BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Structural and Biochemical Characterization of the Bilin Lyase CpcS from *Thermosynechococcus elongatus*

Christina M. Kronfel,[†] Alexandre P. Kuzin,[‡] Farhad Forouhar,[‡] Avijit Biswas,[†] Min Su,[‡] Scott Lew,[‡] Jayaraman Seetharaman,[‡] Rong Xiao,[§] John K. Everett,[§] Li-Chung Ma,[§] Thomas B. Acton,[§] Gaetano T. Montelione,[§] John F. Hunt,[‡] Corry E. C. Paul,[†] Tierna M. Dragomani,[†] M. Nazim Boutaghou,[∥] Richard B. Cole,[∥,⊥] Christian Riml,[†] Richard M. Alvey,[@] Donald A. Bryant,[@,#] and Wendy M. Schluchter*[,†]

[†]Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, United States

[‡]Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027, United States

[§]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Department of Biochemistry, Robert Wood Johnson Medical School, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States

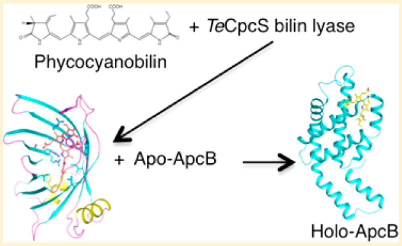
[∥]Department of Chemistry, University of New Orleans, New Orleans, LA 70148, United States

[⊥]Institut Parisien de Chimie Moléculaire, UMR 7201, Université Pierre et Marie Curie (Paris 6), 4 Place Jussieu, 75252 Paris, France

[@]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, United States

[#]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, United States

Supporting Information



ABSTRACT: Cyanobacterial phycobiliproteins have evolved to capture light energy over most of the visible spectrum due to their bilin chromophores, which are linear tetrapyrroles that have been covalently attached by enzymes called bilin lyases. We report here the crystal structure of a bilin lyase of the CpcS family from *Thermosynechococcus elongatus* (*Te*CpcS-III). *Te*CpcS-III is a 10-stranded β barrel with two alpha helices and belongs to the lipocalin structural family. *Te*CpcS-III catalyzes both cognate as well as noncognate bilin attachment to a variety of phycobiliprotein subunits. *Te*CpcS-III ligates phycocyanobilin, phycoerythrobilin, and phytochromobilin to the alpha and beta subunits of allophycocyanin and to the beta subunit of phycocyanin at the Cys82-equivalent position in all cases. The active form of *Te*CpcS-III is a dimer, which is consistent with the structure observed in the crystal. With the use of the UnaG protein and its association with bilirubin as a guide, a model for the association between the native substrate, phycocyanobilin, and *Te*CpcS was produced.

Light harvesting in cyanobacteria is accomplished by phycobilisomes, supramolecular structures principally comprised of phycobiliproteins (PBPs) with covalently attached chromophores.[1] The chromophores, termed bilins, arise from the oxidative cleavage of heme.[2−5] Biliverdin IXα is subsequently reduced by ferredoxin-dependent reductases to form phycocyanobilin (PCB), a blue-colored chromophore present in most cyanobacteria and phycoerythrobilin (PEB), a red-colored chromophore present in some but not all cyanobacteria.[6,7] In plants, biliverdin IXα is reduced by the enzyme HY2 to form phytochromobilin (PΦB),[6] a blue-green bilin which attaches to the light-sensor phytochrome.[8]

Bilins have energy-transfer properties that are greatly enhanced when their conformations are extended and made more rigid by their association with an appropriate PBP.[1,9] Bilins are covalently attached via thioether bonds to conserved cysteine residues on the alpha and beta subunits of the PBPs.[10] A large linker PBP, which serves as the scaffold for phycobilisome core assembly and which anchors the phycobilisome to the thylakoid membrane, is the only protein in the cyanobacterial phycobilisome that can self-ligate its PCB chromophore.[11−13] All other PBPs require bilin lyase enzymes for ligation of their bilin chromophores.[14−20] Bilin lyases are loosely classified according to the substrate PBP subunit (either α or β), the Cys residue to which the bilin is attached, and their amino acid sequences.[17,21] Although their specificity for bilin and PBP substrates appears to be highly specific in situ, heterologous expression studies in *Escherichia coli* have revealed that bilin lyases can ligate diverse bilins to an individual PBP subunit.[13,17,19,22−24]

Received: August 28, 2013
Revised: November 8, 2013
Published: November 9, 2013

dx.doi.org/10.1021/bi401192z | *Biochemistry* 2013, 52, 8663−8676

Table 1. Summary of Plasmids Used in This Study

plasmid name	recombinant proteins produced[a]	parent vector	antibiotic[b]	reference
pApcAB	*Synechococcus* sp. PCC 7002 HT-ApcA and ApcB	pET100	Ap	18
pApcDB	*Synechococcus* sp. PCC 7002 HT-ApcD and ApcB	pET100	Ap	86
pApcF	*Synechococcus* sp. PCC 7002 HT-ApcF	pET100	Ap	86
pApcF/CpcS	*Synechococcus* sp. PCC 7002 HT-ApcF and *Thermosynechococcus elongatus* CpcS	pCDF Duet	Sp	this study
pPcyA	*Synechocystis* sp. PCC 6803 HO1 and *Synechococcus* sp. PCC 7002 HT-PcyA	pACYC Duet	Cm	13
pCpcSU	*Synechococcus* sp. PCC 7002 CpcS-I and CpcU	pCOLA Duet	Km	13
pCpcBA	*Synechocystis* sp. PCC 6803 HT-CpcB and CpcA	pCDF Duet	Sp	13
pPebS	Myovirus HO1 and HT-PebS	pACYC Duet	Cm	87
pPebS2	Myovirus HO1 and HT-PebS	pCOLA Duet	Km	this study
p*Te*CpcS	*Thermosynechococcus elongatus* CpcS	pCOLA Duet	Km	this study
pTER13-21	*Thermosynechococcus elongatus* HT-CpcS	pET21c	Ap	this study
pTER13−30	*Thermosynechococcus elongatus* HT-CpcS	pET30c	Km	this study
pTER13(R151G)	*Thermosynechococcus elongatus* HT-CpcS (R151G)	pTER13−30	Km	this study
pTER13(S155G)	*Thermosynechococcus elongatus* HT-CpcS (S155G)	pTER13−30	Km	this study
pTER13(C2S)	*Thermosynechococcus elongatus* HT-CpcS (C2S)	pTER13−30	Km	this study
pTER13(C169S)	*Thermosynechococcus elongatus* HT-CpcS (C169S)	pTER13−30	Km	this study
pTER13(C2S/C169S)	*Thermosynechococcus elongatus* HT-CpcS (C2S/C169S)	pTER13−30	Km	this study
pHy2	*Synechocystis* sp. PCC 6803 HO1 and *Arabidopsis thaliana* HT-Hy2	pACYC Duet	Cm	24

[a] Proteins that would be produced as fusions are indicated as HT, meaning His tagged. [b] Antibiotic resistance used to select for the presence of the plasmid (Ap: ampicillin, Cm: chloramphenicol, Km: kanamycin, and Sp: spectinomycin).

The first bilin lyase to be described in detail was the CpcE/CpcF lyase of *Synechococcus* sp. PCC 7002. This heterodimeric lyase is representative of the CpcE/CpcF lyase family and attaches PCB to CpcA (α-phycocyanin) at Cys84.[15,20,25−27] PecE/PecF and CpeY/CpeZ, which are paralogs of CpcE/CpcF, are responsible for bilin ligation on the analogous cysteine residues of the α-subunits of other PBP (PecA or α-phycoerythrocyanin and CpeA or α-phycoerythrin, respectively).[28−30] It is postulated that the bilin lyase/isomerases RpcG and MpeZ arose as a fusion of genes encoding a CpcE/CpcF-type lyase,[23,31] and some initial modeling of MpeZ suggests that it adopts a structure that contains primarily α-helices.[31]

Members of the two other lyase families, denoted the CpcT and the CpcS/CpcU families, were also initially identified and characterized from *Synechococcus* sp. PCC 7002.[16,18,32,33] The CpcT bilin lyase attaches PCB at Cys153 of CpcB and PecB,[18,34] and members of this family are probably distantly related to those present in the CpcS/CpcU family. There are members of this CpcS/CpcU family that ligate PCB to CpcB or PecB (at Cys82 position) and to the α and β subunits of allophycocyanin (AP);[13,16,19,33] these members are typically given the designation of "CpcS" or "CpcU". Other members of this group are given the designation "CpeS" or "CpeU" because they are encoded by genes which cluster together in operons with genes that encode phycoerythrin subunits,[33] and some of these CpeS-type lyases have been shown to attach PEB to CpeB (β-phycoerythrin) at Cys80.[28,35] A heterodimer composed of CpcS-I and CpcU targets Cys82 of CpcB (β-phycocyanin) and the equivalent Cys on both subunits of AP.[16] In some cyanobacterial species the S-type lyase is a homodimer,[35] or it functions as a monomer.[17,36]

Here we report the crystal structure of CpcS from *Thermosynechococcus elongatus* BP-1 (*Te*CpcS), which was determined as part of a structural genomics initiative.[37,38] *T. elongatus* does not synthesize PEB and contains only PC and AP as major PBPs.[39] Because of strong sequence similarity among *Te*CpcS, CpcS-I, and CpcU from *Synechococcus* sp. PCC 7002 and CpcS of *Nostoc* sp. PCC 7120, *Te*CpcS was postulated to be a bilin lyase. In accordance with the CyanoLyase database (http://cyanolyase.genouest.org/), both *Te*CpcS and CpcS from the *Nostoc* sp. PCC 7120 belong to the CpcS-III subfamily.[40] Using a heterologous plasmid coexpression system in *E. coli*, we establish here that the *Te*CpcS protein is a bilin lyase. We show that *Te*CpcS is a homodimer, has highly flexible substrate specificity, and can attach any of three bilin substrates, PEB, PCB, and PΦB, to suitable apo-PBP substrates in *E. coli*. *Te*CpcS was able to attach bilin chromophores to Cys82-equivalent position of ApcA, ApcB, and CpcB but was unable to ligate any bilin to CpcA.

■ MATERIALS AND METHODS

Construction of Expression Vectors. Some of the plasmids used in this study have previously been described and are listed in Table 1. The *cpcS* gene from *T. elongatus* (accession number Q8DI91) was amplified by polymerase chain reaction (PCR) using primers TEcpcSF (5′-tccccattagCATATGtgcataggtatggacatccgc-3′, added NdeI site in capital letters) and TEcpcSR (5′-gaaaaaCTCGAGggagttggcgggttgcgtc-3′, added XhoI site in capital letters), digested with NdeI and XhoI, and ligated into similarly digested pCOLADuet-1 (Novagen, Madison, WI). Recombinant *Te*CpcS expressed from this plasmid contains a C-terminal S-tag. For crystallization, mass spectrometry and size-exclusion experiments, the *cpcS* gene was cloned in pET21c to create the clone pTER13-21 with a C-terminal His-tag (described below). For mutagenesis, the *cpcS* gene was also cloned in a similar way in pET30c to create pTER13-30. The pTER13-30 was also used to produce site-specific variants using standard methods as described[28] and using the following primers: *Te*CpcS(R151G): 5′-[Phos]ccccaatttaggtctgcgcacca-3′; *Te*CpcS(S155G): 5′-[Phos]tctgcgcaccggtattctcaagc-3′; *Te*CpcS(C169S): 5′-[Phos]-ggcctccttctcctcggaaattcg-3′; pET30c(XbaI)del: 5′-[Phos]gtgagcggataacaattcccctctacaaataattttg-3′; *Te*CpcS(C2S).F.Nde: 5′-gatataCATATGtccataggtatggacatccgcg-3′; *Te*CpcS(C2S).R.Xho: 5′-ggtgCTCGAGggagttggcgggttg-3′. The five plasmids produced were pTER13(R151G), pTER13(S155G), pTER13(C2S), pTER13(C169S), and pTER13(C2S/C169S) (see Table 1). All constructs were verified by standard DNA sequence analyses.

