



HAL
open science

Crystal structure and molecular mechanism of an E/F type bilin lyase-isomerase

Indika Kumarapperuma, Kes Lynn Joseph, Cong Wang, Linta Biju, Irin Tom, Kourtney Weaver, Théophile Grébert, Frédéric Partensky, Wendy Schluchter, Xiaojing Yang

► **To cite this version:**

Indika Kumarapperuma, Kes Lynn Joseph, Cong Wang, Linta Biju, Irin Tom, et al.. Crystal structure and molecular mechanism of an E/F type bilin lyase-isomerase. *Structure*, 2022, 30 (4), pp.564-574.E3. 10.1016/j.str.2022.01.007 . hal-03631913

HAL Id: hal-03631913

<https://hal.science/hal-03631913>

Submitted on 6 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1

2 **Crystal Structure and Molecular Mechanism**

3 **of an E/F type Bilin Lyase-isomerase**

4 Indika Kumarapperuma^{1§}, Kes Lynn Joseph^{2§}, Cong Wang¹, Linta M. Biju¹, Irin P. Tom¹,
5 Kourtney D. Weaver², Théophile Grébert³, Frédéric Partensky³, Wendy M. Schluchter^{2*}, and
6 Xiaojing Yang^{1,4*}

7 ¹Department of Chemistry, University of Illinois Chicago, Chicago, IL 60607, USA

8 ²Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA

9 ³Ecology of Marine Plankton (ECOMAP) Team, Station Biologique, Sorbonne Université,
10 CNRS, 29680 Roscoff, France

11 ⁴Department of Ophthalmology and Vision Sciences, University of Illinois at Chicago, Chicago,
12 IL 60607, USA

13

14 [§]These authors contributed equally.

15 ^{*}Corresponding authors (xiaojing@uic.edu; wsluch@uno.edu)

16 **Lead Contact:** Xiaojing Yang (xiaojing@uic.edu)

17

18 **Summary**

19 Chromophore attachment of the light-harvesting apparatus represents one of the most
20 important post-translational modifications in photosynthetic cyanobacteria. Extensive pigment
21 diversity of cyanobacteria critically depends on bilin lyases that covalently attach chemically
22 distinct chromophores to phycobiliproteins. However, how bilin lyases catalyze bilin ligation
23 reactions and how some lyases acquire additional isomerase abilities remain elusive at the
24 molecular level. Here we report the crystal structure of a representative bilin lyase-isomerase
25 MpeQ. This structure has revealed a “question-mark” protein architecture that unambiguously
26 establishes the active site conserved among the E/F type bilin lyases. Based on structural,
27 mutational and modeling data, we demonstrate that stereoselectivity of the active site plays a
28 critical role in conferring the isomerase activity of MpeQ. We further advance a tyrosine-
29 mediated reaction scheme unifying different types of bilin lyases. These results suggest that the
30 lyase and isomerase actions of bilin lyases arise from two coupled molecular events of distinct
31 origin.

32 **Introduction**

33
34 A critical step in biogenesis of the light-harvesting complexes in photosynthetic
35 cyanobacteria involves post-translational modifications that covalently attach bilin pigments to
36 specific cysteine residues in phycobiliproteins (Schluchter et al., 2010). Such reactions are
37 catalyzed by an important family of enzymes called bilin lyases. Phycobiliprotein α and β
38 subunits incorporated with 1-3 bilins each are assembled as donut-shaped trimers or hexamers
39 in megadalton phycobilisomes (PBS) where the rods project out from the core to allow light
40 harvesting and directional energy transfer to the reaction centers of photosystems (Glazer, 1989)
41 (Fig. S1A). Bilin lyases are directly responsible for the extensive pigment diversity of the PBS
42 and ubiquity of cyanobacteria on Earth by enabling them to adapt and thrive in different light
43 niches (Flombaum et al., 2013; Grébert et al., 2018; Sanfilippo et al., 2019a).

44 Among three phylogenetically distinct types of bilin lyases (Bretau et al., 2012;
45 Fairchild et al., 1992; Scheer and Zhao, 2008; Schluchter et al., 2010; Shen et al., 2006, 2008; Zhou
46 et al., 1992), the S/U and T types have the β -barrel folds while the E/F type adopts an all-helical
47 structure (Kronfel et al., 2013; Overkamp et al., 2014; Zhao et al., 2017; Zhou et al., 1992, 2014).
48 These enzymes differ in their protein scaffolds and confer different substrate specificities for
49 pigments, acceptor proteins and target residues. In essence, they all catalyze the same chemical
50 reaction resulting in formation of a thioether linkage between the sulfhydryl group of a cysteine
51 residue in phycobiliprotein and the C3¹ atom in the A-ring of a bilin pigment (Scheer and Zhao,
52 2008; Schirmer et al., 1987) (Fig. S1B). It has been proposed that this reaction involves transient
53 association between the bilin pigment and lyase mediated by a nucleophilic residue such as
54 histidine or cysteine from the lyase via the C10 atom (Stumpe et al., 1993; Tu et al., 2009; Zhao et
55 al., 2000). However, the chemical identity of catalytic residue(s) and their roles in the enzymatic
56 action remain elusive (Gasper et al., 2017; Overkamp et al., 2014; Scheer and Zhao, 2008; Tu et
57 al., 2009; Zhao et al., 2017, 2000; Zhou et al., 2014). For the E/F type bilin lyases, there has been
58 only one structure solved for heterodimeric CpcE/F (Zhao et al., 2017).

59 To understand the molecular mechanism underlying the bilin ligation reaction, we
60 investigate a single-chain E/F type bilin lyase MpeQ that incorporates PUB at the Cys83 site of

61 the phycoerythrin-II α -subunit, a phycobiliprotein encoded by the *mpeBA* operon in marine
62 *Synechococcus* (Grébert et al., 2021). The newly identified MpeQ is involved in Type IV
63 chromatic acclimation (CA4) process, a widespread phenomenon where marine *Synechococcus*
64 vary the molar ratio between the blue-light absorbing phycourobilin (PUB) and the green-light
65 absorbing phycoerythrobilin (PEB) in phycobiliproteins depending on the ambient light color
66 (Everroad et al., 2006; Grébert et al., 2018, 2021; Palenik, 2001; Sanfilippo et al., 2019a; Shukla et
67 al., 2012). The CA4 phenomenon is attributed to light-regulated expression of a PEB lyase or
68 lyase-isomerase encoded in a small CA4-conferring genomic island and competing with a
69 constitutively expressed homolog (lyase-isomerase or lyase) encoded in the large genomic
70 region involved in the biosynthesis of phycobilisome rods (Grébert et al., 2021; Mahmoud et al.,
71 2017; Sanfilippo et al., 2016, 2019a; Shukla et al., 2012). Compared to PEB lyases, bilin lyase-
72 isomerases such as MpeQ have acquired an additional ability to isomerize PEB to PUB during
73 the ligation reaction (Blot et al., 2009; Grébert et al., 2021; Shukla et al., 2012) (Fig. S1C).
74 However, the molecular basis for this isomerase action is not known. MpeQ thus offers a superb
75 model system not only for dissecting the molecular mechanism of bilin lyases but also for
76 addressing how these highly homologous enzymes attach different pigments to the same site of
77 phycobiliproteins using the same substrate (Grébert et al., 2021; Humily et al., 2013; Sanfilippo
78 et al., 2019b).

79 This work reports the crystal structure of MpeQ, a representative bilin lyase-isomerase from
80 the E/F family. This structure has revealed a hitherto unknown “question-mark” architecture
81 and active site geometry that call for a major revision of the current model for the E/F bilin
82 lyases (Zhao et al., 2017). Based on structural analyses, site-directed mutagenesis and enzyme
83 assays, we identify key residues responsible for the lyase and isomerase activities, respectively.
84 We propose a general tyrosine-mediated reaction scheme for bilin lyases in the E/F family and
85 beyond. We further advance a model for the enzyme-substrate complex that elucidates how
86 MpeQ accommodates both the bilin and protein substrates to catalyze the ligation reaction
87 while employing stereoselectivity to confer the isomerase activity. These findings provide a
88 structural framework and mechanistic insights into one of the most important enzymatic
89 reactions in photosynthesis.

90 Results

91 *MpeQ adopts a “question-mark” architecture*

92 We have determined the crystal structure of MpeQ at 2.5 Å resolution by the single-wavelength
93 anomalous diffraction (SAD) method using Se-methionine-derivatized crystals (Table 1). In
94 addition to the N-terminal histidine tag, all 398 residues of MpeQ have been accounted for in
95 the electron density map with two protein molecules in an asymmetric unit. The α -solenoid
96 structure of MpeQ adopts a “question-mark” scaffold in which a total of 23 α -helices are super-
97 coiled in a right-handed manner (Fig. 1A; Fig. 1D). The N-terminal (aa. 14-255) and C-terminal
98 (aa. 274-398) domains connected by an extended linker are denoted the E and F domains,
99 respectively, corresponding to the CpcE and CpcF subunits of a prototypical CpcE/F bilin lyase
100 from *Nostoc sp.* PCC 7120 (Zhao et al., 2017). However, the overall architecture of MpeQ starkly
101 contrasts with the crescent-shaped structure proposed for CpcE/F (Zhao et al., 2017) (Fig. 2A,B).
102 Interestingly but not surprisingly, a revised CpcE/F model based on an alternative arrangement
103 of the E and F subunits presents a heterodimeric structure highly comparable to that of MpeQ
104 (Fig. 1B; Fig. 2). Remarkably, four out of five strictly conserved residues (Tyr76, Pro108, Arg112
105 and Trp151; numbering in CpcF) between heterodimeric CpcE/F and single-chain E/F lyases
106 such as MpeQ line up in the tight turns between helices facing an enclosed cavity in the
107 question-mark architecture in contrast to the published CpcE/F structure. This suggests that the
108 MpeQ structure, but not the reported structure of CpcE/F (Zhao et al., 2017), represents the
109 biological relevant protein framework for bilin lyases in the E/F family.

110

111 *Active site and key residues responsible for lyase activity*

112 The E and F domains of MpeQ are roughly perpendicular to each other in terms of the
113 overall helical direction (Fig. 1C). The E domain constitute the “stroke” part of the question
114 mark while the F domain corresponds to the “dot” part (Fig. 1A, Fig. 1D). Besides the linker,
115 the E and F domains are bridged by a long N-terminal extension (aa. 1-13) enclosing a large
116 interior chamber (Fig. 1A, Fig. 3). In the crystal lattice, two molecules of MpeQ dimerize via
117 their N-terminal extensions tethered as two anti-parallel β -strands traversing each other’s

118 chamber (Fig. 1E). The enclosed chamber is open on both sides although one side is notably
119 wider than the other (Fig. 1C). Arg/Lys residues pointing towards the chamber interior form
120 large positively charged surface patches on the ceiling and wall areas (Fig. 3A; Fig. 4A). Along
121 a shallow cleft between the E and F domains, some highly conserved residues as well as those
122 hallmark residues that distinguish the lyases and lyase-isomerases are clustered at the chamber
123 floor constituting a putative active site (Fig. 4B; Fig. S2). This site marked by Tyr318, Lys353 and
124 Glu285 is exposed to the wide side of the chamber (Fig. 3B). In crystal lattice, Tyr318 and Lys353
125 engage close interactions with the E' domain of a neighboring molecule via an interface with
126 buried surface area of 2426 Å² (PISA) (Fig. S3A). In solution, however, MpeQ is monomeric
127 according to the elution profile of size exclusion chromatography (Fig. S3B), which makes this
128 site fully accessible for substrate binding.

129 To confirm the active site, we made several single mutations in this cluster (Fig. 4A, B) and
130 examined how they affect the lyase and/or isomerase activities in MpeQ. By co-expressing
131 MpeQ in *E. coli* with two other plasmids that produce the MpeA and PEB substrates, we
132 measured the normalized enzyme activities by detecting the chromophorylated products,
133 MpeA-PUB or MpeA-PEB, using Zn-fluorescence (Berkelman and Lagarias, 1986) and/or
134 absorption spectra of the purified MpeA (Fig. 4C-E; Fig. 5). The lyase activity was completely
135 abolished in the single mutants of Y318F, Y318A and K353A, with no detectable
136 chromophorylation of MpeA (Fig. 4C; Table S1). The E285A mutant showed significantly
137 reduced lyase activity while the S222A, S224P and W383F variants retained the lyase-isomerase
138 activities with MpeA-PUB produced at varying levels lower than WT (Fig. 4C; Fig. 5). These
139 mutational data corroborate our structural findings, supporting that Tyr318 and Lys353 play
140 essential roles in catalysis and/or substrate binding (Fig. 3B).

141 We also made single mutants on several Arg residues lining the chamber interior (Fig. 4A).
142 When Arg71, Arg79, Arg135 or Arg143 was individually mutated to Asp, no MpeA
143 chromophorylation was detected (Fig. 5C). Similar results were observed for the R228A and
144 R198A variants (Fig. 4C; Fig. S2; Table S1). The R71A, R79A, R228D, and R198K variants
145 showed low-level production of MpeA-PUB (Fig. 5; Table S1). These results suggest that

146 although Arg71, Arg79, Arg198 and Arg228 are not directly involved in catalysis, they are likely
147 to act as protein anchors that help stabilize the bilin pigment via interactions with the
148 negatively charged propionates.

149

150 *Steric factors underlying isomerase activity*

151 To determine the molecular basis of the isomerase activity, we analyzed a collection of
152 protein sequences of highly homologous E/F type lyases and their counterpart lyase-isomerases.
153 Specifically, sequence comparisons of the MpeW and MpeY lyases with the MpeQ and MpeZ
154 lyase-isomerases retrieved from 106 marine *Synechococcus* strains or single-amplified genomes
155 (Dataset S1) highlighted nine residues swapped between the lyases and lyase-isomerases (Fig.
156 S2 and Dataset S2). To examine whether these “swap” residues contribute to the gain of the
157 isomerase function, we individually substituted five such residues at the active site of MpeQ
158 with their counterparts in the PEB lyase MpeW (Fig. S4). All five single mutants (A100T, S224P,
159 V319G, T320A, and Y323Q) largely retained their lyase activities (Fig. 5; Fig. S4). Only the
160 V319G variant produced an additional absorption peak near 550 nm indicating the formation of
161 MpeA-PEB (Fig. S4B). In other words, the Val→Gly substitution led to partial loss of the
162 isomerase activity in MpeQ (Fig. 4D, Fig. S4B).

163 Val319 is located right next to the catalytic Tyr318. We systematically explored possible
164 steric effects on the lyase-isomerase activity by replacing Val319 with Gly, Ala, Leu and Phe,
165 individually (Fig. 4C, 4D). With increasingly bulkier side chains at the position of Val319, these
166 mutants exhibited a striking color tuning effect on the MpeA product (Fig. 5A). The V319G
167 variant with the smallest side chain showed a reddish color indicating a low PUB:PEB ratio in
168 MpeA. The V319A variant with a slightly larger side chain has an orange color suggesting a
169 higher PUB:PEB ratio. With valine or leucine present at the position 319, the yellow-colored
170 sample suggests exclusive formation of MpeA-PUB. Larger side chains at this position appear
171 to interfere with the bilin attachment. As such, the lyase activity of the V319L variant was
172 reduced by about 50%, while the V319F variant barely showed any detectable
173 chromophorylated MpeA (Fig. 4C,D; Fig. 5).

174 It is not surprising that the single swap mutations (A100T, S224P, T320A and Y323Q) did
175 not directly affect the isomerase function as these sites are further away from the catalytic
176 Tyr318 (Fig. 4B; Fig. S4B). However, when these substitutions were made in combination with
177 V319G, they significantly altered the PUB:PEB ratio in the MpeA products (Fig. 4E). The single
178 mutant V319G exhibited a PUB:PEB ratio of 10:1 while the double mutant V319G/T320A
179 lowered the PUB:PEB ratio to 3:1. This ratio was further reduced to 1.5:1 in the triple mutant
180 V319G/T320A/Y323Q. Two quintuple mutants (MpeQ5.1: A100T/V319G/T320A/Y323Q/T352A;
181 MpeQ5.2: S224P/V319G/T320A/Y323Q/T352A) evidently switched to make more MpeA-PEB
182 than MpeA-PUB resulting in the PUB:PEB ratios of 1:1.5 and 1:3, respectively (Fig. 4E). It is
183 noteworthy that higher PUB:PEB ratios coincide with bulkier side chains at these swap sites
184 (Fig. 4A), suggesting that the formation of MpeA-PEB requires a larger reaction volume at the
185 active site. To this end, exclusive attachment of PEB to MpeA-Cys83 by MpeW arises from
186 collective steric effects from multiple swap sites beyond Val319 as exemplified by the sequence
187 differences between MpeQ and its counterpart lyase MpeW (Fig. S2 and Fig. S5).

