino acids at their C-terminal end and instead of methionine, two histidine residue (H^{308,310}) present at C end are necessary for activity. Sulphur (C⁴¹⁹, M⁴²¹) loop is present at external end of the tryptophan e⁻ transfer pathway. PKL protrusion motif (Y²⁴⁰, L²⁵⁸) which is characteristic of 6-4 Phr, was also noticed in Group II Phr (Fig. 4). In Group III Phrs some unique residue were present which were absent in other Groups shown in red boxes (Fig. 4). Six arginine residues shared are by Group III and Phr Groups I at DNA binding region of protein. Group IV Phr are smaller in length in comparison to other Group of Phrs. Group V Phrs possesses H-H bridge (Fig. 4) and four conserved cysteine residue which play role in coordination with [Fe-S] cluster, which are absent in other Phr Groups. Conserved tyrosine residue present at Y⁴³⁹ position play important role in enzyme activity.

Analysis of physicochemical parameters

The physiochemical properties analysis of these five Groups of Phrs showed that they have versatile physiological properties though they share many things in common. Group II has the highest molecular weight (50372.23-64018.06 Da) (Table 1) and maximum numbers of amino acids (428-554) (Table 1). Group IV has minimum number of amino acids (276-385) and the lowest molecular weight (31473.12-46200.21 Da) (Table 1). Protein pI was calculated using pKa values of amino acids. The pKa value of amino acids depends on its side chains. It has an important role in defining the pH-dependent characteristics of a protein. All Groups of Phrs have pI ranges from neutral to basic (Table 1) but Group I has broadest pI ranging from 5.02 (acidic) to 9.19 (basic) (Table 1). The extinction coefficient illustrates how much light a protein absorbs at a certain wavelength. Group V has the highest extinction coefficient (80330-144270) (Table 1)

and lowest in Group IV (64743-80433) (Table1). The instability index provides an estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. Instability index was lowest in Group IV (27.18-48.21) (Table 1) and highest in Group V (32.82-63.42) (Table 1). The aliphatic index of a protein is described as the relative volume occupied by the amino acids such as alanine, valine, isoleucine, and leucine. It is the measure of protein thermostability, a higher value means higher stability. Group I and III have the highest aliphatic index in comparison to II, IV and V (Table 1). The grand average of hydropathicity (GRAVY) value for a protein or a peptide is calculated by adding the hydropathy values (Kyte and Doolittle 1982) of each amino acid residue and dividing them by the number of residues in the sequence or length of the sequence. Increasing positive score indicates a greater hydrophobicity. All five Groups of Phrs have quite negative GRAVY value, showing the hydrophilic nature of Phrs (Table 1).

Structural analysis of different Phrs proteins in cyanobacteria

Further, we have also studied the structural features of all five Groups of Phrs. We selected one representative of each Group and searched in a biological database to find out whether the structures are available or not. We did not find an exact match for our representative structures. So, we generated structures based on the homology modeling approach by using templates, which are selected on the basis of sequence identity and query cover as shown in Table 2. All these structures were verified through Ramachandran plot and Procheck analysis, where more than 98% of residue falls under the allowed region (Fig. 5 and Table 3). *Nostoc punctiforme* 73102, WP_012408666 (Group I), *Nostoc punctiforme* PCC 73102, WP_012413103, (Group II) *Synechocystis* sp. PCC 6803, WP_014407097 (Group III) *Nostoc punctiforme* PCC 73102, WP_012412678 (Group IV) *Nostoc punctiforme* PCC 73102, ACC82051 (Group V).

Each group of Phrs have unique arrangement α and β chains in their 3D structure as shown in topology diagram generated by PDBsum (Fig. 6). All Phr Groups have 5 β chains except Group V where 7 β chains were observed.

All these structures possess specific fold which is common to all Phrs. All Phrs have Rossmann-like fold, this fold consists of parallel β -sheets and initial $\beta\alpha\beta$ fold is a most conserved segment and C-terminal multi-helical fold consist of two sub-helical α -subdomains. Figure 5 shows modeled structures along with Ramachandran plot analysis. Group I Phrs on average consist of 24 α -helices and 5 β -sheets, Group II Phrs have 27 α -helices and 5 β -sheets, Group III consists of 24 α -helices and 5 β -sheets, Group IV have 16 α -helices and no β -sheets, and in Group V Phrs 25 α -helices and 5 β -sheets were observed (Fig. 6). The structures were analyzed by using different parameters and differences at a FAD, nucleic acid, and HDF or DLZ sites were noted down as these are the important sites for Phrs functioning. All groups of Phrs show variable surface binding residues except FAD and HDF binding region. Phrs of Group I, IV, and V have strict conservation in their FAD site whereas Group II and III have a bit of variability. Lesser conservation was found at nucleic acid binding sites (Fig. 7). A higher degree of conservation was also seen in HDF binding site but conservation was higher within the group but among the group it was found to be lower. The Group I Phrs protein surface was dominated by negative electrostatic potential. Positive electrostatic potential was present at DNA binding sites region. FAD binding site was formed mainly by negative electrostatic potential followed by positive along with some neutral residues. A similar pattern was also found in the HDF binding pocket (Fig. 7). Positive electrostatic potential charges form a major proportion followed by the negative electrostatic potential in Group II Phrs protein. The FAD-binding pocket was formed mainly by positive electrostatic potential followed by neutral and negative electrostatic potential.

HDF binding site is mainly dominated by negative electrostatic potential followed by positive and neutral electrostatic potential (Fig. 7). Group III Phrs have both positive and negative electrostatic potential with some neutral electrostatic surface potential. Negative electrostatic potential forms the main portion of FAD-binding pocket followed by positive and neutral electrostatic potential. HDF binding pocket was found to be rich in negative electrostatic potential followed by neutral, whereas positive electrostatic potential was completely missing (Fig. 7). In Group IV, positive electrostatic potential was dominant on the surface of protein followed by neutral and positive electrostatic potential charges (Fig. 7). In Group V negative electrostatic potential was dominant on the surface of the protein with positive and neutral electrostatic potential. Positive electrostatic potential showed a prominent presence in the FAD binding region with negative electrostatic potential. DLZ binding pocket was mainly formed by negative electrostatic potential followed by some positive and neutral electrostatic potential charges (Fig. 7). The hydrophobic surface view analysis showed that all 5 Groups of Phrs proteins have mainly hydrophilic residues on their surface followed by some patchy hydrophobic distribution on protein surfaces (Fig. 7). FAD-binding pocket in Group I is dominated by mainly hydrophilic amino acid residue apart from these hydrophobic and neutral amino acids residue also present in this region. At HDF binding sites both hydrophilic and hydrophobic residues are present in equal proportion with a few neutral residues (Fig. 7). Group II FAD-binding pocket is rich in hydrophobic amino acid residues followed by some hydrophilic residues. In HDF binding region both hydrophilic and hydrophobic residues were present equally. In Group III at both FAD and HDF binding region hydrophilic nature was dominant over hydrophobic nature. A similar pattern was also seen in Group IV. In Group V at the FAD-binding region, hydrophilic

nature was dominant over hydrophobic nature but at DLZ binding region hydrophobic nature was slightly higher than hydrophilic nature (Fig. 7).

CPD and 6-4 Phrs interaction with CPD and 6-4 photoproducts

Molecular docking was performed by Auto-dock 4.0 suit between selected ligands (CPD and 6-4 pps) and various groups of Phrs proteins. Coach and metapocket were used for finding active sites. For the docking studies of the Group I Phr, Trp 287 was chosen as central atom and docked with CPD. The ligand interacted with 13 residues of active binding sites (Table 4). From these, three residues, namely Ser 242, Leu 244 and Trp 281 formed H-bonds with CPD (Table 3) having estimated free energy of binding -6.62 kcal/mol. In Group II CPD Phr, Trp 321 was selected as the central atom which was in close proximity of 12 residues in the active binding sites (Table 4). Three residues (Lys 439, Asp 466, Arg 468) were found to form H-bonds with ligand CPD, having bond length 3.044, 1.848, 2.892 respectively (Table 4), when the same protein was docked with (6-4) photoproduct by using the same central atom. Twelve residues were present in its active binding sites (Table 4). This shows 75% similar binding sites of CPD but the number of H-bonds increased from three to five. Having bond length 1.883, 2.103, 1.786, 2.871, 2.756, Group II Phrs showed better interaction with (6-4) photoproduct than CPD. In Group III Phr, central atom selected was Trp 329 for docking study with CPD. The ligand interacted with 14 residues in active binding site (Table 4; Fig. 8). Two residues were found (Asn 202 and Asn 348) to form H-bonds with a ligand having bond length 2.108, 2.727 and 1.951 respectively. In Group IV Phr, protein was docked with both CPD and (6-4) photoproduct ligands by choosing His170 as the central atom. CPD was found to be interacting with and 15 amino acid residues of helices α -2-5, α -10 in active binding groove of protein (Table 4). Three residues were found (Thr 67, Arg 59, Glu 103) which interacted with CPD and formed five H-

bonds having bond length 2.988, 3.087, 2.764, 3.077 and 2.109 respectively (Table 4; Fig. 8). The Group IV Phr protein was docked with (6-4) photoproduct by retaining the same central atom. It showed interaction with 13 amino acid residues of helices α -2-5, α -10 present in the active binding groove of Phr protein (Table 4; Fig. 8). All binding site residues were similar except Ile 83, which was not present in Phr protein docked with CPD. Three interacting residues (Table 4; Fig. 8) interacted with (6-4) photoproducts, and formed six H-bonds having bond lengths of 1.795, 2.670, 2.889, 2.166, 3.139, and 3.147 respectively. The Group V Phr docked with (6-4) photoproduct ligands by choosing Arg 139 as central atom. Four residues of active binding sites formed six H-bonds (Table 4; Fig. 8). From the above docking study, it was found that Group IV showed the minimum estimated free binding energy and maximum number of H-bonds which indicates better interaction of ligands with protein in docking study of all Groups of cyanobacterial Phr. Group V showed a slightly higher minimum estimated free binding energy but it also formed maximum of six H-bonds showing better interaction in the docking study. We found that enzymes would phylogenetically cluster together according to their function (i.e., substrate and reaction specificities).

Discussion

Cyanobacteria in their niche are subjected to high light fluctuations, UVR, and other stresses which affects their fitness and survival (Latifi et al. 2009; Rajneesh et al. 2019). In this study, the distribution of cyanobacterial Phrs was studied to explore crucial family of Phrs proteins. Several classes of CPF proteins exist and they reverse UV-induced DNA damage using blue light, act as photoreceptors in various organisms and aids in regulating the mammalian circadian clock. Earlier, Kavakli et al. (2017) studied 250 sequences of different organisms and reported seven groups of CPF members. In the present work, we studied exclusively 237 Phrs

from 85 fully sequenced cyanobacteria. It was found that five different groups of Phr sequences exist in cyanobacteria. The five major types of Phrs are CPD Group I, 6-4 Phr/cryptochrome, Cry-DASH, Fe-S bacteria Phrs, and a group having fewer numbers of amino acids (276-385) in length. Fe-S bacteria Phrs are exclusively found in bacteria and CPD Group II are found in higher plants, and in some bacteria, have been traced in cyanobacterium *Acaryochloris marina* in this study. Group II sequences are putative cryptochrome /6-4 Phrs. The Group I Phr, present in cyanobacterium *Acaryochloris marina* MBIC 11017, has a higher resemblance to eukaryotic CPD Phr may be because of its position in between cyanobacteria and a green alga. Phrs are soluble proteins, which are evident from their negative gravity score. Aliphatic index value indicates their thermal stability. Instability index values indicate that these are stable proteins.

Every group of Phrs has light harvesting domain and catalytic FAD binding domain. There are plenty of unique sequences which were also noticed in this study. The presence of different types of Phrs in cyanobacteria is one of various reasons of their higher UVR tolerance. In the cyanobacterium *Gloeobacter violaceus*, sequence resembling 6-4 Phr was observed, which was further confirmed by *in silico* generated 3-D structural study. The distribution of Group I Phr is even because of CPDs which are formed in the majority under UVR whereas other groups were unevenly distributed. Habitats may also affect the distribution of Phrs in cyanobacteria. In the order, Synechococcales, habitat wise trend was observed. *Prochlorococcus marinus*, e.g., MIT 9313, CCMP1375 and MIT 9303 were found to lack Phrs, maybe because their habitat was in the low light zone. *Prochlorococcus* which numerically dominates vast oceanic areas is one of the most abundant oxygenic phototrophs on Earth. High-light-adapted genotypes occupy the upper, well illuminated but nutrient-poor 100 m layer of the water column, whereas low-light

adapted genotypes preferentially thrive at the bottom of the 80-200 m depth at dimmer light but in a nutrient-rich environment (Dufresne et al. 2003).

