Crystal structure determination and analysis of chromophore-protein interaction of phycocyanin

Part of this chapter has been published as:
Gupta GD, Sonani RR, Sharma M, Patel K, Rastogi RP, Madamwar D, Kumar V. Crystal structure analysis of phycocyanin from chromatically adapted Phormidium rubidum A09DM. RSC Advances. 6 (2016) 77898-907
4.1 Introduction

Phycocyanin (PC), the intermediate energy-radiation ($\lambda_{\text{Amax}} \sim 620$ nm) absorbing protein is believed to be evolved before high-energy absorbing PE ($\lambda_{\text{Amax}} \sim 560$ nm) and after low-energy absorbing APC ($\lambda_{\text{Amax}} \sim 653$ nm) in phycobilisome (PBS) (Six et al., 2007). It resides in PBS-rod in very specific manner (proximal to the PBS-core) that allows it to perform an important function of grabbing high-energy radiation and its efficient transfer to the comparatively low-energy absorbing PBS-core (Gantt & Lipschultz, 1973). The PC protein contains covalently attached phycocyanobilin (PCB) chromophore(s), which is mainly responsible for PC’s unique photophysical properties and function (Singh et al., 2015). The spectral nature of chromophore is a combined manifestation of both the conformation of chromophore tetrapyrrol rings and its interaction with surrounding protein microenvironment (Adir, 2005; Adir, 2008; Adir et al., 2006). These both factors vary marginally among diverse cyanobacterial species due to variation at the level of amino acid sequence. Therefore, the variation in structural traits and associated spectral features of cyanobacterial PC-analogues need to be revealed to develop complete understanding of light harvesting principle.

Several crystal structures of PC, ranging from 3.00 to 1.35 Å resolution have been reported from marine as well as fresh water cyanobacteria (PDB ID: 3o2c, 3o18, 310f, 4h0m, 4f0t, 4n6s, 4l1e, 4q70, 4z8k, 4ziz) (Conrad et al., 2015; Fromme et al., 2015; Kumar Singh et al., 2014; Marx & Adir, 2013; Marx & Adir, 2014). The cyanobacterium P. rubidum is type III chromatic adapter (Parmar et al., 2013), whose PE and PC content could be increased upon provision of green and red light, respectively. The crystal structure of PC, obtained from both the normal (white light) and chromatically adapted (red light) P. rubidum conditions are solved and analysed for any possible alteration in chromophore-binding site. Moreover, the geometry of the chromophores have been analysed to probe the role of surrounding residues in energy transfer mechanics.

4.2 Material and methods

4.2.1 Phycocyanin preparation

The P. rubidum was grown under white light and red light as described in Section 2.2.1. The PC, from both white and red light grown cultures was isolated as described in Section 2.2.2.
4.2.2 Crystallization

Both, white and red light PC was concentrated to 10 mg mL\(^{-1}\) in 10 mM Tris-Cl buffer (pH 8.1). Both the proteins was screened for crystallization using several pre-formulated commercial crystallization screens including JCSG-plus, PACT-premier, Morpheus, MIDAS, Structure screens by sitting drop vapour diffusion method in 96-well plate. The white light-PC gave reasonable size crystals (150 x 80 x 80 microns) within 6 days in conditions E4 (Precipitant mixture 4; Ethylene glycols, Buffer system 1, pH 6.5) (Table 4.1) and F4 (Precipitant mixture 4; Carboxylic acids, Buffer system 1, pH 6.5) (Table 4.1) of Morpheus screen. Both the conditions, E4 and F4 were further optimized at larger volume 24-well plate and diffraction-quality crystals of 300 x 100 x 100 microns were obtained after 20 days with the conditions detailed in Table 4.2. Red light-PE gave rod shape crystals with conditions containing PEG 3350 and citrate buffer using hanging-drop method as detailed in Table 4.2.

