ORIGINAL PAPERS

Evidence for an early green/red photocycle that precedes the diversification of GAF domain photoreceptor cyanobacteriochromes

Nibedita Priyadarshini[1,2] · Niklas Steube[3] · Dennis Wiens[3] · Rei Narikawa[4] · Annegret Wilde[1] ·
Georg K. A. Hochberg[3,5,6] · Gen Enomoto[1,7]

Abstract

Phytochromes are linear tetrapyrrole-binding photoreceptors in eukaryotes and bacteria, primarily responding to red and far-red light signals reversibly. Among the GAF domain-based phytochrome superfamily, cyanobacteria-specific cyanobacteriochromes show various optical properties covering the entire visible region. It is unknown what physiological demands drove the evolution of cyanobacteriochromes in cyanobacteria. Here, we utilize ancestral sequence reconstruction and biochemical verification to show that the resurrected ancestral cyanobacteriochrome proteins reversibly respond to green and red light signals. pH titration analyses indicate that the deprotonation of the bound phycocyanobilin chromophore is crucial to perceive green light. The ancestral cyanobacteriochromes show only modest thermal reversion to the green light-absorbing form, suggesting that they evolved to sense the incident green/red light ratio. Many cyanobacteria can utilize green light for photosynthesis using phycobilisome light-harvesting complexes. The green/red sensing cyanobacteriochromes may have allowed better acclimation to changing light environments by rearranging the absorption capacity of the phycobilisome through chromatic acclimation.

Graphical abstract
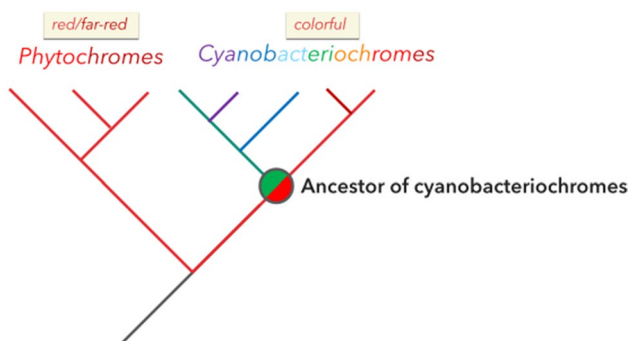
Keywords Ancestral sequence reconstruction · Bilin-based photoreceptor · Cyanobacteriochromes · Photobiology · GAF domain

---

This publication is dedicated to Prof. Silvia E. Braslavsky, a pioneer in photobiology and photobiophysics, on the occasion of her 80th birthday.

Nibedita Priyadarshini and Niklas Steube have contributed equally to this work and are listed in alphabetical order.

Extended author information available on the last page of the article

1 Introduction

Most light-dependent cellular responses are controlled by photoreceptors which sense light and then trigger downstream signal transduction events [1]. Members of the phytochrome superfamily of photoreceptors covalently bind a

linear tetrapyrrole (bilin) molecule as a chromophore to a cysteine (Cys) residue of the protein [2, 3]. The configuration of the bound bilin chromophore reversibly interconverts between *15Z* and *15E*, corresponding to the two isomers at the C15=C16 double bond (Figure S1) [4]. These two states of the chromophore often result in different optical properties, enabling the proteins to sense two different colors of light, in most cases red and far-red. The reversible photochromicity allows the photoreceptor to perceive the ratio of two wavelengths of the incident light. Many phytochromes show thermal reversion (dark reversion), reverting from *15E* to *15Z* without light absorption. Thermal reversion is a temperature-dependent process, and therefore the same photoreceptor integrates light and temperature signals [5, 6]. A fast dark reversion of a photoreceptor indicates that the protein senses the intensity of the incident light rather than the ratio of the two wavelengths [7–10].