The presence of isoforms of Phrs suggest gene duplication event for these cyanobacterial Phrs classes. The low bootstrap values obtained in the phylogenetic analysis of Phrs are due to the highly diverse nature of the amino acid sequences of cyanobacterial Phrs. This indicates that cyanobacterial Phrs have undergone a high rate of amino acid substitution during their course of evolution. Thus, a high substitution rate in cyanobacterial Phrs could be associated with diversity in their biological function similar to other organisms. However, the overall 3-D structure of Phr is highly conserved in all groups. All five Groups of Phrs have versatile physiological properties showing variation with other Phrs Groups. *Atelocyanobacterium thalassa* is a commonly found cyanobacterial species throughout the world's oceans (Thompson et al. 2012). It is unable to perform photosynthesis, unlike many other cyanobacteria. It forms a symbiotic relationship with photosynthetic algae and also lacked Phr. Phrs have lower level of conservation at primary structure but show high conservation in 3-D structures (Kavakli et al. 2017). *Escherichia coli* CPD Phrs lack interdomain loop but large interdomain loop was present in cyanobacterial CPD Phrs. Both *E. coli* and *Nostoc punctiforme* PCC 73102 have higher conservation at their FAD and HDF binding sites. *E. coli* Phrs have higher hydrophobic residue than *N. punctiforme* PCC 73102 Phrs and share a similar electrostatic surface potential.

All Phrs showed common characteristics such as conserved catalytic site residues, mainly residues at N-terminal (Phr binding site) and FAD binding site at C-terminal. The HMM logo diagrams suggest that the amino acid residue of the catalytic site are highly conserved in all cyanobacterial Phrs. However, variation in the FAD and HDF binding sites arises due to the substitutions of acid residues in different classes of Phrs (Fig. 4). Similarly, differences in

charges at FAD and HDF binding sites were also observed in all groups in comparison to their template structures. Not all predicted residues of Phrs showed interaction with CPD/6-4 photoproduct, and furthermore, novel residues in Phrs were found which showed interaction with CPD/6-4 photoproduct. Docking results of present study need to be taken further by developing the point mutants for identified residues to further strengthen their involvement in binding interactions in unique identified Phr sequences. Hence, it can be predicted that presence of various types of Phrs in cyanobacteria might enable them to survive and sustain even in higher intensities of harmful UVR in their natural habits. Cryptochrome/Phr has aroused great research interests, and numerous works on the photoreaction have been reported recently (Kavakli et al. 2017; Rajneesh et al. 2017; Ramírez et al. 2021).

Despite the various advantages and high activity, of cyanobacterial Phr, their application as bioactive compounds in skin products may be limited by the lack of long-term stability and sensitivity to environmental conditions (Ramírez et al. 2021). The search of extremophilic enzymes of pharmaceutical interest in extreme ecosystems has yielded several important enzymes (Lloyd et al. 2004). Extremophilic proteins endowed with high thermal stability and activity open up the opportunity to be used as catalysts in different reactions of commercial interest (Littlechild 2015; Santiago et al. 2016). Phrs from ecosystems which are highly exposed to UVR have shown enhanced DNA photorepair capacity. Therefore, the study of cyanobacteria that have endured the gradual increase in UV intensity caused by continuous depletion of the ozone layer will help to find novel and robust Phrs.

Acknowledgments

Rajneesh is grateful to the Department of Biotechnology, Govt. of India, (DBT-JRF/13/AL/143/2158), for the grant in the form of senior research fellowship. Prashant R. Singh

(09/013(0795)/2018-EMR-I) is thankful to Council of Scientific and Industrial Research, New Delhi for the financial support in the form of junior research fellowship. Soumila Mondal acknowledges Science and Engineering Research Board, Department of Science and Technology, Govt. of India (ECR/2016/000578) for the research grant. The authors declare no conflict of interest.

References

- Asimgil, H., and Kavakli, I.H. 2012. Purification and characterization of five members of photolyase/cryptochrome family from *Cyanidioschyzon merolae*. *Plant Sci.* **185**: 190-198. doi:10.1016/j.plantsci.2011.10.005.
- Batista, L.F.Z., Kaina, B., Meneghini, R., and Menck, C.F.M. 2009. How DNA lesions are turned into powerful killing structures: Insights from UV-induced apoptosis. *Mutat. Res. Rev.* **681**: 197-208. doi:10.1016/j.mrrev.2008.09.001.
- Baxter, J. 1981. Local optima avoidance in depot location. *J. Oper. Res. Soc.* **32**: 815-819. doi:10.1016/0377-2217(84)90186-3.
- Blum, C., Blesa, M.J., Roli, A., and Sampels, M. 2008. Hybrid metaheuristics: An emerging approach to optimization. *Studies in computational intelligence*. Springer-Verlag, Berlin Heidelberg.
- Brudler, R., Hitomi, K., Daiyasu, H., Toh, H., Kucho, K.I., Ishiura, M., Kanehisa, M., Roberts, V.A., Todo, T., Tainer, J.A., and Getzoff, E.D. 2003. Identification of a new cryptochrome class: structure, function, and evolution. *Mol. Cell.* **11**: 59-67. doi:10.1016/s1097-2765(03)00008-x.

- Cashmore, A. R., Jarillo, J. A., Wu, Y. J., and Liu, D. 1999. Cryptochromes: blue light receptors for plants and animals. *Science*. **284**: 760-765. doi:10.1126/science.284.5415.760.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. 2004. WebLogo: a sequence logo generator. *Genome Res.* **14**: 1188-1190. doi:10.1101/gr.849004.
- Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., Duprat, S., Galperin, M.Y., Koonin, E.V., Gall, F.Le., Makarova, K.S., Ostrowski, M., Oztas, S., Robert, C., Rogozin, I.B., Scanlan, D.J., Tandeau de Marsac, N., Weissenbach, J., Wincker, P., Wolf, Y.I., and Hess, W.R. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci.* **100**: 10020-10025. doi:10.1073/pnas.1733211100.
- Eddy, S.R. 1996. Hidden markov models. *Curr. Opin. Struct. Biol.* **6**(3): 361-365. doi:10.1016/s0959-440x(96)80056-x.
- Guex, N., and Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modelling. *Electrophoresis*. **18**(15): 2714-2723. doi:10.1002/elps.1150181505.
- Hitomi, K., Okamoto, K., Daiyasu, H., Miyashita, H., Iwai, S., Toh, H., Ishiura, M., and Todo, T. 2000. Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp. PCC6803. *Nucleic Acids Res.* **28**: 2353-2362. doi:10.1093/nar/28.12.2353.
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**: 518-522. doi:10.1093/molbev/msx281.

- Kallberg, M., Wang, H.P., Wang, S., Peng, J., Wang, Z.Y., Lu, H., and Xu, J.B. 2012. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**: 1511-1522. doi:10.1038/nprot.2012.085.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods.* **14**: 587-589. doi:10.1038/nmeth.4285.
- Kanai, S., Kikuno, R., Toh, H., Ryo, H., and Todo, T. 1997. Molecular evolution of the photolyase-blue-light photoreceptor family. *J. Mol. Evol.* **45**: 535-548. doi:10.1007/PL00006258.
- Kato Jr, T., Todo, T., Ayaki, H., Ishizaki, K., Morita, T., Mitra, S., and Ikenaga, M. 1994. Cloning of a marsupial DNA photolyase gene and the lack of related nucleotide sequences in placental mammals. *Nucleic Acids Res.* **22**: 4119-4124. doi:10.1093/nar/22.20.4119.
- Kavakli, I.H., Baris, I., Tardu, M., Gül, Ş., Öner, H., Çal, S., Bulut, S., Yarpavar, D., Berkel, Ç., Ustaoglu, P., and Aydın, C. 2017. The photolyase/cryptochrome family of proteins as DNA repair enzymes and transcriptional repressors. *Photochem. Photobiol.* **93**: 93-103. doi:10.1111/php.12669.
- Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**: 1870-187. doi:10.1093/molbev/msw0544.
- Kyte, J., and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132. doi:10.1016/0022-2836(82)90515-0.

- Laskowski, R.A. 2001. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **29**: 221-222. doi:10.1093/nar/29.1.221.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283-291. doi:10.1107/S0021889892009944.
- Latifi, A., Ruiz, M., and Zhang, C. 2009. Oxidative stress in cyanobacteria. *FEMS Microbiol. Rev.* **33**: 258-278. doi:10.1111/j.1574-6976.2008.00134.x.
- Littlechild, J.A. 2015. Enzymes from extreme environments and their industrial applications, *Front. Bioeng. Biotechnol.* **3**: 161. doi:10.3389/fbioe.2015.00161.
- Lloyd, R.C., Lloyd, M.C., Smith, M.E., Holt, K.E., Swift, J.P., Keene, P.A., Taylor, S.J., and McCague, R. 2004. Use of hydrolases for the synthesis of cyclic amino acids. *Tetrahedron* **60**: 717-728. doi:10.1016/j.tet.2003.10.116.
- Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* **50**: 437-450. doi:10.1002/prot.10286.
- Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., de Weese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., and Gwadz, M. 2010. CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**: D225-D229. doi:10.1093/nar/gkq1189.

- Mees, A., Klar, T., Gnau, P., Henneck, U., Eker, A.P.M., Carell, T., and Essen, L.O. 2004. Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science* **306**: 1789-1793. doi:10.1126/science.1101598.
- Müller, M., and Carell, T. 2009. Structural biology of DNA photolyases and cryptochromes. *Curr. Opin. Struct. Biol.* **19**: 277-285. doi:10.1016/j.sbi.2009.05.003.
- Nguyen L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32**: 268-274. doi:10.1093/molbev/msu300.
- Öztürk, N., Kao, Y.T., Selby, C.P., Kavaklı, I.H., Partch, C.L., Zhong, D., and Sancar, A. 2008. Purification and characterization of a type III photolyase from *Caulobacter crescentus*. *Biochemistry* **47**: 10255-10261. doi:10.1021/bi801085a.
- Pathak, J., Singh, P.R., Häder, D.P., and Sinha, R.P. 2019. UV-induced DNA damage and repair: A cyanobacterial perspective. *Plant Gene* **19**: 100194. doi:10.1016/j.plgene.2019.100194.
- Rajneesh, Chatterjee, A., Ahmed, H., Singh, V., Singh, D.K., Pathak, J., and Sinha, R.P. 2018. Ultraviolet radiation induced DNA damage and mechanisms of repair in cyanobacteria: An overview. *In Trends in Life Sciences. Edited by R.P.Sinha, U.P.Srivastava.* Nova Science Publishers, NY, USA. pp. 169-218.
- Rajneesh, Pathak, J., Häder, D.P., and Sinha, R.P. 2019. Impacts of ultraviolet radiation on certain physiological and biochemical processes in cyanobacteria inhabiting diverse habitats. *Environ. Exp. Bot.* **161**: 375-387. doi:10.1016/j.envexpbot.2018.10.037.

- Rajneesh, Pathak, J., Kannaujiya, V.K., Singh, S.P., and Sinha, R.P. 2017. Codon usage analysis of photolyase encoding genes of cyanobacteria inhabiting diverse habitats. 3 Biotech **7**: 1-14. doi:10.1007/s13205-017-0826-2.
- Ramírez, N., Serey, M., Illanes, A., Piumetti, M., and Ottone, C. 2021. Immobilization strategies of photolyases: Challenges and perspectives for DNA repairing application. J. Photochem. Photobiol. B: Biol. **215**: 112113. doi:10.1016/j.jphotobiol.2020.112113.
- Rastogi, R.P., Singh, S.P., Häder, D.-P., and Sinha, R.P. 2011. Ultraviolet-B-induced DNA damage and photorepair in the cyanobacterium *Anabaena variabilis* PCC 7937. Environ. Exp. Bot. **74**: 280-288. doi:10.1016/j.envexpbot.2011.06.010.
- Šali, A., and Blundell, T.L. 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**: 779-815. doi:10.1006/jmbi.1993.1626.
- Sancar, A. 1994. Structure and function of DNA photolyase. Biochemistry. **33**: 2-9. doi:10.1021/bi00167a001.
- Sancar, A. 2003. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. Chem. Rev. **103**: 2203-2238. doi:10.1021/cr0204348.
- Santiago, M., Ramírez-Sarmiento, C.A., Zamora, R.A., and Parra, L.P. 2016. Discovery, molecular mechanisms, and industrial applications of cold-active enzymes. Front. Microbiol. **7**: 1408. doi:10.3389/fmicb.2016.01408.
- Sinha, R.P., and Häder, D.-P. 2002. UV induced DNA damage and repair: a review. Photochem. Photobiol. Sci. **1**: 225-236. doi:10.1039/b201230h.