188

189 *Tripartite model and reaction mechanism*

190 Identification of the active site allowed us to explore the modes of substrate binding to MpeQ.
191 We first examined structural flexibility of both MpeQ and MpeA using molecular dynamics
192 (MD) simulations (Fig. S6). A 72-ns MD simulation on MpeQ revealed two distinct dispositions
193 of the long N-terminal extension (~aa. 1-15) (Fig. S6A). In a “latch-on” position similar to that in
194 the crystal structure, the N-terminal extension encloses the active site chamber while the “latch-
195 off” position allows better accessibility to the active site. The RMSD distance matrix derived
196 from 100 simulated structures on the trajectory reveals three rigid bodies in the α -solenoid
197 structure (aa. 30-97 and 117-237 in the E domain; aa. 250-398 in the F domain), which roughly
198 correspond to the ceiling, wall and floor sections of the chamber, respectively (Fig. S6B; Fig. 3B).
199 The MD trajectory shows close coupling between the wall and floor sections while the ceiling
200 section exhibits higher mobility likely due to its proximity to the flexible N-terminal extension.
201 We also performed the MD simulations on an apo-MpeA model (Swiss-model) (Biasini et al.,

202 2014), which also reveals two distinct clusters of conformations in the cap-like loop (aa. 63-81)
203 shielding MpeA-Cys83 from the molecular surface (Fig. S6C; Fig. S1B). The “cap-on”
204 conformation resembles a well-folded α -subunit while the “cap-off” structure exposes MpeA-
205 Cys83 for docking at the active site of MpeQ.

206 Based on surface complementarity, protein flexibility and active site geometry, we present a
207 tripartite model in which MpeQ is complexed with the PEB and MpeA substrates (Fig. 6). First
208 of all, the active site in monomeric MpeQ is fully accessible for substrate binding (Fig. 6A). Apo-
209 MpeA is expected to approach the active site chamber from the wide side. This docking is likely
210 facilitated by induced fit involving both MpeA and MpeQ (Fig. 6A; Fig. S6). Specifically, the N-
211 terminal extension of MpeQ would adopt a “latch-off” position allowing proper positioning of
212 MpeA-Cys83 at the active site (Fig. 6B). With the bilin binding pocket of MpeA open to the
213 chamber interior, the active site of MpeQ is accessible to the PEB chromophore from the narrow
214 side (Fig. 6C). This model features PEB in its final binding cleft of MpeA (Fig. 6B,C) with its A-
215 ring sandwiched between MpeQ-Tyr318 and MpeA-Cys83.

216 At the active site, the A-ring assumes an extended *anti*-conformation as in the final product.
217 This disposition presents the C3=C3¹ double bond, the target site for ligation, to the hydroxyl
218 group of Tyr318 while its lactam group is positioned to engage direct interactions with
219 Lys353/Glu285 (Fig. 6D; Fig. S7A). This active site geometry fully agrees with the mutagenesis
220 data (Fig. 4C) supporting that Tyr318 is critical for catalyzing the ligation reaction while
221 Lys353/Glu285 are important for conferring specific orientation of the A-ring. In addition, this
222 model places Val319 right next to the *R*-configured C2 atom, which perfectly explains the
223 remarkable steric effects of Val319 on the lyase and isomerase activities (Fig. 4D).

224 This tripartite model clearly points to distinct chemical origins for the lyase and isomerase
225 actions of MpeQ. To address whether the PEB→PUB isomerization occurs before or after bilin
226 ligation to MpeA-Cys83, we performed a simple experiment by co-expressing MpeQ with a
227 PEB-producing plasmid and the MpeA-C83A variant. We reason that since MpeA-C83A is
228 unable to covalently attach the bilin pigment, the ligation reaction cannot complete thereby
229 allowing us to detect the isomerization product PUB if it is formed in the absence of ligation.

230 We measured absorption spectra from the load, flowthrough and eluted fractions of the affinity
231 chromatography column that was used to separate bulk proteins and free pigments from the
232 His-tagged MpeQ and MpeA (Fig. S8A). Both the load and flowthrough samples showed the
233 absorption band around 560 nm supporting that the PEB substrate was sufficiently available in
234 this co-expression system. While PUB was barely detectable in the bulk samples of load and
235 flowthrough, the eluted fraction exhibits a small yet clear peak at ~495 nm (Fig. S8A),
236 suggesting that PUB is formed in the absence of ligation to MpeA-Cys83 and remains associated
237 with MpeQ and/or MpeA.

238 Taken together, we advance a reaction scheme to elucidate a tyrosine-mediated bilin
239 ligation and isomerization catalyzed by MpeQ. In essence, the net reaction catalyzed by a bilin
240 lyase is a nucleophilic addition of the sulfhydryl (-SH) group of cysteine in the phycobiliprotein
241 substrate to the C3=C3¹ double bond of the bilin substrate (Fig. 7). In the final product of MpeQ,
242 a thioether bond is formed between the C3¹ atom and the sulfur atom of MpeA-Cys83 while the
243 C3=C3¹ double bond is reduced. Such a nucleophilic addition to alkene requires a catalytic
244 group to polarize the C=C double bond. Our structural and mutagenesis data strongly suggests
245 that the hydroxyl group of MpeQ-Tyr318 serves this role. We propose that in the first step of the
246 reaction, Tyr318 activates the bilin substrate resulting in a polarized C3=C3¹ double bond (Fig.
247 7). This bilin activation step is critical as no lyase activities were detected in the Y318F or Y318A
248 mutants when this hydroxyl group is absent, or in V319F where the catalytic hydroxyl group
249 cannot get close enough to the C3 atom (Fig. 4). In MpeW where the active site poses no major
250 steric hinderance, the bilin activation would proceed to the nucleophilic addition reaction
251 giving rise to the MpeA-PEB product. In MpeQ, on the other hand, the MpeA-PEB formation is
252 not allowed due to incompatibility between the active site residue (Val319) and the chiral
253 centers at the C3 and/or C2 positions in PEB. The addition reaction cannot take place unless
254 such a conflict is resolved in a rearrangement reaction where PEB→PUB isomerization
255 effectively shifts the chiral centers from the C2/C3 to C4/C5 positions (Fig. 6; Fig. 7). Selected by
256 evolution, the side chain of Val319 is small enough to permit the catalytic action of Tyr318, yet
257 bulky enough to exclude the PEB attachment. Similar to retinoid isomerase RPE65 (Kiser et al.,

258 2015), MpeQ exemplifies a general molecular strategy to confer the isomerase activity via
259 stereoselectivity of the active site geometry (Fig. 7).

260

261 **Discussion**

262 Prior to this study, no crystal structure has been reported for any bilin lyase-isomerase; and
263 CpcE/F (a PCB lyase from *Nostoc sp.* PCC 7120) was the only structure available for the E/F type
264 bilin lyases (Zhao et al., 2017). However, the interface between CpcE and CpcF proposed by
265 Zhao et al. is not present in the structure of single-chain MpeQ (Fig. 1E; Fig. 2B)(Zhao et al.,
266 2017). In fact, no consistent pattern or interface other than the question-mark arrangement is
267 found between the crystal structures of MpeQ and CpcE/F (Fig. 2D). Therefore, it is unlikely
268 that the crescent-shaped architecture (PDB ID: 5N3U) represents the biological assembly or
269 functional unit of CpcE/F despite its high resolution. Because the crystal lattice of CpcE/F
270 presents two crystallography-equivalent yet distinct dimeric arrangements (Fig. 2), it is
271 susceptible to misinterpretation of the biologically relevant heterodimer. The single-chain
272 MpeQ, on the other hand, is not prone to such ambiguity especially when all residues have been
273 accounted for in the electron density map. Given the similarities in protein sequence and
274 secondary structure (Fig. 1, Fig. S5), we posit that the E/F family bilin lyases including the
275 single-chain and heterodimeric enzymes share a common structural framework represented by
276 the question-mark architecture MpeQ.

277 We also note remarkable similarities in the active site geometry between MpeQ and other
278 types of bilin lyases despite their distinct protein scaffolds (Fig. S7A). The catalytic triad of
279 MpeQ is not only found in the revised structure of CpcE/F but also in CpcT/CpeT (PDB ID:
280 4O4O, 4O4S, 5HI8) and CpcS/CpeS (PDBID: 3BDR and 4TQ2) (Gasper et al., 2017; Kronfel et al.,
281 2013; Overkamp et al., 2014; Zhao et al., 2017; Zhou et al., 2014). Specifically, they all feature a
282 conserved tyrosine at the rim of a large protein cavity where the phycobiliprotein and bilin
283 substrates are supposed to bind. And CpcF-Tyr76 is perfectly aligned with MpeQ-Tyr318 (Fig.
284 S5; Fig. S7A). Beyond its locality, this conserved active-site tyrosine seems to play a similar role.
285 Just like MpeQ-Y318F, no bilin attachment was detected in the single mutants of CpcF-Y76A

286 and CpcT-Y65F (Zhao et al., 2017; Zhou et al., 2014). We speculate that the proposed tyrosine-
287 mediated reaction mechanism also applies to bilin lyases in the T and S/U clans (Fig. 7).

288 Regarding the origin of the isomerase activity, we have considered two scenarios. *First*, the
289 E/F lyase and isomerase activities are resulted from two distinct enzymatic reactions involving
290 different catalytic residues, and these swap residues simply play a role in tipping off the balance
291 between these two reactions. *Second*, the isomerase activity is an “add-on” function associated
292 with the lyase reaction catalyzed by Tyr318, which arises from collective steric effects from the
293 swap residues near the active site. The first scenario implies that the catalytic residues
294 responsible for the isomerase activity are near the C4=C5 double bond of PEB, and they are
295 distinct from Tyr318. However, our current docking model shows that the only residues within
296 a 5-Å radius of the C4=C5 double bond of PEB are MpeQ-Tyr318 (3.5 Å) and MpeA-Cys83 (4.4
297 Å) (Fig. 6). In addition, it is unlikely that MpeA-Cys83 is directly involved in isomerization (Fig.
298 S8A). On the other hand, a bioinformatic study on isomerases (Martinez Cuesta et al., 2014;
299 Martínez Cuesta et al., 2016) suggested that the isomerase and lyase activities are evolutionarily
300 related due to their similarity in chemistry, which goes along the second scenario. We propose
301 that highly homologous MpeQ and MpeW share the similar reaction mechanism for their lyase
302 actions while the substrate transformation between the activation and ligation steps is directly
303 influenced by the geometry and steric factors of the active site (Fig. 7).

304 Our mutagenesis data show that the steric factor at the position of Val319 directly influences
305 the isomerase activities (Fig. 4C,D, Fig. S8B,C). Consistently, among the 106 sequences of the
306 MpeQ/W/Y/Z family retrieved from marine *Synechococcus* strains or SAGs (Grébert et al., 2021;
307 Sanfilippo et al., 2019a; Shukla et al., 2012), all enzymes with the isomerase function have a
308 valine residue at the this position, whereas all lyases have a glycine (Fig. S2 and Dataset S2).
309 This includes MpeZ, a lyase-isomerase shown to be involved in the CA4 strain RS9916
310 (Sanfilippo et al., 2019a; Shukla et al., 2012), which features Val next to the catalytic Tyr321
311 while its counterpart PEB lyase MpeY has Gly instead. However, the Val→Gly substitution
312 alone is not sufficient (even in combination with some swap sites) to confer exclusive PEB
313 attachment in MpeQ (Fig. 4E). Evidently, a complete switch from MpeQ to a MpeW-like PEB
314 lyase requires contributions from other swaps sites, which collectively confer an active site

315 spacious enough to accommodate the C2/C3 chiral centers of a non-planar A-ring during the
316 ligation reaction (Fig. 4E; Fig. S2).

317 It is worth noting that the heterodimeric lyase-isomerase PecE/F lacks tyrosine
318 corresponding to Tyr318. Homology modeling of PecE/F places its signature HC motif right at
319 the space occupied by MpeQ-Tyr318 (SwissModel) (Fig. S7B). In addition, the single mutations
320 in the HC motif abolished the lyase activity of PecE/F (Zhao et al., 2017). It is possible that
321 PecE/F employs histidine instead of tyrosine as a catalytic residue for substrate activation (Tu et
322 al., 2009) (Fig. S7B). This exception lends further support to the active site geometry proposed
323 for the E/F bilin lyases.

324 The active site geometry as illustrated in our tripartite model not only supports the
325 respective roles of Tyr318 and Val319 but also has other implications. It is plausible that
326 Lys353/Glu285 from the catalytic triad contribute to stabilizing the A-ring in a *5,anti*-
327 conformation via hydrogen bond interactions with the lactam group (Fig. S7A; Fig. 6D). This
328 feature hints on the hitherto unknown structural basis for the *5,anti*-conformation of bilins
329 exclusively found in phycobiliproteins, in contrast to the *5,syn*-conformation observed in
330 phytochromes (30, 31). In other words, the characteristic *5,anti*-conformation arises from
331 substrate selectivity of the bilin lyases. As such, altered protein-chromophore interactions in the
332 K353A and E285A mutants resulted in poor bilin attachment (Fig. 4C). Furthermore, positively
333 charged residues (e.g. Arg79/Arg228) are positioned close to the bilin propionates in the active
334 site chamber (Fig. 6D); such ionic interactions are reminiscent of the bilin binding modes
335 observed in phytochromes and other bilin-binding proteins (Wagner et al., 2005; Yang et al.,
336 2008). Not surprisingly, the R79D and R228D mutants adversely affected the lyase activities of
337 MpeQ (Fig. 5).

338 Needless to say, further studies are needed to validate the atomic details of the proposed
339 reaction scheme, which are beyond the scope of this work. To this end, our attempts to directly
340 introduce substrates to the MpeQ crystals by soaking have been unsuccessful because the
341 current crystal form blocks any substrate binding to the active site (Fig. 1E, Fig. S3A). MpeQ
342 complexed with pigment alone is also deemed unfeasible for co-crystallization due to the large

343 size difference between the active site chamber and PEB. We reason that both substrates (MpeA
344 and PEB) are needed to assemble stable complexes for structural studies by crystallography or
345 cryoEM. To accumulate homogeneous species for static studies, the ligation reaction can be
346 arrested by using either an inactive enzyme such as MpeQ-Y318F or a disabled substrate such
347 as MpeA-C83A. For example, MpeQ-Y318F can be used to capture the “pre-activation” state,
348 while MpeA-C83A allows trapping of later “pre-ligation” intermediate(s). It will be intriguing
349 to see whether bilin activation mediated by Tyr318 involves any formation of a tyrosine radical
350 and/or transient covalent intermediates.

351 Why do some E/F lyases such as CpcE/F employ two polypeptide chains while other
352 homologous enzymes such as MpeQ achieve the same function using a single protein
353 chain? Biswas et al. suggested that the CpcE/F family evolved after the CpcS/U family of lyases
354 because some E/F members have the ability to remove bilins (Biswas et al., 2011). We speculate
355 that E/F heterodimers have advantages in allowing mix-and-match of different E-like and F-like
356 subunits. It is noteworthy that the gene fusion events seem to occur more frequently in
357 phycoerythrin-containing organisms. For example, RpcG (a PEB lyase-isomerase) is a fusion
358 protein with its N-terminal and C-terminal domains closely related to PecE and PecF,
359 respectively (Blot et al., 2009) while CpeY (a PEB lyase for α -phycoerythrin I) appears to result
360 from the fusion of CpcE-like and CpcF-like domains (Biswas et al., 2011). These examples
361 provide clues about how the E/F clan shifted from the mainly heterodimers to single-chain
362 enzymes, the latter conformation being the rule for all phycoerythrin-related lyases and lyases-
363 isomerases.