Within the phytochrome superfamily, cyanobacteriochromes (CBCRs) are a distinct class of minimal photoreceptors [11, 12], which only need a single GAF (cGMP phosphodiesterase, adenylyl cyclase, and FhlA) domain to sense light genuinely. This contrasts with other phytochrome members that strictly require at least another neighboring PHY domain for genuine light perception [2, 3]. The functional light sensing module of canonical phytochromes features a typical PAS-GAF-PHY tridomain architecture, with the exception of some members lacking the PAS domain (knotless phytochromes) that are closely related to CBCRs [2, 3]. Phytochromes are widespread among eukaryotes and bacteria, whereas CBCRs are found exclusively in cyanobacteria, a group of photoautotrophic bacteria performing oxygenic photosynthesis. Through a process of gene duplication and domain shuffling, CBCRs have evolved a remarkable diversity in their absorption characteristics and thermal reversion kinetics [7, 13–16], making them a promising scaffold to develop a new generation of optogenetic tools [10, 17, 18]. Depending on their properties, CBCRs control a diverse range of physiological processes in cyanobacteria [19]. Green/red sensing CBCRs with slow reversion kinetics, including the first discovered CBCR RcaE, are used to adjust the relative amounts of red and green absorbing photosynthetic pigments (phycocyanin and phycoerythrin, respectively) in phycobilisomes during chromatic acclimation by sensing the ratio of green and red wavelengths [15, 20–22]. Blue/green sensing CBCRs, on the other hand, are considered to be used to detect shading by other cells in cyanobacterial mats [23, 24].

However, the original function of CBCRs remains unknown. We have previously speculated that blue/green perceiving CBCR-mediated cell shade sensing might be the ancestral function of these photoreceptors [23] because blue/green photochemistry is unique to CBCRs and should be more efficient than red/far-red phytochromes in an upper region of a microbial mat, where blue light diminishes while green, red, and far-red light are still available [25]. Further, early-branching cyanobacteria such as *Gloeobacter violaceus* PCC 7421 and *Anthocerotibacter panamensis* [26] only possess potential relatives of this kind of blue/green perceiving CBCRs based on sequence similarity, although they have not yet been characterized biochemically. However, the phylogenetic history of CBCRs is very complex, including frequent gene and domain duplications, making this question hard to resolve. It is difficult to make unambiguous predictions about the properties of the last common ancestor (LCA) of all CBCR GAF domains using existing phylogenies, because not enough GAF domains have been characterized biochemically, and their relative branching order remains uncertain [27].

Here, we used ancestral sequence reconstruction [28] to experimentally determine the photochemistry of the LCA of all extant CBCRs. We show that ancient CBCR proteins most likely sensed the ratio of green/red incident light, but not blue/green light. This inference is robust to alternative hypotheses about the exact branching order within CBCR GAF domains that is hard to resolve. Our results suggest that the first CBCR was likely used by cyanobacteria to tune the relative abundances of red and green light-absorbing pigments in response to changes in the incident light. The stunning diversity of colors sensed by extant CBCRs nowadays, therefore, may have evolved from an ancient CBCR most likely used for chromatic acclimation.

2 Results

2.1 Ancestral sequence reconstruction of cyanobacteriochromes

To investigate the characteristics of the earliest CBCRs, we first used maximum likelihood (ML) phylogenetics and ancestral sequence reconstruction to infer the most likely GAF domain sequence of the LCA of all extant CBCRs. To do this, we used HMMER to identify all CBCR GAF domains in 30 cyanobacterial species that span the entire known species diversity. We inferred a maximum likelihood phylogeny of 575 CBCR GAF domains. Although it is not yet clear which family of phytochromes evolved first, it is uncontroversial that knotless phytochromes form a closely related sister group to CBCRs [29]. Thus, we used 45 cyanobacterial knotless phytochrome GAF domains as the outgroup to root our tree.

The tree clearly separates the GAF domains of all cyanobacterial knotless phytochromes from the ones of all CBCRs on our tree (Fig. 1A, Figure S2). Beyond that, the phylogeny of the CBCR domains was extremely difficult to resolve. Our maximum likelihood tree did not contain any well-supported