- Terai, Y., Sato, R., Matsumura, R., Iwai, S., and Yamamoto, J. 2020. Enhanced DNA repair by DNA photolyase bearing an artificial light-harvesting chromophore. *Nucleic Acids Res.* **48**: 10076-10086. doi:10.1093/nar/gkaa719.
- Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaultot, D., Kuypers, M.M., and Zehr, J.P. 2012. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science*. **337**: 1546-1550. doi:10.1126/science.1222700.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680. doi:10.1093/nar/22.22.4673.
- Todo, T. 1999. Functional diversity of the DNA photolyase/blue light receptor family. *Mutat. Res.* **434**: 89-97. doi:10.1016/S0921-8777(99)00013-0.
- Todo, T., Takemori, H., Ryo, H., Ihara, M., Matsunaga, T., Nikaido, O., Sato, K., and Nomura, T. 1993. A new photoreactivating enzyme that specifically repairs ultraviolet light-induced (6-4) photoproducts. *Nature*. **361**: 371-374. doi:10.1038/361371a0.
- Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., and Minh, B.Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**(W1): W232-W235. doi:10.1093/nar/gkw256.
- Wang, J., Du, X., Pan, W., Wang, X., and Wu, W. 2015. Photoactivation of the cryptochrome/photolyase superfamily. *J. Photochem. Photobiol. C: Photochem. Rev.* **22**: 84-102. doi:10.1016/j.jphotochemrev.2014.12.001.

- Wheeler, T.J., Clements, J., and Finn, R.D. 2014. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinform.* **15**: 7. doi:10.1186/1471-2105-15-7.
- Yasui, A., Eker, A.P., Yasuhira, S., Yajima, H., Kobayashi, T., Takao, M., and Oikawa, A. 1994. A new class of DNA photolyases present in various organisms including a placental mammals. *EMBO J.* **13**: 6143-6151. doi:10.1002/j.1460-2075.1994.tb06961.x.
- Zhu, S., and Zheng, G. 2018. Dynamic kinetic resolution of Vince lactam catalyzed by γ -lactamases: a mini-review. *J. Ind. Microbiol. Biotechnol.* **45**: 1017-1031. doi:10.1007/s10295-018-2093-6.

Draft

Table 1. Physiological parameter of protein sequences of different groups of cyanobacterial photolyase proteins.

Group 1								
Accession Number	Number of Amino Acids	Theoretical pI	Molecular Weight	Formula	Extinction coefficient	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
WP_023172026 <i>Gloeobacter kilaueensis</i>	478	8.24	54769.21	C ₂₄₇₂ H ₃₇₉₃ N ₇₀₃ O ₆₉₃ S ₁₀	94100	44.03	78.26	-0.442
NP_923730 <i>Gloeobacter violaceus</i> PCC 7421	479	7.76	54620.96	C ₂₄₅₅ H ₃₇₈₁ N ₇₀₉ O ₆₉₂ S ₁₀	106340	40.69	78.68	-0.476
WP_049691578 <i>Synechococcus</i> sp. WH 8103	477	5.55	54128.38	C ₂₄₂₅ H ₃₇₂₁ N ₆₈₉ O ₆₉₃ S ₁₆	122380	50.74	80.84	-0.402
WP_038551530 <i>Synechococcus</i> sp. KORDI 52	477	5.87	54021.38	C ₂₄₂₆ H ₃₇₂₈ N ₆₈₈ O ₆₈₈ S ₁₅	122380	48.51	82.29	-0.384
WP_011359075 <i>Synechococcus</i> sp. CC9902	477	5.82	54159.50	C ₂₄₃₁ H ₃₇₅₀ N ₆₉₂ O ₆₉₀ S ₁₄	120890	49.51	84.93	-0.384
WP_011363244 <i>Synechococcus</i> sp. CC9605	477	5.35	53934.21	C ₂₄₂₁ H ₃₇₁₉ N ₆₈₁ O ₆₉₃ S ₁₅	123870	55.85	82.08	-0.384
WP_006851735 <i>Synechococcus</i> sp. WH 8109	477	5.62	53949.20	C ₂₄₂₂ H ₃₇₂₄ N ₆₈₂ O ₆₉₄ S ₁₄	122380	47.89	83.31	-0.398
WP_043694469 <i>Synechococcus</i> sp. KORDI 49	477	5.11	54068.28	C ₂₄₂₁ H ₃₇₁₇ N ₆₈₅ O ₆₉₆ S ₁₆	120890	53.24	83.90	-0.378
WP_011932191 <i>Synechococcus</i> sp. WH 7803	492	5.40	54826.25	C ₂₄₄₇ H ₃₇₇₆ N ₇₀₀ O ₇₀₃ S ₁₈	119400	44.49	83.01	-0.309
WP_011618233 <i>Synechococcus</i> sp. CC9311	492	5.24	55425.95	C ₂₄₇₇ H ₃₈₂₉ N ₇₀₃ O ₇₁₄ S ₁₇	119400	48.50	84.35	-0.344
WP_038542864 <i>Synechococcus</i> sp. KORDI 100	476	5.58	54149.44	C ₂₄₃₀ H ₃₇₄₀ N ₆₉₄ O ₆₉₁ S ₁₃	120890	49.61	83.68	-0.424

WP_015109684.1 <i>Cyanobium gracile</i>	504	5.92	54626.16	C ₂₄₄₈ H ₃₇₉₂ N ₇₁₀ O ₆₉₀ S ₁₃	117910	43.43	87.36	-0.145
WP_011934643 <i>Synechococcus</i> sp. RCC307	467	6.05	52869.29	C ₂₃₈₃ H ₃₆₅₉ N ₆₇₅ O ₆₆₄ S ₁₅	123410	48.14	83.23	-0.387
WP_041424986 <i>Synechococcus</i> sp. CC9902	495	6.66	57278.68	C ₂₅₉₅ H ₃₉₄₇ N ₇₁₃ O ₇₁₁ S ₂₄	88810	32.99	83.54	-0.351
WP_042849890 <i>Prochlorococcus</i> sp. MIT 0604	496	9.04	58462.61	C ₂₆₈₆ H ₄₁₀₁ N ₆₈₉ O ₇₃₂ S ₂₂	80330	37.69	85.26	-0.350
WP_011823130 <i>Prochlorococcus marinus</i> NATLIA	504	7.55	58180.14	C ₂₆₁₄ H ₄₀₅₄ N ₇₀₂ O ₇₇₂ S ₁₆	118830	30.16	85.36	-0.502
WP_011819636 <i>Prochlorococcus marinus</i> 9515	478	9.15	55891.04	C ₂₅₄₅ H ₃₉₁₃ N ₆₈₁ O ₇₁₂ S ₁₄	115850	32.84	84.29	-0.536
WP_012007084 <i>Prochlorococcus marinus</i> 9215	478	9.19	55904.95	C ₂₅₅₁ H ₃₈₉₁ N ₆₇₅ O ₇₁₃ S ₁₅	118830	31.71	82.22	-0.501
WP_011862325 <i>Prochlorococcus marinus</i> 9301	478	8.86	55982.92	C ₂₅₄₈ H ₃₈₇₇ N ₆₇₅ O ₇₁₇ S ₁₇	118830	34.08	79.16	-0.554
WP_42849782 <i>Prochlorococcus</i> sp. MIT 0604	478	8.73	56117.96	C ₂₅₅₅ H ₃₈₇₀ N ₆₇₈ O ₇₁₈ S ₁₇	118830	32.03	76.92	-0.568
WP_011817777 <i>Prochlorococcus marinus</i> str. AS9601	477	9.03	55816.69	C ₂₅₃₈ H ₃₈₆₇ N ₆₇₇ O ₇₁₅ S ₁₆	118830	28.83	79.12	-0.561
WP_011375850 <i>Prochlorococcus marinus</i> 9312	478	8.48	55669.44	C ₂₅₂₅ H ₃₈₆₆ N ₆₇₂ O ₇₂₂ S ₁₅	115850	32.87	83.03	-0.514
WP_011131922 <i>Prochlorococcus marinus</i> 1986	478	8.63	55858.52	C ₂₅₄₁ H ₃₈₅₇ N ₆₇₇ O ₇₂₀ S ₁₄	111840	32.78	80.19	-0.522
WP_015167526 <i>Synechococcus</i> sp. PCC 7502	481	7.13	54835.61	C ₂₄₇₈ H ₃₈₁₉ N ₆₇₁ O ₇₀₇ S ₁₆	99350	40.45	84.03	-0.394
WP_015163355 <i>Pseudanabaena</i> sp. PCC 7367	489	5.71	55242.39	C ₂₄₆₂ H ₃₇₈₄ N ₆₉₀ O ₇₂₆ S ₁₈	94880	42.09	75.36	-0.453

WP_011431873 <i>Synechococcus</i> sp. JA-2-3B'a(2-13)	487	6.34	55540.12	C ₂₄₉₈ H ₃₈₁₅ N ₇₀₉ O ₇₀₁ S ₁₇	132350	54.06	80.02	-0.401
WP_011429671 <i>Synechococcus</i> sp. JA-3-3Ab	479	6.06	54811.13	C ₂₄₆₃ H ₃₇₅₁ N ₇₀₃ O ₆₉₃ S ₁₆	134870	53.09	78.68	-0.446
WP_015187610 <i>Gloeocapsa</i> sp. PCC 7428	479	5.53	54677.65	C ₂₄₅₂ H ₃₇₅₆ N ₆₈₂ O ₇₁₉ S ₁₂	107830	37.41	80.33	-0.463
WP_068509788 <i>Leptolyngbya</i> sp. O-77	479	8.07	54386.64	C ₂₄₄₁ H ₃₇₅₅ N ₆₈₉ O ₇₀₁ S ₁₃	117340	38.19	76.89	-0.508
WP_011377448 <i>Synechococcus elongatus</i> PCC 7942	484	6.15	54475.75	C ₂₄₅₇ H ₃₇₆₂ N ₆₈₈ O ₆₉₉ S ₁₁	114360	53.89	84.05	-0.377
WP_011377448 <i>Synechococcus</i> sp. 2973	484	6.15	54475.75	C ₂₄₅₇ H ₃₇₆₂ N ₆₈₈ O ₆₉₉ S ₁₁	114360	53.89	84.05	-0.377
WP_011243704 <i>Synechococcus elongates</i> PCC 6301	484	6.28	54461.77	C ₂₄₆₀ H ₃₇₆₀ N ₆₈₈ O ₆₉₆ S ₁₁	119860	53.96	84.05	-0.365
WP_015149490 <i>Oscillatoria acuminate</i>	479	5.02	55097.25	C ₂₄₇₇ H ₃₇₆₅ N ₆₆₇ O ₇₂₇ S ₁₈	124330	42.90	77.10	-0.482
NP_681215 <i>Thermosynechococcus elongates</i> BP-1	480	7.26	54594.07	C ₂₄₆₆ H ₃₇₆₄ N ₇₀₀ O ₆₈₅ S ₁₃	115850	49.72	81.44	-0.383
WP_015229213 <i>Dactylococcopsis salina</i>	476	5.39	55007.88	C ₂₄₇₇ H ₃₇₆₂ N ₆₇₈ O ₇₂₆ S ₁₁	122840	38.50	77.54	-0.607
WP_015225625 <i>Halothece</i> sp. PCC 7418	477	5.49	54757.72	C ₂₄₆₇ H ₃₇₅₆ N ₆₇₆ O ₇₂₀ S ₁₁	121350	37.15	79.22	-0.567
WP_028947705 <i>Synechocystis</i> sp. PCC 6714	479	6.32	55302	C ₂₄₉₈ H ₃₈₀₆ N ₆₉₄ O ₇₁₀ S ₁₂	126850	50.30	80.31	-0.548
WP_014407099 <i>Synechocystis</i> sp. PCC 6803	479	5.67	54969.28	C ₂₄₈₇ H ₃₇₈₁ N ₆₇₉ O ₇₁₀ S ₁₃	126850	52.30	80.75	-0.496
WP_015133579 <i>Leptolyngbya</i> sp. PCC 7376	475	5.70	54637.4	C ₂₄₆₆ H ₃₇₈₉ N ₆₅₉ O ₇₀₈ S ₂₀	114360	42.62	78.88	-0.419
WP_012308383 <i>Synechococcus</i> sp. PCC 7002	477	6.05	54728.36	C ₂₄₈₂ H ₃₇₉₀ N ₆₇₀ O ₇₀₂ S ₁₅	115850	36.28	81.68	-0.406