364 Taken together, results from this study have revealed distinct chemical origins for the lyase
365 and isomerase actions of bilin lyase-isomerase in the E/F family. Our structural and
366 mutagenesis data strongly support that stereoselectivity of the active site geometry plays a
367 critical role in conferring the isomerase activity of MpeQ. We further propose a tyrosine-
368 mediated reaction scheme shared among different types of bilin lyases. This work thus presents
369 a structural and mechanistic framework that has advanced the fundamental understanding of
370 critical biochemical reactions underlying the phycobilisome biogenesis.

371 **Acknowledgement**

372 We thank the staff of the Life Science Consortium Access Team (LS-CAT) at the Advanced
373 Photo Source (APS) for support in X-ray diffraction data collection. Use of the LS-CAT Sector 21
374 is supported by the Michigan Economic Development Corporation and the Michigan
375 Technology Tri-Corridor under Grant 085P1000817. Use of the Advanced Photon Source is
376 supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences,
377 under Contract No. DE-AC02-06CH11357. This work is supported by the National Science
378 Foundation grant MCB 2017171 to WS and MCB 2017274 to XY, the National Institutes of
379 Health grant R01EY024363 to XY and the French national agency for research (ANR) program
380 EFFICACY (ANR-19-CE02-0019) to FP.

381

382 **Author contributions**

383 Conceptualization, XY and WMS; Methodology, IK, KLJ, WMS and XY; Investigation, IK, KLJ,
384 LB, IT, KDW, CW, WMS and XY; Analysis, IK, KLJ, TG, FP, WMS and XY; Writing – Original
385 Draft, XY; Writing – Review and Editing, XY, WMS, TG and FP; Funding Acquisition, XY, WMS
386 and FP; Resources, IK and KLJ; Supervision, XY and WMS.

387

388 **Declaration of interests**

389 The authors declare no competing interests.

390

391 **Inclusion and Diversity**

392 One or more of the authors of this paper self-identifies as an underrepresented ethnic minority
393 in science.

394

395 **Figure Legends**

396 **Figure 1. Protein architecture, topology and crystal packing of MpeQ.** A) Ribbon diagram of
397 MpeQ shows an α -solenoid “question-mark”-shaped structure with distinct E (aa. 14-255) and F
398 (aa. 274-398) domains. A long N-terminal extension loops back to the F domain as a gate to the
399 large interior chamber. The structure is colored in rainbow from the N-terminus (blue) to C-

400 terminus (red). **B**) The revised CpcE/F heterodimer (PDBID: 5n3u) shows a protein architecture
401 similar to that of MpeQ. **C**) A side view of MpeQ relative to panel A shows the wide and
402 narrow openings to the chamber at opposite sides of MpeQ. The openings are about 8.5 Å on
403 the narrow side (between Lys81 and Lys238) and 17 Å on the wide side (between Arg285 and
404 Ile132), respectively. **D**) Topological diagram of the MpeQ structure is shown in a perspective
405 to facilitate comparisons with CpcE/F (Zhao et al 2017). A total of 23 helices shown in circles
406 and rods highlight roughly perpendicular directions of helices in the E and F domains (shaded
407 areas). **E**) Crystal packing of MpeQ shows two distinct packing interfaces. Interface I (buried
408 surface area of 2426 Å²) is much larger than interface II (buried surface area of 400 Å²).

409 **Figure 2. Crystal packing and dimer assemblies.** **A**) Crystal packing of the CpcE/F structure
410 shows two possible dimer assemblies of the E (blue) and F (green) domains, namely, the E/F
411 assembly or E/F' assembly. In crystallography, E and F are equivalent to E' and F', respectively.
412 **B**) The CpcE/F assembly proposed by Zhao et al. (PNAS, 2017) (top) shows a crescent-shaped
413 protein architecture (bottom). **C**) The revised CpcE/F' assembly reveals a question-mark-
414 shaped protein architecture (bottom). **D**) Comparisons between single-chain MpeQ (in rainbow)
415 and two CpcE/F assemblies (E in blue; F in green). Magenta spheres mark the conserved
416 residues (Tyr318, Pro349, Lys353 and Trp383: numbering in MpeQ; Tyr76, Pro108, Arg112 and
417 Trp151: numbering in CpcF). See also Fig. S5.

418 **Figure 3. The active site chamber of MpeQ.** **A**) The electrostatic surface viewed from the
419 narrow (left) and wide (right) sides. The chamber interior is largely positively charged where
420 Arg/Lys residues are clustered in the ceiling, wall and floor regions of the chamber. **B**) The
421 catalytic triad at the floor consists of Tyr318, Lys353 and Glu285 (green spheres) located in the
422 loops between helices in the F domain.

423 **Figure 4. Site-directed mutagenesis in the active site.** **A**) Sites of mutation are highlighted in
424 stick models: arginines (cyan), swap residues (magenta) and catalytic residues (green spheres).
425 **B**) Conserved (green) and swap (magenta) residues form a surface patch at the chamber floor at
426 the E/F interface. **C**) Coomassie blue staining (top) and Zinc fluorescence (bottom) of the same
427 SDS-PAGE gel of MpeA obtained via co-expression with MpeQ. **D**) Normalized absorption

428 spectra of MpeA obtained from the MpeQ mutants at Val319. **E)** MpeQ variants carrying one or
429 more swap mutations show varying PEB:PUB ratios. Q2: V319G/T320A; Q3:
430 V319G/T320A/Y323Q; Q5.1: A100T/V319G/T320A/Y323Q/T352A and Q5.2:
431 S224P/V319G/T320A/Y323Q/T352A. MpeW is the PEB lyase counterpart of MpeQ.

432 **Figure 5. Lyase and lyase-isomerase activities of the MpeQ variants in an *E. coli* co-**
433 **expression system. A)** The distinct sample color of MpeA co-expressed with MpeQ serves as a
434 good indicator of the lyase and lyase-isomerase activities of MpeQ. The yellow color of MpeA-
435 PUB is attributed to the lyase-isomerase activity while the red/orange color of MpeA-PEB is a
436 result of the lyase activity as shown for V319G. **B)** Normalized absorption spectra of MpeA co-
437 expressed with various single mutants of MpeQ. **C)** Relative activities of the MpeQ variants.
438 Absorbance at 495 nm normalized by relative band intensities of MpeA (Fig. 4C) is used to
439 quantify the lyase activity. All relative activities are obtained in reference to that of wild type
440 (WT) MpeQ (set to 100%).

441 **Figure 6. A docking model for substrates at the active site. A)** The active site of MpeQ (green)
442 is accessible from both sides of the chamber. MpeA (gray) approaches the active site from the
443 wide opening while the PEB substrate enters from the narrow side (blue arrow). **B)** Viewed
444 from the narrow side, the bound bilin (PEB in magenta; PUB in yellow) is docked at the final
445 binding cleft in the MpeA structure (rainbow ribbon diagram). **C)** Viewed from the wide side,
446 the catalytic Tyr318 is poised to activate the A-ring, which causes a steric clash between PEB
447 and Val319. MpeQ resolves this conflict via PEB→PUB isomerization, which presents the ring A
448 to MpeA-C83 for ligation. **D)** In this front view, PEB is docked with its ring A/B side lying in the
449 cleft between the E and F domains while the propionate groups are stabilized by Arg79 and
450 Arg228.

451 **Figure 7. Proposed reaction scheme of bilin addition reaction catalyzed by bilin lyase MpeW**
452 **and lyase-isomerase MpeQ.** In the first step of the reaction, the bilin substrate PEB is activated
453 by a catalytic tyrosine in both MpeW and MpeQ resulting in a polarized C3=C3¹ double bond.
454 In MpeW, Gly319 next to the catalytic Tyr318 is compatible with the non-planar A-ring
455 conformation of PEB. The activated PEB is thus allowed to proceed to the nucleophilic addition

456 reaction forming a thioether bond between MpeA-Cys83 (blue) and the A-ring C3¹ atom. In
 457 MpeQ, the side chain of Val319 renders a steric conflict with the A-ring of PEB. The ligation
 458 reaction only proceeds when this conflict is resolved via the PEB→PUB isomerization resulting
 459 in the MpeA-PUB product.

460 **Table 1. Crystallography Data Collection and Refinement Statistics.**

461

Structure	MpeQ SeMet	MpeQ SeMet
Space group	C2	C222 ₁
Cell parameters		
<i>a, b, c</i> (Å)	83.84, 173.25, 114.42	136.67, 173.05, 113.7
α, β, γ (°)	90, 126.08, 90	90, 90, 90
Residues	1-398	1-398
Chain	A, B	A, B
Water		
Diffraction data		
X-ray source	21-ID-F, APS	21-ID-F, APS
Methods	Monochromatic	Monochromatic
	100 K	100 K
Wavelength (Å)	0.97872	0.97872
Resolution (Å)	56-2.5 (2.54-2.50)	50-2.90 (2.95-2.90)
<i>R</i> _{merge}	0.101 (0.605)	0.082 (0.761)
Completeness (%)		
All	99.5 (99.7)	99.8 (95.8)
Anomalous	97.2 (84.7)	99.8 (95.5)
Redundancy	3.7 (3.5)	3.9 (1.6)
<i>I</i> / σ (<i>I</i>)	7.4 (1.3)	4.7 (1.0)
Refinement		
Resolution (Å)	20-2.50 (2.79-2.50)	39.3-2.95 (3.01-2.95)
<i>R</i>	0.239 (0.279)	0.242 (0.361)
<i>R</i> _{free}	0.284 (0.329)	0.298 (0.303)
RMSD		
Bond length (Å)	0.009	0.008
Bond angle (°)	1.013	1.248
Average <i>B</i> (Å ²)	28.63	98.9
Ramachandran		
Favored (%)	96.6	96.1
Allowed (%)	2.0	3.3
Disallowed (%)	0.4	0.55
PDB entry	7MC4	7MCH

463 **STAR METHODS**

464 **RESOURCE AVAILABILITY**

465 *Lead Contact:*

- 466 • Xiaojing Yang (xiaojing@uic.edu).

467 *Material availability*

- 468 • Further information and requests for resources and reagents should be directed to and will
469 be fulfilled by the Lead Contact, Xiaojing Yang (xiaojing@uic.edu).

470 *Data and code availability*

- 471 • The coordinates and structure factor amplitudes of MpeQ in the space group C2 and C222₁
472 have been deposited in the Protein Data Bank under the accession numbers 7MC4 and
473 7MCH, respectively.
- 474 • This paper does not report original code.
- 475 • Any additional information required to reanalyze the data reported in this paper is available
476 from the lead contact upon request.

477

478 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

479 The *mpeQ* and *mpeW* genes from the CA-B strain *Synechococcus* A15-62 were cloned into the
480 pCDF-Duet 1 and pET-Duet 1 vectors using the restriction sites SacI/NotI (or BglII/EcoRV) and
481 NcoI/HindIII, respectively. MpeQ carrying a N-terminal 6xHis affinity tag was over-expressed
482 in *E. coli* BL21(DE3) by isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM) induction followed
483 by overnight shaking at 18°C.

484 **METHOD DETAILS**

485 *Protein purification and site-directed mutagenesis*

486 After cell harvesting and cell lysis, His-tagged MpeQ protein was extracted by Co²⁺-affinity
487 chromatography and further purified by anion exchange chromatography (HiTrap Q- HP
488 column). The selenomethionine (SeMet)-derivatized MpeQ protein was prepared according to
489 the standard protocols (Doublé, 1997) and purified using the same protocols as for the native
490 MpeQ.

491 Site-directed mutagenesis was carried out using combined overlapping polymerase Chain
492 Reaction (COE-PCR) as previously described with modifications (Hussain and Chong, 2016).
493 Mutagenic primers were created using the Thermo Fisher Scientific GeneArt primer and
494 construct design tool (<https://www.thermofisher.com/order/oligoDesigner/mutagenesis>) or NEB
495 primer design tool (<http://nebasechanger.neb.com>) (Table S2, Table S3). Quintuple mutants
496 were created using the GeneArt Seamless Cloning and Assembly Enzyme Mix (Invitrogen). All
497 mutants of MpeQ were over-expressed and purified using the same protocols as for the wild
498 type (Table S4).

499 *Crystallization, data collection and structure determination*

500 Purified MpeQ protein was crystallized using the hanging drop vapor diffusion method by
501 mixing the protein sample (5 mg/mL) with the crystallization solution (2.1 M DL-Malic acid, pH
502 7.0) in 1:1 ratio. Single crystals obtained by macro-seeding were cryoprotected in the mother
503 liquor containing 25% glycerol for X-ray diffraction experiments. All diffraction images were
504 collected at the Life-Science Consortium Access Team (LS-CAT) beam stations at the Advanced
505 Photon Source, Argonne National Laboratory. The diffraction datasets were indexed, integrated
506 and scaled using HKL2000 (Otwinowski and Minor, 1997) and/or xia2 in CCP4 (Winter, 2010).

507 The crystal structure of SeMet-MpeQ was determined using the single-wavelength
508 anomalous dispersion method. The initial SAD phasing was carried out using the MpeQ dataset
509 of 2.9 Å resolution in the space group of C222₁ (Shelx97) (Sheldrick, 2008). An initial model was
510 used as a search model to determine the MpeQ structure in the space group C2 by molecular
511 replacement (Phaser) (McCoy et al., 2007). The final model was refined at 2.5 Å resolution with
512 the *R*-factor and free *R*-factor of 0.232 and 0.285, respectively (Phenix.refine) (Adams et al.,
513 2010). The final model in the space group C222₁ was refined at 2.95 Å resolution with the *R*-
514 factor and free *R*-factor of 0.245 and 0.307, respectively. Model building and molecular graphics
515 were done using Coot (Emsley and Cowtan, 2004) and/or PyMOL(DeLano, 2002).

516 *Sequence analysis and identification of switch residues*

517 Sequence alignment of 106 bilin lyases and lyase-isomerases in the MpeQWYZ family were
518 aligned using MUSCLE (Edgar, 2004) (Dataset S1). Each sequence was assigned a lyase or lyase-
519 isomerase function based on the published biochemical data (Grébert et al., 2021; Sanfilippo et al.,
520 2019b; Shukla et al., 2012). Positions identical or differing between lyases and isomerases were
521 identified using Biopython 1.77 (Cock et al., 2009) (Dataset S2). Sequence logo was generated
522 using WebLogo (Crooks, 2004).

523 *Enzyme activity assays based on a co-expression system*

524 A three-plasmid co-expression system was used to test the lyase activities of MpeQ and
525 MpeW. Specifically, MpeQ or MpeW was co-expressed in *E. coli* BL21(DE3) cells with MpeA
526 (pCOLA-Duet vector) along with a pACYC-Duet vector carrying heme oxygenase (HO1) and
527 PEB synthase (PebS) genes (Kronfel et al., 2019). His-tagged MpeQ or non-tagged MpeW was
528 then co-purified with MpeA using Co²⁺-affinity column (Talen). Eluted fractions were then
529 examined by SDS polyacrylamide electrophoresis (SDS-PAGE) and absorption spectroscopy,
530 and/or fluorescence spectroscopy. The bilin lyase activity was measured by the
531 chromophorylated MpeA detected either by Zn-fluorescence (Berkelman and Lagarias, 1986),
532 absorption spectroscopy (Shimadzu UV-2600 UV-Vis spectrophotometer), and/or fluorescence
533 spectroscopy (Perkin Elmer LS55 fluorescence spectrophotometer). MpeA-PUB and MpeA-PEB
534 have characteristic absorption peaks at 495 nm and 550 nm, respectively, while their

535 corresponding emission peaks are around 500 nm and 565 nm. The lyase activities of the wild
 536 type MpeQ and mutants were quantified and compared by the absorbance readings at 495 nm
 537 normalized by the corresponding MpeA band intensities detected by SDS-PAGE (ImageJ). The
 538 relative activity of a specific mutant (in percentage) was calculated in comparison to the wild
 539 type for which the lyase activity was set to 100%.

540 *Molecular dynamics (MD) simulations*

541 72-ns MD simulations were performed on the structures of apo-MpeA and MpeQ (PDB ID:
 542 7MC4). The starting structure for apo-MpeA was based on a homology model of MpeA
 543 obtained by SwissModel (Biasini et al., 2014) with all the pigments removed. All minimization
 544 and MD simulation steps were performed using NAMD (Phillips et al., 2020) with
 545 CHARMM36m force field (Huang et al., 2017). The simulation was performed on 32 processors
 546 of a Cray T3E parallel supercomputer at University of Illinois Chicago High-Performance
 547 Computing. Frames were collected at 100-ps intervals for the simulation length of 72 ns, giving
 548 720 frames for conformational analysis.