Fig. 1 Ancestral CBCR GAF domain reconstruction on ML phylogenies. **A–C** Maximum Likelihood phylogenetic trees of cyanobacterial GAF domains used for ancestral sequence reconstruction. Numbers labeling clades denote the quantity of taxa. Colored squares highlight biochemically characterized domains and the colors they sense. "Ins-Cys" and "DXCIP" denote families sensing various colors. "M" indicates the extant early branching CBCR GAF domain of *Microcoleus* sp. FACHB1 MBD2125673 that we characterized. The clade of 19 first branching sequences shown in red was deleted for tree B. Node support is shown as approximate likelihood test statistics in italics. Scale bar: 0.2 average substitutions per site. Consensus neighbor and output domains of corresponding full-length proteins are shown to the right of the trees with domains that only appear in most of the proteins with dashed outlines. *var:* variable domains. *other:* conserved domains other than PAS (Per-Arnt-Sim), PHY (phytochrome-specific domain) or HK (histidine kinase/HATPase). **D** Amino acid sequences of the extant (M) and reconstructed ancestral GAF domains (Anc1-3). Arrows point positions important for color sensing in extant CBCRs, and states are red if conserved and blue if not
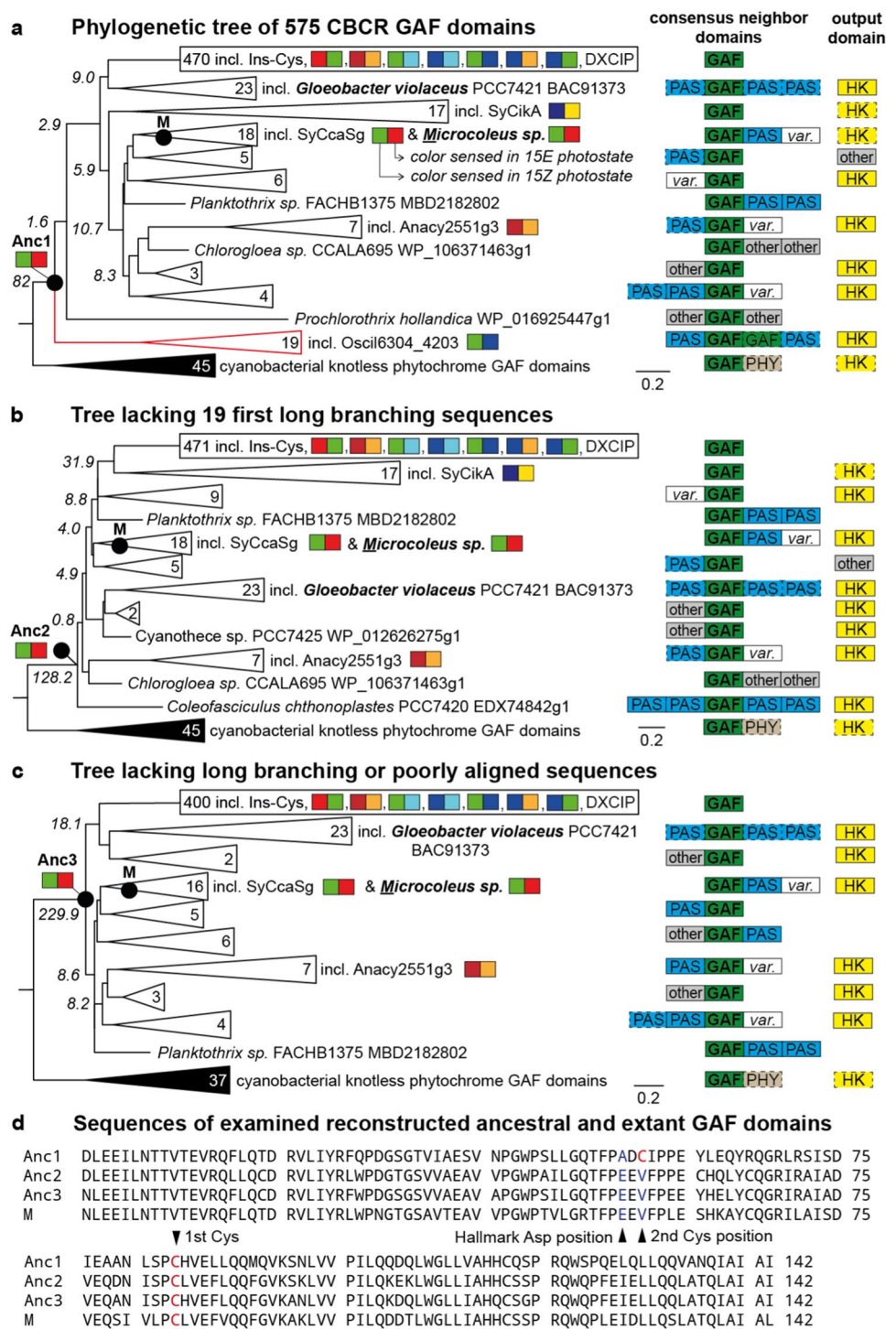
monophyletic groups of CBCR domains that clearly originated from gene duplication or domain-swapping events. CBCR domains are grouped loosely by domain architecture of the full-length proteins they are found in, but even these architectures vary substantially among GAF domains that group closely together. Mapping known CBCR color-sensing characteristics on the tree did not reveal an obvious pattern or a clear inference for the ancestral color. The earliest branching CBCRs on the tree presented here are green/red, red/orange, and green/blue receptors. The clade containing green/blue receptors connects to the root via a long branch, so its placement may result from a long branch attraction artifact (Fig. 1A, Figure S2).