WP_015220137 <i>Cyanobacterium aponinum</i>	473	5.88	54878.35	C ₂₄₈₆ H ₃₇₉₂ N ₆₆₀ O ₇₁₇ S ₁₅	121810	42.30	80.49	-0.528
AFZ48192 <i>Cyanobacterium stanieri</i> PCC 7202	476	5.95	55559.19	C ₂₅₂₄ H ₃₈₂₄ N ₆₆₈ O ₇₁₈ S ₁₇	130290	38.09	77.33	-0.582
WP_044106756 <i>Epithemia turgida</i>	474	5.75	55396.87	C ₂₅₀₁ H ₃₇₉₈ N ₆₆₂ O ₇₂₆ S ₂₀	124790	42.55	75.34	-0.568
WP_015783666 <i>Cyanothece</i> sp. PCC 8802	481	5.10	55702.08	C ₂₅₁₂ H ₃₈₃₇ N ₆₆₇ O ₇₃₈ S ₁₆	121810	45.23	82.14	-0.480
WP_012594771 <i>Cyanothece</i> sp. PCC 8801	481	5.14	55699.08	C ₂₅₁₂ H ₃₈₃₆ N ₆₆₈ O ₇₃₇ S ₁₆	121810	48.21	82.74	-0.470
WP_009544875 <i>Cyanothece</i> sp. ATCC 51142	476	5.28	55442.65	C ₂₅₀₂ H ₃₈₀₀ N ₆₆₂ O ₇₃₆ S ₁₆	126280	43.59	78.09	-0.613
WP_041623646 <i>Oscillatoria nigro viridis</i>	481	6.68	55291.88	C ₂₄₈₂ H ₃₈₁₁ N ₆₈₇ O ₇₁₃ S ₁₉	122840	45.11	79.17	-0.459
WP_011610862 <i>Trichodesmium erythraeum</i>	474	7.99	55327.06	C ₂₅₀₆ H ₃₈₄₀ N ₆₆₈ O ₇₁₈ S ₁₆	130290	40.45	78.80	-0.582
WP_015185922 <i>Microcoleus</i> sp. PCC 7113	478	5.58	54614.64	C ₂₄₆₃ H ₃₇₄₃ N ₆₆₉ O ₇₁₇ S ₁₃	122840	41.98	78.72	-0.513
WP_046661007 <i>Microcystis aeruginosa</i> NIES 2549	474	6.20	54666.06	C ₂₄₆₅ H ₃₇₈₀ N ₆₇₄ O ₇₁₀ S ₁₄	104850	37.97	80.74	-0.517
WP_015173466 <i>Geitlerinema</i> sp.	481	5.64	54344.40	C ₂₄₃₉ H ₃₇₂₅ N ₆₈₃ O ₇₀₅ S ₁₄	132350	39.11	79.65	-.0409
WP_012264286 <i>Microcystis aeruginosa</i> NIES843	474	6.25	54606.06	C ₂₄₆₄ H ₃₇₇₆ N ₆₇₂ O ₇₀₇ S ₁₅	104850	38.16	80.74	-0.502
AKV66280 <i>Microcystis panniformis</i> FACHB 1757	474	5.95	54612.95	C ₂₄₆₂ H ₃₇₇₃ N ₆₇₁ O ₇₁₂ S ₁₄	104850	37.63	80.13	-0.528
WP_013321644 <i>Cyanothece</i> sp. PCC 7822	476	6.97	55123.75	C ₂₄₉₅ H ₃₈₂₇ N ₆₇₃ O ₇₁₄ S ₁₄	120320	46.09	80.19	-0.526
WP_012597497 <i>Cyanothece</i> sp. PCC 7424	475	5.37	55007.33	C ₂₄₇₇ H ₃₇₉₉ N ₆₆₅ O ₇₂₇ S ₁₅	121810	42.88	79.73	-0.589
WP_015195349 <i>Stanieria cyanosphaera</i> PCC 7437	475	5.53	55171.21	C ₂₄₇₅ H ₃₇₇₅ N ₆₇₁ O ₇₃₃ S ₁₆	127310	47.75	74.19	-0.633
WP_015203979	474	6.05	53923.07	C ₂₄₃₀ H ₃₇₂₂ N ₆₆₂ O ₇₀₆ S ₁₃	117340	37.72	81.81	-0.395

<i>Crinalium epipsammum</i>								
WP_015142809 <i>Pleurocapsa</i> sp. PCC7327	477	5.70	54794.09	C ₂₄₆₂ H ₃₇₉₂ N ₆₇₀ O ₇₂₃ S ₁₄	103820	46.61	80.27	-0.506
AKV65962 <i>Microcystis panniformis</i> FACHB 1757	485	6.62	57155.22	C ₂₆₂₀ H ₃₉₅₈ N ₆₈₂ O ₇₃₁ S ₁₄	126740	41.76	82.62	-0.473
WP_012625739 <i>Cyanothece</i> sp. PCC 7425	475	5.8	54288.39	C ₂₄₄₆ H ₃₇₂₉ N ₆₇₅ O ₇₀₅ S ₁₃	121350	44.77	77.94	-0.439
WP_015123016 <i>Synechococcus</i> sp. PCC 6312	486	6.61	54913.59	C ₂₄₇₅ H ₃₈₀₁ N ₆₈₅ O ₇₀₁ S ₁₇	110350	38.00	82.84	-0.336
WP_053538182 <i>Anabaena</i> sp. wa102	478	5.23	54753.06	C ₂₄₈₆ H ₃₇₈₉ N ₆₄₉ O ₇₂₅ S ₁₂	121810	41.95	83.77	-0.400
WP_015083582 <i>Anabaena</i> sp. 90	478	5.52	54586	C ₂₄₈₄ H ₃₇₈₀ N ₆₅₂ O ₇₁₄ S ₁₂	121810	39.06	84.81	-0.347
WP_015213341.1 <i>Anabaena cylindrica</i>	478	5.42	54336.45	C ₂₄₅₄ H ₃₇₄₁ N ₆₅₇ O ₇₁₇ S ₁₃	117340	42.14	81.51	-0.384
WP_013190227 <i>Trichormus azollae</i>	479	6.19	54567.03	C ₂₄₅₉ H ₃₇₇₉ N ₆₆₃ O ₇₁₄ S ₁₆	117340	37.03	82.73	-0.378
WP_035151531 <i>Calothrix</i> sp. 336/3	476	5.64	54361.53	C ₂₄₄₇ H ₃₇₄₄ N ₆₆₈ O ₇₁₂ S ₁₄	113330	45.39	79.81	-0.450
WP_015200442 <i>Calothrix</i> sp. PCC 6303	476	5.93	54315.43	C ₂₄₃₇ H ₃₇₃₈ N ₆₇₆ O ₇₁₀ S ₁₄	110810	38.57	82.04	-0.443
WP_015121066 <i>Rivularia</i> sp. PCC 7116	482	5.26	55350.73	C ₂₄₉₂ H ₃₈₂₀ N ₆₆₆ O ₇₃₃ S ₁₆	118830	39.02	78.80	-0.521
WP_015111284 <i>Nostoc</i> sp. PCC 7107	479	5.79	54679.86	C ₂₄₅₃ H ₃₇₇₄ N ₆₇₂ O ₇₂₂ S ₁₄	111840	49.31	78.52	-0.463
WP_042341371 <i>Calothrix</i> sp. PCC 7507	478	5.20	54540.45	C ₂₄₄₆ H ₃₇₃₃ N ₆₇₁ O ₇₂₂ S ₁₄	117340	45.46	77.47	-0.512
WP_012408666 <i>Nostoc punctiforme</i>	481	5.27	55083.11	C ₂₄₆₉ H ₃₇₈₇ N ₆₇₇ O ₇₃₂ S ₁₃	108860	45.02	79.13	-0.515
WP_015139123 <i>Nostoc</i> sp. PCC 7524	479	5.39	54764.09	C ₂₄₆₄ H ₃₇₇₈ N ₆₆₆ O ₇₂₀ S ₁₆	115850	44.77	79.31	-0.450
ABA23891 <i>Anabaena variabilis</i> ATCC 29413	479	6.16	54726.97	C ₂₄₆₀ H ₃₇₇₅ N ₆₇₅ O ₇₁₇ S ₁₄	115850	38.55	78.06	-0.479
WP_067775770 <i>Nostoc</i> sp. NIES 3756	479	5.80	54669.87	C ₂₄₅₈ H ₃₇₈₄ N ₆₇₂ O ₇₂₁ S ₁₂	110350	45.29	80.10	-0.463
WP_010996878	479	5.89	54770.94	C ₂₄₆₃ H ₃₇₈₃ N ₆₇₅ O ₇₂₁ S ₁₂	110350	44.33	78.50	-0.506

<i>Nostoc</i> sp. PCC 7120								
WP_068388061 <i>Leptolyngbya</i> sp. NIES 3755	477	5.36	54238.09	C ₂₄₂₂ H ₃₇₁₅ N ₆₆₉ O ₇₂₂ S ₁₅	92820	37.40	74.74	-0.556
Group 2								
Accession number	Number of Amino Acids	Theoretical pI	Molecular Weight	Formula	Extinction coefficient	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
WP_068507646 <i>Leptolyngbya</i> sp. O-77	554	7.20	63496.64	C ₂₈₇₂ H ₄₄₄₄ N ₈₀₆ O ₇₉₉ S ₁₄	108290	46.99	88.99	-0.358
WP_068384449 <i>Leptolyngbya</i> sp. NIES 3755	550	8.77	64018.06	C ₂₈₈₆ H ₄₄₆₃ N ₈₁₅ O ₈₁₀ S ₁₅	113790	49.51	85.49	-0.469
WP_023175030 <i>Gloeobacter kilaueensis</i>	468	8.31	53064.48	C ₂₃₇₅ H ₃₆₈₀ N ₆₉₂ O ₆₆₆ S ₁₅	100380	42.95	86.32	-0.278
WP_012413103 <i>Nostoc punctiforme</i>	516	8.92	61205.35	C ₂₇₈₂ H ₄₂₂₀ N ₇₄₆ O ₇₉₄ S ₁₂	130180	36.79	74.48	-0.667
WP_083782514 <i>Nostoc punctiforme</i>	428	9.29	50372.23	C ₂₂₉₀ H ₃₄₆₉ N ₆₂₁ O ₆₄₃ S ₁₂	101760	35.95	74.77	-0.548
WP_067776699 <i>Nostoc</i> sp. NIES 3756	514	8.58	61008.17	C ₂₇₆₇ H ₄₂₁₃ N ₇₅₁ O ₇₈₉ S ₁₂	124220	42.52	78.54	-0.647
WP_015121656 <i>Rivularia</i> sp. PCC 7116	519	8.62	61631.10	C ₂₇₉₅ H ₄₂₃₅ N ₇₄₇ O ₇₉₃ S ₂₀	124220	36.54	70.46	-0.650
WP_011817854 <i>Prochlorococcus marinus</i> str. AS9601	498	9.35	59374.06	C ₂₆₉₃ H ₄₁₄₆ N ₇₃₀ O ₇₅₃ S ₁₈	122840	40.63	76.99	-0.712
WP_011862398 <i>Prochlorococcus marinus</i> 9301	498	9.44	59292.09	C ₂₆₈₇ H ₄₁₅₆ N ₇₃₂ O ₇₅₀ S ₁₈	122840	39.13	78.15	-0.715
WP_042849845 <i>Prochlorococcus</i> sp. MIT 0604	498	9.31	59462.04	C ₂₆₉₇ H ₄₁₅₄ N ₇₃₀ O ₇₅₉ S ₁₆	121350	42.88	76.59	-0.731
WP_012007154 <i>Prochlorococcus marinus</i> 9215	497	9.38	59181.95	C ₂₆₈₈ H ₄₁₅₀ N ₇₂₆ O ₇₅₀ S ₁₇	124330	40.40	76.90	-0.723