549 **QUANTIFICATION AND STATISTICAL ANALYSIS**

550 Statistical data in Table 1 for the crystallography data and structural analysis were obtained
 551 from the outputs of HKL2000 (Otwinowski and Minor, 1997) and Phenix (Adams et al., 2010).

552

553 **KEY RESOURCES TABLE**

554

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>Escherichia coli</i> : DH5 α	Invitrogen	CAT#:18-258-012
<i>Escherichia coli</i> : BL21(DE3)	Invitrogen	CAT#:C600003
Chemicals, peptides, and recombinant proteins		
Selenomethionine	Acros Organics	CAT#: 3211-76-5
TALON resin	Takara	CAT#: 635504
DL-Malic acid	Sigma Aldrich	LOT#: MKBW2323V
<i>Synechococcus</i> sp. A1562 MpeQ	This study	Genbank: AGW21721
<i>Synechococcus</i> sp. A1562 MpeW	This study	Genbank: AGW21717
<i>Synechococcus</i> sp. A1562 MpeA	This study	Genbank: AVH76704
See Table S2 for more recombinant proteins	This study	
Critical Commercial Assays		
GeneArt Seamless Cloning and Assembly Enzyme Mix	Invitrogen	CAT#: A14606
Deposited data		
MpeQ crystal structure in space group C2	This paper	PDB:7MC4
MpeQ crystal structure in space group C222 ₁	This paper	PDB:7MCH
Recombinant DNA		

pCDF Duet-1 vector	Novagen	CAT#:71340-3
pET Duet-1 vector	Novagen	CAT#:71146-3
Plasmid: A1562 Nus-MpeQ/pET44b	(Grébert et al., 2021)	N/A
Plasmid: A1562 Nus-MpeW/pET44b	(Grébert et al., 2021)	N/A
Plasmid: A1562 HT-MpeA/pCOLA	(Grébert et al., 2021)	N/A
Plasmid: NT-PebS/NT-HOI/pACYC	(Kronfel et al., 2019)	N/A
Plasmid: A1562 NT-MpeQ/pCDF	This study	N/A
Plasmid: A1562 HT-MpeQ/pCDF	This study	N/A
Plasmid: A1562 NT-MpeW/pET	This study	N/A
See Table S4 for more recombinant DNA		
Software and algorithms		
HKL2000	(Otwinowski & Minor, 1997)	https://hkl-xray.com/hkl-2000
XIA2 (CCP4)	(Winter, 2010)	https://www.ccp4.ac.uk/
Shelx97 (CCP4)	(Sheldrick, 2008)	https://www.ccp4.ac.uk/
Phaser (Phenix)	(McCoy et al., 2007)	http://www.phenix-online.org/
Phenix.refine	(Adams et al., 2010)	http://www.phenix-online.org/
Coot	(Emsley and Cowtan, 2004)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PyMOL	(DeLano, 2002)	https://pymol.org/2/
MUSCLE	(Edgar, 2004)	https://www.ebi.ac.uk/Tools/msa/muscle/
BioPython	(Cock et al., 2009)	https://biopython.org/
WebLOGO	(Crooks, 2004)	http://weblogo.threeplusone.com/
SwissMODEL	(Biasini et al., 2014)	https://swissmodel.expasy.org/
NAMD	(Phillips et al., 2020)	https://www.ks.uiuc.edu/Research/namd/

555

556

557 **SUPPLEMENTAL TABLES AND DATASETS**

558

559 **Table S1.** Summary of MpeQ mutants (related to Fig. 4 and Fig. 5).

560 **Table S2.** Supplemental STAR methods Oligonucleotides (related to the STAR Methods **Method Details**).

561 **Table S3.** Supplemental STAR methods Recombinant DNAs (related to the STAR Methods **Experimental Model and Subject Details**).

563 **Table S4.** Supplemental STAR methods Recombinant Proteins (related to the STAR Methods **Method Details**).

565 **Dataset S1.** Sequences of members of the MpeQWYZ protein family used to draw Fig. S2.

566 Related to Fig. 4. Genomes or protein sequences were retrieved from the Genbank and single

567 amplified genomes (SAGs) from the Joint Genome/Institute Integrated microbial genomes &
568 microbiomes (IMG/M) databases.

569 **Dataset S2.** Summary of the residues found in 106 sequences MpeQ/W/Y/Z from marine *Synechococcus*
570 isolates and single-cell amplified genomes (**related to Fig. 4**).

571 **References**

- 572 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd,
573 J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a
574 comprehensive Python-based system for macromolecular structure solution. *Acta*
575 *Crystallogr. D Biol. Crystallogr.* *66*, 213–221.
- 576 Berkelman, T.R., and Lagarias, J.C. (1986). Visualization of bilin-linked peptides and
577 proteins in polyacrylamide gels. *Anal. Biochem.* *156*, 194–201.
- 578 Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F.,
579 Cassarino, T.G., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein
580 tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*
581 *42*, W252–W258.
- 582 Biswas, A., Boutaghou, M.N., Alvey, R.M., Kronfel, C.M., Cole, R.B., Bryant, D.A., and
583 Schluchter, W.M. (2011). Characterization of the Activities of the CpeY, CpeZ, and CpeS
584 Bilin Lyases in Phycoerythrin Biosynthesis in *Fremyella diplosiphon* Strain UTEX 481*.
585 *J. Biol. Chem.* *286*, 35509–35521.
- 586 Blot, N., Wu, X.-J., Thomas, J.-C., Zhang, J., Garczarek, L., Böhm, S., Tu, J.-M., Zhou, M.,
587 Plöscher, M., Eichacker, L., et al. (2009). Phycourobilin in Trichromatic Phycocyanin
588 from Oceanic Cyanobacteria Is Formed Post-translationally by a Phycocourobilin
589 Lyase-Isomerase. *J. Biol. Chem.* *284*, 9290–9298.
- 590 Bretaudeau, A., Coste, F., Humily, F., Garczarek, L., Le Corguillé, G., Six, C., Ratin, M.,
591 Collin, O., Schluchter, W.M., and Partensky, F. (2012). CyanoLyase: a database of
592 phycobilin lyase sequences, motifs and functions. *Nucleic Acids Res.* *41*, D396–D401.
- 593 Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I.,
594 Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available Python
595 tools for computational molecular biology and bioinformatics. *Bioinformatics* *25*, 1422–
596 1423.
- 597 Crooks, G.E. (2004). WebLogo: A Sequence Logo Generator. *Genome Res.* *14*, 1188–1190.
- 598 DeLano, W.L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter*
599 *On Protein Crystallography.* 82–92.
- 600 Doublíé, S. (1997). [29] Preparation of selenomethionyl proteins for phase
601 determination. In *Methods in Enzymology*, Jr. Charles W. Carter, ed. (Academic Press),
602 pp. 523–530.

603 Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
604 throughput. *Nucleic Acids Res.* *32*, 1792–1797.

605 Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics.
606 *Acta Crystallogr. D Biol. Crystallogr.* *60*, 2126–2132.

607 Everroad, C., Six, C., Partensky, F., Thomas, J.-C., Holtzendorff, J., and Wood, A.M.
608 (2006). Biochemical Bases of Type IV Chromatic Adaptation in Marine *Synechococcus*
609 spp. *J. Bacteriol.* *188*, 3345–3356.

610 Fairchild, C.D., ZHAOt, J., ZHOuTl, J., Colson, S.E., BRYANTt, D.A., and Glazer, A.N.
611 (1992). Phycocyanin a-subunit phycocyanobilin lyase. *89*, 7017–7021.

612 Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincon, J., Zabala, L.L., Jiao, N., Karl, D.M.,
613 Li, W.K.W., Lomas, M.W., Veneziano, D., et al. (2013). Present and future global
614 distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc.*
615 *Natl. Acad. Sci.* *110*, 9824–9829.

616 Gasper, R., Schwach, J., Hartmann, J., Holtkamp, A., Wiethaus, J., Riedel, N., Hofmann,
617 E., and Frankenberg-Dinkel, N. (2017). Distinct Features of Cyanophage-encoded T-type
618 Phycobiliprotein Lyase Φ CpeT: THE ROLE OF AUXILIARY METABOLIC GENES. *J.*
619 *Biol. Chem.* *292*, 3089–3098.

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

622 Grébert, T., Doré, H., Partensky, F., Farrant, G.K., Boss, E.S., Picheral, M., Guidi, L.,
623 Pesant, S., Scanlan, D.J., Wincker, P., et al. (2018). Light color acclimation is a key
624 process in the global ocean distribution of *Synechococcus cyanobacteria*. *Proc. Natl. Acad.*
625 *Sci.* *115*, E2010–E2019.

626 Grébert, T., Nguyen, A.A., Pokhrel, S., Joseph, K.L., Ratin, M., Dufour, L., Chen, B.,
627 Haney, A.M., Karty, J.A., Trinidad, J.C., et al. (2021). Molecular bases of an alternative
628 dual-enzyme system for light color acclimation of marine *Synechococcus cyanobacteria*.
629 *Proc. Natl. Acad. Sci.* *118*, e2019715118.

630 Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H.,
631 and MacKerell, A.D. (2017). CHARMM36m: an improved force field for folded and
632 intrinsically disordered proteins. *Nat. Methods* *14*, 71–73.

633 Humily, F., Partensky, F., Six, C., Farrant, G.K., Ratin, M., Marie, D., and Garczarek, L.
634 (2013). A Gene Island with Two Possible Configurations Is Involved in Chromatic
635 Acclimation in Marine *Synechococcus*. *PLoS ONE* 8, e84459.

636 Hussain, H., and Chong, N.F.-M. (2016). Combined Overlap Extension PCR Method for
637 Improved Site Directed Mutagenesis. *BioMed Res. Int.* 2016, 1–7.

638 Kiser, P.D., Zhang, J., Badiee, M., Li, Q., Shi, W., Sui, X., Golczak, M., Tochtrop, G.P.,
639 and Palczewski, K. (2015). Catalytic mechanism of a retinoid isomerase essential for
640 vertebrate vision. *Nat. Chem. Biol.* 11, 409–415.

641 Kronfel, C.M., Kuzin, A.P., Forouhar, F., Biswas, A., Su, M., Lew, S., Seetharaman, J.,
642 Xiao, R., Everett, J.K., Ma, L.-C., et al. (2013). Structural and Biochemical
643 Characterization of the Bilin Lyase CpcS from *Thermosynechococcus elongatus*.
644 *Biochemistry* 52, 8663–8676.

645 Kronfel, C.M., Biswas, A., Frick, J.P., Gutu, A., Blensdorf, T., Karty, J.A., Kehoe, D.M.,
646 and Schluchter, W.M. (2019). The roles of the chaperone-like protein CpeZ and the
647 phycoerythrobilin lyase CpeY in phycoerythrin biogenesis. *Biochim. Biophys. Acta BBA*
648 - *Bioenerg.* 1860, 549–561.

649 Mahmoud, R.M., Sanfilippo, J.E., Nguyen, A.A., Strnat, J.A., Partensky, F., Garczarek,
650 L., Abo El Kassem, N., Kehoe, D.M., and Schluchter, W.M. (2017). Adaptation to Blue
651 Light in Marine *Synechococcus* Requires MpeU, an Enzyme with Similarity to
652 Phycoerythrobilin Lyase Isomerases. *Front. Microbiol.* 8.

653 Martínez Cuesta, S., Furnham, N., Rahman, S.A., Sillitoe, I., and Thornton, J.M. (2014).
654 The evolution of enzyme function in the isomerases. *Curr. Opin. Struct. Biol.* 26, 121–
655 130.

656 Martínez Cuesta, S., Rahman, S.A., and Thornton, J.M. (2016). Exploring the chemistry
657 and evolution of the isomerases. *Proc. Natl. Acad. Sci.* 113, 1796–1801.

658 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and
659 Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.

660 Otwinowski, Z., and Minor, W. (1997). [20] Processing of X-ray diffraction data collected
661 in oscillation mode. In *Methods in Enzymology*, Jr. Charles W. Carter, ed. (Academic
662 Press), pp. 307–326.

663 Overkamp, K.E., Gasper, R., Kock, K., Herrmann, C., Hofmann, E., and Frankenberg-
664 Dinkel, N. (2014). Insights into the Biosynthesis and Assembly of Cryptophycean
665 Phycobiliproteins. *J. Biol. Chem.* 289, 26691–26707.

666 Palenik, B. (2001). Chromatic Adaptation in Marine *Synechococcus* Strains. *Appl.*
667 *Environ. Microbiol.* 67, 991–994.

668 Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R.,
669 Fiorin, G., Hénin, J., Jiang, W., et al. (2020). Scalable molecular dynamics on CPU and
670 GPU architectures with NAMD. *J. Chem. Phys.* 153, 044130.

671 Sanfilippo, J.E., Nguyen, A.A., Karty, J.A., Shukla, A., Schluchter, W.M., Garczarek, L.,
672 Partensky, F., and Kehoe, D.M. (2016). Self-regulating genomic island encoding tandem
673 regulators confers chromatic acclimation to marine *Synechococcus*. *Proc. Natl. Acad. Sci.*
674 113, 6077–6082.

675 Sanfilippo, J.E., Garczarek, L., Partensky, F., and Kehoe, D.M. (2019a). Chromatic
676 Acclimation in Cyanobacteria: A Diverse and Widespread Process for Optimizing
677 Photosynthesis. *Annu. Rev. Microbiol.* 73, 407–433.

678 Sanfilippo, J.E., Nguyen, A.A., Garczarek, L., Karty, J.A., Pokhrel, S., Strnat, J.A.,
679 Partensky, F., Schluchter, W.M., and Kehoe, D.M. (2019b). Interplay between
680 differentially expressed enzymes contributes to light color acclimation in marine
681 *Synechococcus*. *Proc. Natl. Acad. Sci.* 116, 6457–6462.

682 Scheer, H., and Zhao, K.-H. (2008). Biliprotein maturation: the chromophore
683 attachment: Biliprotein chromophore attachment. *Mol. Microbiol.* 68, 263–276.

684 Schirmer, T., Bode, W., and Huber, R. (1987). Refined three-dimensional structures of
685 two cyanobacterial C-phycoyanins at 2.1 and 2.5 Å resolution. *J. Mol. Biol.* 196, 677–
686 695.

687 Schluchter, W.M., Shen, G., Alvey, R.M., Biswas, A., Saunée, N.A., Williams, S.R., Mille,
688 C.A., and Bryant, D.A. (2010). Phycobiliprotein Biosynthesis in Cyanobacteria: Structure
689 and Function of Enzymes Involved in Post-translational Modification. In *Recent*
690 *Advances in Phototrophic Prokaryotes*, P.C. Hallenbeck, ed. (New York, NY: Springer
691 New York), pp. 211–228.

692 Sheldrick, G.M. (2008). A short history of *SHELX*. *Acta Crystallogr. A* 64, 112–122.

693 Shen, G., Saunée, N.A., Williams, S.R., Gallo, E.F., Schluchter, W.M., and Bryant, D.A.
694 (2006). Identification and Characterization of a New Class of Bilin Lyase THE *cpcT*

695 GENE ENCODES A BILIN LYASE RESPONSIBLE FOR ATTACHMENT OF
696 PHYCOCYANOBILIN TO CYS-153 ON THE β -SUBUNIT OF PHYCOCYANIN IN
697 SYNECHOCOCCUS SP. PCC 7002. *J. Biol. Chem.* 281, 17768–17778.

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

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

706 Stumpe, H., Müller, N., and Grubmayr, K. (1993). The addition of methyl-2-
707 mercaptoacetate to phycocyanobilin dimethyl ester: A model reaction for biliprotein
708 biosynthesis? *Tetrahedron Lett.* 34, 4165–4168.

709 Tu, J.-M., Zhou, M., Haessner, R., Plöscher, M., Eichacker, L., Scheer, H., and Zhao, K.-
710 H. (2009). Toward a Mechanism for Biliprotein Lyases: Revisiting Nucleophilic
711 Addition to Phycocyanobilin. *J. Am. Chem. Soc.* 131, 5399–5401.