Protein Expression and Purification for Crystallization. The production of *Te*CpcS-III protein was carried out as part of the high-throughput protein production process of the Northeast Structural Genomics Consortium (NESG).[41,42] *Te*CpcS-III corresponds to NESG Target Ter13. The full-length *cpcS-III* (*ycf58*) gene from *T. elongatus* (strain BP-1) was cloned into a pET21 (Novagen) derivative, generating plasmid pTER13-21. The resulting recombinant protein contains eight non-native residues (LEHHHHHH) at the C-terminus. *E. coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pTER13-21. A single isolate was cultured in MJ9 minimal media[43] supplemented with selenomethionine, lysine, phenylalanine, threonine, isoleucine, leucine, and valine for the production of selenomethionine-labeled *Te*CpcS-III.[44] Initial growth was carried out at 37 °C until the $OD_{600\ nm}$ of the culture reached 0.6−0.8. The incubation temperature was then decreased to 17 °C, and protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Following overnight incubation, the cells were harvested by centrifugation.

Selenomethionyl *Te*CpcS-III was purified by standard methods. Cell pellets were resuspended in lysis buffer [50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol] and disrupted by sonication. The resulting lysate was clarified by centrifugation at 26000*g* at 4 °C for 45 min. The supernatant was loaded onto a nickel nitrolotriacetic acid (Ni-NTA) column (Qiagen, Inc., Chatsworth, CA) and eluted in lysis buffer containing 250 mM imidazole. Fractions containing partially purified *Te*CpcS-III were pooled and buffer conditions providing monodisperse samples were optimized by analytical gel filtration detected by static light scattering, following the protocol described elsewhere.[41,42] Preparative gel filtration (Superdex 75, GE Healthcare, Piscataway, NJ) was then performed using a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM dithiothreitol (DTT), and 0.02% NaN_3. The purified *Te*CpcS-III protein was concentrated to 7.6 mg/mL, flash frozen in aliquots, and used for crystallization screening. Sample purity (>97%) and molecular mass (21.914 kDa) were verified by SDS-PAGE and MALDI-TOF mass spectrometry, respectively. The yield of purified protein was approximately 5 mg/L.

Protein Crystallization. The *Te*CpcS-III free enzyme was crystallized at 20 °C by the hanging-drop vapor diffusion method. For the free enzyme crystals, aliquots (2 μL) of protein solution containing *Te*CpcS-III (7.6 mg/mL in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM DTT, 0.02% NaN_3) were mixed with an equal volume of the reservoir solution consisting of 0.1 M *N*-cyclohexyl-3-aminopropanesulfonic acid, (CAPS, pH 10.0), 18% polyethylene glycol 20000, and 0.1 M NH_4H_2PO4. The crystals were cryo-protected by transferring them to a crystallization solution that was supplemented by 20% (v/v) glycerol. The crystals were flash-frozen in liquid propane for data collection at 100 K.

The free enzyme crystals belong to space group $P4_12_12$, with cell parameters of $a = b = 75.01$ Å and $c = 83.36$ Å. There is one protomer of *Te*CpcS-III in the crystallographic asymmetric unit.

Data Collection and Structure Determination. A single-wavelength anomalous diffraction data set to 2.8 Å resolution was collected on a single crystal of the *Te*CpcS-III free enzyme at the X4A beamline of the National Synchrotron Light Source. The diffraction images were processed with the HKL package.[45] The data processing statistics are summarized in Table 2.

Table 2. X-ray Data-Quality and Structure-Refinement Statistics[a]

data collection	
space group	$P4_12_12$
cell dimensions	
a, b, c (Å)	75.014, 75.014, 83.359
α, β, γ (°)	90, 90, 90
wavelength (Å)	0.979
resolution range (Å)	50.0−2.80 (2.90−2.80)[b]
R_{sym} or R_{merge}	0.081 or 0.059 (0.691 or 0.634)
total reflections	362419
observed reflections	11252
$I/\sigma I$	28.2 (4.6)
completeness (%)	99.9 (100.0)
redundancy	10.1 (9.7)
refinement	
resolution (Å)	20.0−2.80
no. reflections	10,246
R_{work}/R_{free}	0.236/0.262 (0.312/0.395)
no. atoms	1231
protein	1225 (subunit A: 4−76, 84−107, 118−176)
ligand	5
water	1
B factors (Å^2)	
protein	66.5
ligand/ion	97
water	56.8
r.m.s. deviations	
bond lengths (Å)	0.008
bond angles (deg)	1.40
Ramachandran distribution (%)	
favored	86.6
allowed	13.4
generously allowed	0.0
disallowed	0.0
PDB id	3BDR

[a]Standard parameter definitions were used.[88] [b]Highest-resolution shell is shown in parentheses. Statistics belong to one crystal.

The selenium sites were located with the program SnB.[46] SOLVE/RESOLVE [47] was used for phase calculation, phase improvement, and automated model building, but only about 8% of the residues were placed. The complete atomic models were built with XtalView,[48] followed by structure refinement using CNS (Crystallography and NMR System).[49] The refinement statistics are summarized in Table 2.

Structure Analysis. Visualization and comparison of protein structures and manual docking of ligand molecule were performed using PyMol.[50] XtalView [48] and CNS [49] were used for the protein docking exercises and minimization of the steric clashes between the ligand and the protein. The overlaid structures of *Te*CpcS and UnaG bound to bilirubin (BR) reveal that BR resides nicely inside the barrel of *Te*CpcS. We also realized that the structure of UnaG provides insight into the conformation of two unstructured loops (residues 77−83 and 108−117) in the structure of *Te*CpcS. We therefore used the structure of UnaG (PDB id: 4I3B) as a template for modeling PCB and the two missing loops for *Te*CpcS. Subsequent to the manual building of the two missing loops and docking of the PCB in *Te*CpcS, the resulting model was subject to energy minimization using CNS.

Expression and Purification of Recombinant Proteins for Enzyme Analyses. Plasmids were cotransformed into *E. coli* BL21 (DE3), and cells were grown on Luria−Bertani (LB) medium containing the appropriate antibiotics for selection as listed in Table 1 at the following concentrations: ampicillin, 100 μg mL^{-1}; chloramphenicol, 34 μg mL^{-1}; kanamycin, 50 μg mL^{-1}; and/or spectinomycin, 100 μg mL^{-1}. Isolated colonies were used to inoculate cultures (50 mL of LB medium amended with necessary antibiotics combinations at the aforementioned concentrations), which were grown at 37 °C with shaking at 225 rpm. This starter culture was used to inoculate 1 L of LB medium containing antibiotics for incubation until the culture reached an optical density $OD_{600\ nm}$ = 0.6 (approximately 4 h). The temperature was lowered to 30 °C before induction of T7 RNA polymerase by addition of 1 mM IPTG, and the culture was incubated with shaking for an additional 3 h. Cells were collected by centrifugation at 10000*g* for 10 min and stored at −20 °C until required. For PEB and PΦB production using pPebS and pHy2 expression plasmids, respectively, cultures were induced at 18 °C with 1 mM IPTG and incubated for 16 h prior to harvesting cells by centrifugation.

Frozen cell pellets were thawed and resuspended in a buffer of 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 at 2.5 mL buffer per gram wet-weight of cells, and the resulting cell suspension was lysed by three passages through a chilled French pressure cell at 138 MPa. Inclusion bodies, cell debris, and unbroken cells were removed by centrifugation for 20 min at 13000*g*. For purification of hexa-histidine-tagged proteins, the supernatant containing soluble proteins was applied to a Ni-NTA superflow affinity column (Qiagen, Inc., Chatsworth, CA), as previously described.[18] After elution of the desired protein with imidazole, the purified protein fraction was dialyzed overnight in imidazole-free suspension buffer containing 10 mM 2-mercaptoethanol.

Fluorescence Emission and Absorbance Spectra. Absorbance spectra were acquired using a lambda 35, dual-beam UV/vis spectrophotometer, and fluorescence emission and excitation spectra were recorded with a Perkin-Elmer LS55 fluorescence spectrophotometer (Waltham, MA), as described.[13] The excitation wavelength was 490 nm for recombinant proteins carrying PEB chromophores and 590 nm for those carrying PCB or PΦB chromophores.

Protein and Bilin Analysis. Polypeptides were resolved by polyacrylamide gel electrophoresis (PAGE; 15% w/v) in the presence of sodium dodecyl sulfate (SDS), and polypeptides were visualized by staining with Coomassie blue, as described.[18] To detect proteins containing bound chromophores (PEB, PCB, or PΦB), gels were soaked in 10 mM $ZnSO_4$.[51] The resulting enhanced fluorescence produced by chelation of the bilin by Zn was visualized using an FX imaging system (BioRad, Hercules, CA). Excitation at 532 nm was used to detect all three bilins, but excitation at 635 nm was used to detect PCB and PΦB.

Size-Exclusion Chromatography (SEC). SEC was performed using the protocol described earlier.[16] The molecular mass of native *Te*CpcS was calculated from a standard curve derived from a set of molecular mass standards (Bio-Rad).

Mass Spectrometric Analysis. Purified *Te*CpcS was digested with trypsin, and the resulting tryptic peptides were analyzed using MALDI mass spectrometer (MS) and tandem MALDI MS/MS on a 4800 MALDI-ToF/ToF (AB Sciex, Concord, Ontario), following the procedures described previously.[28]

■ RESULTS

Structural Analysis of *Te*CpcS. The X-ray crystal structure of CpcS-III (hereafter *Te*CpcS) from the *T. elongatus* strain BP-1 (tll1699) was determined to 2.8 Å resolution; the coordinates and the structure factors were released by the protein data bank (PDB id: 3BDR).[38] The structure reveals that *Te*CpcS contains two α-helices (Figure 1) and one 10-stranded, antiparallel β-sheet that adopts a beta-barrel fold, which is capped by α1 at one side and is widely open at the other side, presumably for the proper delivery of the substrate to the target phycobiliprotein.

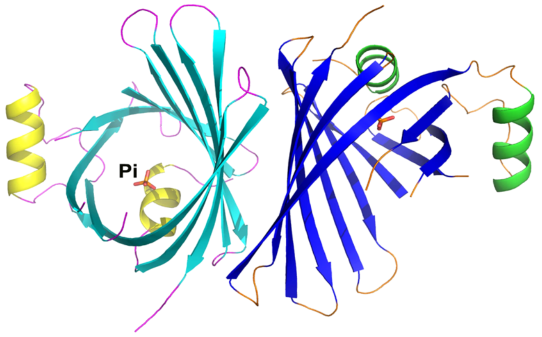


Figure 1. Structure of *Te*CpcS from *T. elongatus* BP-1 (PDB id: 3BDR). The crystal structure of the homodimer *Te*CpcS. The two α-helices, 10 β-strands, and the associated loops of each subunit are depicted in yellow/cyan/magenta for subunit A and in green/blue/orange for subunit B, respectively. A phosphate ion cocrystallized with each subunit is shown as a ball-and-stick model.

*Te*CpcS is found as a homodimer in the crystal, and it belongs to the lipocalin structural family (Figure 1). Various lipocalins exhibit different oligomeric states and occur as monomers, homodimers, heterodimers, or tetramers. Other members of this protein family also bind ligands, including fatty acids, retinol, carotenoids, pheromones, prostaglandins, biliverdin, and bilirubin.[52−62] In the crystal structure, a phosphate ion forms a hydrogen bond with the side chain of invariant Arg-151 located near the bottom of the funnellike cavity. This interaction suggests that Arg151 is possibly involved in substrate recognition.