WP_011819716 <i>Prochlorococcus marinus</i> 9515	503	9.48	60182.35	C ₂₇₃₅ H ₄₂₂₇ N ₇₃₉ O ₇₅₅ S ₂₀	121810	40.68	75.63	-0.745
WP_038651402 <i>Prochlorococcus</i> sp. MIT 0801	499	9.64	59297.22	C ₂₆₈₄ H ₄₁₆₃ N ₇₅₅ O ₇₃₄ S ₁₇	113330	46.92	81.46	-0.641
WP_011935602 <i>Synechococcus</i> sp. RCC 307	503	9.55	57115.93	C ₂₅₂₄ H ₃₉₂₉ N ₇₇₅ O ₇₀₉ S ₂₀	113900	52.86	76.70	-0.519
WP_015108500 <i>Cyanobium gracile</i>	531	9.81	59473.62	C ₂₆₃₁ H ₄₁₁₄ N ₈₂₀ O ₇₃₃ S ₁₆	124440	56.05	81.68	-0.413
WP_011363710 <i>Synechococcus</i> sp. CC9605	493	9.34	56633.63	C ₂₅₀₀ H ₃₈₈₃ N ₇₆₁ O ₆₉₈ S ₂₇	114930	57.85	74.24	-0.501
WP_038556279 <i>Synechococcus</i> sp. KORDI 52	493	9.65	56684.82	C ₂₅₁₁ H ₃₉₀₄ N ₇₆₄ O ₆₉₃ S ₂₅	120430	58.58	74.22	-0.534
WP_006851713 <i>Synechococcus</i> sp. WH 8109	493	9.64	56648.76	C ₂₅₀₄ H ₃₉₀₀ N ₇₆₂ O ₆₉₆ S ₂₆	116420	52.16	73.43	-0.539
WP_049692361 <i>Synechococcus</i> sp. WH 8103	490	9.21	56366.14	C ₂₄₈₃ H ₃₈₆₂ N ₇₅₈ O ₇₀₂ S ₂₅	116420	58.65	72.88	-0.578
WP_011360420 <i>Synechococcus</i> sp. CC 9902	496	9.76	57305.56	C ₂₅₄₆ H ₃₉₆₈ N ₇₆₈ O ₇₀₄ S ₂₂	116420	57.66	75.91	-0.572
WP_011932548 <i>Synechococcus</i> sp. WH 7803	490	9.41	56227.13	C ₂₄₈₀ H ₃₈₅₇ N ₇₆₇ O ₆₈₈ S ₂₅	113440	60.11	75.53	-0.492
WP_011619922 <i>Synechococcus</i> sp. CC9311	504	9.47	57782.03	C ₂₅₇₄ H ₃₉₉₂ N ₇₇₄ O ₇₀₈ S ₂₁	117910	54.24	80.75	-0.435
WP_038543674 <i>Synechococcus</i> sp. KORDI100	492	9.19	56751.54	C ₂₅₁₄ H ₃₈₇₇ N ₇₆₇ O ₆₉₈ S ₂₃	117910	47.52	75.00	-0.500
WP_043695663 <i>Synechococcus</i> sp. KORDI 49	496	9.35	57333.37	C ₂₅₃₄ H ₃₉₃₂ N ₇₇₆ O ₇₀₄ S ₂₅	114930	60.73	73.21	-0.575
WP_023172734 <i>Gloeobacter kilaueensis</i>	493	6.66	56083.94	C ₂₅₄₀ H ₃₈₆₈ N ₇₀₈ O ₇₀₃ S ₁₆	102790	43.59	82.62	-0.284

NP_924695 <i>Gloeobacter violaceus</i> PCC 7421	490	6.82	5574.53	C ₂₅₂₀ H ₃₈₆₁ N ₇₀₇ O ₇₀₄ S ₁₄	109780	44.75	82.06	-0.391
Group 3								
Accession number	Number of Amino Acids	Theoretical pI	Molecular Weight	Formula	Extinction coefficient	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
WP_023172734 <i>Gloeobacter kilaueensis</i>	493	6.66	56083.94	C ₂₅₄₀ H ₃₈₆₈ N ₇₀₈ O ₇₀₃ S ₁₆	102790	43.59	82.62	-0.284
NP_924695 <i>Gloeobacter violaceus</i> PCC 7421	490	6.82	55774.53	C ₂₅₂₀ H ₃₈₆₁ N ₇₀₇ O ₇₀₄ S ₁₄	109780	44.75	82.06	-0.391
WP_015109316 <i>Cyanobium gracile</i>	413	7.27	45840.94	C ₂₀₇₆ H ₃₁₇₅ N ₆₀₅ O ₅₇₁ S ₃	85940	52.30	83.10	-0.312
WP_011429214 <i>Synechococcus</i> sp. JA-3-3Ab	488	8.82	55429.28	C ₂₅₂₁ H ₃₈₆₂ N ₇₁₀ O ₆₈₉ S ₉	102790	42.17	86.25	-0.352
WP_011432441 <i>Synechococcus</i> sp. JA-2-3B'a(2-13)	486	8.42	55648.32	C ₂₅₁₂ H ₃₈₆₇ N ₇₀₉ O ₇₀₆ S ₁₁	106800	41.06	83.33	-0.442
WP_014407097 <i>Synechocystis</i> sp. PCC 6803	489	8.82	57039.95	C ₂₅₉₉ H ₃₉₃₃ N ₇₁₃ O ₇₁₆ S ₁₃	120320	48.46	78.16	-0.531
WP_028948235 <i>Synechocystis</i> sp. PCC 6714	489	8.29	56589.25	C ₂₅₇₅ H ₃₈₉₅ N ₇₀₃ O ₇₁₉ S ₁₂	127310	45.39	79.00	-0.453
NP_923781 <i>Gloeobacter violaceus</i> PCC 7421	500	7.82	57263.13	C ₂₆₁₁ H ₃₉₇₄ N ₇₂₆ O ₇₁₉ S ₇	113330	46.38	83.34	-0.391
WP_023173156 <i>Gloeobacter kilaueensis</i>	496	7.23	56815.49	C ₂₅₇₉ H ₃₉₄₅ N ₇₂₇ O ₇₁₈ S ₆	109320	47.31	85.40	-0.415
WP_015164987 <i>Pseudanabaena</i> sp. PCC 7367	500	5.88	57383.92	C ₂₆₀₄ H ₃₉₄₁ N ₇₀₃ O ₇₄₄ S ₁₂	115280	33.77	80.32	-0.424
WP_015133030 <i>Leptolyngbya</i> sp. PCC 7376	518	9.44	60080.30	C ₂₇₀₇ H ₄₁₈₀ N ₇₆₆ O ₇₆₇ S ₁₁	114820	50.42	76.62	-0.638

WP_083764472 <i>Synechococcus</i> sp. PCC 7002	516	8.68	59690.90	C ₂₇₀₄ H ₄₁₃₇ N ₇₅₃ O ₇₅₇ S ₁₂	127310	45.15	78.08	-0.576
AFZ47259 <i>Cyanobacterium stanieri</i> PCC 7202	486	9.01	57120.37	C ₂₆₁₄ H ₃₉₉₃ N ₆₇₃ O ₇₃₉ S ₁₄	115740	35.95	75.23	-0.590
WP_041922652 <i>Cyanobacterium aponinum</i>	486	6.39	57088.19	C ₂₆₂₂ H ₃₉₆₅ N ₆₇₅ O ₇₃₃ S ₁₃	119750	39.01	82.08	-0.442
WP_012266221 <i>Microcystis aeruginosa</i> NIES843	485	7.60	57057.06	C ₂₆₁₇ H ₃₉₄₈ N ₆₈₄ O ₇₂₈ S ₁₃	122730	40.31	80.00	-0.500
WP_046662052 <i>Microcystis aeruginosa</i>	485	7.14	56986.09	C ₂₆₁₅ H ₃₉₆₃ N ₆₈₃ O ₇₂₇ S ₁₂	119750	39.71	83.61	-0.476
WP_015155962 <i>Chroococcidiopsis thermalis</i>	477	5.30	55152.12	C ₂₄₆₇ H ₃₇₇₄ N ₆₈₈ O ₇₂₇ S ₁₄	107830	46.57	76.96	-0.576
WP_015226252 <i>Halotheca</i> sp. PCC 7418	483	7.09	56089.91	C ₂₅₄₄ H ₃₈₉₄ N ₆₇₈ O ₇₂₇ S ₁₅	115280	38.61	80.08	-0.501
WP_015229255 <i>Dactylococcopsis salina</i>	485	6.25	56648.24	C ₂₅₅₉ H ₃₉₀₈ N ₆₈₈ O ₇₄₁ S ₁₅	109780	37.02	77.55	-0.574
WP_012162902 <i>Acaryochloris marina</i>	484	6.30	56040.61	C ₂₅₃₇ H ₃₈₅₁ N ₆₉₅ O ₇₁₇ S ₁₅	110350	47.62	75.97	-0.500
WP_015169557 <i>Synechococcus</i> sp. PCC 7502	501	8.49	57799.17	C ₂₆₃₈ H ₄₀₄₃ N ₆₉₅ O ₇₄₁ S ₁₄	108290	40.88	85.89	-0.346
WP_011613840 <i>Trichodesmium erythraeum</i>	498	9.23	58243.92	C ₂₆₈₁ H ₄₀₉₅ N ₇₀₉ O ₇₂₉ S ₁₀	119290	31.74	83.55	-0.507
WP_015149163 <i>Oscillatoria acuminata</i>	486	6.21	55841.50	C ₂₅₅₉ H ₃₈₅₀ N ₆₇₀ O ₇₁₆ S ₁₂	113790	37.60	80.47	-0.352
WP_015194171 <i>Stanieria cyanosphaera</i> PCC 7437	488	8.72	57389.52	C ₂₆₄₂ H ₄₀₀₀ N ₆₈₆ O ₇₃₃ S ₉	128230	33.87	81.95	-0.537
WP_015144761 <i>Pleurocapsa</i> sp. PCC7327	488	9.08	57326.83	C ₂₆₂₉ H ₄₀₂₃ N ₆₉₉ O ₇₂₀ S ₁₂	111270	35.46	84.71	-0.483
WP_068514184 <i>Leptolyngbya</i> sp. O-77	505	8.43	57657.69	C ₂₆₂₄ H ₄₀₀₄ N ₇₂₆ O ₇₂₆ S ₁₀	107830	44.48	83.31	-0.389

WP_009543627 <i>Cyanothece</i> sp. ATCC 51142	491	8.78	57534.93	C ₂₆₃₆ H ₄₀₂₃ N ₆₈₅ O ₇₃₆ S ₁₄	117230	31.83	79.00	-0.568
WP_015213341.1 <i>Anabaena cylindrica</i>	478	5.42	54336.45	C ₂₄₅₄ H ₃₇₄₁ N ₆₅₇ O ₇₁₇ S ₁₃	117340	42.14	81.51	-0.384
WP_012597761 <i>Cyanothece</i> sp. PCC 7424	488	8.47	57070.17	C ₂₆₁₄ H ₃₉₇₁ N ₆₉₅ O ₇₂₄ S ₁₁	112760	39.49	81.54	-0.523
WP_012594061 <i>Cyanothece</i> sp. PCC 8801	488	8.64	57019.09	C ₂₆₁₁ H ₃₉₆₂ N ₆₈₂ O ₇₃₁ S ₁₃	112760	32.82	79.53	-0.506
WP_015783278 <i>Cyanothece</i> sp. PCC 8802	488	8.64	56978.04	C ₂₆₁₂ H ₃₉₅₅ N ₆₈₁ O ₇₂₉ S ₁₃	112760	33.52	78.93	-0.494
WP_012266221 <i>Microcystis aeruginosa</i> NIES843	485	7.60	57057.06	C ₂₆₁₇ H ₃₉₄₈ N ₆₈₄ O ₇₂₈ S ₁₃	122730	40.31	80.00	-0.500
WP_068388952 <i>Leptolyngbya</i> sp. NIES 3755	514	8.41	59674.87	C ₂₇₁₆ H ₄₁₅₃ N ₇₃₇ O ₇₆₆ S ₉	120780	33.84	79.32	-0.541
WP_015173466 <i>Geitlerinema</i> sp.	498	6.91	56803.43	C ₂₅₈₄ H ₃₈₂₁ N ₇₁₁ O ₇₂₃ S ₉	107803	43.74	78.01	-0.445
Group 4								
Accession Number	Number of Amino Acids	Theoretical pI	Molecular Weight	Formula	Extinction coefficient	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
WP_011818976.1 <i>Prochlorococcus marinus</i> AS9601	384	7.75	46317.55	C ₂₁₄₇ H ₃₁₉₁ N ₅₄₅ O ₅₉₅ S ₅	89840	29.31	78.18	-0.599
WP_042850913 <i>Prochlorococcus</i> sp. MIT 0604	384	7.04	46216.39	C ₂₁₃₇ H ₃₁₈₈ N ₅₄₂ O ₅₉₉ S ₅	89840	32.70	79.19	-0.585
WP_011863473 <i>Prochlorococcus marinus</i> 9301	382	8.57	46105.42	C ₂₁₃₈ H ₃₁₉₁ N ₅₄₁ O ₅₉₂ S ₅	91330	27.19	78.85	-0.616
WP_012008240 <i>Prochlorococcus marinus</i> 9215	384	9.06	46230.64	C ₂₁₄₄ H ₃₂₀₆ N ₅₄₆ O ₅₉₀ S ₅	89840	32.43	78.93	-0.601
WP_011377002 <i>Prochlorococcus marinus</i>	384	7.70	46053.58	C ₂₁₄₄ H ₃₂₁₅ N ₅₃₅ O ₅₉₀ S ₄	91330	30.71	87.58	-0.466