712 Wagner, J.R., Brunzelle, J.S., Forest, K.T., and Vierstra, R.D. (2005). A light-sensing knot
713 revealed by the structure of the chromophore-binding domain of phytochrome. *Nature*
714 438, 325–331.

715 Winter, G. (2010). xia2: an expert system for macromolecular crystallography data
716 reduction. *J. Appl. Crystallogr.* 43, 186–190.

717 Yang, X., Stojković, E.A., Kuk, J., and Moffat, K. (2007). Crystal structure of the
718 chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals
719 residues that modulate photoconversion. *Proc. Natl. Acad. Sci.* 104, 12571–12576.

720 Yang, X., Kuk, J., and Moffat, K. (2008). Crystal structure of *Pseudomonas aeruginosa*
721 bacteriophytochrome: Photoconversion and signal transduction. *Proc. Natl. Acad. Sci.*
722 105, 14715–14720.

723 Zhao, C., Höppner, A., Xu, Q.-Z., Gärtner, W., Scheer, H., Zhou, M., and Zhao, K.-H.
724 (2017). Structures and enzymatic mechanisms of phycobiliprotein lyases CpcE/F and
725 PecE/F. *Proc. Natl. Acad. Sci.* 114, 13170–13175.

726 Zhao, K.-H., Deng, M.-G., Zheng, M., Zhou, M., Parbel, A., Storf, M., Meyer, M.,
727 Strohmann, B., and Scheer, H. (2000). Novel activity of a phycobiliprotein lyase: both
728 the attachment of phycocyanobilin and the isomerization to phycoviolobin are
729 catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon.
730 FEBS Lett. 469, 9–13.

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

734 Zhou, W., Ding, W.-L., Zeng, X.-L., Dong, L.-L., Zhao, B., Zhou, M., Scheer, H., Zhao,
735 K.-H., and Yang, X. (2014). Structure and Mechanism of the Phycobiliprotein Lyase
736 CpcT. J. Biol. Chem. 289, 26677–26689.