The structure of *Te*CpcS was originally undertaken as part of a large-scale effort of the NIH Protein Structure Initiative to provide structural representatives from large domain families for which no structures were yet available. The aim of this program was to provide structural templates which could be leveraged by large-scale homology modeling.[37,63] *Te*CpcS is the first structure of a protein from PFAM family PF09367, which includes more than 209 proteins and protein domains. This number is conservative because PFAM only considers select proteomes. A Hidden Markov model (HMM) analysis of the 21.4 M distinct UniProt sequence using the PF09367 HMM provided by PFAM shows that there are 288 distinct sequences that have this domain signature. On the basis of the Uniprot version 2013_04 (with ∼21.4 M unique sequences), this structure has a total modeling-level coverage (E val 1e^{-10}) of 317 proteins and a novel modeling-level coverage [defined by ref 64] of 317 proteins. Hence, the *Te*CpcS structure provides an important template for structural studies of this biologically important domain family, both by homology modeling and X-ray crystallography using molecular replacement.

In accordance with an analysis using DALI,[65] *Te*CpcS now has numerous structural homologues in the PDB, the top 55 of which have Z-scores ranging from 12.4 to 10.0. Most of these proteins are fatty-acid-binding proteins that are functionally unrelated to *Te*CpcS. In contrast, the closest structural homologue that is functionally related to *Te*CpcS is the fluorescent protein UnaG [62] (PDB id: 4I3B, Z-score 10.1, RMSD 2.8 over 109 residues). UnaG is a fluorescent protein recently discovered in freshwater eel (unagi) muscle that becomes fluorescent via noncovalent but tight binding to unconjugated bilirubin. This protein is hypothesized to aid in muscle metabolism of juvenile eels, which must travel long distances and may benefit from the known antioxidant effects of bilirubin.[62] Interestingly, when the structure of UnaG bound to a bilirubin is overlaid onto the structure of *Te*CpcS, the substrate can readily be docked into the cavity of *Te*CpcS. This alignment only produced negligible clashes with the side chains of a few residues that are inside the funnel. This remarkable fit is the consequence of both proteins having similar β-barrel topologies, although they only share 11% sequence identity. More importantly, the structure of UnaG provides insight into the conformation of two unstructured loops (residues 77−83 and 108−117) in the structure of *Te*CpcS. The structure of UnaG suggests that these two loops become ordered only in the presence of the substrate because they interact with the substrate while partially capping the funnel. We therefore used the structure of UnaG (PDB id: 4I3B) as a template for modeling PCB and the two missing loops for *Te*CpcS (Figure 2A).

With the use of the structure of UnaG and its interaction with bilirubin as a guide, manual docking of PCB to *Te*CpcS was performed (Figure 2A) by XtalView.[48] We performed the docking exercises by taking advantage of two important factors: (1) most bilins are negatively charged at neutral pH and (2) a significant number of residues lying inside the barrel are highly conserved and predominantly positively charged (see Figure 3). Once a preliminary model was achieved, a refinement step using CNS was performed[49] to minimize possible polar and nonpolar clashes between the protein and the substrate. The manual docking of UnaG in complex with bilirubin (BR) (PDB id: 4I3B) onto the structure of *Te*CpcS reveals that BR fits well in the cavity of *Te*CpcS (Figure 2B). More importantly, the overlaid structures show that several residues that recognize the substrate BR in UnaG have close counterparts in *Te*CpcS, namely Arg112 and Arg132, in UnaG are closely aligned to Arg22 and Arg153 in *Te*CpcS, respectively. Additionally, there are several hydrophobic residues that interact with BR in UnaG, which have representatives in *Te*CpcS. We therefore took advantage of these striking similarities between the two proteins and modeled PCB based on what we observed in the overlaid structures.

As seen in Figure 2A, the D ring of the model substrate is buried at the bottom of the funnel, where its carbonyl group is hydrogen bonded to the side chain of invariant Arg151. The position of this carbonyl group is similar to that of the phosphate ion in the crystal structure of the substrate-free *Te*CpcS. While the hydrophobic moieties of PCB interact with side chains of several conserved hydrophobic amino acids, namely Phe10, Phe11, Ile41, Leu90, and Leu130, the propionate group on the C ring of PCB forms hydrogen bonds with invariant Trp73 and Ser155. Interestingly, the modeling of two missing loops suggests that the invariant

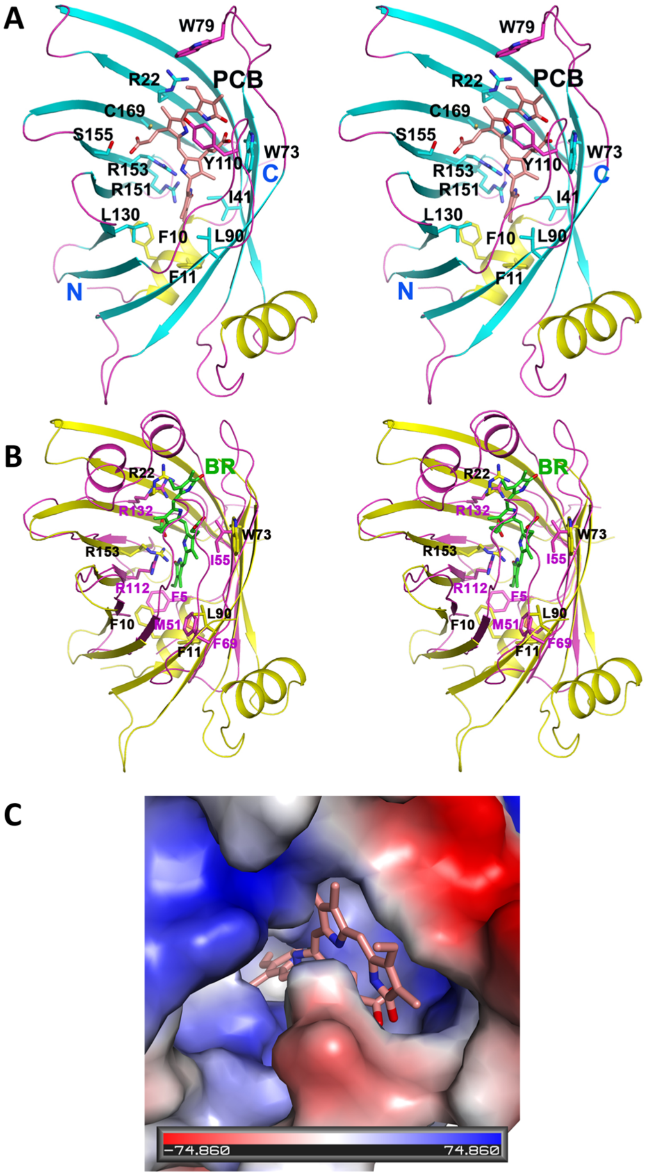


Figure 2. Structures of *Te*CpcS modeled with PCB. (A) Stereo pair view of a *Te*CpcS protomer with PCB modeled in the β-barrel. The modeled PCB and the side chains of strictly conserved and conservatively substituted residues are shown in ball-and-stick models. The two loops comprising residues 77−83 and 108−117, which are disordered in the *Te*CpcS structure are also modeled. The N- and C-termini of *Te*CpcS are labeled. (B) Stereo pair view of the structural superposition of UnaG bound to bilirubin (PDB id: 4I3B; in magenta) and *Te*CpcS (in yellow). The side chains of identical and similar residues, and the bilirubin (BR) bound to UnaG are depicted as stick models and are labeled. (C) View of a PCB model docked into the cavity of *Te*CpcS, which is shown with a surface-charged map representation. The stick model of the docked PCB is shown in pink oriented such that the ethylidene group on ring A is close to the surface and ring D is buried in the cavity/barrel, while the *Te*CpcS surface is depicted in color with gray/white representing neutral/hydrophobic amino acids, blue representing positively charged amino acids, and red representing negatively charged amino acids.

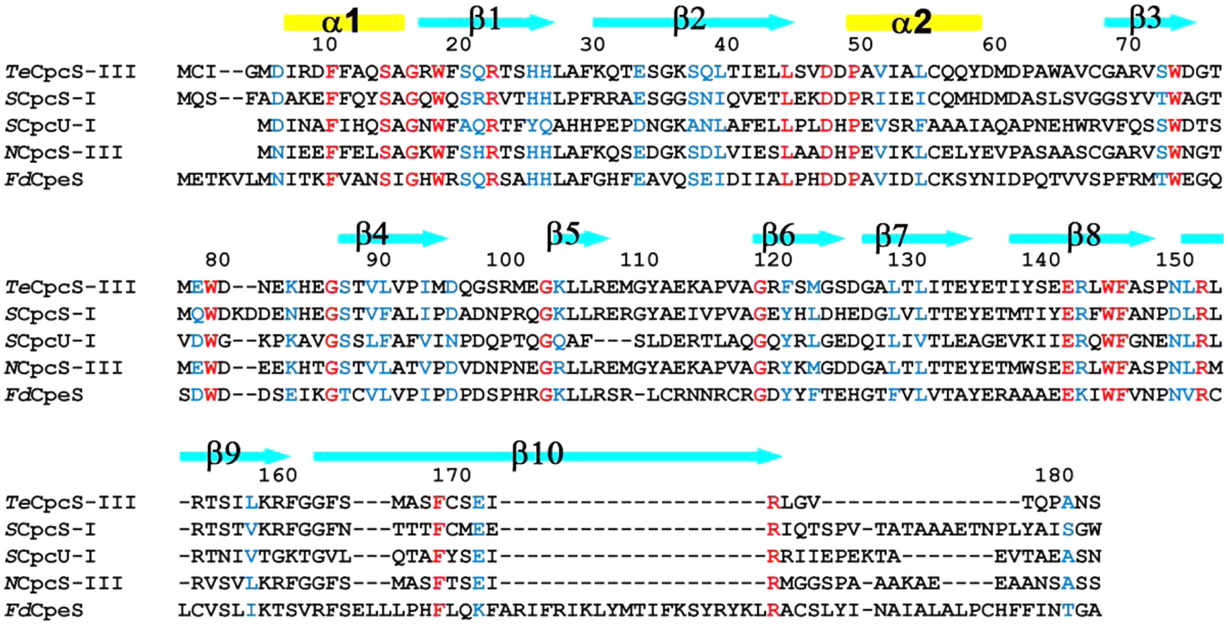


Figure 3. Sequence alignment of representative CpcS-type proteins performed using T-COFFEE.[89] For each sequence, the organism name appears as italic letters before the protein: *T. elongatus* BP-1 (*Te*), *Synechococcus* sp. PCC 7002 (*S*), *Nostoc* sp. PCC 7120 (*N*), and *Fremyella diplosiphon* (*Fd*). Identical and conservatively replaced residues in all sequences are shown in red and blue, respectively. Secondary structural elements, observed in the crystal structure of *Te*CpcS, are shown above the alignment with α-helices and β-strands represented by rectangles and arrows, respectively.

Trp79 possibly caps the A ring and Tyr110 forms $\pi-\pi$ interactions with the A ring. The structure of the PCB model docked into the cavity of *Te*CpcS is shown with the surface charge representation in Figure 2C. The ethylidene group on ring A of PCB is exposed and projects into the solvent, and there is a charge distribution difference along the surface of *Te*CpcS which may help to explain the substrate preference for CpcB, ApcA, ApcB, ApcD, and ApcF, while excluding CpcA. Although these substrates are very similar at the structural level, one difference in charge near the chromophore attachment site may allow one subunit to bind, while precluding CpcA from binding to *Te*CpcS. In all of the substrates for *Te*CpcS, the PCB binding motif CxRDx is followed by a negatively charged amino acid, either E or D, whereas in CpcA, the CxRDx motif is followed by an uncharged G residue. A positively charged region containing K or R residues precedes the CxRDx motif in all phycobiliprotein subunits.