9312								
WP_011132994 <i>Prochlorococcus marinus</i> 1986	379	8.70	45540.90	C ₂₀₉₉ H ₃₁₆₇ N ₅₃₇ O ₅₈₇ S ₇	94310	27.22	83.56	-0.558
WP_011820825 <i>Prochlorococcus marinus</i> 9515	378	8.90	45543.85	C ₂₁₁₀ H ₃₁₅₂ N ₅₃₈ O ₅₈₁ S ₆	89840	33.24	79.15	-0.567
WP_038653645 <i>Prochlorococcus</i> sp. MIT 0801	371	9.24	44190.25	C ₂₀₂₄ H ₃₀₆₄ N ₅₃₄ O ₅₆₈ S ₇	92250	37.41	88.06	-0.512
WP_011294817 <i>Prochlorococcus marinus</i> NATL2A	371	9.02	44167.36	C ₂₀₁₅ H ₃₀₆₃ N ₅₃₁ O ₅₆₈ S ₁₁	89270	36.30	88.84	-0.489
WP_011824136 <i>Prochlorococcus marinus</i> NATL1A	371	8.98	44109.03	C ₂₀₁₁ H ₃₀₄₁ N ₅₃₃ O ₅₇₁ S ₉	89270	38.34	86.74	-0.505
WP_011359081 <i>Synechococcus</i> sp. CC9902	297	8.60	34149.18	C ₁₅₀₀ H ₂₂₉₃ N ₄₅₉ O ₄₄₀ S ₁₁	69440	48.39	62.53	-0.759
WP_081858946 <i>Synechococcus</i> sp. KORDI100	341	9.84	39126.90	C ₁₇₀₇ H ₂₆₄₅ N ₅₅₁ O ₄₉₁ S ₁₂	69440	50.25	65.84	-0.762
WP_043694479 <i>Synechococcus</i> sp. KORDI49	300	7.80	33653.71	C ₁₄₈₉ H ₂₂₆₄ N ₄₅₀ O ₄₂₉ S ₁₀	70930	53.00	65.50	-0.634
WP_049691582 <i>Synechococcus</i> sp. WH8103	299	7.21	33587.56	C ₁₄₈₄ H ₂₂₅₄ N ₄₄₈ O ₄₃₁ S ₁₀	70930	47.90	64.41	-0.659
WP_038551512 <i>Synechococcus</i> sp. KORDI52	303	7.23	34286.21	C ₁₅₀₉ H ₂₂₉₂ N ₄₆₀ O ₄₄₃ S ₁₀	70930	41.70	64.85	-0.708
WP_006851392 <i>Synechococcus</i> sp. WH 8109	304	9.48	34434.65	C ₁₅₁₆ H ₂₃₂₈ N ₄₇₀ O ₄₃₆ S ₁₀	70930	49.77	67.83	-0.676
WP_011363249 <i>Synechococcus</i> sp. CC9605	303	8.94	34348.38	C ₁₅₀₉ H ₂₃₀₆ N ₄₆₈ O ₄₃₉ S ₁₀	70930	52.68	64.85	-0.721
WP_011618238 <i>Synechococcus</i> sp. CC9311	315	8.53	36575.78	C ₁₅₉₆ H ₂₄₅₃ N ₄₉₅ O ₄₇₆ S ₁₂	72420	58.64	64.48	-0.847

WP_041428739 <i>Synechococcus</i> sp. 7803	299	7.78	34200.08	C ₁₅₀₀ H ₂₂₈₆ N ₄₆₂ O ₄₄₃ S ₁₀	76430	51.54	62.74	-0.822
WP_043738407 <i>Synechococcus</i> sp. RCC307	288	8.61	32920.13	C ₁₄₄₇ H ₂₂₂₅ N ₄₄₁ O ₄₁₇ S ₁₄	68410	51.01	69.86	-0.627
WP_011378226 <i>Synechococcus elongatus</i> PCC 7942	293	8.32	34674.57	C ₁₅₃₄ H ₂₃₁₂ N ₄₆₆ O ₄₄₆ S ₈	96370	52.78	63.04	-0.935
WP_011378226 <i>Synechococcus</i> sp. 2973	293	8.32	34674.57	C ₁₅₃₄ H ₂₃₁₂ N ₄₆₆ O ₄₄₆ S ₈	96370	52.78	63.04	-0.935
NP_681342 <i>Thermo Synechococcus</i> <i>elongatus</i> BP-1	275	6.24	31473.39	C ₁₄₁₁ H ₂₁₂₆ N ₄₀₂ O ₄₀₄ S ₉	82850	44.76	79.24	-0.391
WP_041431768 <i>Synechococcus</i> sp. PCC 6312	284	8.80	32695.77	C ₁₄₇₁ H ₂₂₁₇ N ₄₂₁ O ₄₁₇ S ₇	82850	38.83	71.55	-0.637
WP_015175588 <i>Oscillatoria nigro viridis</i>	286	8.74	33212.19	C ₁₄₈₃ H ₂₂₄₉ N ₄₂₇ O ₄₃₃ S ₇	80330	37.85	64.83	-0.780
WP_041781292 <i>Microcoleus</i> sp. PCC 7113	280	7.05	32583.54	C ₁₄₅₇ H ₂₁₉₇ N ₄₁₅ O ₄₂₃ S ₉	73340	37.13	68.00	-0.694
WP_015200773 <i>Calothrix</i> sp. PCC 6303	275	8.52	32266.43	C ₁₄₄₉ H ₂₁₉₀ N ₄₀₂ O ₄₁₇ S ₁₁	70360	30.26	63.85	-0.697
WP_015157008 <i>Chroococcidiopsis</i> <i>thermalis</i>	335	9.01	37916.08	C ₁₆₆₂ H ₂₅₃₀ N ₅₀₀ O ₅₀₃ S ₁₁	80330	46.61	53.31	-0.929
WP_044292594 <i>Rivularia</i> sp. PCC 7116	285	8.66	33798.09	C ₁₅₁₆ H ₂₃₀₄ N ₄₂₄ O ₄₄₀ S ₉	74830	40.67	62.67	-0.905
WP_035152880 <i>Calothrix</i> sp. 336/3	283	8.71	33274.62	C ₁₅₀₆ H ₂₂₇₁ N ₄₁₇ O ₄₂₅ S ₈	75860	40.22	68.98	-0.723
WP_042465464 <i>Anabaena cylindrica</i>	279	6.76	32851.94	C ₁₄₈₂ H ₂₂₃₃ N ₄₀₉ O ₄₂₈ S ₇	80330	37.01	72.72	-0.712
WP_053540055 <i>Anabaena</i> sp. wa102	279	8.25	32908.96	C ₁₄₈₁ H ₂₂₂₈ N ₄₂₀ O ₄₂₃ S ₇	75860	48.27	70.36	-0.770
WP_015113149 <i>Nostoc</i> sp. PCC 7107	281	8.24	32839.89	C ₁₄₈₁ H ₂₂₂₅ N ₄₁₃ O ₄₂₅ S ₇	80330	31.16	69.15	-0.754
WP_012412678 <i>Nostoc punctiforme</i>	295	6.67	34644.69	C ₁₅₄₃ H ₂₃₃₀ N ₄₄₀ O ₄₅₇ S ₉	80330	32.24	62.85	-0.870

WP_015130562 <i>Calothrix</i> sp. PCC 7507	278	8.18	32625.65	C ₁₄₆₉ H ₂₂₀₃ N ₄₁₃ O ₄₂₀ S ₈	80330	39.10	70.22	-0.696
WP_015140806 <i>Nostoc</i> sp. PCC 7524	278	7.68	32796.75	C ₁₄₆₉ H ₂₂₁₆ N ₄₁₄ O ₄₂₉ S ₈	80330	34.96	65.65	-0.854
WP_067767917 <i>Nostoc</i> sp. NIES 3756	278	9.23	32870.11	C ₁₄₈₁ H ₂₂₄₃ N ₄₁₅ O ₄₂₂ S ₈	81820	33.28	66.01	-0.831
WP_015130562 <i>Calothrix</i> sp. PCC 7507	278	8.18	32625.65	C ₁₄₆₉ H ₂₂₀₃ N ₄₁₃ O ₄₂₀ S ₈	80330	39.10	70.22	-0.696
WP_010997117 <i>Nostoc</i> sp. PCC 7120	278	8.68	32760.89	C ₁₄₇₆ H ₂₂₂₈ N ₄₀₈ O ₄₂₆ S ₈	81820	34.89	66.73	-0.812
WP_041918742 <i>Gloeocapsa</i> sp. PCC 7428	286	8.85	33865.06	C ₁₅₁₂ H ₂₂₉₇ N ₄₃₅ O ₄₃₈ S ₉	81820	39.30	66.50	-0.838
Group 5								
Accession Number	Number of Amino Acids	Theoretical pI	Molecular Weight	Formula	Extinction coefficient	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
WP_038654986 <i>Prochlorococcus</i> sp. MIT 0801	493	7.55	56970.74	C ₂₅₆₂ H ₃₉₇₄ N ₆₉₀ O ₇₅₅ S ₁₄	117340	30.48	86.47	-0.504
WP_012161330 <i>Acaryochloris marina</i> MBIC 11017	465	6.51	53762.10	C ₂₃₉₉ H ₃₇₀₃ N ₆₆₇ O ₇₀₀ S ₂₁	84230	47.74	82.00	-0.460
WP_012007253 <i>Prochlorococcus marinus</i> 9215	496	9.08	58575.81	C ₂₇₀₁ H ₄₁₀₈ N ₆₈₆ O ₇₃₀ S ₂₂	81820	37.68	84.48	-0.344
WP_042849890 <i>Prochlorococcus</i> sp. MIT 0604	496	9.04	58462.61	C ₂₆₈₆ H ₄₁₀₁ N ₆₈₉ O ₇₃₂ S ₂₂	80330	37.69	85.26	-0.350
WP_011862447 <i>Prochlorococcus marinus</i> 9301	496	9.03	58687.88	C ₂₇₀₂ H ₄₁₂₀ N ₆₉₀ O ₇₃₄ S ₂₁	81820	37.50	85.24	-0.365
WP_011817934 <i>Prochlorococcus marinus</i> str. AS9601	496	9.18	58554.83	C ₂₆₉₆ H ₄₁₁₇ N ₆₈₇ O ₇₃₁ S ₂₂	81820	37.41	85.44	-0.338
WP_011375983 <i>Prochlorococcus marinus</i> 9312	496	8.96	85702.00	C ₂₇₀₇ H ₄₁₂₆ N ₆₈₄ O ₇₃₄ S ₂₂	83310	37.84	86.83	-0.308

WP_011132061 <i>Prochlorococcus marinus</i> 1986	496	8.97	58695.74	C ₂₆₉₈ H ₄₁₀₄ N ₆₈₆ O ₇₄₀ S ₂₂	92250	42.67	83.89	-0.374
WP_011819804 <i>Prochlorococcus marinus</i> 9515	496	9.15	58813.22	C ₂₇₁₂ H ₄₁₃₁ N ₆₈₅ O ₇₃₂ S ₂₄	90760	37.15	84.27	-0.348
WP_011294815 <i>Prochlorococcus marinus</i> NATL2A	496	9.01	59429.07	C ₂₇₁₃ H ₄₁₁₉ N ₇₁₃ O ₇₅₄ S ₂₀	108750	43.58	79.21	-0.577
WP_011824134 <i>Prochlorococcus marinus</i> NATL1A	496	8.97	59416.08	C ₂₇₁₁ H ₄₁₂₀ N ₇₀₈ O ₇₅₇ S ₂₁	108750	43.49	78.02	-0.584
WP_038653639 <i>Prochlorococcus</i> sp. MIT 0801	496	9.11	59431.08	C ₂₇₁₁ H ₄₁₂₅ N ₇₁₁ O ₇₅₇ S ₂₀	108750	44.54	78.02	-0.591
WP_011619284 <i>Synechococcus</i> sp. CC9311	468	6.52	54126.23	C ₂₄₄₀ H ₃₇₄₄ N ₆₆₄ O ₆₈₀ S ₂₇	82850	42.54	83.74	-0.306
WP_011933177 <i>Synechococcus</i> sp. WH 7803	495	6.47	57334.52	C ₂₅₉₁ H ₃₉₁₅ N ₇₂₅ O ₇₀₇ S ₂₅	89840	42.96	78.06	-0.342
WP_041424986 <i>Synechococcus</i> sp. CC9902	495	6.66	57278.68	C ₂₅₉₅ H ₃₉₄₇ N ₇₁₃ O ₇₁₁ S ₂₄	88810	32.99	83.54	-0.351
WP_049692099 <i>Synechococcus</i> sp. WH8103	495	6.93	56749.18	C ₂₅₅₉ H ₃₉₂₈ N ₇₂₀ O ₇₀₀ S ₂₄	86860	38.00	84.51	-0.244
WP_043691106 <i>Synechococcus</i> sp. KORDI49	495	7.13	56913.34	C ₂₅₆₂ H ₃₉₂₄ N ₇₃₄ O ₆₉₄ S ₂₅	89840	38.39	84.00	-0.282
WP_038544691 <i>Synechococcus</i> sp. KORDI100	495	6.19	56665.96	C ₂₅₆₁ H ₃₉₀₃ N ₇₀₅ O ₇₀₆ S ₂₅	84340	42.92	83.60	-0.244
WP_006851478 <i>Synechococcus</i> sp. WH8109	498	6.31	56814.11	C ₂₅₆₂ H ₃₉₂₃ N ₇₁₉ O ₇₀₅ S ₂₃	88350	41.51	86.81	-0.203
WP_038554883 <i>Synechococcus</i> sp. KORDI52	498	6.44	56808.22	C ₂₅₆₀ H ₃₉₂₅ N ₇₁₉ O ₇₀₂ S ₂₅	88350	41.24	86.02	-0.184