737

Fig. 1

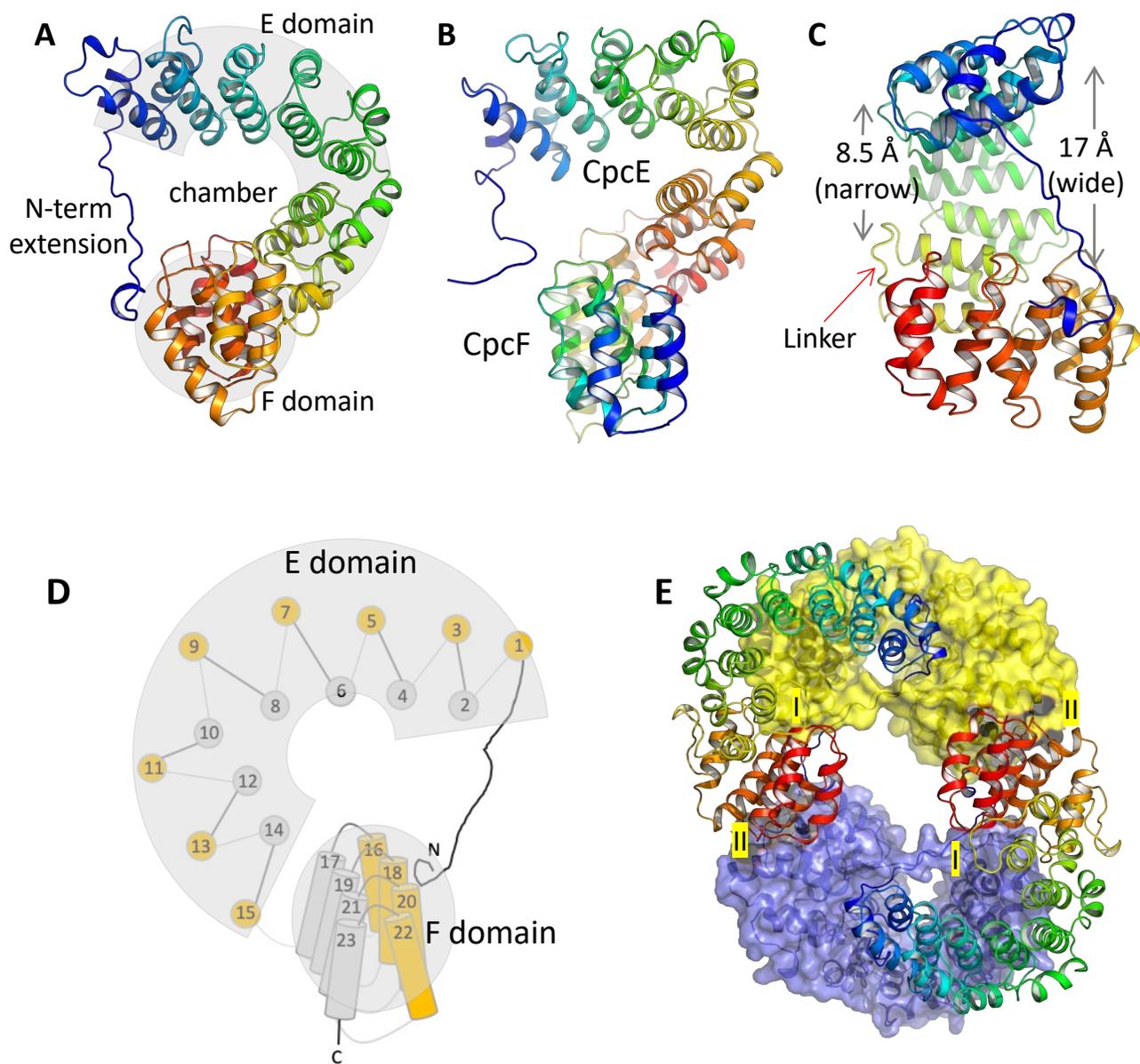


Fig. 2

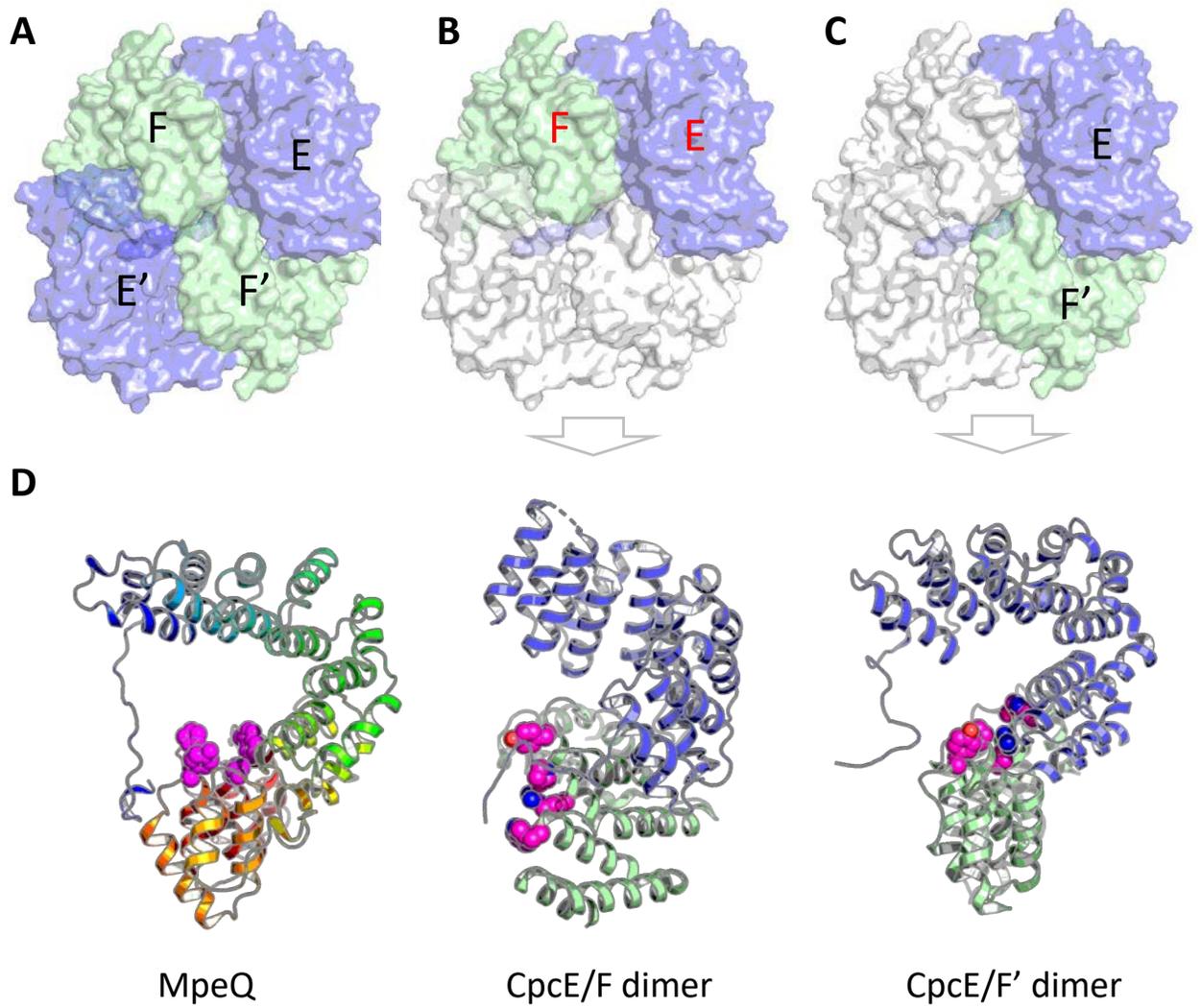


Fig. 3

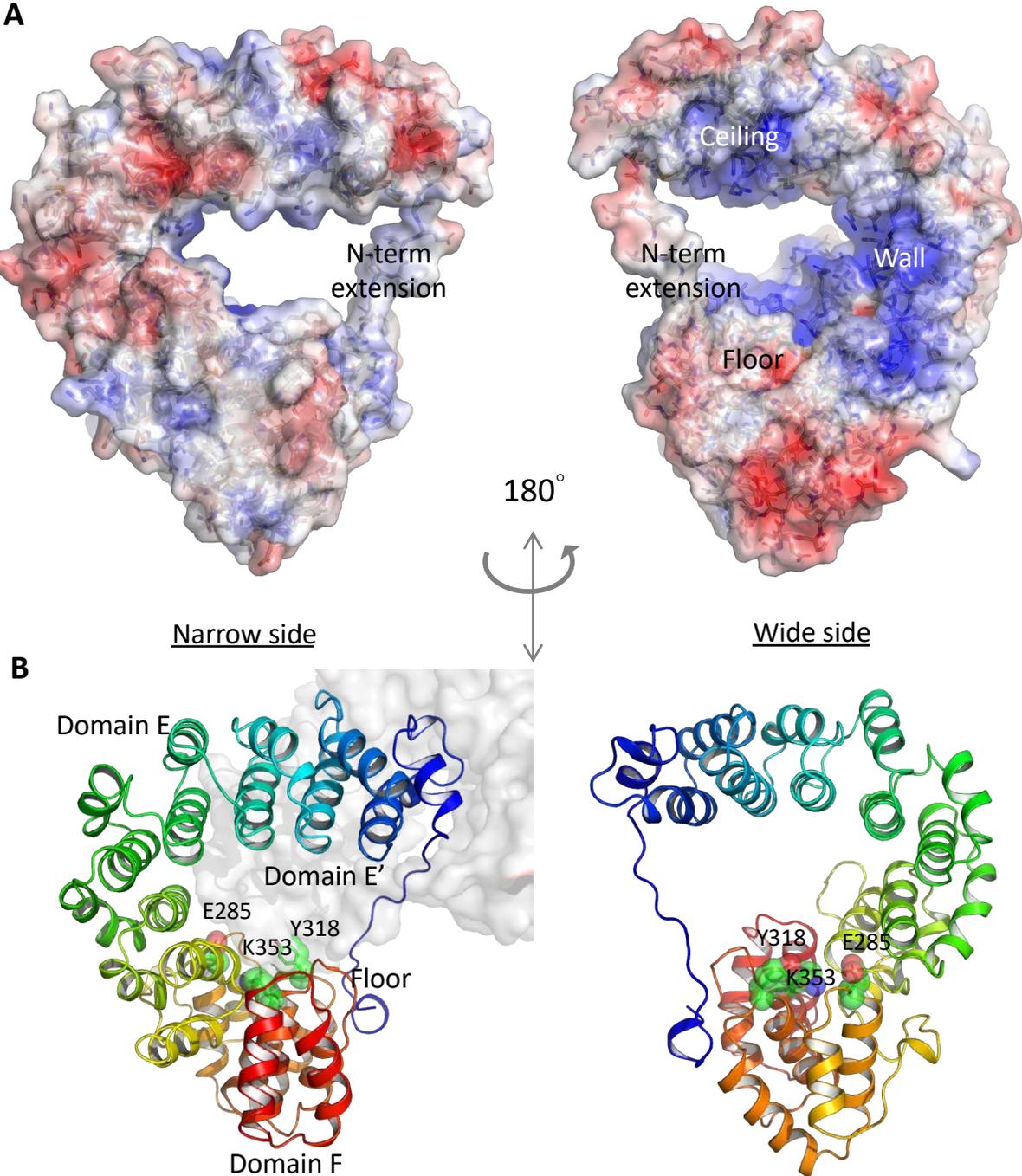


Fig. 4

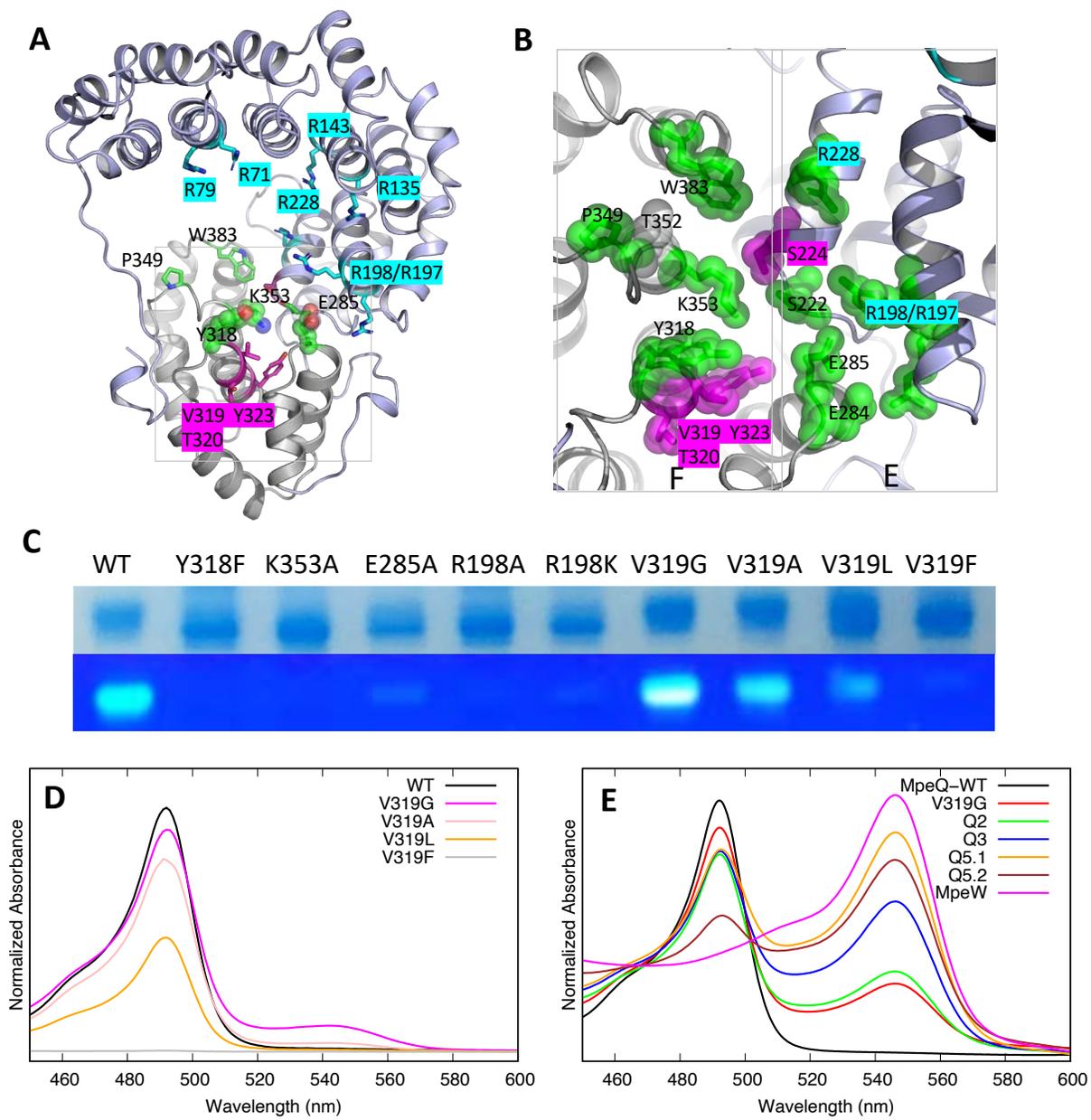
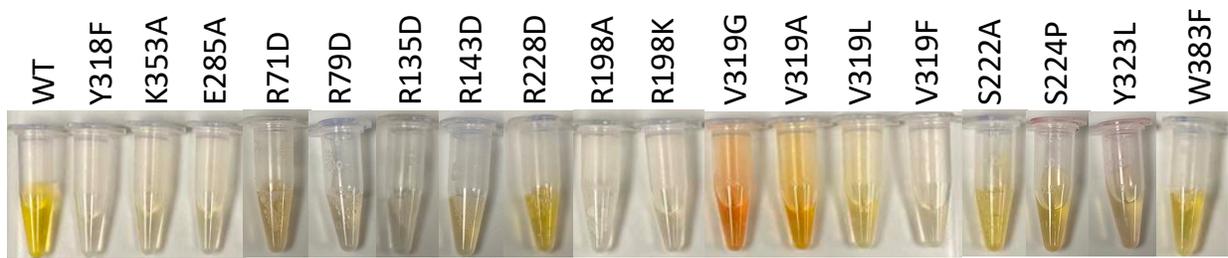
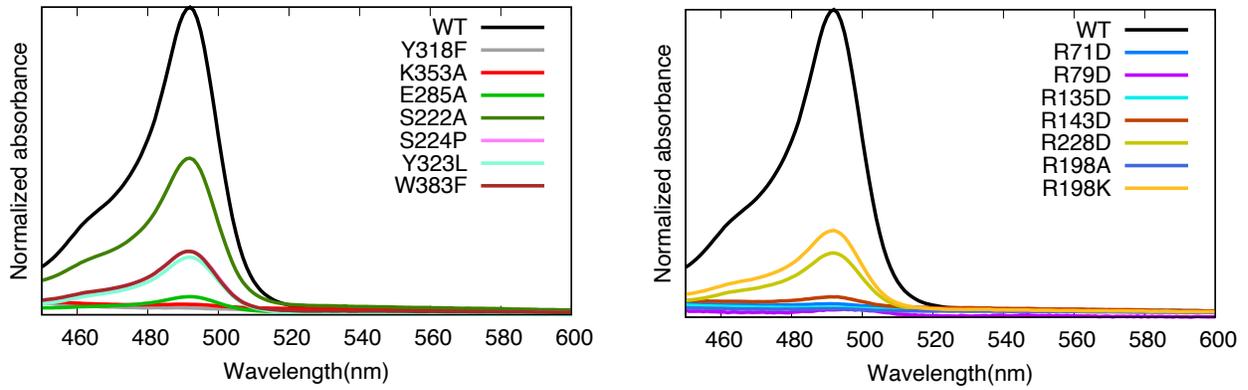


Fig. 5

A



B



C

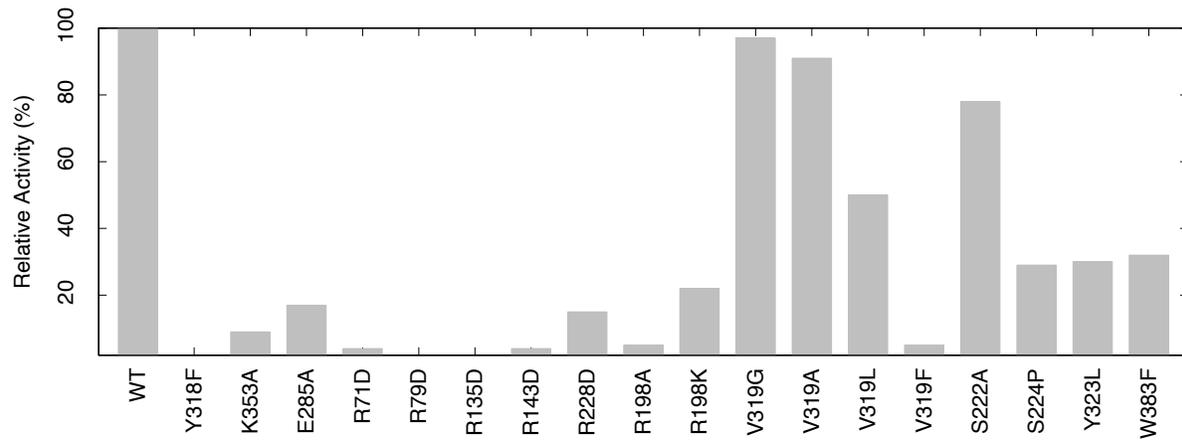


Fig. 6

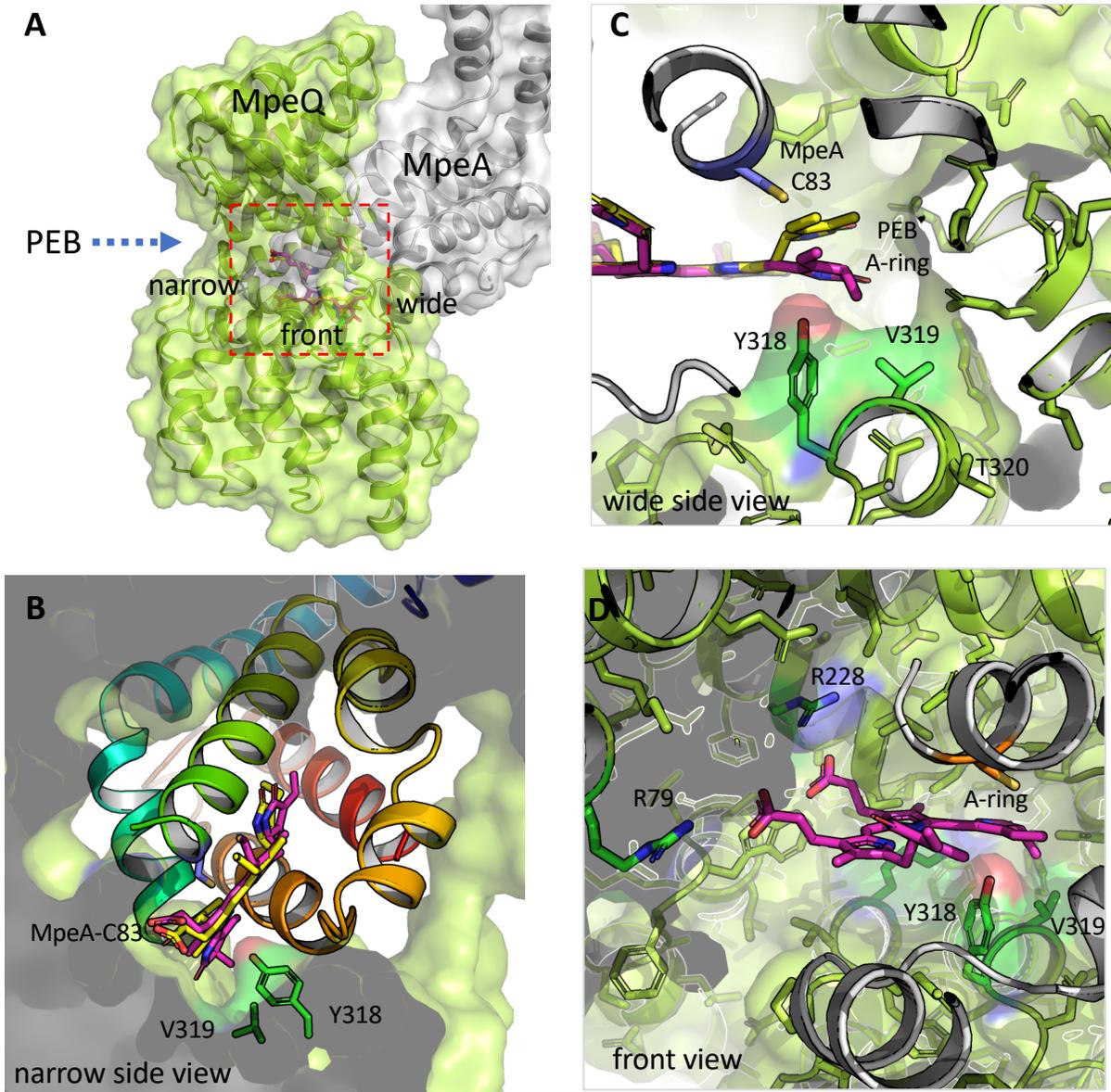
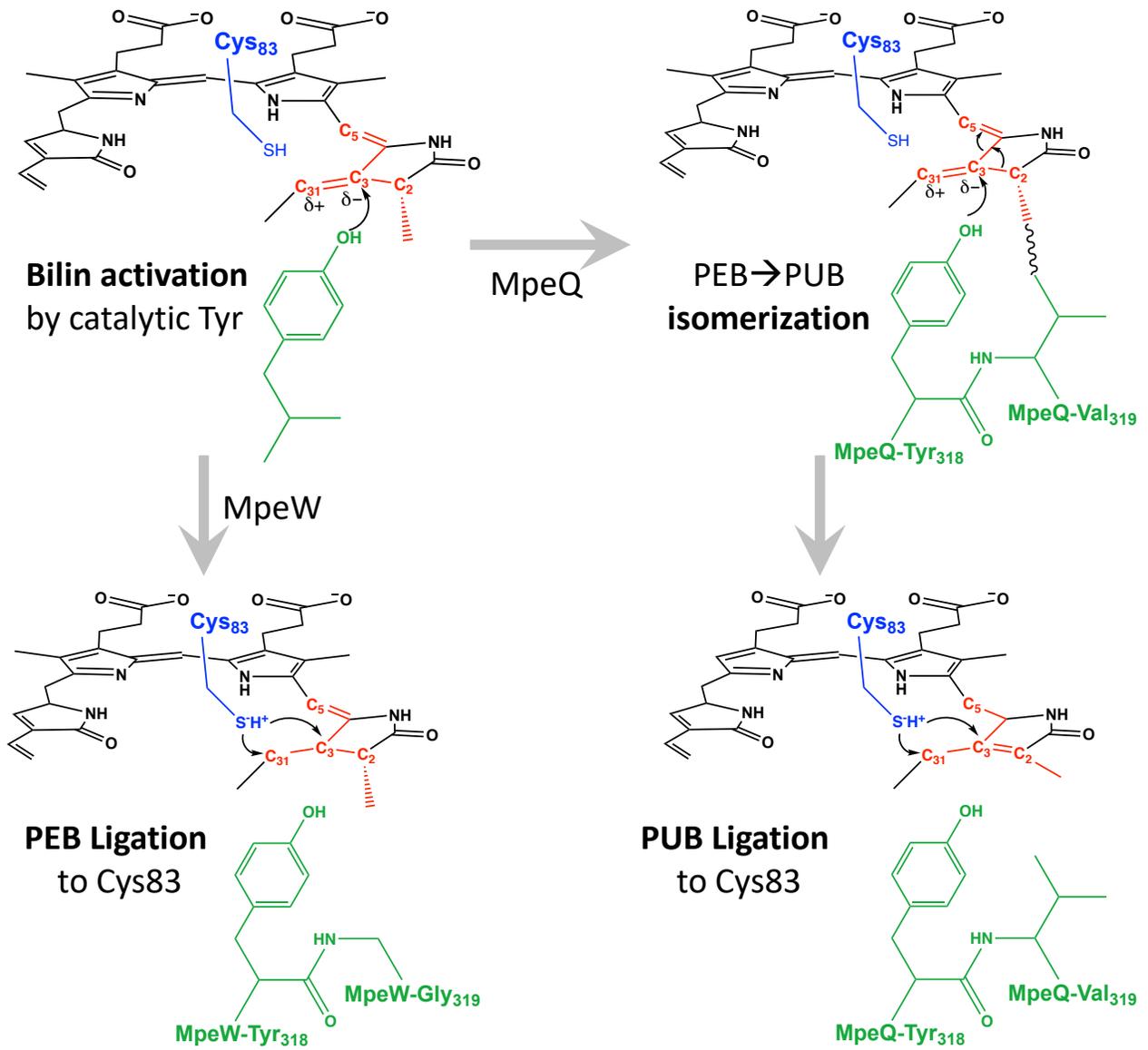


Fig. 7



Supporting Information

Crystal Structure and Molecular Mechanism of an E/F type Bilin Lyase-isomerase

Indika Kumarapperuma^{1§}, Kes Lynn Joseph^{2§}, Cong Wang¹, Linta M. Biju¹, Irin P. Tom¹,
Kourtney D. Weaver², Théophile Grébert³, Frédéric Partensky³, Wendy M. Schluchter^{2*},
and Xiaojing Yang^{1*}

¹ Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA

² Department of Biological Sciences, University of New Orleans, New Orleans, LA
70148, USA

³ Ecology of Marine Plankton (ECOMAP) Team, Station Biologique, Sorbonne Université,
CNRS, 29680 Roscoff, France

§ Joint first author

*Corresponding authors

SI content:

- Materials and Methods
- 8 SI figures
- 4 SI Tables
- 2 SI Datasets

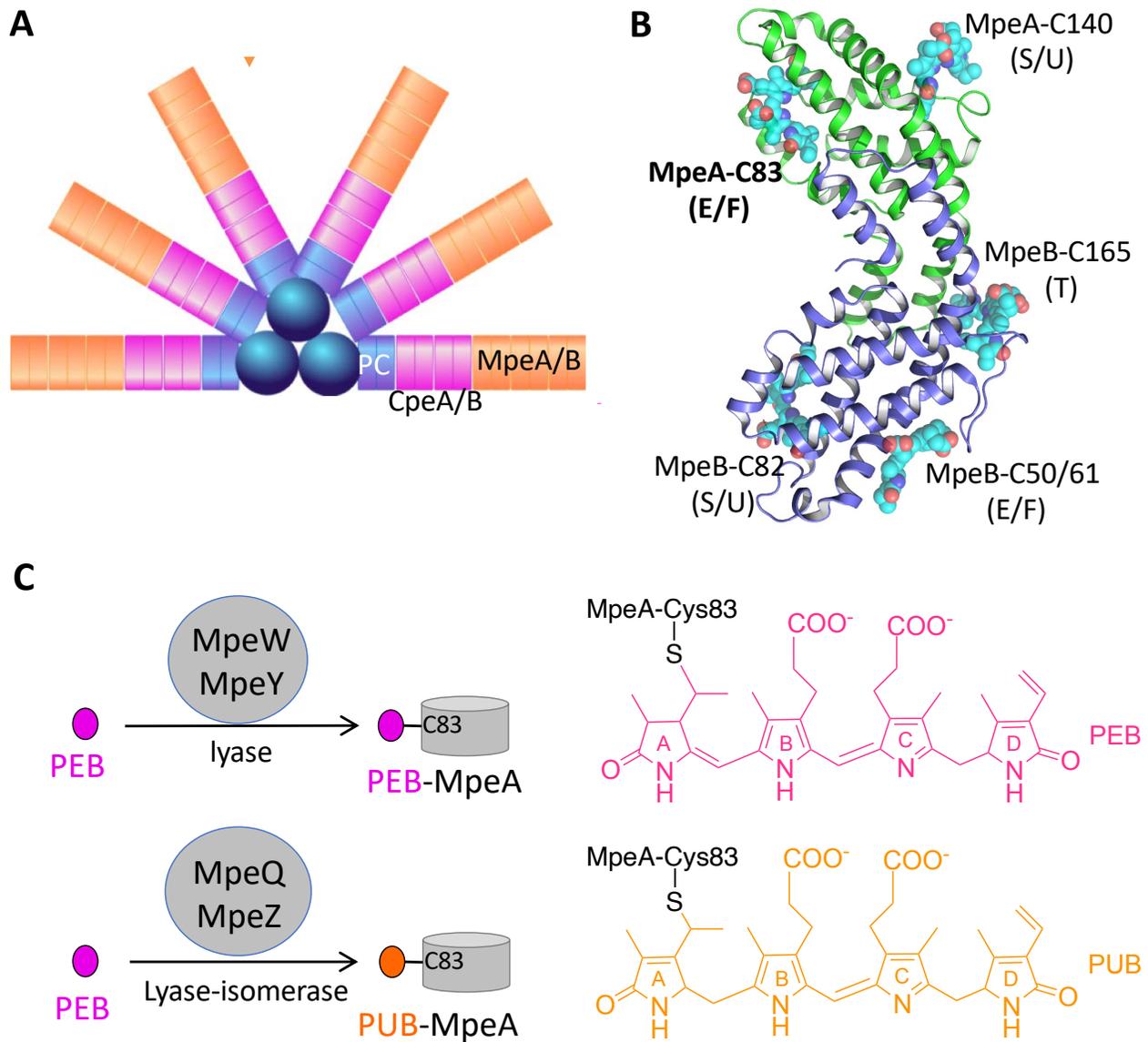


Figure S1. Bilin lyases catalyze bilin attachment to phycobilisome (Related to Fig. 1, Fig. 6-7). **A**) A schematic structure of light harvesting phycobilisome (PBS). Compared to the phycocyanins (PC) and CpeA/B dimers, the MpeA/B dimers are located on the remote ends of the PBS rods radiating from the PBS core (blue sphere). **B**) Ribbon diagram of a homology model for the MpeA/B dimer (Swiss-Model, Biasini et al 2014). Distinct attachment sites and the corresponding lyase types are labeled and the bilin pigments are shown in cyan spheres. MpeA-Cys83 is the target site of MpeQ. **C**) Schematic reactions (left) and products (right) catalyzed by PEB lyases (MpeQ/MpeY) and their counterpart lyase-isomerases (MpeQ/MpeZ) involved on type IV chromatic acclimation.