*Te*CpcS is 182 amino acids in length and is most related in sequence to the CpcS-III lyase from *Nostoc* sp. PCC 7120 (64.1% identity, 74% similarity) and to CpcS-I (45.5% identity, 60.6% similarity) and CpcU (28.9% identity; 44.7% similarity) from *Synechococcus* sp. PCC 7002; it is also similar in sequence to other S-type bilin lyases which attach PEB to phycoerythrins from *Prochlorococcus* spp. and *Fremyella diplosiphon* (see Table 1 of the Supporting Information). CpcS-I and CpcU form a heterodimeric bilin lyase that can ligate PCB to Cys82 (or the equivalent position) on four types of AP subunits (ApcA, ApcB, ApcD, and ApcF) and CpcB.[13,16] When *Te*CpcS was aligned with the sequences from *Synechococcus* sp. PCC 7002 CpcS-I and CpcU, the sequences were highly conserved across the entire sequence until the extreme C-terminus (Figure 3). Although the gene encoding *Te*CpcS is the only open reading frame that shows similarity to a CpcS-type bilin lyase in the *T. elongatus* genome, its functionality as a bilin lyase had not been demonstrated.

Analysis of Lyase Activity of *Te*CpcS with CpcB. The bilin lyase activity of *Te*CpcS toward various apo-PBP subunits was tested using a previously developed heterologous coexpression system in *E. coli*.[13] Recombinant histidine-tagged proteins were purified using Ni-NTA affinity chromatography. The purified PBPs obtained from the *E. coli* cells coproducing CpcB and CpcA subunits (CpcBA), *Te*CpcS, and enzymes to synthesize PCB from heme were analyzed by absorbance and fluorescence spectroscopy. Figure 4A shows the absorbance spectrum (solid blue line, absorbance maximum at 620 nm) and fluorescence emission spectrum (dotted blue line, emission maximum at 644 nm) of PCB ligated to CpcB by *Te*CpcS (see Table 3). The black solid and dotted lines correspond to absorbance and fluorescence spectra, respectively, for purified CpcB/CpcA heterodimers obtained from cells producing PCB but in the absence of *Te*CpcS. The activity of the *Te*CpcS lyase was very similar to that of the CpcS-I/CpcU lyase, and red solid and dotted lines correspond to absorbance and fluorescence spectra, respectively, of CpcB chromophorylated at Cys82 by CpcS-I/CpcU (Figure 4A; Table 3). The CpcB protein contains two PCB attachment sites, but the spectral properties of each ligated PCB are distinct: PCB ligated at Cys153 has maximal absorbance at 592 nm and maximal fluorescence emission at 624 nm,[13,18] whereas PCB ligated at Cys82 has an absorbance maximum at 621 nm and maximal fluorescence emission at 644 nm[13,16] (see Table 3). Therefore, we conclude that *Te*CpcS is a PCB lyase with the same specificity as CpcS-I/CpcU.

To exclude the possibility that *Te*CpcS was attaching PCB to Cys84 of CpcA, the purified, PCB-containing CpcB/CpcA complex was resolved by SDS–PAGE (Figure 4B). Covalent bilin addition to proteins was verified by fluorescence emission of the same gel after incubation with Zn sulfate, as shown in Figure 4C. The strong Zn-enhanced fluorescence emission observed for CpcB in lane 3 indicates that *Te*CpcS ligated PCB to CpcB but not to CpcA. A similar result was obtained when

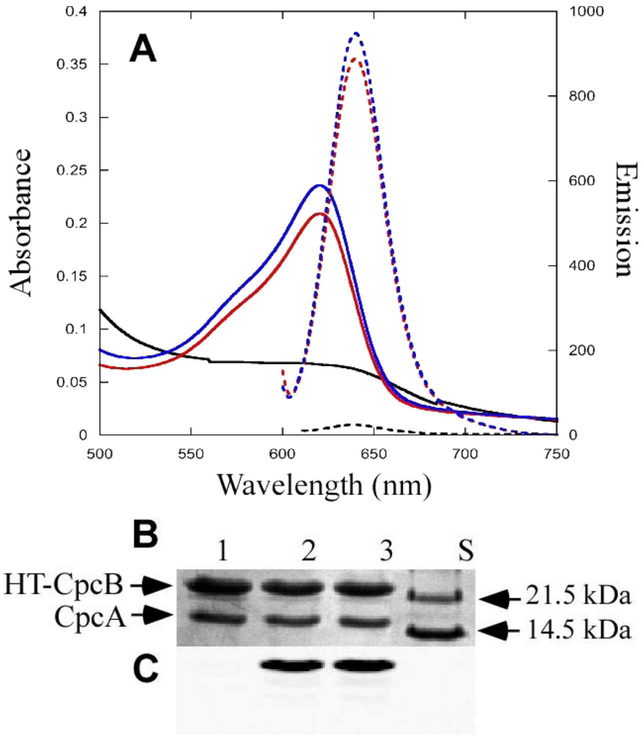


Figure 4. Chromophorylation of CpcB by *Te*CpcS versus CpcS/CpcU. (A) Absorbance (solid) and fluorescence emission (dashed) spectra of HT-CpcBA purified from recombinant *E. coli* cells containing pCpcBA with pPcyA and expressing either *Te*CpcS (blue), CpcSU (red), or no lyase (black). (B) Coomassie-stained SDS−polyacrylamide gel of HT-CpcBA purified from cells containing pCpcBA, pPcyA, and expressing either CpcSU (lane 2), *Te*CpcS (lane 3), or no lyase (lane 1). Molecular mass standards were loaded in lane S. (C) Zn-enhanced bilin fluorescence (excitation at 635 nm) of the gel in panel B.

Table 3. Spectral Properties for PC and AP Subunits Chromophorylated with PCB, PEB, or PΦB

holo recombinant PBPs (plasmid present)	λ_{max} (nm) ($Q_{vis/UV}$)	fluorescence emission, λ_{max} (nm)
HT-CpcB (pCpcBA + pPcyA)[a]	620/394 (4.3)	644
HT-CpcB (pCpcBA + pPcyA)[b]	621/393 (4.7)	644
HT-CpcB (pCpcBA + pPebS)[a]	557/372 (4.95)	573
HT-CpcB (pCpcBA + pHY2)[a]	635/347 (2.25)	647
HT-ApcA/ApcB (pApcAB + pPcyA)[a]	614/392 (5.3)	632
HT-ApcA/ApcB (pApcAB + pPebS)[a]	560/376 (8.3)	571
HT-ApcA/ApcB (pApcAB + pHY2)[a]	629/391 (4.8)	648
HT-ApcD (pApcDB + pPcyA)[a]	615, 672[c]/ 370 (1.2)	635, 675
HT-ApcD (pApcDB + pPebS)[a]	572/371 (2.8)	571
HT-ApcD (pApcDB + pHY2)[a]	629/391 (2.4)	630
HT-ApcF (pApcF + pPcyA)[a]	615/393 (4.2)	632
HT-ApcF (pApcF + pPebS)[a]	560/376 (8.1)	572
HT-ApcF (pApcF + pHY2)[a]	630/387 (3.5)	648

[a]The construct was coexpressed with p*Te*CpcS. [b]The construct was coexpressed with pCpcSU. [c]Indicates a second peak.

the lyase was CpcS-I/CpcU (Figure 4C, lane 2). In the absence of any lyase, no Zn-enhanced fluorescence emission was detected (Figure 4C, lane 1). There was no bilin addition to CpcA in any of the lanes shown in Figure 4C. From these experiments, it can be concluded that *Te*CpcS acts as an S-type lyase that ligates PCB to CpcB at Cys82.

From the standpoint of potential biotechnological applications, it was also interesting to test the *Te*CpcS lyase activity with the noncognate bilin substrates, PEB and PΦB. CpcB/CpcA and *Te*CpcS were produced in cells harboring pPebS or pHy2, which direct the production of PEB and PΦB from heme, respectively. As shown in Figure 1A of the Supporting Information, *Te*CpcS could attach PEB (solid red lines for absorbance, dotted red lines for fluorescence emission) and PΦB (solid green lines for absorbance and dotted green lines for fluorescence) to CpcB to produce highly fluorescent products with interesting spectral properties (see Table 3). These purified proteins were resolved by SDS−PAGE and stained with Coomassie Blue (Figure 1B of the Supporting Information). The bilin content of CpcA and CpcB was examined by Zn-enhanced fluorescence of the same gel with excitation at 532 nm (Figure 1C of the Supporting Information) or at 635 nm (Figure 1D of the Supporting Information), which detects PEB and PΦB, respectively. The results indicate that CpcB carries covalently bound PEB (lane 1) and PΦB (lane 2) when CpcB and CpcA are coproduced with *Te*CpcS in the presence of the appropriate bilin, but no bilin ligation was detected to CpcA in either case. These experiments established that *Te*CpcS is a versatile bilin lyase that is capable of attaching both cognate and noncognate bilins to CpcB.

Analyzing Activity of *Te*CpcS Lyase on Major Allophycocyanin Subunits ApcA/ApcB. The heterodimeric CpcS-I/CpcU lyase can chromophorylate four AP subunits at Cys81,[16,19] so the *Te*CpcS was tested for its ability to chromophorylate ApcA and ApcB. ApcA and ApcB were coproduced with *Te*CpcS in cells that were also producing enzymes to synthesize PCB, PEB, and PΦB (see Table 1). The expressed proteins were purified using Ni-NTA affinity chromatography and analyzed by absorbance and fluorescence emission spectroscopy, and the data are shown in Figure 2 of the Supporting Information. The solid lines show the absorbance spectra, and the dotted lines show the fluorescence emission spectra of the purified ApcA and ApcB produced in the presence of PCB (Figure 2A of the Supporting Information), PEB (Figure 2B of the Supporting Information), or PΦB (Figure 2C of the Supporting Information). The spectra in Figure 2A of the Supporting Information are consistent with the correct addition of PCB as established in published studies.[16,19] No significant bilin ligation occurred in the absence of the *Te*CpcS lyase (data not shown). The strong absorbance and fluorescence emission of purified ApcA/ApcB complexes coproduced with PEB (Figure 2B of the Supporting Information) and PΦB (Figure 2C of the Supporting Information) was strong evidence that these bilins were also attached to Cys81 residues. The purified proteins were separated on SDS−PAGE and sequentially stained with Zn sulfate and Coomassie blue (Figure 2D of the Supporting Information), which verified that both ApcA and ApcB subunits were present in equal amounts. In addition, as shown in Figure 2E of the Supporting Information, both ApcA and ApcB subunits carried covalently bound, fluorescent PCB (lane 1), PEB (lane 2), or PΦB (lane 3). These data establish that *Te*CpcS can ligate PCB, PEB, and PΦB to both ApcA and ApcB.