Table 4.1. Details of Morpheus crystallization screens

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitant mix 4</td>
<td>25% MPD + 25% PEG 1000 + 25% PEG 3350</td>
</tr>
<tr>
<td>Ethylene glycols</td>
<td>0.3M Diethylene glycol; 0.3M Triethylene glycol; 0.3M Tetraethylene glycol</td>
</tr>
<tr>
<td>Carboxylic acid</td>
<td>0.2M D-Glucose; 0.2M D-Mannose; 0.2M D-Galactose; 0.2M L-Fucose; 0.2M D-Xylose; 0.2M N-Acetyl-D-Glucosamine</td>
</tr>
<tr>
<td>Buffer system 1</td>
<td>1.0 M Imidazole + 1.0M MES monohydrate, pH 6.5</td>
</tr>
</tbody>
</table>

Table 4.2. Optimized crystallization conditions for white light-PC and red light-PC

<table>
<thead>
<tr>
<th></th>
<th>White light PC</th>
<th>Red light PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (K)</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td>Protein concentration</td>
<td>10 mg mL(^{-1})</td>
<td>10 mg mL(^{-1})</td>
</tr>
<tr>
<td>Buffer</td>
<td>0.1 M Imidazole + 0.1 M MES monohydrate, pH 6.5</td>
<td>0.1 M Sodium Citrate (pH 6.0)</td>
</tr>
<tr>
<td>Precipitants</td>
<td>13.5 % MPD, 13.5 % PEG 1000, 13.5 % PEG 3350</td>
<td>Precipitants</td>
</tr>
<tr>
<td>Salt/Ligands</td>
<td>30 mM Diethylene glycol; 30 mM Triethylene glycol; 30 mM Tetraethylene glycol; 30 mM Pentaethylene glycol</td>
<td>Salt/Ligands</td>
</tr>
<tr>
<td>Drop composition</td>
<td>Drop volume 10 µL with protein and reservoir in 1:1 ratio</td>
<td>None</td>
</tr>
<tr>
<td>Reservoir volume</td>
<td>600 µL</td>
<td></td>
</tr>
</tbody>
</table>
4.2.3 Data collection

White light-PE crystals were cooled at 100 K by spraying moisture free liquid nitrogen stream before and during the data collection. The white light PC crystals (already cryoprotected in condition) were diffracted up to ~1.6-1.7 Å resolution during testing at in-house fine focused X-ray source. From which, best crystals were subjected to data collection on the proteins crystallography beam line I04 at Diamond Light Source (DLS) synchrotron, UK. Red light-PC crystals were transferred to cryo-protectant solution having 10% PEG 400 along with crystallization mother liquor prior to freezing in liquid nitrogen stream. Diffraction intensity data were recorded at 100 K on CCD detector (TITAN) using in-house Agilent supernova system having a micro-focus sealed tube X-rays source with multilayer optics operated at 50 kV and 0.8 mA (Cu Kα).

4.2.4 Data processing and structure refinement

The intensity data of both proteins were indexed and integrated using the XDS program (Kabsch, 2010) and were scaled, merged and truncated using Aimless and Ctruncate programs in CCP4 suite (Winn et al., 2011). Initial phases for model were determined with molecular replacement method (Vagin & Teplyakov, 2010). Crystal structure of PC from *Spirulina plantensis* (PDB ID: 3o18) (David et al., 2011) was used as search model for red light PC data. Atomic coordinates of αβ monomer having alanine at non-conserved/mismatch positions, after removing PCB-chromophores and water molecules, were used for molecular replacement solutions. Refined red light-PC model was used as a search model for white light PC. Two αβ monomers were observed in the asymmetric unit. The phases were reasonably accurate as a clear electron density could be observed for PCB chromophores not used in search model. The coordinates were subsequently refined by REFMAC5 (Murshudov et al., 1997) with intermittent model building using COOT (Emsley et al., 2010). Refinement of the model was monitored by R_work and R_free. The geometry and stereochemistry of the models were analyzed throughout model building and refinement with the program PROCHECK (Laskowski et al., 1993). Exact sequences of protein was obtained by structure-based protein sequencing from high resolution white light-PC diffraction data, and the same were verified by molecular mass of α- and β-subunits, obtained from MALDI-ToF profile.
4.2.5 Sequence and phylogenetic analysis

The homologous sequences for PC α and β-subunits were retrieved from the Protein Data Bank (PDB) or Uniprot database using BLAST. The sequences were aligned with multiple sequence alignment tool clustal-omega (McWilliam et al., 2013). The phylogenetic analyses were conducted in MEGA6 (Tamura et al., 2013) using the Neighbour-Joining Method. The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992). The sequence alignment figures were prepared with ESPript (Robert & Gouet, 2014).