WP_011364333 <i>Synechococcus</i> sp. CC9605	498	6.75	56679.05	C ₂₅₅₇ H ₃₉₁₄ N ₇₂₀ O ₆₉₈ S ₂₄	86860	36.88	85.06	-0.183
WP_015109315 <i>Cyanobium-gracile</i>	514	6.33	58623.06	C ₂₆₃₄ H ₄₀₅₆ N ₇₅₄ O ₇₂₉ S ₂₁	101870	59.05	84.84	-0.303
WP_015124969 <i>Synechococcus</i> sp. PCC6312	493	7.31	57112.73	C ₂₆₀₁ H ₃₉₇₇ N ₇₁₁ O ₇₀₄ S ₂₀	102330	42.96	88.64	-0.221
WP_012165858.1 <i>Acaryochloris marina</i>	486	6.41	56765.46	C ₂₅₈₅ H ₃₈₉₅ N ₆₉₁ O ₇₂₇ S ₁₅	99240	39.23	82.08	-0.379
WP_011613287 <i>Trichodesmium erythraeum</i>	476	8.97	57072.61	C ₂₆₄₁ H ₄₀₀₁ N ₆₆₃ O ₇₂₆ S ₁₃	113220	32.82	84.98	-0.513
WP_015168673 <i>Synechococcus</i> sp. PCC7502	510	6.44	59088.46	C ₂₆₈₈ H ₄₁₀₃ N ₇₀₁ O ₇₆₇ S ₁₈	88700	43.50	87.65	-0.325
WP_011935620 <i>Synechococcus</i> sp. RCC307	508	5.83	58091.07	C ₂₆₁₆ H ₃₉₇₉ N ₇₃₇ O ₇₃₁ S ₂₀	122380	45.42	82.46	-0.375
WP_011429960 <i>Synechococcus</i> sp. JA-33Ab	513	8.84	59645.06	C ₂₇₀₀ H ₄₀₈₃ N ₇₇₃ O ₇₂₇ S ₂₀	119290	54.88	78.21	-0.452
WP_011377897 <i>Synechococcus elongatus</i> PCC 7942	503	6.17	58330.25	C ₂₆₄₄ H ₃₉₈₈ N ₇₂₆ O ₇₃₈ S ₁₈	136820	44.85	82.13	-0.392
WP_011377897 <i>Synechococcus</i> sp. 2973	503	6.17	58330.25	C ₂₆₄₄ H ₃₉₈₈ N ₇₂₆ O ₇₃₈ S ₁₈	136820	44.85	82.13	-0.392
WP_011242691 <i>Synechococcus elongatus</i> PCC 6301	503	6.17	58292.24	C ₂₆₄₂ H ₃₉₉₀ N ₇₂₆ O ₇₃₇ S ₁₈	136820	45.15	82.90	-0.392
WP_015128955 <i>Calothrix</i> sp. PCC 7507	498	7.34	58308.80	C ₂₆₄₉ H ₄₀₁₈ N ₇₁₈ O ₇₂₈ S ₂₃	123300	53.12	83.27	-0.401
WP_041565504 <i>Nostoc punctiforme</i>	499	6.39	58754.18	C ₂₆₆₆ H ₄₀₄₁ N ₇₂₁ O ₇₃₉ S ₂₃	127310	56.03	84.19	-0.390
WP_067765227 <i>Nostoc</i> sp. NIES 3756	499	8.52	58855.48	C ₂₆₇₃ H ₄₀₆₀ N ₇₂₂ O ₇₃₆ S ₂₄	130290	53.90	78.40	-0.486
WP_015120345 <i>Rivularia</i> sp. PCC 7116	497	5.68	58501.66	C ₂₆₅₃ H ₄₀₁₃ N ₆₉₉ O ₇₅₄ S ₂₃	116770	47.34	80.48	-0.521

WP_011433543 <i>Synechococcus</i> sp. JA-2-3B'a (2-13)	525	8.20	60790.60	C ₂₇₅₉ H ₄₁₈₇ N ₇₇₅ O ₇₄₄ S ₂₁	123300	55.30	83.30	-0.381
WP_012629348 <i>Cyanothece</i> sp. PCC 7425	507	7.01	59248.66	C ₂₆₉₁ H ₄₀₅₅ N ₇₃₅ O ₇₃₈ S ₂₃	141290	52.86	80.61	-0.405
WP_015155170 <i>Chroococcidiopsis</i> <i>thermalis</i>	512	6.22	60468.09	C ₂₇₄₇ H ₄₁₂₁ N ₇₃₅ O ₇₅₈ S ₂₈	144270	52.78	77.01	-0.457
WP_015196918 <i>Calothrix</i> sp. PCC 6303	495	6.13	57892.93	C ₂₆₃₆ H ₃₉₄₉ N ₆₉₃ O ₇₃₈ S ₂₃	112760	45.80	77.84	-0.426
WP_035157668 <i>Calothrix</i> sp. 336/3	505	6.70	59096.67	C ₂₆₈₉ H ₄₀₆₇ N ₇₁₇ O ₇₄₃ S ₂₄	126280	43.40	82.85	-0.368
WP_010872781 <i>Synechocystis</i> sp. PCC 6803	504	6.28	58142.09	C ₂₆₃₆ H ₃₉₇₂ N ₇₁₆ O ₇₃₈ S ₂₀	138310	46.57	81.31	-0.371
WP_028946479_ <i>Synechocystis</i> sp. PCC 6714	503	6.50	58367.59	C ₂₆₅₉ H ₄₀₂₃ N ₇₁₁ O ₇₄₀ S ₁₈	143810	45.62	85.33	-0.329
AFZ47414 <i>Cyanobacterium stanieri</i> PCC 7202	519	6.52	61264.83	C ₂₇₈₆ H ₄₂₁₃ N ₇₂₉ O ₇₉₀ S ₂₂	123760	44.17	78.55	-0.533
WP_015218314 <i>Cyanobacterium aponinum</i>	500	8.43	59360.96	C ₂₇₀₄ H ₄₁₃₁ N ₇₁₅ O ₇₅₈ S ₁₈	122730	51.77	85.42	-0.555
WP_015177117 <i>Oscillatoria nigroviridis</i>	498	6.65	57984.36	C ₂₆₃₇ H ₃₉₉₀ N ₇₀₆ O ₇₂₉ S ₂₃	129830	46.81	79.74	-0.357
WP_012265680 <i>Microcystis aeruginosa</i> NIES843	507	6.41	59237.95	C ₂₆₉₁ H ₄₁₀₂ N ₇₂₄ O ₇₄₄ S ₂₃	127310	46.48	88.17	-0.340
WP_046661706 <i>Microcystis aeruginosa</i> NIES 2549	504	6.54	58702.39	C ₂₆₆₅ H ₄₀₇₃ N ₇₁₇ O ₇₃₈ S ₂₃	127310	45.49	90.62	-0.312
WP_009544563 <i>Cyanothece</i> sp. ATCC 51142	501	6.33	59204.34	C ₂₆₈₇ H ₄₀₆₃ N ₇₁₃ O ₇₆₃ S ₂₀	122270	56.27	79.62	-0.565
WP_015956435 <i>Cyanothece</i> sp. PCC 7424	506	5.91	59579.73	C ₂₆₉₁ H ₄₀₉₆ N ₇₂₆ O ₇₇₀ S ₂₀	123300	53.42	81.72	-0.555
WP_012307894 <i>Synechococcus</i> sp. PCC	511	8.41	59183.57	C ₂₆₉₆ H ₄₀₉₆ N ₇₄₆ O ₇₃₆ S ₁₄	127310	40.34	85.62	-0.402

7002								
WP_015224925 <i>Halothece</i> sp. PCC 7418	502	5.37	58393.15	C ₂₆₄₂ H ₃₉₇₃ N ₇₀₇ O ₇₅₅ S ₂₁	125820	44.70	80.28	-0.393
WP_015230902 <i>Dactylococcopsis salina</i>	502	5.87	58435.17	C ₂₆₄₄ H ₃₉₉₉ N ₇₁₁ O ₇₅₉ S ₁₇	121810	38.76	81.35	-0.466
WP_084762058 <i>Leptolyngbya</i> sp. NIES 3755	530	5.83	62690.17	C ₂₈₃₂ H ₄₂₄₉ N ₇₆₇ O ₇₉₉ S ₂₇	144270	63.42	71.21	-0.606
WP_068513283 <i>Leptolyngbya</i> sp. O-77	505	6.16	58598.95	C ₂₆₇₀ H ₄₀₂₄ N ₇₁₈ O ₇₃₄ S ₂₁	131780	49.12	82.34	-0.364

Draft

Table 2. Template selected for modelling protein structure in the study.

Group I		
PDB Id	Query Cover	Sequence Identity
1qnf (<i>Anacystis nidulans</i>)	0.98	59.45
1tez (<i>A. nidulans</i>)	0.97	59.74
4u63 (<i>Agrobacterium tumefaciens</i>)	0.95	38.02
1dnp (<i>E. coli</i>)	0.93	40.09
Group II		
4k03 (<i>Drosophila melanogaster</i>)	0.89	30.85
4gu5 (<i>D. melanogaster</i>)	0.89	30.85
4jzy (<i>D. melanogaster</i>)	0.87	32.00
3cvx (<i>D. melanogaster</i>)	0.86	30.77
Group III		
1np7 (<i>Synechocystis</i> sp. PCC 6803)	1.0	100
2vtb (<i>Arabidopsis thaliana</i>)	0.93	50.88
2ijg (<i>A. thaliana</i>)	0.93	50.88
2j4d (<i>A. thaliana</i>)	0.93	50.88
Group IV		
4u63 (<i>Agrobacterium tumefaciens</i>)	0.76	27.23
1np7 (<i>Synechocystis</i> sp. PCC 6803)	0.76	27.23
1dnp (<i>E. coli</i>)	0.74	27.85
3zxs (<i>Rhodobacter sphaeroides</i>)	0.78	18.18
Group V		
5kcm (<i>Agrobacterium fabrum</i>)	1.0	43.38
5lfa (<i>A. fabrum</i>)	1.0	43.45
4dja (<i>A. tumefaciens</i>)	1.0	42.66
5kcm (<i>A. fabrum</i>)	0.98	43.38

Table 3. Ramachandran plot analysis for structures of different groups of cyanobacterial photolyases.

	Group 1	Group 2	Group 3	Group 4	Group 5
Plot statistics Residues in most favoured regions [A,B,L]	91.2 %	85.8%	85.8%	88.6%	90.6%
Residues in additional allowed regions [a,b,l,p]	8.1%	12.7%	12.7%	10.9%	7.6%
Residues in generously allowed regions [~a,~b,~l,~p]	0.5%	0.8%	0.8%	00.0%	1.6%
Residues in disallowed regions	0.2 %	0.6%	0.6%	0.5%	0.2%

Draft

Table 4. Binding energy information obtained from the docking study of photolyase (Phr)-CPD/6-4 photoproduct interactions. *Nostoc punctiforme* 73102, WP_012408666 (Group I), *N. punctiforme* PCC 73102, WP_012413103, (Group II) *Synechocystis* sp. PCC 6803, WP_014407097 (Group III) *N. punctiforme* PCC 73102, WP_012412678 (Group IV), *N. punctiforme* PCC 73102, ACC82051 (Group V), respectively, were taken for this study.