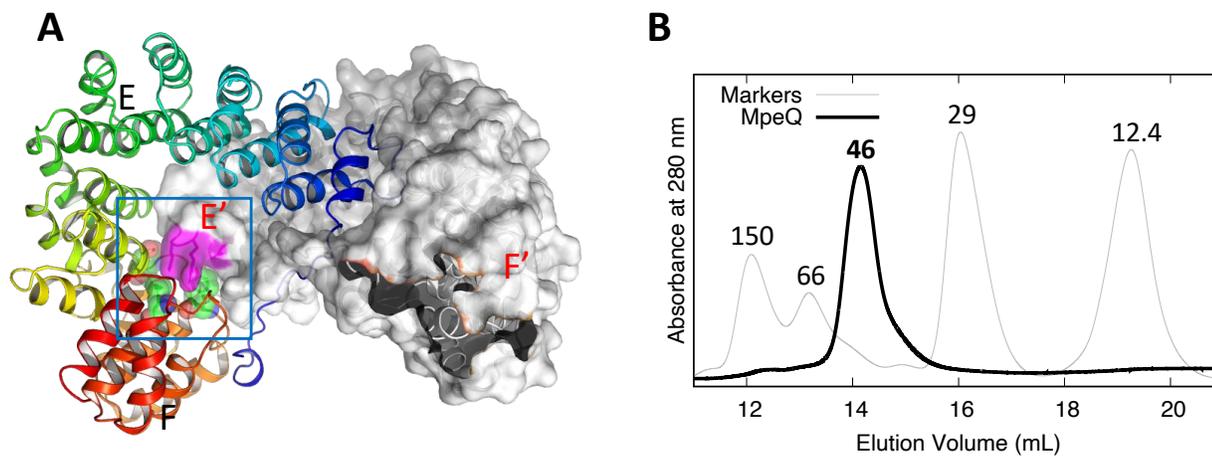


Figure S3. MpeQ is monomeric in solution. (Related to Fig. 1, Fig. 3). **A**) A zoom-in view of interface I highlights close intermolecular interactions between the active site (green) and residues (magenta) from the E' domain of a neighboring molecule (surface rendering). **B**) Elution profiles of size exclusion chromatography of MpeQ (black) and four marker (gray) proteins show that MpeQ is monomeric in solution. The number above each peak marks the corresponding molecular weight in kilo-Dalton.

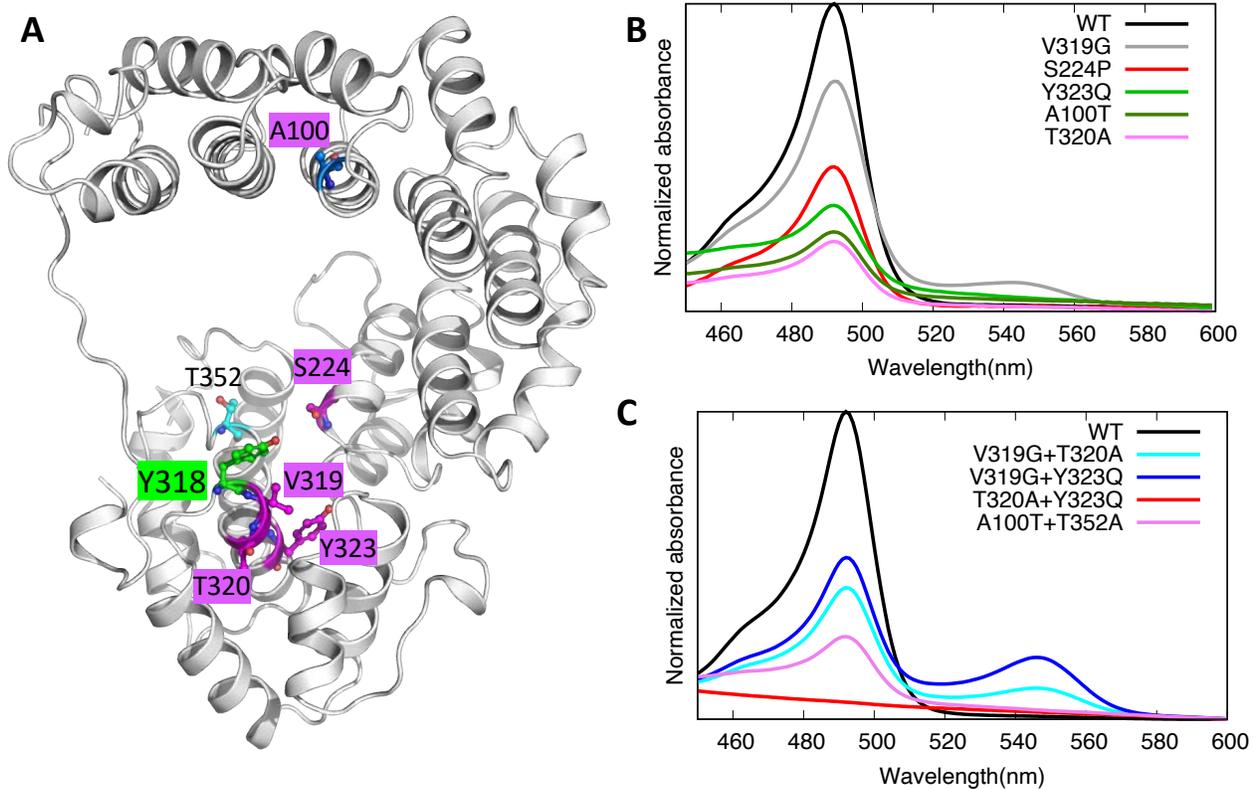


Figure S4. Effect of residue swapping on isomerase activity. (Related to Fig. 4-5) **A)** Locations of selected swap mutation sites shown in the MpeQ structure. **B)** Normalized absorption spectra of the MpeA product obtained from co-expression in *E. coli* with different MpeQ variants, each carrying a single mutation. **C)** Normalized absorption spectra of MpeA co-expressed with the MpeQ variants that carry double mutations.

MpeQ*	----MAERFDNLVE--GLTEERAMAVILADPDSLERPVDKYMAATRLGASNSEE--SLDVL	53
MpeW	-MVPSSQRFNDLFS--WMGEQQAIELLSQNVSTLDNPGIKYIAATRLGACSSRD--SLDAL	56
MpeZ*	----MAERFDLVE--GISEAEALLLFAKTADVPRPSDRYFAATRLGLCSSGA--TLDAL	53
MpeY	----MAERFDILFE--GLSEEKALELLRTNPEDLSNPVEKYTAATRLAASRSEQ--SLEGL	53
CpcE	MIEPSVEEFPANQPQLTPELAIANL-----QSSDLSLRYAAWWLGKYRVKESAVDAL	55
PecE*	-----MTAEPILSPETAIAAL-----SGEDNQIRYYAAWWLGKHNVA--GF'TAL	43
MpeQ*	IQA-A-ELDPEHLFNRI TRRKAIDALGRRKSPKALPSLFFKALKCSDEAAVINSVEAITKI	111
MpeW	VVA-A-TGDRDNIFDSITRRKSIEALGRRRDLSTLPTIYEAMHSSDEQTI VNSVDTLIHF	114
MpeZ*	IRATY-ELKIDELFDRI TRRKAIEALGRRKDAKAVSALVDVLGCTDTEAVINAISALIRI	112
MpeY	IEA-I-QLDPENLFNRI TRRKVLEALGRRRDARALPALFTALAFDDETSVINAVDSIAQI	111
CpcE	IAALEADRTELGGYPLRRNAARALGKLGNRKAVPGLINCLECPD-FYVREAAQS--L	112
PecE*	CVALFDERYRIPSGGYPLRRQAARALGLLKNPQAVPALIAALECEDLRLREAVICS--L	101
MpeQ*	DAPL TEADHEKLEALKGED-----I-QKRAVIQAFCLRGVPGV--INSISPL	156
MpeW	GIPLDAQFKSSLLKIIQNGS-----DVLKRVAIQCFTRLEMHEF--NGIIREQ	160
MpeZ*	GWSPSQVEHDQLLRLLDGED-----VTQSRAVIQAFGRLSIKSAASKSRIREL	160
MpeY	GAALNSNQSQKLLCALRGSD-----N-QKRSVIQTHTRLGLAGG--EEEIAAL	156
CpcE	EMLKDKTAAPALIKLLDGGVAQAVQVTGRPHLVQPYEAVLEALGATDA--IPLIQPF	170
PecE*	AAIGDKRAVTPLLNLLQS-----SQAQPYEALIEALATLQVWSA--RPQIEPF	147
MpeQ*	QDDSNPLVAGAAARAYMSKVALQPDGLEVLIPQLVDP IAGRRRSAVIDLG DAGDVTRLEAL	216
MpeW	ESNPNILVNGASIAYSIRVEDDKRKLHILENHLNLDLNVIHRRSSVIDIGDAGDPSLLSSI	220
MpeZ*	CDHESALVSGAARSCLSKLYGEKHLMQPLLSQLTDLVAGKRRSAVIDIGDGDHSLFLDL	220
MpeY	EQENPLVAGAAARAYAAARVHGRDLRVLPHIQLTDSIAGRRAAVIDLG DAGDVSMPLPHL	216
CpcE	LEHPVSRVQCAARAMYQLTQEPVYVELLVKVLAGNDLNLRRVALGDLGATGYLAAAEAI	230
PecE*	LQHYSERVQCAARYMYLLTQESEYIERIVKNLNDNMYL RWA AVFDLGA VAHQAVQAI	207
MpeQ*	VTAPVMSLRARSFQLVD-----PDKTCQVPEKYA---ELITQLLQ-----DN	257
MpeW	SKAAVSMPLRAKSAFKMTS-----KIK----TKSQR---TLINQMLK-----DD	257
MpeZ*	VRAPVMSLRAKSAFLQV-D-----EDGHLKTEKYQ---KLFKRLLL-----DD	260
MpeY	VTCPVSMPLRAKSAFQLVD-----PEKTGTVP'TAHS---DLIIQLLQ-----DD	257
CpcE	ANAKAENSFKLIALKGLELHQMSAESNALSISDQAIRVMNLMDSLL-----276	
PecE*	LTAQVPNSLKLNLKRIEAMLNNSVNHE---KAALLFGAIDDLLIRQL-----253	
MpeQ*	PQQ-LKLRKEWICDI----EPTIEIENLQHRDEARQYGGASSLMAMPKAERMILINEIKE	312
MpeW	SRH-LYFTQKFDAPN----DLEKVCDFLKHRDEERQYFGAKALFSWPASTVTDAIKTIWD	312
MpeZ*	PLL-LDLDEWKCTL----NPEDIERNLSHRDEARQYGAASLMQIDRRECLELIDSMQE	315
MpeY	PAM-LSLQKDWICSS----NLTEIEKNLQHRDEGKQYGGALSMLNMPQELQIETIDQLRS	312
CpcF	-PEKLVKAVQELALAKDVAAIPTL IAVFGYNNPTAAA IASTALVQLGEVAVPQLLTQIDD	73
PecF*	-----MNQ---ASLSVDAITNLIEAFHHHPAVRSAAVDEL IKLGSITVNLIIAAYDD	59
MpeQ*	KLWSDYVTHYYLTAVVGLQG----LEERSDLIRLALAE T I P Q Y T K S R I A-----AAWGC	362
MpeW	NHGSDYGVHYQVNCVLSQLG----LKELSYI IKESLLEPAPQYAKSKIA-----ATWGC	362
MpeZ*	RLWSDYVTHYYLTCVIGLRA----LHEKSYLVRSA LAETTPQYTKSRIA-----AAWAC	365
MpeY	RYWSDYGANYYLTAVVGLKN----VEERSDVVRTALAE T L P Q Y A K S R V A-----AAWAC	362
CpcF	--Y-NYGARAYSIRTLAAIADPRALDVLIDAAAT---DFAP-SVR-RAAAKGLGNLHWHK	125
PecF*	--SQDQGFQAQIIQVLAQIGDAKALKLLAEVVGTSVA N H C G N V R -RIAARGLGKIASTT	107
MpeQ*	LRLGLVD-----QKPLLEELSVSAFWLPLKWTCQRVLKQLS-----398	
MpeW	LTLGLME-----CRSDIENLLKSTWEPLRWTCAEVLRREL-----397	
MpeZ*	VELQLDD-----QLELLYELSNSSPWFPLRWACQQAFAQLVDQQOTSDDLALR-----413	
MpeY	LRLGLTD-----QVELLRDIQNTATW I P L K W S C E H A L K G L S-----398	
CpcF	LEFPDNQTAPKKAETLLF-ISQDAEW-SIRYAAIVGLQGLVNI PDLQQP I H T R L K E M L A	183
PecF*	SNTEI----INNAQEKLIIWALLTPEDW-GLRYAAAVSLQEIATPKA-----KAALQQAIA	157
CpcF	SDAEKAVRARILLAQSQ	200
PecF*	QETDPVVRSMIAIALS-	173

Figure S5. Sequence alignment of six representative E/F type blin lyases. (Related to Fig. 2-3, Fig. 7). MpeQ/W/Z/Y are single-chain lyases while CpcE/F and PecE/F are heterodimeric lyases. Lyase-isomerases are marked with *. MpeW/Q and MpeY/Z are equivalent pairs from the CA4 strains *Synechococcus* A15-62 and RS9916, respectively. The sequences of CpcE/F and PecE/F are from *Nostoc sp.* 7120 and *M. laminosus*, respectively. Residues highlighted in cyan mark the conserved positions between MpeQ and CpcE/F.