4.3 Results and discussion

4.3.1 Crystallographic analysis

The white light and red light PC crystals were diffracted the X-rays up to 2.7 Å and 1.17 Å resolution, respectively. Both proteins crystals belong to P63 space group with unit cell parameters a = b = 102.35 Å, c = 109.02 Å for red light PC, and a = b = 106.32 Å, c = 58.67 Å for white light PC. Data statistics is shown in Table 4.3. MALDI-ToF analysis confirmed the presence of covalently linked chromophores with the purified PC protein. Two major lines at 18026 Da and 19335 Da were observed in MALDI spectrum (Fig. 4.1A). The first peak corresponds to mass of chromophorylated α-subunit (theoretical mass of 17 344.7 Da deduced from amino acid sequence and PCB-chromophore expected mass of 588.7 Da). The second MALDI-ToF peak matched well with combined mass of β-subunit (18 097.5 Da deduced from sequence including the mass of additional methyl group at βAsn-72 due to post translational modification, also confirmed by electron density) and two covalently linked PCB-chromophores of 588.7 Da each.

In red light PC structure, each asymmetric unit contains two αβ-dimers, whereas that in white light PC contains only one αβ-dimer that is justified by shortening of third (c-axis) axis in unit cell. The phases were reasonably accurate as clear electron density was observed for PCB-chromophores (CYC), not used in search model (Fig. 4.1B & 4.1C). The g-N-methyl asparagine (MEN) residues were built instead of βAsn-72 in the model. The model exhibits good stereochemistry, as evaluated with MOLPROBITY (Chen et al., 2010). All residues were observed in the allowed region of Ramachandran's plot (98% in favoured region and 1.7% in allowed region), except for βThr75 of both the αβ monomers.
Table 4.3. Summary of data-collection and atomic model refinement statistics. Values given in the brackets are for the highest resolution shell

<table>
<thead>
<tr>
<th></th>
<th>Red light PC</th>
<th>White light PC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unit Cell</strong></td>
<td>102.35, 102.35, 109.02 (Å), 90, 90, 120(°)</td>
<td>106.32, 106.32, 58.67 (Å), 90, 90, 120(°)</td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>$P \ 6_3$</td>
<td>$P \ 6_3$</td>
</tr>
<tr>
<td><strong>Solvent content (%)</strong></td>
<td>43.9</td>
<td>52.4</td>
</tr>
<tr>
<td><strong>Resolution limits (Å)</strong></td>
<td>19.35 – 2.7</td>
<td>92.08 – 1.17</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>17802</td>
<td>120881</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>6.1</td>
<td>14.0</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>99.9 (100)</td>
<td>99.9 (100)</td>
</tr>
<tr>
<td><strong>Rmerge</strong></td>
<td>0.132</td>
<td>0.062</td>
</tr>
<tr>
<td><strong>Mean I/ mean σ(I)</strong></td>
<td>3.3</td>
<td>1.5</td>
</tr>
</tbody>
</table>

**Refinement statistics**

| Resolution range (Å)     | 19.35 – 2.7                                      | 49.48 – 1.17                                       |
| Wilson B (Å²)            | 24.5                                             | 11.7                                              |
| Final $R_{work}$/$R_{free}$ | 0.207/0.258                                     | 0.1103/0.1436                                     |
| Non-hydrogen atoms       | 5327                                             | 3224                                              |
| Ramachandran plot        | 98.0/1.7/0.3                                     | 98.5/1.5/0.0                                      |
| RMSD bond lengths (Å)    | 0.011                                            | 0.035                                              |
| RMSD bond angles (°)     | 2.08                                             | 3.38                                               |