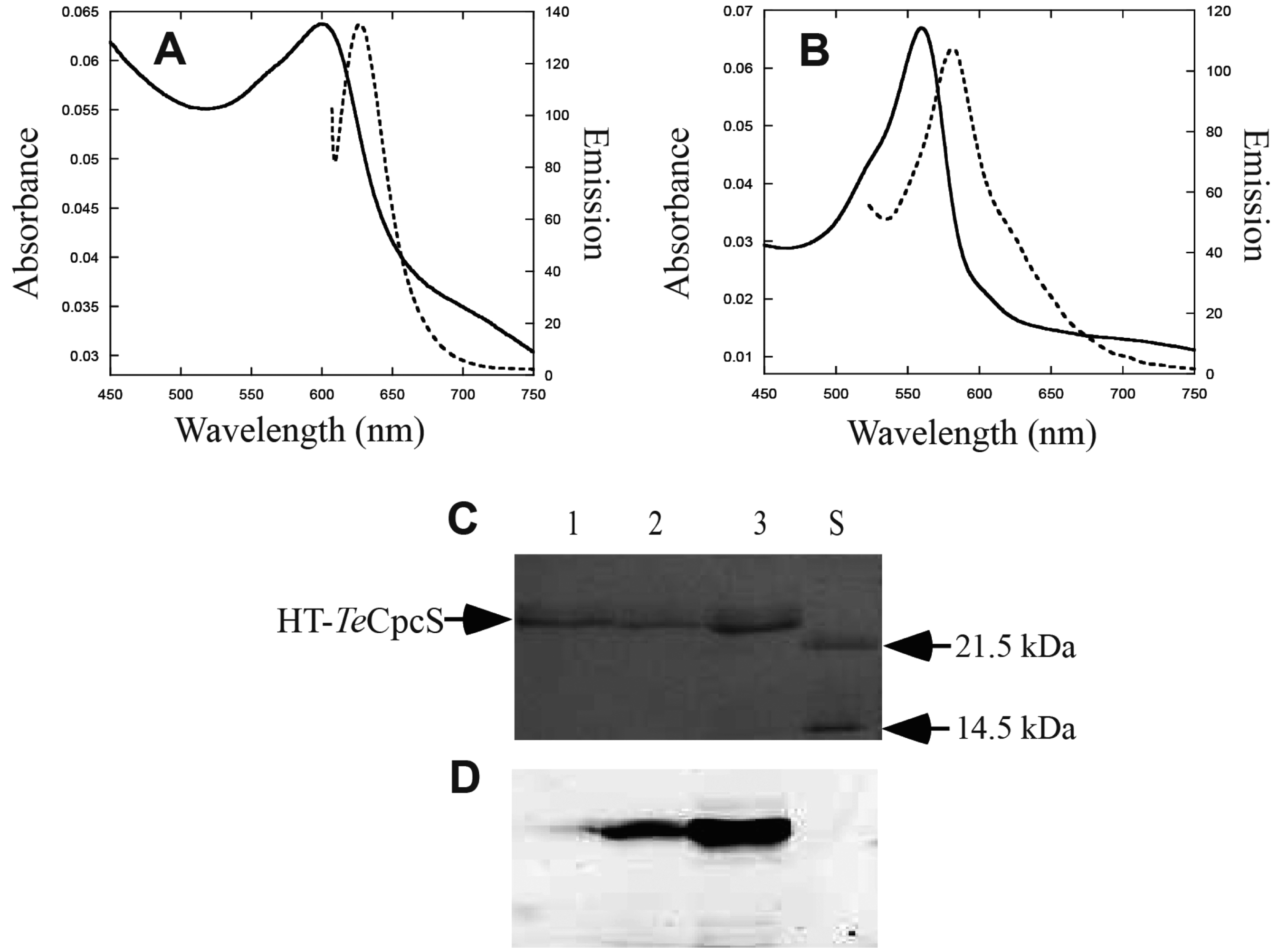


Figure 5. Bilin binding to the lyase *Te*CpcS. (A) Absorbance (solid) and fluorescence emission (dashed) spectra of HT-*Te*CpcS purified from recombinant *E. coli* cells containing pTER13-21 and pPcyA. (B) Absorbance (solid) and fluorescence emission (dashed) spectra of purified HT-*Te*CpcS from cells purified from recombinant *E. coli* containing pTER13-21and pPebS. (C) Coomassie-stained SDS–polyacrylamide gel of HT-*Te*CpcS purified from cells containing pTER13-21 alone (lane 1), pTER13-21 and pPcyA (lane 2), or pTER13-21 and pPebS (lane 3). Molecular mass standards were loaded in lane S. (D) Zn-enhanced bilin fluorescence of the gel (excitation at 532 nm) in panel C.

TeCpcS Activity on AP α-Like Subunit ApcD. AP-B is an important terminal emitter of the PBS that transfers energy to photosystem I and is composed of ApcB along with ApcD, a variant of the AP alpha subunit ($\alpha^{\text{AP-B}}$).[66−69] CpcS-I/CpcU lyases ligate PCB to apo-ApcD when coproduced with PCB in *E. coli*. When ApcD is coproduced with ApcB, the solubility of ApcD is improved, and energy transfer from the bilin on ApcB to that on ApcD can be observed in the recombinant protein.[13]

To test the bilin lyase activity of *Te*CpcS on ApcD, *Te*CpcS was coexpressed with ApcD/ApcB in the presence of one of the three bilins (PCB, PEB, or PΦB). These resulting cells were intensely colored (data not shown), which suggested that covalent bilin ligation had occurred. The His-tagged PBPs were purified from cells using Ni-NTA affinity chromatography and characterized as described above using absorbance and fluorescence emission spectroscopy. The absorbance (solid lines) and fluorescence emission (dashed lines) in Figure 3 (panels A, B, and C) of the Supporting Information show the results for ApcD/ApcB ligated with three different bilins: PCB (Figure 3A of the Supporting Information), PEB (Figure 3B of the Supporting Information), and PΦB (Figure 3C of the Supporting Information) (see Table 3). Interestingly, only when PCB was ligated to these proteins did the longer wavelength absorption band characteristic of native ApcD appear. To show that the chromophores were covalently bound to both ApcD and ApcB, the purified proteins were separated on SDS–PAGE and stained with Zn sulfate (Figure 3E, lanes 1–3, of the Supporting Information) and subsequently with Coomassie blue (Figure 3D of the Supporting Information). Zn-enhanced fluorescence showed that bilin chromophores were covalently attached to both ApcB and ApcD. Therefore, *Te*CpcS can attach PCB, PEB, and PΦB to ApcD.

TeCpcS Activity on ApcF. ApcF is a variant type of the AP β subunit (also known as β^{18}) that partners with the amino-terminal domain of ApcE (the core membrane linker PBP subunit), the second terminal emitter of PBS. In *Synechococcus* sp. strain PCC 7002, the loss of ApcF affects the energy transfer from PBS to the photosystem II.[67,69−71] A previous study showed that the heterodimeric CpcS-I/CpcU lyase can ligate PCB on ApcF.[13] *Te*CpcS was coproduced with ApcF together with enzymes to produce three bilins (PCB, PEB, or PΦB and a combination of both PCB and PEB). The resulting cells were intensely colored (data not shown), which suggested that efficient ligation of chromophores to ApcF had occurred. The proteins were purified using Ni-NTA affinity chromatography and analyzed. In Figure 4A of the Supporting Information, the solid lines represent absorbance and dotted lines represent the fluorescence emission spectra of ApcF carrying PCB (blue) or PEB (red; Table 3). Next, a competition experiment with *Te*CpcS was performed in which cells coexpressed ApcF, *Te*CpcS, and the genes required to make both PCB and PEB. Even though PCB is the cognate bilin in *T. elongatus*, *Te*CpcS was unable to discriminate between PCB vs PEB, as the absorbance spectrum shown in purple shows that both bilins were attached to ApcF (see Figure 4B of the Supporting Information). Fluorescence emission from PEB (red dotted line) and PCB (blue dotted line) was observed as well, indicating that both bilins had been ligated to ApcF. *Te*CpcS was also capable of attaching PΦB to ApcF (Figure 4C of the Supporting Information). The purified proteins were separated on SDS−PAGE and stained with Zn sulfate to confirm bilin addition to ApcF (Figure 4E of the Supporting Information, lanes 1 through 4). *Te*CpcS was able to attach PCB, PEB, and PΦB to ApcF, even though ApcF naturally carries only PCB.

Intrinsic Bilin Binding of *Te*CpcS. The purified HT-*Te*CpcS obtained by producing the protein together with PCB or PEB was fluorescent (Figure 5A and 5B). However, the fluorescence intensity of *Te*CpcS was much lower than that obtained with PBP subunits. After electrophoresis in the presence of SDS, both *Te*CpcS-PCB and *Te*CpcS-PEB had readily detected Zn-enhanced fluorescence emission bands (Figure 5D, lanes 2 and 3, respectively). To understand the nature of the interaction of the bilin to *Te*CpcS, the HT-*Te*CpcS-PCB and HT-*Te*CpcS-PEB adducts were analyzed by trypsin digestion and mass spectrometry. HT-*Te*CpcS bound both PEB and PCB at Cys2 and Cys169 (Figure 5 of the Supporting Information). On the basis of the X-ray structure and modeling (Figures 2A and 3), the bilin is covalently bound to *Te*CpcS within the cavity of the β-barrel at Cys169 (location labeled in Figure 2A) and at the N-terminal loop at Cys2 near the first helix. Neither of these Cys residues is conserved within the CpcS-I/U family (see Figure 3), which suggests that covalent ligation to these Cys residues may occur adventitiously during the expression and/or purification process. In the cyanobacterial cytoplasm, it is unlikely that a PCB molecule would have a long association with the *Te*CpcS enzyme. The function of *Te*CpcS is to ligate PCB to apo-phycobiliprotein substrates, and these substrate proteins are the most abundant proteins in cyanobacterial cells under most growth conditions. Therefore, it seems unlikely that the covalent addition is a normal occurrence in cyanobacteria or a normal part of the reaction cycle.

To test whether these two Cys residues were involved in catalysis, site-specific variants within *Te*CpcS were produced (C2S, C169S, and C2S/C169S). These HT-*Te*CpcS variants were produced in *E. coli* by coexpressing them with pPcyA, purified, and analyzed by absorbance and fluorescence emission spectroscopy followed by SDS−PAGE and Zn-enhanced bilin fluorescence. Wild-type *Te*CpcS and all of the variants retained the ability to bind PCB as judged by their absorbance and fluorescence spectra (Figure 6A of the Supporting Information; absorbance maximum at 605 nm and fluorescence emission maximum at 633 nm). However, when these protein/bilin complexes were separated by SDS−PAGE (Figure 6C of the Supporting Information) and assayed for their covalent binding of PCB by Zn-enhanced fluorescence of *Te*CpcS, the C2S and the C2S/C169S variants had no detectable PCB fluorescence associated with the *Te*CpcS protein (Figure 6D, lanes 1−4, of the Supporting Information). This indicates that covalent addition of PCB to *Te*CpcS mainly occurs through Cys2.

In order to assay whether these site-specific variants retained their PCB ligation activity, the genes encoding the *Te*CpcS variants were coexpressed with genes encoded by pPcyA and pCpcBA. In this experiment both *Te*CpcS and CpcB were His-tagged and copurified. The absorbance and fluorescence emission properties of the purified proteins from these coexpressions are consistent with those of HT-CpcB with PCB covalently attached at Cys82 (see Figure 6B of the Supporting Information) with an absorbance maxima at 621 nm and fluorescence maxima at 640 nm.[16] These samples containing HT-CpcB were also diluted approximately 4-fold in comparison to those shown in Figure 6A of the Supporting Information. Although the masses of HT-*Te*CpcS and HT-CpcB are very similar and they comigrate in SDS−PAGE (see Figure 6, panels C and D, lanes 5−8, of the Supporting Information), the large amount of bilin fluorescence seen in the samples containing all of the variants indicates that they are capable of PCB ligation to HT-CpcB. Therefore, we conclude that neither Cys2 nor Cys169 is required for the activity of *Te*CpcS.

In the *Te*CpcS model with bound PCB, PCB is suggested to have hydrophobic interactions as well as hydrogen bonding with numerous residues. According to the model, both Arg151 and Ser155 interact with PCB, forming hydrogen bonds to the D and C rings, respectively (see Figure 2A). Therefore, site-specific variants R151G and S155G were generated within *Te*CpcS and tested for their ability to bind PCB. These HT-*Te*CpcS variants were produced in *E. coli* by coexpressing them with pPcyA, purified, and analyzed by absorbance and fluorescence emission spectroscopy, which was followed by SDS−PAGE and Zn-enhanced bilin fluorescence. Both the wild-type *Te*CpcS and the S155G variant retained the ability to bind PCB as seen in Figure 7 (panels A, B, and C) of the Supporting Information. The S155G variant has less PCB attached to *Te*CpcS than the wild-type, suggesting the strength of the interaction with PCB may be weaker (compare Figure 7C, lanes 1 and 3, of the Supporting Information). However, the R151G variant had no detectable PCB fluorescence (Figure 7A of the Supporting Information, orange spectra), and covalent PCB attachment to *Te*CpcS was not detected by Zn-enhanced fluorescence (Figure 7C, lane 2, of the Supporting Information). Therefore, we conclude that both residues appear to be involved in the binding of PCB to *Te*CpcS, consistent with the model in Figure 2A. However, when these variants were coexpressed with pPcyA and pCpcBA to test for PCB ligation activity, both variants were capable of ligating PCB to HT-CpcB (Figure 7, panels D, E, and F, of the Supporting Information). Therefore, even though both *Te*CpcS variants S155G and R151G appear to contain less PCB than wild-type after SDS−PAGE, these variants still bind PCB sufficiently well to allow attachment to CpcB to occur in *E. coli*.