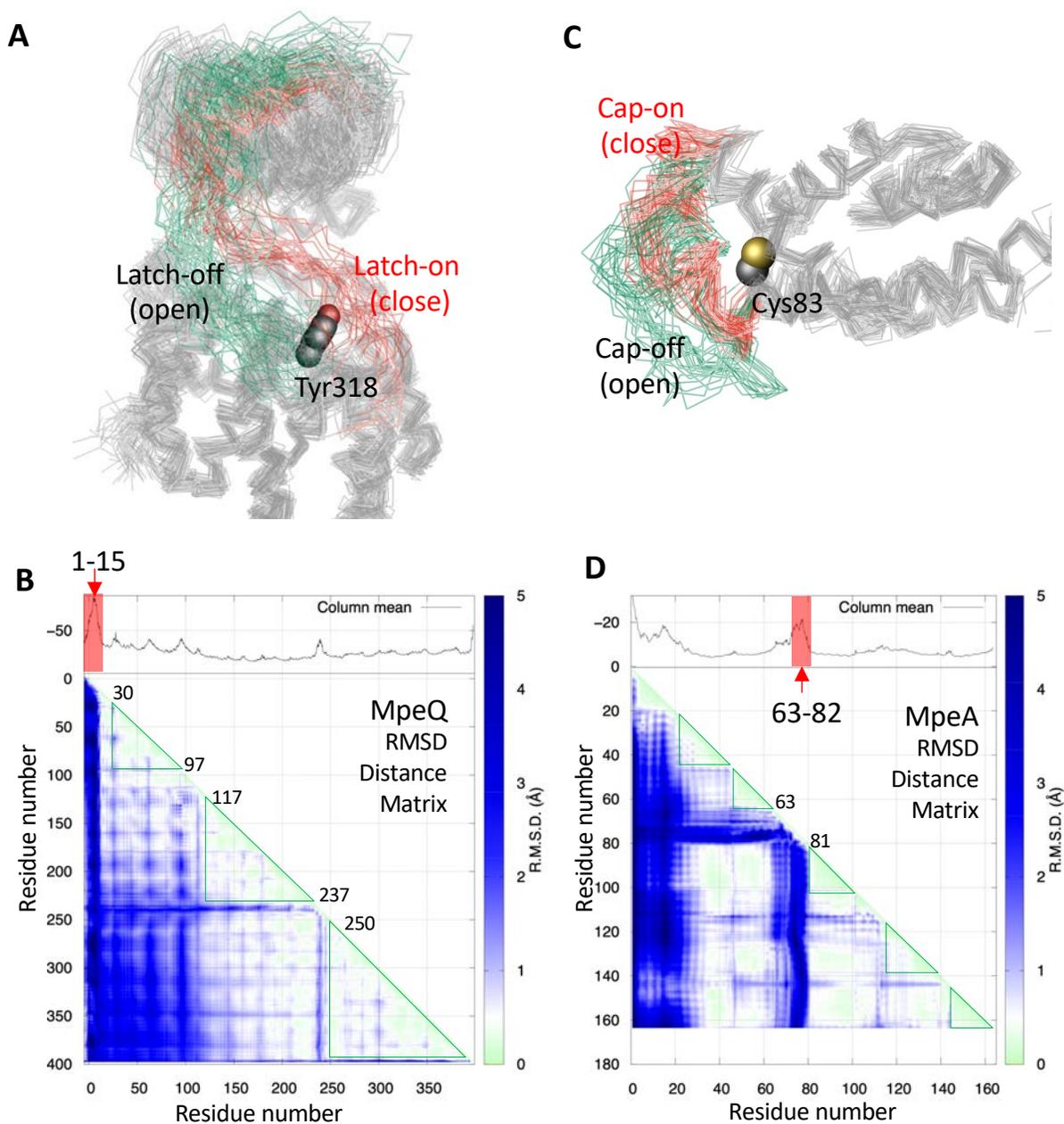


Figure S6. Conformational flexibility of MpeQ and MpeA. (Related to Fig. 6 and STAR methods). **A)** Molecular dynamics (MD) simulations of MpeQ reveal two clustered conformations in the N-terminal latch region. The “close” conformations (red) correspond to that of the crystal structure while the “open” conformations (green) expose the catalytic Tyr318 (spheres). **B)** RMSD distance matrix of the simulated MpeQ structures reveals the highly flexible N-terminus along with three rigid frameworks (green triangles). **C)** MD simulations of MpeA without pigments reveals a flexible segment (aa. 63-82) in two distinct conformations. In the “close” conformation, this “cap” helps shield Cys83 from the molecular surface while the “open” conformation makes Cys83 accessible for the bilin ligation. **D)** RMSD distance matrix calculated from the simulated MpeA structures shows the flexible region (shaded box) showing large RMSD values while green triangles mark those rigid regions of small RMSD values. High mobility observed in the N-terminus in MpeA is likely due to the absence of β subunit in MD simulations.

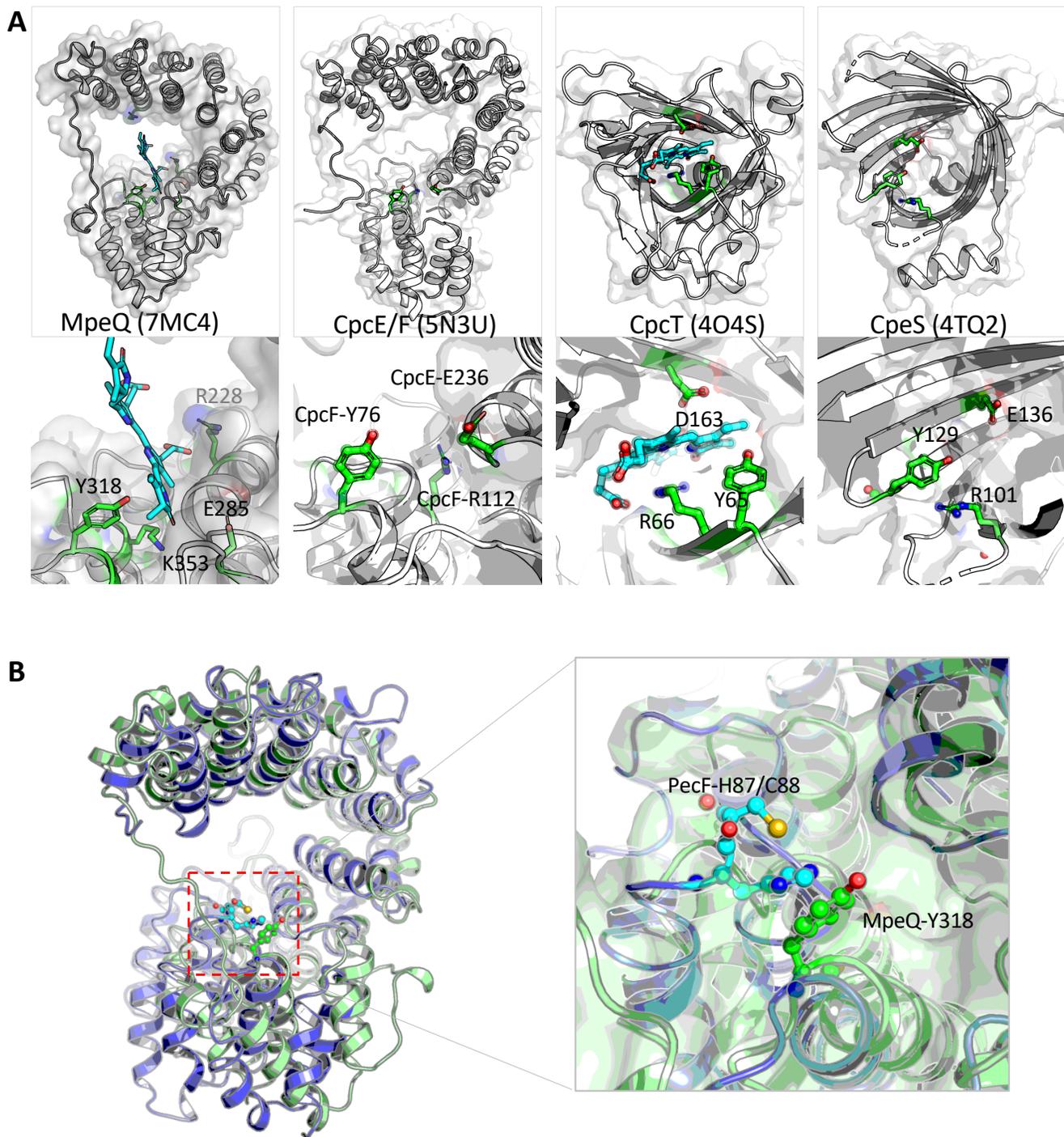


Figure S7. Active sites in representative bilin lyases. (Related to Fig. 6). **A)** The protein architecture (top panel) and active site (bottom panel) in MpeQ, CpcE/F, CpcT and CpeS. The bilin chromophore in MpeQ is shown based on the docking model presented in Fig. 4. **B)** Superposition of the MpeQ structure (green) and homology model of heterodimer PecE/F (blue/cyan) places the HC motif of PecE/F right next to the catalytic Tyr318 in MpeQ.

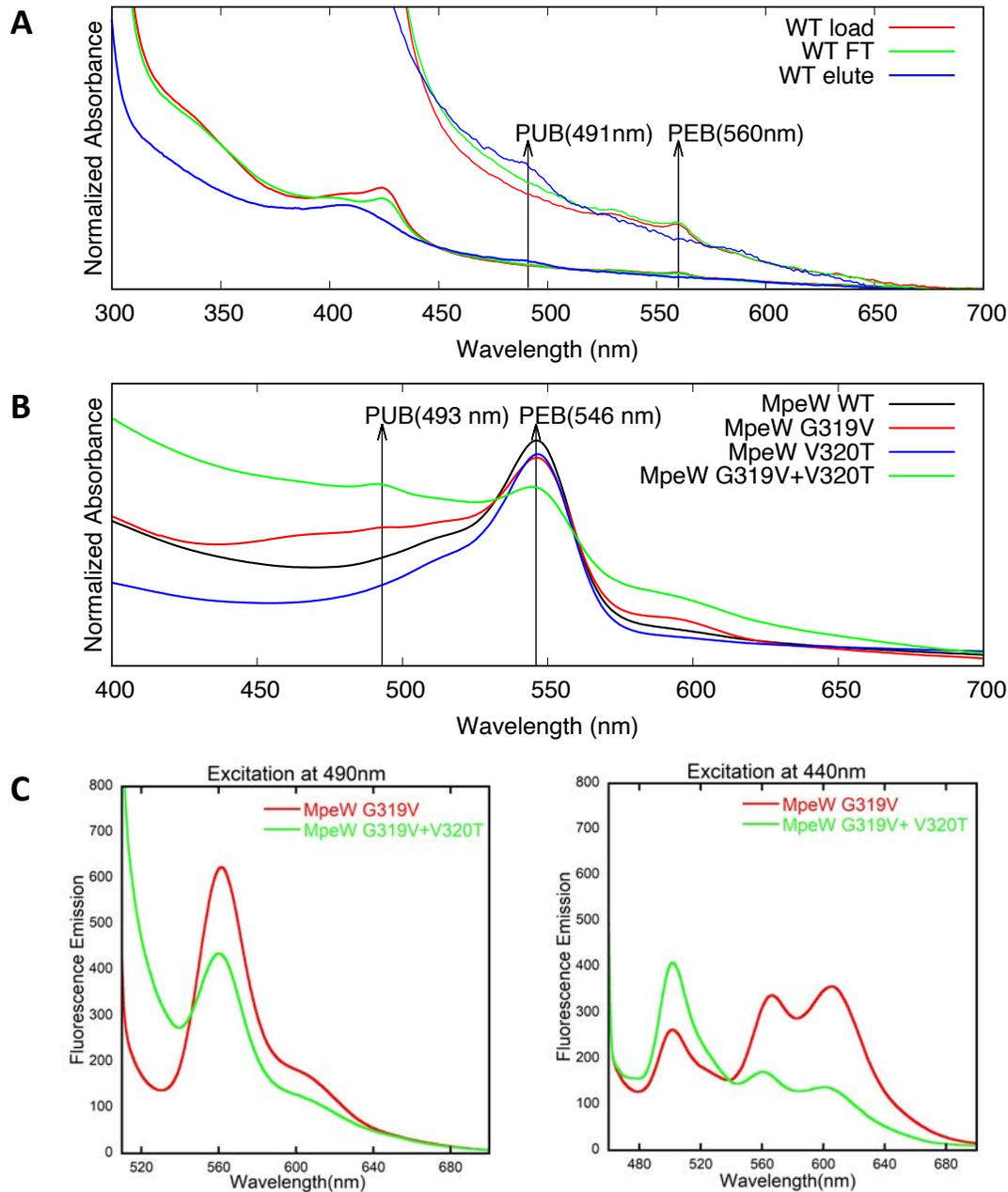


Figure S8. Factors that affect PUB formation in MpeQ and MpeW. (Related to Fig. 6,7). **A)**

To examine how bilin ligation to MpeA affects the PUB formation, we performed a co-expression experiment with MpeQ and MpeA-C83A along the PEB-producing construct (see STAR methods). The absorption spectra of the load (red), flowthrough (FT; green) and eluted peak fraction (blue) from Co^{2+} affinity chromatography were used to detect PEB and PUB. A 491-nm peak in the elute suggests the PUB formation in the absence of ligation to MpeA. Spectra between 450 and 700 nm were amplified by 5x fold to highlight the characteristic peaks of PEB and PUB. **B)** In a co-expression experiment of MpeA with MpeW mutants, PUB formation was detected in the G319V and G319V+V320T mutants of MpeW. **C)** Fluorescence emission spectra of MpeA-PUB and MpeA-PEB were obtained using excitation at 440 nm and 490 nm, respectively. Both G319V and double mutant G319V+V320T of MpeW exhibited emission signals associated with PUB. The PUB signal is higher in G319V+V320T than that of the single mutant G319V.

Table S1. Summary of MpeQ mutants

	MpeA-PUB*	MpeA-PEB	PUB:PEB	Lyase vs. isomerase	Comments
Wild Type (WT)	High	not detected	10:0	lyase-isomerase	
Y318A or Y318F	not detected	not detected	n/a	abolished/no switch	catalytic
K353A	very low	not detected	n/a	abolished/no switch	A-ring interaction
E285A	very low	not detected	0.5:0	reduced/no switch	A-ring interaction
W383F	medium	not detected	2:0	reduced/no switch	MpeA binding
R71A or R71D	very low	not detected	n/a	abolished/no switch	MpeA binding
R79A	very low	not detected	n/a	reduced/no switch	Propionate interaction
R79D	not detected	not detected	n/a	abolished/no switch	Propionate interaction
R135D	not detected	not detected	n/a	abolished/no switch	MpeA binding
R143D	very low	not detected	0.1:0	reduced/no switch	MpeA binding
R198A	very low	not detected	n/a	abolished/no switch	Substrate binding
R198K	low	not detected	3:0	reduced/no switch	Substrate binding
R228A	not detected	not detected	n/a	abolished/no switch	propionate interaction
R228D	low	not detected	0.7:0	reduced/no switch	propionate interaction
R287A	high	not detected		reduced/no switch	substrate binding
V319G	high	low	10:1	high/partial switch	Swap/active site
V319A	high	Very low	10:0.5	high/partial switch	Swap/active site
V319L	medium	not detected	5:0	reduced/partial switch	Swap/active site
V319F	very low	not detected	n/a	abolished/no switch	Swap/active site
A100T	high	not detected	4:0	high/no switch	Swap/remote site
S222A	high	not detected	3:0	reduced/no switch	Swap site
S224P	medium	not detected	6:0	high/no switch	Swap site
T320A	high	not detected	4:0	high/no switch	Swap site
Y323Q	high	not detected	4:0	high/no switch	Swap site
Y323L	medium	not detected	2:0	high/no switch	Swap site
T352A	high	not detected		high/no switch	Swap site
V319G/T320A	high	low	3:1	high/partial switch	Double swap sites
V319G/Y323Q	medium	low	3:1	medium/partial switch	Double swap sites
T320A/Y323Q	very low	not detected	n/a	abolished/no switch	Double swap sites
A100T/T352A	high	not detected	3:0	high/no switch	Double swap sites
V319G/T320A/Y323Q	medium	low	3:2	high/partial switch	Triple swap sites
V319G/A100T/T320A/Y323Q/T352A (Q5.1)	medium	medium	2:3	medium/switch	Five swap sites
V319G/S224P/T320A/Y323Q/T352A (Q5.2)	medium	high	1:3	medium/switch	Five swap sites

Note: *Compared to WT, High: >75%; medium: 25-50%; low: <25%; very low: <10%