Figure 4.1. (A) MALDI-TOF spectra of PC. (B, C) Electron density map fit of a phycocyanobilin (PCB) chromophore ($\beta$CYC-202, covalently attached to $\beta$Cys-153) in ($2F_o-F_c$) map of red light (B) and white light (C) PC (drawn at 1.5σ contour level)

The strained geometry of $\beta$Thr-75 has also been observed in other PC structures (David et al., 2011; Duerring et al., 1991; Sonani et al., 2015). The atomic coordinates of red light PC have been submitted in the protein data bank (PDB ID: 4yjj) and used to analyse structural features, whereas white light PC structure was discussion only at the places where it differs from red light PC.
4.3.2 Structural analysis

The tertiary folds of α and β-subunits of (both red and white light) PC are similar to that of other known PC structures. The superposition of *Phormidium* PC α/β-subunit with PC of *F. diplosiphon* (PDB ID: 1cpc) and *T. vulcanus* (PDB ID: 3o18) yielded RMSD between Cα atoms of 0.45 and 0.55 Å, respectively. In brief, both α and β-subunits adopt very similar helical globin-like fold (Fig. 4.2A) each comprising of eight helices. The RMSD between Cα atoms of α- and β-subunits is 1.9 Å. The two N-terminal helices of each subunit extend away from the rest of the subunit and make significant contacts with the other subunit to form a stable αβ-monomer. These inter-subunit contacts are well conserved in all PC structures. This results in a burial of nearly 6380 Å² (30%) of the solvent accessible area with an estimated gain of nearly 72.6 kcal M⁻¹ of solvation free energy in the formation of αβ-monomer, as calculated from PISA server (Krissinel & Henrick, 2007). Further, analysis of the crystal contacts suggested that the red light PC adopted hexameric [(αβ)₃]₂, whereas white light PC adopted trimeric (αβ)₃ assembly as a crystallographic symmetry (Fig. 4.2B). Three of αβ monomers associate to form a disk like (αβ)₃ trimeric structure in which α-subunits of one monomer interact with the β-subunit of neighboring monomer. This trimeric structure is further extended into the [(αβ)₃]₂ hexamer in red light PC by the face to face interaction of two such trimeric disks, through their α-subunits, such that α-subunits are sandwiched between β-subunits of both the trimmers (Fig. 4.2B). The total buried surface area in formation of hexamer is 62 070 Å² (49%) with an estimated gain of nearly 543.9 kcal M⁻¹ of solvation free energy. There are two such [(αβ)₃]₂ hexamers in the unit cell, placed one over another with only partial overlap such that it cannot further extend to form rod-like assemblies (Fig. 4.2B).

4.3.3 Chromophore conformation

Three PCB-chromophores (CYC), one with α subunit and two with β-subunit, are covalently linked to the αCys-84, βCys-83 and βCys-153, respectively through thioether bonds and are named here as αCYC-201, βCYC-201 and βCYC-202. Each chromophore is comprised of four pyrrole rings (A-D) and was observed to have a curved structure.
Figure 4.2. (A) Cartoon representation of αβ monomer of red light PC. The α-subunit is colored green, while β-subunit is colored cyan. The chromophores are labelled and shown as ball-stick representation. (B) Cartoon presentation of two [(αβ)₃]₂ hexamers observed in the unit cell in the crystal structure of red light PC (left) and two trimers in the unit cell in the crystal structure of white light-PC (right)