Investigation of the Native Molecular Weight of the *Te*CpcS Protein. In order to determine the oligomeric status of *Te*CpcS in its native state, purified recombinant HT-*Te*CpcS was subjected to size-exclusion chromatography (Figure 6). The protein complex had a retention time of 26.9 min, and the molecular weight of this complex was calculated to be 45600 (Figure 6, inset graph). The calculated molecular mass of the HT-*Te*CpcS polypeptide is 21.9 kDa. These data suggest that the native protein is stable and active as a homodimer, which is

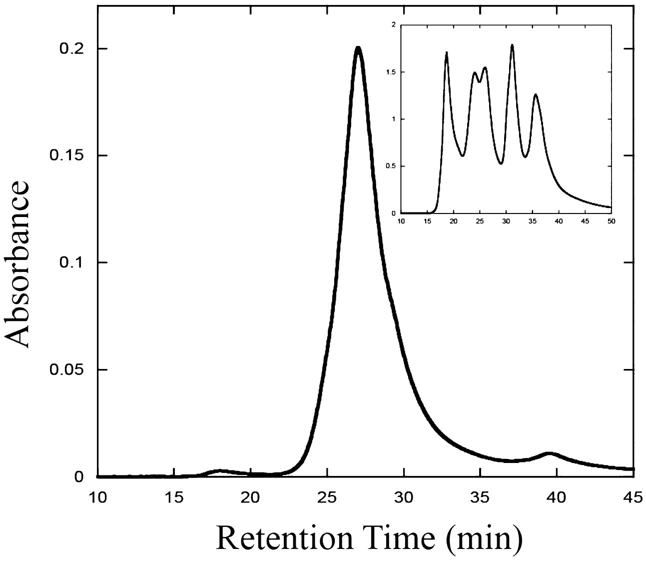


Figure 6. Size-exclusion chromatography of recombinant *Te*CpcS. Recombinant HT-*Te*CpcS was injected onto the HPLC column and protein eluates were monitored at 280 nm. The calculated M_r of the fraction eluting at 26.9 min was 45600, which suggests that HT-*Te*CpcS is a dimer under the conditions of this experiment. The inset graph shows the chromatogram for the molecular weight standards.

consistent with the observation that *Te*CpcS crystallizes as a dimer (Figure 1).

DISCUSSION

We describe here the first structure for a bilin lyase of the CpcS/CpcU family and demonstrate its functionality and substrate capabilities in a recombinant, heterologous expression system. *Te*CpcS is a member of a diverse family of proteins known as the lipocalins. All members of the lipocalin family are composed of similar beta-barrels, and they bind a diverse array of ligands, most of which are hydrophobic. Two members of the lipocalin family of proteins bind bilins, UnaG, and the bilin-binding protein of insects that binds biliverdin IXγ. Interestingly, the motivation for solving the structure of the bilin-binding protein of insects in 1987 was to compare it to the structure of PBPs because both of these proteins are associated with bilins. However, they were not structurally related in any way: PBPs are most similar to hemoglobin and myoglobin, while the insect bilin-binding protein was most similar to the retinol-binding protein. Very recently, the structure of UnaG, the first fluorescent protein found in vertebrates, was solved.[62] This protein is most similar to fatty acid-binding proteins and is fluorescent because it noncovalently binds bilirubin. UnaG is only fluorescent when bilirubin is bound and is not fluorescent when the similar bilin, biliverdin IXα, is bound.[62] UnaG and *Te*CpcS share structural similarities and both proteins bind bilins. Our modeling suggests that the association of PCB with *Te*CpcS (Figure 2) is similar to that shown for UnaG and bilirubin. The structure of UnaG allowed us to predict the conformation of two unstructured loops (residues 77–83 and 108–117) that were missing in the electron density map for *Te*CpcS. These two loops may become ordered only in presence of the bilin because they interact with the bilin while partially capping the funnel of the β-barrel. The location of these loops near the bilin and near the wide-end of the "funnel" also suggests that one or both of these loops may also be involved in phycobiliprotein substrate docking. An examination of the alignment in Figure 3 reveals some substantial differences in the region of 108–117 between the CpcS-type lyases and the CpeS lyase. The CpcS-type lyases in the alignment ligate PCB to allophycocyanin and phycocyanin β-subunits, whereas the CpeS lyase ligates PEB to phycoerythrin β-subunits (CpeB). It is also possible that the structures of these loops in *Te*CpcS are very different from those in UnaG. The function of UnaG in eel is not currently known, but it does not appear to have a catalytic role in bilin ligation as we show here for *Te*CpcS.

The CpcE/CpcF bilin lyase family members are not phylogenetically related to the CpcS/CpcU family.[17,21,32,40] All CpcE/CpcF-type lyases contain 5–6 HEAT-repeat motifs,[72,73] which are also found in diverse proteins from various eukaryotic organisms and are thought to facilitate protein–protein interactions.[74,75] No structures have been solved for these CpcE/CpcF lyase proteins, but Phyre^2 analyses[76] of a fusion protein MpeZ from this group suggests that their structures are primarily α-helical.[31]

Phyre^2 analyses of the proteins within the CpcT lyase family suggest that they are distantly related to the CpcS/CpcU proteins.[40] It was hypothesized that lyases of the CpcT-type were necessary in order to form the *S*-stereoisomer at C3^1 of the bilin located at Cys153 on CpcB because bilins attached by the CpcS/CpcU-type lyases have an R-configuration at C3^1.[18] Presumably, if members of the CpcT family are also members of the same structural family as *Te*CpcS, then the bilin would be held in a different conformation within the beta-barrel to allow the *S-stereoisomer* to form. Evolution of the CpcE/CpcF family of lyases may have occurred later, as suggested by Biswas et al.,[28] in order to create a set of lyases that was highly specific for the alpha subunits only, and these lyases later evolved an isomerase activity to allow a more diverse array of bilins to be produced with these antenna proteins.

*Te*CpcS is a member of the CpcS/CpcU family of bilin lyases, and we show here that in *E. coli*, it can ligate a diverse array of bilins to its PBP substrates CpcB and allophycocyanin subunits ApcA, ApcB, ApcD, and ApcF at the equivalent position to Cys82. Members of the CpcS/CpcU lyase family all seem to bind bilins rapidly and tightly and then catalyze a slow thioether bond formation to the PBPs. This appears to be very important to prevent spontaneous but incorrectly ligated products from forming.[35,77−79] These lyases play a chaperone-like role in which they quickly and specifically bind the bilin and deliver it to the PBP substrate in an appropriate conformation for ligation.[77,80,81]

The high reactivity of bilins toward cysteines precluded initial attempts to obtain cocrystals of PCB with *Te*CpcS, probably due to the mixed population of covalent adducts that was formed. This problem could possibly be overcome by mutating the cysteine residues or by using a bilin substrate analog without substituent groups on the A and D rings. As we demonstrate here, bilins were covalently bound to Cys2 and Cys169 on *Te*CpcS (Figure 5 of the Supporting Information), but neither of these Cys is highly conserved, as seen in the alignment of Figure 3, and both are very far from the target ethylidene group. We also show here that site-specific variants C2S, C169S, and C2S/C169S variants were capable of PCB ligation to HT-CpcB, indicating that neither Cys is essential for the catalytic mechanism.

In the structure of *Te*CpcS modeled with PCB, the target ethylidene group is exposed in a shallow cavity on the surface of the protein (see Figure 2C). *Te*CpcS holds it in a precise geometry ready for catalysis, but the active site would need to be completed by the binding of the phycobiliprotein subunit that brings its target Cys residue into the groove on the surface of *Te*CpcS to allow nucleophilic attack by the cysteinyl sulfur on the ethylidene group on ring A of PCB. This would mean the enzyme active site would actually be divided between *Te*CpcS and the Cys on the target phycobiliprotein subunit, and this could explain why chromophore ligation is much slower than bilin binding.

A histidine was suggested as a candidate amino acid involved in the binding/ligation reaction mechanism in phytochromes.[82] Tu et al. suggested that His residues were the most likely candidate for the strong association of PCB with the lyase because they were unable to detect covalent addition with Cys after trypsin digestion in their mass spectrometry experiments.[83] The results of Kupka et al. suggested when these histidines in CpcS-III from *Nostoc* sp. PCC 7120 were mutated, the conformation and or protonation state of the chromophore was affected.[77] Cytochrome *c* biogenesis requires the heme chaperone CcmE, which was shown to bind heme covalently through an association with histidine residues.[84] However, in the modeling with PCB shown in Figure 2, no histidines occur in locations to allow a strong interaction with the bilin. Studies performed on the *Nostoc* sp. PCC 7120 CpcS-III protein showed that two arginines (Arg18 and Arg149, numbering from 7120 CpcS-III) were essential for lyase activity. In our model with *Te*CpcS and PCB, Arg151 (equivalent to Arg149 in the *Nostoc* sp. protein; see Figure 3) was modeled to hydrogen bond to the carbonyl group of ring D.[77] When Arg151 was mutated to glycine, *Te*CpcS did not retain PCB after purification with the enzyme, unlike the wild-type protein, confirming its importance in the stable association between *Te*CpcS and PCB. However, this variant retained the ability to ligate PCB to HT-CpcB, indicating that PCB can still bind to this variant of *Te*CpcS.

Like some other lyases, under the heterologous conditions employed here (i.e., in *E. coli*), *Te*CpcS displayed an ability to attach both cognate as well as noncognate bilins to various PBP subunits.[23,24] In vivo, it was also demonstrated that coexpression of the genes, required to make PEB in *Synechococcus* sp. PCC 7002, resulted in the attachment of the noncognate PEB to CpcA (by the CpcE/CpcF lyase), while expression of the genes required to produce PΦB resulted in many phycobiliproteins containing the PΦB chromophore.[85] *T. elongatus* is a cyanobacterial species that only contains PCB, so its lyases do not need to have a strong discrimination between the isomeric bilins PEB and PCB. Other lyases, like CpeS from *Fremyella diplosiphon*, are very specific with respect to their bilin substrate, specifically attaching PEB but not PCB at Cys80 on CpeB,[28] and this ability to discriminate presumably evolved to ensure that the appropriate bilin is attached to the correct protein so that energy transfer within PBS is efficient and unidirectional. *Te*CpcS was shown to ligate three different bilins (its cognate bilin, PCB, as well as PEB and PΦB) to Cys82 of five different apo-PBP substrates (CpcB, ApcA, ApcB, ApcD, and ApcF). This capability allows the production of a diverse array of fluorescent, natural, and unnatural phycobiliproteins for protein tagging and bioimaging purposes. Thus, the findings presented here illustrate the utility of this thermostable enzyme for producing fluorescent proteins for future biotechnological applications.

CONCLUSION

*Te*CpcS is a bilin lyase of the CpcS family and has a beta-barrel structure that is similar to those of diverse members of the lipocalin family. *Te*CpcS forms homodimers in solution and can ligate the cognate bilin, PCB, as well as two noncognate bilins PEB and PΦB to Cys82 of five different PBP substrates, including CpcB but notably not CpcA. The crystal structure and an energy-minimized model with bound substrate provide the structural basis for future studies to understand how specificity is generated within this class of enzymes.

ASSOCIATED CONTENT

Supporting Information

One table and seven figures. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The Protein Database ID for *Te*CpcS is 3BDR.

AUTHOR INFORMATION

Corresponding Author

*Address: Department of Biological Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148. E-mail: wschluch@uno.edu. Tel: (504) 280-7194. Fax: (504) 280-6121.

