Structure Based Computational Analysis and Molecular Phylogeny of C- Phycocyanin Gene from the Selected Cyanobacteria

N. Reehana, A. Parveez Ahamed, D. Mubarak

Ali, A. Suresh, R. Arvind Kumar, N. Thajuddin

Abstract—Cyanobacteria play a vital role in the production of phycobiliproteins that includes phycocyanin and phycoerythrin pigments. Phycocyanin and related phycobiliproteins have wide variety of application that is used in the food, biotechnology and cosmetic industry because of their color, fluorescent and antioxidant properties. The present study is focused to understand the pigment at molecular level in the Cyanobacteria Oscillatoria terebriformis NTRI05 and Oscillatoria foreaui NTR06. After extraction of genomic DNA, the amplification of C-Phycocyanin gene was done with the suitable primer PCβF and PCαR and the sequencing was performed. Structural and Phylogenetic analysis was attained using the sequence to develop a molecular model.

Keywords—Cyanobacteria, C-Phycocyanin gene, Phylogenetic analysis, Structural analysis.

I. INTRODUCTION

CYANOBACTERIA are prokaryotic microorganisms with the unique ability to fix atmospheric nitrogen. They obtain their name from the bluish pigment phycocyanin, which are helpful to capture light for photosynthesis. In some Cyanobacteria, the color of light influences the composition of phycobilisomes. In green light, the cells accumulate more phycocerythrin, whereas in red light they produce more phycocyanin. Thus the Cyanobacteria appear green in red light and red in green light which is a complementary chromatic adaptation [1].

Molecular methods have become an indispensable tool for the characterization of Cyanobacteria. Utilizing the bioinformatic tools, the development of genomics and molecular technologies combined together to obtain more information related to molecular biology of these organisms [2]. Molecular assessment of Cyanobacterial biodiversity frequently uses markers like 16S rDNA, phycocyanin locus, nif gene, rpo gene, ITS region, introns, STRR, RAPD, M13 etc. [3]-[12]. A BLAST (Basic Local Alignment Search Tool) search enables a researcher to compare a query sequence with a library above certain threshold. Molecular phylogenetics is used to gain information on an organism's evolutionary relationships based on the database of taxonomy and biogeography using MEGA software. Protein has three main structures in which primary structure is the linear amino acid sequence usually represented by a one letter notation. α-helices, β - sheets and loops are formed in secondary structure, when the sequences of primary structures tend to arrange themselves in regular conformations. Protein folding is the process that results in a compact structure in which secondary structure elements are packed against each other in a stable configuration. This three dimensional structure of the protein is known as the protein tertiary structure [13]. Molecular modelling is helpful to know the assumed structure of molecules and also used to investigate the dynamics and thermodynamics of inorganic, biological, and polymeric systems [14].

II. MATERIALS AND METHODS

A. Morphological Identification of Cyanobacteria

The Fresh and Marine water Cyanobacterial strains were obtained as axenic cultures from germplasm of Department of Microbiology, Bharathidasan University, Tiruchirapalli and were examined carefully using light microscope. The strains were maintained with alternative illumination (i.e. 16 hr light and 8 hr dark conditions) in germplasm at 25°C and exposed to 2000 Lux light intensity. The Morphological characters were determined with their cell shape and cell size.

B. Molecular Characterization of Cyanobacteria

Two Cyanobacterial cultures namely Oscillatoria terebriformis NTRI05 and Oscillatoria foreaui NTR06 were harvested at their exponential growth phase used for the extraction of genomic DNA [15].

C. C-Phycocyanin Gene Amplification by Polymerase Chain Reaction (PCR)

PCR amplification was performed for the respective two samples of purified DNA in a thermal cycler (MYGene™ Series Peltier model MG96G) using the universal primers [PCβF - 5’ GGCTGCTTGTATAGCGACA 3’ and PCαR - 5’ CCAGTACCCACGGACACTAA 3’] [6]. The Initial denaturation was achieved at 94°C for 2 minutes and further denaturation was carried out at 94°C for 5sec; annealing at 47°C for 10sec, elongation at 72°C for 30sec and a final elongation at 70°C for 7min for 40 cycles. The

N. Reehana is with the Department of Microbiology, Bharathidasan University, Tiruchirapalli – 620024, Tamilnadu, India (e-mail: reehana@yahoo.in).
A. Parveez Ahamed, D. Mubarak Ali, A. Suresh, and R. Arvind Kumar are with the Department of Microbiology, Bharathidasan University, Tiruchirapalli – 620024, Tamilnadu, India.
N. Thajuddin is with the Department of Microbiology, Bharathidasan University, Tiruchirapalli – 620024, Tamilnadu, India (Phone: 9842379719; e-mail: nthaju2002@yahoo.com).
Amplified products were isolated by electrophoresis on 1.2% agarose gel using 1X TBE buffer and the bands were observed under gel documentation system (Photonyx).