The curvatures of these chromophores (between pyrrole rings) were maintained by aspartates αAsp-87, βAsp-85 and βAsp-39, respectively as observed in other PC structures (Duerring et al., 1991) and these Asp residues are strictly conserved in all the PC sequences. The conformations of αCYC-201 and βCYC-201 are very similar in all available PC structures. Their binding site residues are also well conserved. The βCYC-202 chromophore, which resides on the outer surface of the trimeric ring, however is observed to have a stretched conformation where D-ring of the tetrapyrrole chromophore extends away and forms contacts with α-subunit of adjacent αβ monomer of PC. The detailed structural analysis of available PC structures revealed that βCYC-202 chromophore adopts either of the two preferred conformations in PC proteins. We found this to be dependent upon the microenvironment of the chromophore-binding site. The protein microenvironment alone dictates the conformation of phycobilin chromophores has been observed from the structure of a phycoerythrin protein (Kumar et al., 2016). The conf1 is observed in the present PC as well as PC structures from F. diplosiphon (PDB ID: 1epc), S. elongates (PDB ID: 4h0m), Synechocystis sp. PCC 6803 (PDB ID: 4f0t) and A. platensis (PDB ID: 1gh0).
Whereas, the conf2 is observed in *T. vulcanus* (PDB ID: 3o18), *C. caldarium* (PDB ID: 1phn), *G. violaceus* (PDB ID: 2vjr), *P. urceolata* (PDB ID: 1f99) and *G. chilensis* (PDB ID: 2bv8) (Fig. 4.3). Both the conformations differ from each other with a rotation of ~60° around the bond between ring-C and ring-D of the chromophore (marked in Fig. 4.3), in such a way that delocalization of the conjugated double bonds can be expected to be more effective in conf1, the stretched conformation. This could influence the spectral characteristics of different PC proteins. For instance, deviation from co-planarity of A-B rings in phycoerythrobilin has been shown to correlate with the p-coupling that governs the absorption characteristics of the phycoerythrin protein (Gaigalas et al., 2006; Kumar et al., 2016). Peng et al. (2014) have also reported that enhanced planarity in B, C and D ring of PCB-chromophore in allophycocyanin and allophycocyanin-B leads to red-shift in their absorption maxima.

**Figure 4.3.** Conformation of chromophore βCYC-202 attached to βCys-153 of β-subunit. Chromophore and neighboring residues of α-subunit and β-subunit within a distance of 6 Å from the chromophore atoms are displayed in stick representation. The pyrrole rings of the chromophore are labeled as (A), (B), (C) and (D). The chromophore adopts stretched conformation (conf1) in present structure. The hydrogen bonds stabilizing the D-ring conformation of the chromophores are displayed as dashed lines. Also shown here are the chromophores of superposed structures of *T. vulcanus* PC having conf2 conformation (PDB ID, 3o18; pink) and of *F. diplosiphon* PC having conf1 conformation (PDB ID, 1cpc; yellow) as wire model. The βCYC-202 chromophore was observed to adopt stretched conformation (conf1) in all the structures having αPhe-28, αGln-33 and αAsp-145 residues in their α-subunit.
The binding site analyses reveal that residues αPhe-28, αGln-33 and αAsp-145 present on the α-subunit of the *Phormidium* PC play the key role in the formation of stretched conformation of the chromophore. The conf2 is sterically hindered in the *Phormidium* PC due to the presence of bulkier residue αPhe-28 and the D-ring of the chromophore extends out. This conformation is stabilized by hydrogen bonds with αGln-33 and αAsp-145 of the adjacent αβ monomer at the trimer-trimer interface. These hydrogen bonds are formed between the O-atom of the carbonyl group of D-ring and N-atom of side chain amide group of αGln-33; and between N-atom of D-ring of the chromophore and O-atom of carboxylate group of αAsp-145 (Fig. 4.3). The chromophore at this site (bound to βCys-153) has been earlier suggested to play role in alternate pathways for intra and inter-rod energy transfer (Stec et al., 1999).

The initial sequence and structure analysis using PC sequences from Protein Data Bank showed that the sequences having Phe residue at position 28 in the α-subunit, invariably have αGln-33 and αAsp-145 as co-conserved residues. In all of these structures the chromophore at βCys-153 position adopts the conformation observed in present PC structure (conf1). In the phylogenetic analysis, conducted using MEGA6, α-subunit sequences of PC with Phe-28, Gln-33 and Asp-145 residues are clustered on the same branch of phylogenetic tree (Fig. 4.4 & 4.5).