Funding

This work was funded by grants from the Protein Structure Initiative of the National Institutes of Health to G.T.M. and J.F.H. (Grant U54 GM094597) and from the National Science Foundation to W.M.S. (Grants MCB-0843664 and MCB-1244339), D.A.B. (Grants MCB-0519743 and MCB-1021725), and R.B.C. (Grant CHE-1058764).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Randy Abramowitz and John Schwanof for setting up the X4A beamline; Sergey M. Vorobiev for assistance in crystal screening. The W. M. Keck Foundation provided support for equipment utilized for this study and located in the Keck Conservation and Molecular Genetics lab at the University of New Orleans.

ABBREVIATIONS:

AP, allophycocyanin; DTT, dithiothreitol; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria–Bertani medium; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; Ni-NTA, nickel nitrolotriacetic acid; PAGE, polyacrylamide gel electrophoresis; PBP, phycobiliprotein(s); PBS, phycobilisome(s); PC, phycocyanin; PCB, phycocyanobilin; PCR, polymerase chain reaction; PEB, phycoerythrobilin; PΦB, phytochromobilin; SDS, sodium dodecylsulfate; SEC, size exclusion chromatography; *Te*, *Thermosynechococcus elongatus*; ToF, time-of-flight

REFERENCES

(1) Glazer, A. N. (1989) Light guides: Directional energy transfer in a photosynthetic antenna. *J. Biol. Chem. 264*, 1–4.

(2) Willows, R. D., Mayer, S. M., Foulk, M. S., DeLong, A., Hanson, K., Chory, J., and Beale, S. I. (2000) Phytobilin biosynthesis: The *Synechocystis* sp. PCC 6803 heme oxygenase-encoding *ho1* gene complements a phytochrome-deficient *Arabidopsis thaliana* hy1 mutant. *Plant Mol. Biol. 43*, 113–120.

(3) Rhie, G., and Beale, S. I. (1995) Phycobilin biosynthesis: Reductant requirements and product identification for heme oxygenase from *Cyanidium caldarium*. *Arch. Biochem. Biophys. 320*, 182−194.

(4) Cornejo, J., Willows, R. D., and Beale, S. I. (1998) Phytobilin biosynthesis: Cloning and expression of a gene encoding soluble ferredoxin-dependent heme oxygenase from *Synechocystis* sp. PCC 6803. *Plant J. 15*, 99−107.

(5) Rhie, G., and Beale, S. I. (1992) Biosynthesis of phycobilins: Ferredoxin-supported NADPH-independent heme oxygenase and phycobilin-forming activities from *Cyanidium caldarium*. *J. Biol. Chem. 267*, 16088−16093.

(6) Kohchi, T., Mukougawa, K., Frankenberg, N., Masuda, M., Yakota, A., and Lagarias, J. C. (2001) The *Arabidopsis* HY2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. *Plant Cell 13*, 425−436.

(7) Frankenberg, N., Mukougawa, K. K., and Lagarias, J. C. (2001) Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. *Plant Cell 13*, 965−978.

(8) Rockwell, N. C., Su, Y. S., and Lagarias, J. C. (2006) Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol. 57*, 837−858.

(9) Schirmer, T., Huber, R., Schneider, M., Bode, W., Miller, M., and Hackert, M. L. (1986) Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*: The molecular model and its implications from light-harvesting. *J. Mol. Biol. 188*, 651−676.

(10) Glazer, A. N., and Fang, S. (1973) Chromophore content of blue-green algal phycobiliproteins. *J. Biol. Chem. 248*, 659−662.

(11) Capuano, V., Braux, A. S., Tandeau de Marsac, N., and Houmard, J. (1991) The "anchor polypeptide" of cyanobacterial phycobilisomes. Molecular characterization of the *Synechococcus* sp. PCC 6301 *apcE* gene. *J. Biol. Chem. 266*, 7239−7247.

(12) Zhao, K. H., Ping, S., Bohm, S., Bo, S., Ming, Z., Bubenzer, C., and Scheer, H. (2005) Reconstitution of phycobilisome core-membrane linker, L-CM, by autocatalytic chromophore binding to ApcE. *Biochim. Biophys. Acta, Bioenerg. 1706*, 81−87.

(13) Biswas, A., Vasquez, Y. M., Dragomani, T. M., Kronfel, M. L., Williams, S. R., Alvey, R. M., Bryant, D. A., and Schluchter, W. M. (2010) Biosynthesis of cyanobacterial phycobiliproteins in *Escherichia coli*: Chromophorylation efficiency and specificity of all bilin lyases from *Synechococcus* sp strain PCC 7002. *Appl. Environ. Microbiol. 76*, 2729−2739.

(14) Zhao, K. H., Deng, M. G., Zheng, M., Zhou, M., Parbel, A., Storf, M., Meyer, M., Strohmann, B., and Scheer, H. (2000) Novel activity of a phycobiliprotein lyase: Both the attachment of phycocyanobilin and the isomerization to phycobiliviolin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. *FEBS Lett. 469*, 9−13.

(15) Fairchild, C. D., Zhao, J., Zhou, J., Colson, S. E., Bryant, D. A., and Glazer, A. N. (1992) Phycocyanin α subunit phycocyanobilin lyase. *Proc. Natl. Acad. Sci. U.S.A. 89*, 7017−7021.

(16) Saunée, N. A., Williams, S. R., Bryant, D. A., and Schluchter, W. M. (2008) Biogenesis of phycobiliproteins. II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to Cys-82 of beta -phycocyanin and Cys-81 of allophycocyanin subunits in *Synechococcus* sp. PCC 7002. *J. Biol. Chem. 283*, 7513−7522.

(17) Scheer, H., and Zhao, K. H. (2008) Biliprotein maturation: The chromophore attachment. *Mol. Microbiol. 68*, 263−276.

(18) Shen, G., Saunee, N. A., Williams, S. R., Gallo, E. F., Schluchter, W. M., and Bryant, D. A. (2006) Identification and characterization of a new class of bilin lyase: The *cpcT* gene encodes a bilin lyase responsible for attachment of phycocyanobilin to Cys-153 on the beta subunit of phycocyanin in *Synechococcus* sp. PCC 7002. *J. Biol. Chem. 281*, 17768−17778.

(19) Zhao, K. H., Su, P., Tu, J., Wang, X., Liu, H., Ploscher, M., Eichacker, L., Yang, B., Zhou, M., and Scheer, H. (2007) Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. *Proc. Natl. Acad. Sci. U.S.A. 104*, 14300−14305.

(20) Zhou, J., Gasparich, G. E., Stirewalt, V. L., de Lorimier, R., and Bryant, D. A. (1992) The *cpcE* and *cpcF* genes of *Synechococcus* sp. PCC 7002: Construction and phenotypic characterization of interposon mutants. *J. Biol. Chem. 267*, 16138−16145.

(21) Schluchter, W. M., Shen, G., Alvey, R. M., Biswas, A., Saunée, N. A., Williams, S. R., Miller, C. A., and Bryant, D. A. (2010) Phycobiliprotein biosynthesis in cyanobacteria: structure and function of enzymes involved in post-translational modification, In *Advances in Experimental Medicine & Biology* (Hallenbeck, P. C., Ed.), Springer, NY, NY.

(22) Tang, K., Zeng, X. L., Yang, Y., Wang, Z. B., Wu, X. J., Zhou, M., Noy, D., Scheer, H., and Zhao, K. H. (2012) A minimal phycobilisome: Fusion and chromophorylation of the truncated core-membrane linker and phycocyanin. *Biochim. Biophys. Acta, Bioenerg. 1817*, 1030−1036.

(23) Blot, N., Wu, X. J., Thomas, J. C., Zhang, J., Garczarek, L., Bohm, S., Tu, J. M., Zhou, M., Ploscher, M., Eichacker, L., Partensky, F., Scheer, H., and Zhao, K. H. (2009) Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed post-translationally by a phycoerythrobilin lyase-isomerase. *J. Biol. Chem. 284*, 9290−9298.

(24) Alvey, R. M., Biswas, A., Schluchter, W. M., and Bryant, D. A. (2011) Attachment of noncognate chromophores to CpcA of *Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7002 by heterologous expression in *Escherichia coli*. *Biochemistry 50*, 4890−4902.

(25) Bhalerao, R. P., Kind, L. K., and Gustafsson, P. (1994) Cloning of the *cpcE* and *cpcF* genes from *Synechococcus* sp. PCC 6301 and their inactivation in *Synechococcus* sp. PCC 7942. *Plant Mol. Biol. 26*, 313−326.

(26) Fairchild, C. D., and Glazer, A. N. (1994) Oligomeric structure, enzyme kinetics, and substrate specificity of the phycocyanin alpha subunit phycocyanobilin lyase. *J. Biol. Chem. 269*, 8686−8694.

(27) Swanson, R. V., Zhou, J., Leary, J. A., Williams, T., de Lorimier, R., Bryant, D. A., and Glazer, A. N. (1992) Characterization of phycocyanin produced by *cpcE* and *cpcF* mutants and identification of an intergenic suppressor of the defect in bilin attachment. *J. Biol. Chem. 267*, 16146−16154.

(28) Biswas, A., Boutaghou, M. N., Alvey, R. M., Kronfel, C. M., Cole, R. B., Bryant, D. A., and Schluchter, W. M. (2011) Characterization of the activities of the CpeY, CpeZ, and CpeS bilin lyases in phycoerythrin biosynthesis in *Fremyella diplosiphon* strain UTEX 481. *J. Biol. Chem. 286*, 35509−35521.

(29) Storf, M., Parbel, A., Meyer, M., Strohmann, B., Scheer, H., Deng, M. G., Zheng, M., Zhou, M., and Zhao, K. H. (2001) Chromophore attachment to biliproteins: Specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-cys-alpha 84-phycoviolobilin chromophore of phycoerythrocyanin. *Biochemistry 40*, 12444−12456.

(30) Jung, L. J., Chan, C. F., and Glazer, A. N. (1995) Candidate genes for the phycoerythrocyanin α subunit lyase: Biochemical analysis of *pecE* and *pecF* interposon mutants. *J. Biol. Chem. 270*, 12877−12884.

(31) Shukla, A., Biswas, A., Blot, N., Partensky, F., Karty, J. A., Hammad, L. A., Garczarek, L., Gutu, A., Schluchter, W. M., and Kehoe, D. M. (2012) Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A. 109*, 20136−20141.

(32) Shen, G. Z., Saunee, N. A., Gallo, E., Begovic, Z., Schluchter, W. M., and Bryant, D. A. (2004) Identification of novel phycobiliprotein lyases in cyanobacteria. In *Photosynthesis 2004 Light-Harvesting Systems Workshop* (Niederman, R. A., Blankenship, R. E., Frank, H., Robert, B., and van Grondelle, R., Eds.), pp 14−15, Saint Adele, Quebec, Canada.

(33) Shen, G., Schluchter, W. M., and Bryant, D. A. (2008) Biogenesis of phycobiliproteins. I. *cpcS-I* and *cpcU* mutants of the cyanobacterium *Synechococcus* sp. PCC 7002 define a heterodimeric phycocaynobilin lyase specific for beta -phycocyanin and allophycocyanin subunits. *J. Biol. Chem. 283*, 7503−7512.

(34) Zhao, K. H., Zhang, J., Tu, J. M., Boehm, S., Ploescher, M., Eichacker, L., Bubenzer, C., Scheer, H., Wang, X., and Zhou, M.

(2007) Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. *J. Biol. Chem. 282*, 34093−34103.

(35) Wiethaus, J., Busch, A. W. U., Kock, K., Leichert, L. I., Herrmann, C., and Frankenberg-Dinkel, N. (2010) CpeS is a lyase specific for attachment of 3Z-PEB to Cys82 of ®-phycoerythrin from *Prochlorococcus marinus* MED4. *J. Biol. Chem. 285*, 37561−37569.

(36) Zhao, K. H., Su, P., Li, J., Tu, J. M., Zhou, M., Bubenzer, C., and Scheer, H. (2006) Chromophore attachment to phycobiliprotein beta-subunits: Phycocyanobilin:cystein-beta84 phycobiliprotein lyase activity of CpeS-like protein from *Anabaena* sp. PCC7120. *J. Biol. Chem. 281*, 8573−8581.