D. Sequencing of C-Phycocyanin Gene
The sequencing of C-Phycocyanin gene was done by Ocimum Biosolutions, Hyderabad. Primer sequences were checked for homology to other sequences deposited in the available databases using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

E. Phylogenetic Analysis
Based on the sequenced data, the phylogenetic tree was constructed using bioinformatics tool Mega 5.05 [16], which has been used for aligning the sequences by Neighbor joining method (http://www.megasoftware.net/mega.php).

F. Primary Structure Prediction
The nucleotide sequence was converted to amino acid sequence using transq software and the primary structure was predicted by online software ProtParam, a tool which allows the computation of various physical and chemical parameters (http://www.expasy.org/tools/protparam.html).

G. Secondary Structure Prediction
The translated protein sequences were analyzed for secondary structure prediction. The query sequence was uploaded in alignment box and the query was submitted to GOR secondary structure prediction method version IV for structural analysis. The structure was predicted and compared with their models (npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) [17].

H. Molecular Modelling
The protein was performed for homology modelling by MODELLER9V8 (http://www.salilab.org/modeller/). The constructed model was minimized by CHIMERA. The overall stereochemical properties of the proteins were analysed in the RAMPAGE server. The three dimensional structure were further verified by VERIFY3D. RMS-Z score for bond angles of modeled protein structure was estimated by QMEAN server. The models are viewed in CHIMERA.

III. RESULTS AND DISCUSSION

A. Morphological Identification of Cyanobacteria
The Oscillatoria terebriformis NTRI05 and Oscillatoria foreaui NTRI06 were identified based on their morphological characters with standard Cyanophyta Monograph [18].

B. Molecular Characterization of Cyanobacteria
The Extraction of genomic DNA for Oscillatoria terebriformis NTRI05 and Oscillatoria foreaui NTRI06 were amplified for C-Phycocyanin gene using the primers PCβF and PCαR. The range of C- phycocyanin gene was 700 bp and the sequence similarity was analysed using the online tool BLAST (Fig. 1).

The genetic material of an organism determines the character of a species. The application of DNA based amplification methods for the molecular typing of microorganisms in complex natural populations are making the study of Cyanobacterial systematics definitive rather than descriptive [2].

C. Phylogenetic Analysis
Phylogenetic analysis is the easiest way to depict any evolutionary relationship between the groups of organisms. The phylogenetic tree for the two isolates namely Oscillatoria terebriformis NTRI05 and Oscillatoria foreaui NTRI06 were constructed using C-Phycocyanin gene sequence with that of other Cyanobacterial species by Neighbor Joining [19]. The sum of branch length of the tree is 1.90892857. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site (Fig. 2). The analysis involved 5 nucleotide sequences and codon positions included were 1st+2nd+3rd +Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 70 positions in the final data set. Evolutionary analyses were conducted in MEGA5 [20].
D. Structural Analysis

The obtained nucleotide sequence for both isolates (*Oscillatoria terebriformis* NTRI05 and *Oscillatoria foreaui* NTRI06) were analyzed for primary structure prediction. The physio-chemical parameters such as the molecular weight, isoelectric point (pI), extinction coefficient, half life, aliphatic index, amino acid property, instability index, Grand Average of Hydropathicity (GRAVY) were calculated using the ProtParam tool of the Expasy proteomics server.