Based on the branch length, it can be inferred that these PC sequences have evolved later with time. For detailed analysis of the co-variance of residues in the chromophore-binding site, we retrieved the PC sequences from the UniProtKB/Swiss-Prot sequence database (reviewed sequences only). The sequences showed very high sequence similarity. Sequences having higher than 85% sequence identity were removed to reduce the redundancy. The multiple sequence alignment revealed that Phe-28, Gln-33 and Asp-145 residues in the α-subunit of the PC are co-conserved and co-evolving in a major sub-group of PC sequences (Fig. 4.6) with exception of *Mastigocladus laminosus* (UniProt ID: P00309) and *Cyanophora paradoxa* (Uniprot ID: P05730). The retrieved sequence of *M. laminosus* is a phycoerythrocyanin (PEC) and besides PCB, it is also chromophorylated with phycoviolobilin chromophore (Duerring et al., 1990). The αβ3 trimers of PEC display distinct absorption maximum at ~576 nm, which is due to combined absorbance of PCB and phycoviolobilin (Parbel et al., 1997). The *C. paradoxa* sequence also shares higher sequence similarity with *M. laminosus* and should be checked for its chromophore preference.
Figure 4.4. Multiple sequence alignment of α-subunit of PC sequences retrieved from Protein Data Bank. The alignment was performed using Clustal omega. The residue positions of Phe-28, Gln-33 and Asp-145 of Phormidium PC-alpha sequence are marked with ‘#’

Such co-conservation of functionally related residues is usually considered as adaptive evolution and allows proteins to acquire specific functional modifications while maintaining its overall structural-functional integrity (Chakrabarti & Panchenko, 2010; Sandler et al., 2014).
Figure 4.5. The phylogenetic tree generated using Neighbour-joining method using MEGA6. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method. The PC sequences with three co-conserved residues (Phe-28, Gln-33 and Asp-145) are clustered on the same branch and are boxed.

To further find out whether these residues are helping the species to adapt in different ecological niches, we tried to find the sequences based on their habitat. It was observed that the PC in low-light adapted (deep sea inhabiting) cyanobacteria, including *Planktothrix agardhii* (GenBank Accession No. 653002532), *Planktothrix rubescens* and *Microcystis aeruginosa* (GenBank Accession No. 488837520), harbour Phe-28, Gln-33 and Asp-145 residues in their α-subunits. It would be interesting to see if the stretched conformation of chromophore with more effective π-delocalization is helping the species to adapt in low-light ecological niches. In conclusion, the crystal structure of red light and white light grown *P. rubidium* PC have been analysed and found similar except crystal packing. The PC is observed to exist as a [(αβ)$_3$]$_2$ hexamer and (αβ)$_3$ trimer in red light and white light PC crystals, respectively. The overall tertiary structure of α- and β-subunits and quaternary fold of the *Phormidium* PC resembles the other known PC structures from cyanobacteria and algae. However, the structural and sequence analyses revealed three co-evolving residue positions that determine the conformation of a PCB chromophore suggested earlier to play role in alternate pathways for intra and inter-rod energy transfer.
Figure 4.6. Multiple sequence alignment of α-subunit of PC sequences retrieved from Uniprot/Swissprot database of reviewed sequences. The sequences are named with their Uniprot ID. The residue positions of Phe-28, Gln-33 and Asp-145 of Phormidium PC-alpha (Phormodium-alpha) sequence are marked with ‘#’. The sequence of *Mastigocladus laminosus* (Uniprot ID, P00309) is marked with ‘*’ and is a PEC with phycoviolobilin chromophore. Also shown are PC-alpha sequences of *Planktothrix agardhii* (GenBank Accession No. 653002522) and *Microcystis aeruginosa* (GenBank Accession No. 488837520) found under deep water; and the β-subunit sequence of *Phormidium* PC (Phormodium-beta). The alignment was performed using Clustal omega.
4.4 References


Marx, A., Adir, N. 2014. Structural characteristics that stabilize or destabilize different assembly levels of phycocyanin by urea. Photosynthesis research, 121(1), 87-93.