Group	Central atom	Interaction with helices	Residue present at the binding site	Free Energy of Binding (kcal/mol)	Interacting residue	H-bonds	
						Numbers	Bond length
I	Trp 287	α -8,9, α -11	Phe 221, Tyr 229, Gly 240, THR 241, Ser 242, Gln 243, Leu 244, SER 245, Leu 248, Trp 281, Glu 284, Leu 285, Arg 288	-6.62	Ser-242, Leu-244	2	2.989 2.798
II (CPD)	Trp 321	α -15, α -21, α -24, α -26	Gly 327, Phe 328, Ser 407, Glu 438, Lys 439, Tyr 440, Ile 463, Ser 465, Asp 466, Leu 467, Arg 468, Lys 469	-6.10	Lys 439 Asp 466 Arg 468	3	3.044 1.848 2.892
II (6-4 Phrs)	Trp 321	α -20, α -23, α -25	Phe 328, Asn 405, Ser 407, Glu 438, Tyr 440, Ile 463, Val 464, Ser 465, Asp 466, Leu 467, Arg 468, Lys 469	-5.98	Glu 438 Ser 465 Leu 467 Arg 468	5	1.883 2.103 1.786 2.871 2.756
III	Trp 329	α -3, α -13, α -14, α -16	Pro 57, Trp 58, Gln 201, Ile 202, Asn 203, Phe 204, Asp 205, Arg 331, Leu 347, Asn 348, Leu 349, Thr 350, Trp 379, Cys 383	-6.32	Asn 203 Asn 348	3	2.108 2.727 1.951
IV (6-4 Phrs)	His 170	α -2-5, α -10	Tyr 55, Arg 59, Val 66, Thr 67, Arg 68, Leu 69, Ser 70, Ile 73, Ile 83, Leu 100, Glu 103, Leu 104, Asn 170	-6.62	Thr 67 Glu 103 Arg 59	6	1.795 2.670 2.889 2.166 3.139 3.147
IV (CPD)	His 170	α -2-5, α -10	Tyr 55, Arg 59, Val 66, Thr 67, Arg 68, Leu 69, Ser 70, Ile 73, Leu 100, Glu 103, Leu 104, Trp 106, Arg 107, Trp 110, Asn 170	-6.66	Thr 67 Arg 59 Glu 103	6	2.891 3.027 3.000 3.074 3.493 2.104
V	Arg 139	Found in loop region mainly	Lys 140, Arg 141, Leu 142, Phe 406, asp 408, Gly 409, Val 411, Leu 412, Ala 413, Gln 467	-5.85	Leu 142 Arg 141 Gly 409 Gln 467	6	3.140 3.110 2.720 1.992 2.224 2.895

Figure Captions

Fig. 1. Phylogenetic tree based on photolyase protein sequences of cyanobacteria. Neighbor-Joining trees are drawn to scale, and branch lengths (20.67995730) are based on the number of substitutions per site.

Fig. 2. Distribution of different types of photolyase (Phr) proteins in cyanobacteria. Data have been compiled from 238 amino acid sequences of Phr proteins encoded by 85 cyanobacterial species available in the NCBI genome database. The x-axis represents the number of Phr proteins and the y-axis represents the name of cyanobacterial species. Different colors of bars represent different classes of Phr proteins.

Fig. 3. Domain architecture of the different types of photolyase proteins present in cyanobacteria. The conserved domains were identified by the conserved domains database (CDD) tool.

Fig. 4. Sequence conservation of different classes of cyanobacterial photolyases based on multiple sequence alignment. The WebLogo diagrams of Group I (A; 79 sequences), Group II (B; 31 sequences), Group III (C; 33 sequences), Group IV (D; 43) and Group V (E; 52 sequences) Phrs were built using multiple sequence alignment files of each class. The red box shows catalytic site/unique residues putative binding site residues.

Fig. 5. 3-D Structures of different groups of photolyases (Group I to V) along with Ramachandran plot analysis results.

Fig. 6. Topology diagram of different classes of photolyases (Phrs) protein built up using PDBsum. All cyanobacterial Phr structures were found to possess typical Rossmann-like fold, i.e., conserved five beta-strands followed by alpha-helices. The amino acid sequence of *Nostoc punctiforme* 73102, WP_012408666 (A; Group I), *N. punctiforme* PCC 73102, WP_012413103,

(B; Group II) *Synechocystis* sp. PCC 6803, WP_014407097 (C; Group III) *N. punctiforme* PCC 73102, WP_012412678 (D; Group IV) *Nostoc punctiforme* PCC 73102, ACC82051 (E; Group V), respectively, for this study.

Fig. 7. Structure of different photolyases. Ribbon diagrams show the folds of the N-terminal/domain (*top*) and C-terminal FAD-binding-helical domain (*bottom*) connected by a long loop CRY/Phr structures (A), a cross-species amino acid conservation map of CPDPHRs using sequences highlights DNA binding groove and substrate binding pocket (B), the electrostatic potential surface shows a positively charged DNA binding groove above FAD binding cavity (C), hydrophobic surface view showing hydrophilic residue in blue color and hydrophobic residue in orange color (D).

Fig. 8. Images showing putative CPD/6-4PP binding sites on different classes of cyanobacterial photolyases (Phrs). Molecular docking study was conducted using modelled structures of *Nostoc punctiforme* 73102, WP_012408666 (Group I), *N. punctiforme* PCC 73102, WP_012413103, (Group II) *Synechocystis* sp. PCC 6803, WP_014407097 (Group III) *N. punctiforme* PCC 73102, WP_012412678 (Group IV), *N. punctiforme* PCC 73102, ACC82051 (Group V), putative CPD/6-4PP binding sites. CPD/6-4PP is shown in blue/gray while interacting residues of Phrs are shown in red color. The surface of the protein is shown in green color, and interacting hydrogen bonds between CPD/6-4photoproduct and Phr (< 6 Å zone).

Phylogenetic tree based on photolyase protein sequences of cyanobacteria. Neighbor-Joining trees are drawn to scale, and branch lengths (20.67995730) are based on the number of substitutions per site.

338x190mm (300 x 300 DPI)

	Group 1	Group 2	Group 3	Group 4	Group 5		
	Group 1	Group 2	Group 3	Group 4	Group 5		
Chroococcales	Acaryochloa	0	0	1	0	2	
	Candidatus	0	0	0	0	0	
	Cyanobacterium	1	0	1	0	1	
	Cyanobacterium	1	0	1	0	1	
	Cyanobium	1	1	1	0	1	
	Cyanothece	1	1	1	0	1	
	Cyanothece	1	0	1	0	1	
	Cyanothece	1	0	0	0	1	
	Cyanothece	1	0	1	0	0	
	Cyanothece	1	0	1	0	0	
	Cyanothece	1	0	1	0	0	
	Dactylococcopsis	1	0	1	0	1	
	Gloeocapsa	1	0	0	1	0	
	Halotheca sp.	1	0	1	0	1	
	Microcystis	1	0	1	0	1	
	Microcystis	1	0	1	1	1	
	Microcystis	1	0	1	0	0	
	Synechococcus	1	0	1	0	1	
	Synechococcus	1	0	1	0	1	
	Thermosynechococcus	1	0	0	1	0	
	Gloeobacterium	1	1	2	0	0	
	Gloeobacterium	1	0	2	0	0	
	Anabaena cylindrica	1	0	0	1	0	
	Anabaena sp.	1	0	0	0	0	
	Anabaena sp.	1	0	0	1	0	
Anabaena variabilis	1	0	0	1	0		
Nostocales	Calothrix sp.	1	0	0	1	1	
	Calothrix sp.	1	0	0	1	1	
	Calothrix sp.	1	0	0	0	0	
	Nostoc azolium	1	0	0	0	0	
	Nostoc punctatum	1	2	0	1	1	
	Nostoc sp. l.	1	0	0	1	0	
	Nostoc sp. l.	1	0	0	1	0	
	Nostoc sp. l.	1	0	0	1	0	
	Nostoc sp. l.	1	1	0	1	1	
	Rivularia sp.	1	1	0	1	1	
	Fischerella sp.	1	0	0	1	0	
	Crinalium elegans	1	0	0	0	0	
Oscillatoriales	Geitlerinema	1	0	1	0	1	
	Leptolyngbya	1	0	1	0	0	
	Leptolyngbya	1	1	1	0	1	
	Leptolyngbya	1	1	1	0	1	
	Microcoleus	1	1	0	1	0	
	Oscillatoria	1	0	1	0	0	
	Oscillatoria	1	0	0	1	1	
	Pseudanabaena	1	0	1	0	0	
	Pleurocapsales	Trichodesmium	1	1	0	0	1
		Chroococcoides	1	0	0	1	1

Pleurocapsa	1	0	1	0	0
Stanieria cy	1	0	1	0	0
Prochloroc	0	0	0	0	0
Prochloroc	1	1	0	1	1
Prochloroc	1	1	0	1	1
Prochloroc	1	1	0	1	1
Prochloroc	0	0	0	0	0
Prochloroc	1	0	0	1	1
Prochloroc	1	1	0	1	1
Prochloroc	1	0	0	1	1
Prochloroc	1	0	0	1	1
Prochloroc	0	0	0	0	0
Prochloroc	1	0	0	1	0
Prochloroc	1	1	0	1	0
Prochloroc	1	1	0	1	0
Synechococ	1	0	0	1	0
Synechococ	1	0	0	1	0
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	0	1	0	1
Synechococ	1	0	1	0	1
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	0	0	1	1
Synechococ	1	0	1	0	1
Synechococ	1	0	1	0	1
Synechococ	1	1	0	1	1
Synechococ	1	0	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	1	1	0	1	1
Synechococ	0	1	0	0	1
Synechococ	1	0	1	0	1
Synechococ	1	1	0	0	1
Synechococ	1	1	0	1	1

Draft

<i>Synechococcus</i> sp. <i>RSCCF101</i>	
<i>Synechococcus</i> sp. <i>SynAce01</i>	
<i>Synechococcus</i> sp. <i>PCC 8807</i>	
<i>Synechococcus</i> sp. <i>PCC CB0101</i>	
<i>Synechococcus</i> sp. <i>WH 8109</i>	
<i>Synechococcus</i> sp. <i>WH 8103</i>	
<i>Synechococcus</i> sp. <i>WH 7803</i>	
<i>Synechococcus</i> sp. <i>UTEX 2973</i>	
<i>Synechococcus</i> sp. <i>RCC307</i>	
<i>Synechococcus</i> sp. <i>PCC 7502</i>	
<i>Synechococcus</i> sp. <i>PCC 7002</i>	
<i>Synechococcus</i> sp. <i>PCC 6312</i>	
<i>Synechococcus</i> sp. <i>KORDI-52</i>	
<i>Synechococcus</i> sp. <i>KORDI-49</i>	
<i>Synechococcus</i> sp. <i>KORDI-100</i>	
<i>Synechococcus</i> sp. <i>JA-3-3-Ab</i>	
<i>Synechococcus</i> sp. <i>JA-2-3B'a-(2-13)</i>	
<i>Synechococcus</i> sp. <i>CC9902</i>	
<i>Synechococcus</i> sp. <i>CC9605</i>	
<i>Synechococcus</i> sp. <i>CC9311</i>	
<i>Synechococcus elongatus</i> <i>PCC 6301</i>	
<i>Synechococcus elongatus</i> <i>PCC 7942</i>	
<i>Prochlorococcus</i> sp. <i>MIT 0801</i>	
<i>Prochlorococcus</i> sp. <i>MIT 0604</i>	
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. <i>CCMP1986</i>	
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. <i>CCMP1375</i>	
<i>Prochlorococcus marinus</i> str. <i>NATL2A</i>	
<i>Prochlorococcus marinus</i> str. <i>NATL1A</i>	
<i>Prochlorococcus marinus</i> str. <i>MIT 9515</i>	
<i>Prochlorococcus marinus</i> str. <i>MIT 9312</i>	
<i>Prochlorococcus marinus</i> str. <i>MIT 8202</i>	

Prochlorococcus marinus str. MIT 9303

Prochlorococcus marinus str. MIT 9301

Prochlorococcus marinus str. MIT 9215

Prochlorococcus marinus str. AS9601

Prochlorococcus marinus MIT 9313

Stanieria cyanosphaera PCC 7437

Pleurocapsa sp. PCC 7327

Chroococcidiopsis thermalis PCC 7203

Trichodesmium erythraeum IMS101

Pseudanabaena sp. PCC 7367

Oscillatoria nigro-viridis PCC 7112

Oscillatoria acuminata PCC 6304

Microcoleus sp. PCC 7113

Leptolyngbya sp. NIES-3755

Leptolyngbya sp. O-77

Leptolyngbya sp. PCC 7376

Geitlerinema sp. PCC 7407

Crinalium epipsammum PCC 9333

Fischerella sp. NIES-3754

Rivularia sp. PCC 7116

Nostoc sp. NIES-3756

Nostoc sp. PCC 7524

Nostoc sp. PCC 7120

Nostoc sp. PCC 7107

Nostoc punctiforme PCC 73102

Nostoc azollae 0708

Calothrix sp. PCC 7507

Calothrix sp. PCC 6303

Calothrix sp. 336/3

Anabaena variabilis ATCC 29413

Anabaena sp. wa102

Anabaena sp. 90

Anabaena cylindrica PCC 7122

© The Author(s) or their Institution(s)

© The Author(s) or their Institution(s)

Draft

Draft

Draft

Domain architecture of the different types of photolyase proteins present in cyanobacteria. The conserved domains were identified by the conserved domains database (CDD) tool.

705x529mm (72 x 72 DPI)

C

WebLogo 3.7.4

D

Group 1

Group 2

Group 3

Group 4

Group 5

B

E

Group 4 **Group 5**