Results showed that *Oscillatoria terebriformis* NTRI05 has 295 amino acid residues with the average molecular weight of 32909.6 g mol$^{-1}$. Conversely, the *Oscillatoria foreaui* NTRI06 had an average molecular weight of 30007.1 g mol$^{-1}$. The computed isoelectric point (PI) was above 7 indicating that the proteins are useful for developing buffer systems for purification by isoelectric focusing method. The total number of negatively charged residues (Asp+ Glu) was 25 and positively charged residues (Arg+ Lys) was 45 for *Oscillatoria terebriformis* NTRI05 where as for *Oscillatoria foreaui* NTRI06 (Asp+ Glu) was 18 and (Arg+ Lys) was 27. The Extinction coefficient of *Oscillatoria terebriformis* NTRI05 at 280 nm is 25035 M$^{-1}$ cm$^{-1}$ and for *Oscillatoria foreaui* NTRI06 32860 M$^{-1}$ cm$^{-1}$ with respect to the concentration of Cys residues. According to the Prot Param server, a protein whose instability index is larger than 40 may be unstable [21] therefore, the program predicted for both Cyanobacteria would be unstable in solution. The aliphatic index for the *Oscillatoria terebriformis* NTRI05 was 69.05 and for *Oscillatoria foreaui* NTRI06 90.63, which is regarded as a positive factor for the increase of thermal stability of globular proteins. The GRAVY value for both ranged from -0.464 to -0.068 indicating that the proteins will interact favorably with water (Table I).

### TABLE I

**PRIMARY STRUCTURE OF OSCILLATORIA TEREBRIFORMIS NTRI05 & OSCILLATORIA FOREAUI NTRI06**

<table>
<thead>
<tr>
<th>Physicochemical parameters</th>
<th>Oscillatoria terebriformis NTRI05</th>
<th>Oscillatoria foreaui NTRI06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (total no. of amino acids)</td>
<td>295</td>
<td>268</td>
</tr>
<tr>
<td>Isoelectric point (PI)</td>
<td>9.68</td>
<td>8.83</td>
</tr>
<tr>
<td>Molecular weight (MW)</td>
<td>32909.6</td>
<td>30007.1</td>
</tr>
<tr>
<td>Negatively charged residues (-R)</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>Positively charged residues (+R)</td>
<td>45</td>
<td>27</td>
</tr>
<tr>
<td>Extinction coefficient (EC)</td>
<td>25035</td>
<td>32860</td>
</tr>
<tr>
<td>Instability index (II)</td>
<td>41.96</td>
<td>90.63</td>
</tr>
<tr>
<td>Aliphatic index (AI)</td>
<td>69.05</td>
<td>90.63</td>
</tr>
<tr>
<td>Grand average of hydropathicity (GRAVY)</td>
<td>-0.464</td>
<td>-0.068</td>
</tr>
</tbody>
</table>

Secondary structure predicts the structure of proteins and nucleic acid sequences based only on the knowledge of their primary structure. The secondary structure helps in determining the exact structure of the gene which was predicted using GOR IV software. The alpha helix and beta sheet are made up of two networks: a sequence-to-structure network and a structure-to-structure network. The results of *Oscillatoria terebriformis* NTRI05 revealed that random coil (61.36%) dominated among secondary structure elements and alpha helices (20.00%) and extended strand (18.64%) are presented (Fig. 3). The results of *Oscillatoria foreaui* NTRI06 revealed that random coil (59.70%) dominated among secondary structure elements and alpha helices (14.55%) and extended strand (25.75%) are presented (Fig. 4). The secondary structure was employed to improve the sequence alignment. The secondary structures are more conserved than the nucleotide sequences; their analysis helps in understanding molecular evolution and increases the number of structural characters.

![Fig. 3 GOR4 result for Secondary Structure of Oscillatoria terebriformis NTRI05](image-url)
Fig. 4 GOR4 result for Secondary Structure of Oscillatoria foreaui NTRIO6

The modelling of the three dimensional structure of the protein was performed by MODELLER9V8 for both the isolates Oscillatoria terebriformis NTRIO5 and Oscillatoria foreaui NTRIO6 (Figs. 5 & 6). Molecular modelling methods are now routinely used to investigate the structure, dynamics, surface properties and thermodynamics of inorganic, biological and polymeric systems [4].

Fig. 5 Three dimensional structure of C-Phycocyanin gene - Oscillatoria terebriformis NTRIO5

IV. CONCLUSION

In this study, C-Phycocyanin gene was amplified and sequenced from the Cyanobacterial strains, Oscillatoria terebriformis NTRIO5 and Oscillatoria foreaui NTRIO6. Phylogenetic analysis, primary & secondary structure analysis and modelling of three dimensional structure of the protein were also performed. Our future approach will step forward towards barcoding for taxonomic application.

ACKNOWLEDGMENT

The first author (N. Reehana) is thankful to UGC-MANF (Maulana Azad National Fellowship) for their financial assistance.

REFERENCES