(37) Montelione, G. T. (2012) The Protein Structure Initiative: Achievements and visions for the future. *F1000 Biology Reports 4*, 7.

(38) Kuzin, A. P., Su, M., Seetharaman, J., Forouhar, F., Wang, D., Janjua, H., Cunningham, K., Ma, L.-C., Xiao, R., Liu, J., Baran, M. C., Acton, T. B., Rost, B., Montelione, G. T., Tong, L., and Hunt, J. F. (2007) Crystal structure of fatty acid-binding protein-like Ycf58 from *Thermosynechoccus elongatus* target Ter13, In *pdb ID: 3BDR, DOI:*10.2210/pdb3bdr/pdb Northeast Structural Genomics ConsortiumTarget Ter13.

(39) Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2002) Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Res. 9*, 123−130.

(40) Bretaudeau, A., Coste, F., Humily, F., Garczarek, L., Le Corguillé, G., Six, C., Ratin, M., Collin, O., Schluchter, W. M., and Partensky, F. (2013) CyanoLyase: A database of phycobilin lyase sequences, motifs and functions. *Nucleic Acids Res. 41*, D396−D401.

(41) Xiao, R., Anderson, S., Aramini, J. M., Belote, R., Buchwald, W., Ciccosanti, C., Conover, K., Everett, J. K., Hamilton, K., Huang, Y. J., Janjua, H., Jiang, M., Kornhaber, G. J., Lee, D. Y., Locke, J. Y., Ma, L.-C., Maglaqui, M., Mao, L., Mitra, S., Patel, D., Rossi, P., Sahdev, S., Sharma, S., Shastry, R., Swapna, G. V. T., Tong, S. N., Wang, D., Wang, H., Zhao, L., Montelione, G. T., and Acton, T. B. (2010) The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J. Struct. Biol. 172*, 21−33.

(42) Acton, T. B., Xiao, R., Anderson, S., Aramini, J. M., Buchwald, W., Ciccosanti, C., Conover, K., Everett, J. K., Hamilton, K., Huang, Y. J., Janjua, H., Kornhaber, G. J., Lau, J., Lee, D. Y., Liu, G., Maglaqui, M., Ma, L.-C., Mao, L., Patel, D., Rossi, P., Sahdev, S., Sharma, S., Shastry, R., Swapna, G. V. T., Tang, Y., Tong, S. N., Wang, D., Wang, H., Zhao, L., and Montelione, G. T. (2011) Preparation of protein samples for NMR structure, function, and small molecule screening studies. *Methods Enzymol. 493*, 21−60.

(43) Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, G. T., and Nilsson, B. (1996) High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. *J. Biomol. NMR 7*, 131−141.

(44) Doublié, S., Kapp, U., Åberg, A., Brown, K., Strub, K., and Cusack, S. (1996) Crystallization and preliminary X-ray analysis of the 9 kDa protein of the mouse signal recognition particle and the selenomethionyl-SRP9. *FEBS Lett. 384*, 219−221.

(45) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.

(46) Weeks, C. M., and Miller, R. (1999) The design and implementation of SnB v2.0. *J. Appl. Crystallogr. 32*, 120−124.

(47) Terwilliger, T. C. (2003) Solve and resolve: Automated structure solution and density modification. *Methods Enzymol. 374*, 22−37.

(48) McRee, D. E. (1999) XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol. 125*, 156−165.

(49) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 905−921.

(50) DeLano, W. L. (2002) *The Pymol Manual*, DeLano Scientific, San Carlos, CA.

(51) Berkelman, T. R., and Lagarias, J. C. (1986) Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal. Biochem. 156*, 194−201.

(52) Bishop, R. E. (2000) The bacterial lipocalins. *Biochim. Biophys. Acta 1482*, 73−83.

(53) Charron, J. B. F., Ouellet, F., Pelletier, M., Danyluk, J., Chauve, C., and Sarhan, F. (2005) Identification, expression, and evolutionary analyses of plant lipocalins. *Plant Physiol. 139*, 2017−2028.

(54) Grzyb, J., Latowski, D., and Strzalka, K. (2006) Lipocalins: A family portrait. *J. Plant Physiol. 163*, 895−915.

(55) Hieber, A. D., Bugos, R. C., and Yamamoto, H. Y. (2000) Plant lipocalins: Violaxanthin de-epoxidase and zeaxanthin epoxidase. *Biochim. Biophys. Acta 1482*, 84−91.

(56) Skerra, A. (2000) Lipocalins as a scaffold. *Biochim. Biophys. Acta 1482*, 337−350.

(57) Huber, R., Schneider, M. J., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. (1987) Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol. 198*, 499−513.

(58) Huber, R., Schneidera, M., Eppa, O., Mayra, I., Messerschmidta, A., Pflugratha, J., and Kayser, H. (1987) Crystallization, crystal-structure analysis and preliminary molecular-model of the bilin binding protein from the insect *Pieris brassicae*. *J. Mol. Biol. 195*, 423−434.

(59) Andersen, J. F., Weichsel, A., Balfour, C. A., Champagne, D. E., and Montfort, W. R. (1998) The crystal structure of nitrophorin 4 at 1.5 angstrom resolution: Transport of nitric oxide by a lipocalin-based heme protein. *Structure 6*, 1315−1327.

(60) Flower, D. R., North, A. C. T., and Sansom, C. E. (2000) The lipocalin protein family: Structural and sequence overview. *Biochim. Biophys. Acta 1482*, 9−24.

(61) Newcomer, M. E., and Ong, D. E. (2000) Plasma retinol binding protein: Structure and function of the prototypic lipocalin. *Biochim. Biophys. Acta 1482*, 57−64.

(62) Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T., and Miyawaki, A. (2013) A bilirubin-inducible fluorescent protein from eel muscle. *Cell 153*, 1602−1611.

(63) Montelione, G. T., and Anderson, S. (1999) Structural genomics: Keystone for a human proteome project. *Nat. Struct. Biol. 6*, 11−12.

(64) Liu, J., Montelione, G. T., and Rost, B. (2007) Novel leverage of structural genomics. *Nat. Biotechnol. 25*, 849−851.

(65) Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res. 38*, W545−W549.

(66) Lundell, D. J., and Glazer, A. N. (1981) Allophycocyanin B: A common β subunit in *Synechococcus* allophycocyanin B (λ_{max} 670 nm) and allophycocyanin (λ_{max} 650 nm). *J. Biol. Chem. 246*, 12600−12606.

(67) Ashby, M. K., and Mullineaux, C. W. (1999) The role of ApcD and ApcF in energy transfer from phycobilisomes to PSI and PSII in a cyanobacterium. *Photosynth. Res. 61*, 169−179.

(68) Dong, C. X., Tang, A. H., Zhao, J. D., Mullineaux, C. W., Shen, G. Z., and Bryant, D. A. (2009) ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp PCC 7002. *Biochim. Biophys. Acta, Bioenerg. 1787*, 1122−1128.

(69) Zhao, J., Shen, G., and Bryant, D. A. (2001) Photosystem stoichiometry and state transitions in a mutant of the cyanobacterium *Synechococcus* sp. PCC 7002 lacking phycocyanin. *Biochim. Biosphys. Acta 1505*, 248−257.

(70) Zhao, J., Zhou, J., and Bryant, D. A. (1992) Energy transfer processes in phycobilisomes as deduced from analyses of mutants of

Synechococcus sp. PCC 7002. In *Research in Photosynthesis* (Murata, N., Ed.), pp 25−32, Kluwer, Dordrecht.

(71) Gindt, Y. M., Zhou, J. H., Bryant, D. A., and Sauer, K. (1994) Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores - Assignment of excited-state energies to the L_{CM}, beta^{18} and alpha^{AP-B} chromophores. *Biochim. Biophys. Acta, Bioenerg. 1186*, 153−162.

(72) Morimoto, K., Sato, S., Tabata, S., and Nakai, M. (2003) A HEAT-repeats containing protein, IaiH, stablizes the iron-sulfur cluster bound to the cyanobacterial IscA homologue, IscA2. *J. Biochem. (Tokyo) 134*, 211−217.

(73) Morimoto, K., Nishio, K., and Nakai, M. (2002) Identification of a novel prokaryotic HEAT-repeats-containing protein which interacts with a cyanobacterial IscA homolog. *FEBS Lett. 519*, 123−127.

(74) Takano, H., and Gusella, J. (2002) The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF-kB/Rel/dorsal family transcription factor. *BMC Neurosci. 3*, 15.

(75) Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W., and Bork, P. (2001) Comparison of ARM and HEAT protein repeats. *J. Mol. Biol. 309*, 1−18.

(76) Kelley, L. A., and Sternberg, M. J. E. (2009) Protein structure prediction on the web: A case study using the Phyre server. *Nat. Protoc. 4*, 363−371.

(77) Kupka, M., Zhang, J., Fu, W. L., Tu, J. M., Bohm, S., Su, P., Chen, Y., Zhou, M., Scheer, H., and Zhao, K. H. (2009) Catalytic mechanism of S-type phycobiliprotein lyase chaperone-like action and functional amino acid residues. *J. Biol. Chem. 284*, 36405−36414.

(78) Arciero, D. M., Bryant, D. A., and Glazer, A. N. (1988) *In vitro* attachment of bilins to apophycocyanin I: specific covalent adduct formation at cysteinyl residues involved in phycocyanobilin binding in C-phycocyanin. *J. Biol. Chem. 263*, 18343−18349.

(79) Arciero, D. M., Dallas, J. L., and Glazer, A. N. (1988) *In vitro* attachment of bilins to apophycocyanin: II: Determination of the structures of tryptic bilin peptides derived from the phycocyanobilin adduct. *J. Biol. Chem. 263*, 18350−18357.

(80) Schluchter, W. M., and Glazer, A. N. (1999) Biosynthesis of phycobiliproteins in cyanobacteria, In *The Phototrophic Prokaryotes* (Peschek, G. A., Löffelhardt, W., and Schmetterer, G., Eds.), pp 83−95, Plenum, New York.

(81) Bohm, S., Endres, S., Scheer, H., and Zhao, K. H. (2007) Biliprotein chromophore attachment: Chaperone-like function of the PecE subunit of alpha-phycoerythrocyanin lyase. *J. Biol. Chem. 282*, 25357−25366.

(82) Wu, S.-H., and Lagarias, J. C. (2000) Defining the bilin lyase domain: Lessons from the extended phytochrome superfamily. *Biochemistry 39*, 13487−13495.

(83) Tu, J. M., Kupka, M., Boehm, S., Ploescher, M., Eichacker, L., Zhao, K. H., and Scheer, H. (2008) Intermediate binding of phycocyanobilin to the lyase, CpeS1, and transfer to apoprotein. *Photosynth. Res. 95*, 163−168.

(84) Schulz, H., Hennecke, H., and Thöny-Meyer, L. (1998) Prototype of a heme chaperone essential for cytochrome *c* maturation. *Science 281*, 1197−1200.

(85) Alvey, R. M., Biswas, A., Schluchter, W. M., and Bryant, D. A. (2011) Effects of modified phycobilin biosynthesis in the cyanobacterium *Synechococcus* sp strain PCC 7002. *J. Bacteriol. 193*, 1663−1671.

(86) Miller, C. A., Leonard, H. S., Pinsky, I. G., Turner, B. M., Williams, S. R., Harrison, L., Jr., Fletcher, A. F., Shen, G., Bryant, D. A., and Schluchter, W. M. (2008) Biogenesis of phycobiliproteins. III. CpcM is the asparagine methyltransferase for phycobiliprotein β-subunits in cyanobacteria. *J. Biol. Chem. 283*, 19293−19300.

(87) Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W., and Frankenberg-Dinkel, N. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanabacteria. *Curr. Biol. 18*, 442−448.

(88) Drenth, J. (1999) *Principles of Protein X-Ray Crystallography*, 2nd ed., Springer, NY.

(89) Notredame, C., Higgins, D., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol. 302*, 205−217.