Our phylogenetic tree implies that the exact branching order of CBCR GAF domains is not resolvable with current methods, making inferences about the LCA impossible

by comparing only the absorption/emission spectra of extant CBCRs. We reasoned that we might still gain some insights into its potential properties by ancestral sequence reconstruction, even if the topology of our ML tree could be wrong within the CBCR domains. Ancestral sequence reconstruction infers the likely sequence at internal nodes of the tree, given the tree topology, alignment, and a model of sequence evolution [28]. We reasoned that we could use this technique to test whether our different trees imply any consistent emission/absorption properties that are robust to phylogenetic uncertainty. All basal internal branches on our tree are short and poorly supported. Under such circumstances, it is possible that such errors do not affect reconstructions at functionally important sites (for which the signal should be strong) [30].

To test if there is any phylogenetic signal for a particular color sensing of the LCA on our tree phylogeny, we decided to use ancestral protein resurrection to test biochemically which color our tree implies. To do this, we inferred the sequence of the LCA of all CBCRs on our tree, resurrected the ancient GAF domain, and characterized it biochemically, as reported below. To determine if those characteristics strongly depend on the exact branching order of the tree or if they are robust to slight rearrangements of the poorly resolved branching order within CBCR GAF domains, we decided to infer two additional trees. For one, we only removed the first clade of long branching green/blue receptors and re-inferred the tree (Fig. 1B). For a third tree, we additionally removed sequences that were only poorly aligned or very long branching on our first tree (Fig. 1C). The two additional trees did not improve on the unresolved branching pattern inside the CBCRs, but had slight rearrangements near the root. Notably, in all three trees, the single GAF domain found in *Gloeobacter violaceus* PCC 7421 (the earliest branching cyanobacterial species on our trees) branched near the root. Furthermore, far-red/orange Ancy2551g3 and green/red SyCcaSg always appeared as early branching among the known characterized CBCR GAF domains. All three trees would be incorrect in the exact branching order within CBCR GAF domains. We, therefore, view the ancestral sequences we inferred from them not as a historically accurate inference, but simply as a test for whether there is any residual phylogenetic signal for the color of the LCA of all CBCRs that may be robust to slight rearrangements of branches near the root.

We inferred the most likely amino acid sequences of the LCA of extant CBCR GAF domains on all three topologies (Anc1–Anc3) to an average posterior probability of 0.81, 0.92, and 0.94, respectively (Figure S3). The ancestral sequences all contained the conserved "first" cysteine that binds the bilin chromophore in extant CBCRs as expected but differed at between 37 and 44 out of 142 total residues (Fig. 1D, Figure S4). To further validate our findings, we

attempted to characterize CBCR GAF domains of early branching extant species that only have short evolutionary distance (in branch lengths on our trees) to the reconstructed ancestral CBCR GAF domain sequences and review if their biochemical properties match the suggested ones of the ancestors.

CBCR GAF domains are located on a variety of proteins ranging from single domain up to multi-domain proteins that often contain several GAF domains. Some GAF domains function on their own as a CBCR, and others belong to other phytochromes that are strictly dependent on adjacent domains for genuine light perception. Large evolutionary distances between GAF domains on the same protein indicate early domain duplications or frequent horizontal transfer events between cyanobacterial species (Figure S5). To estimate the most probable domain architecture of the ancestral CBCR protein, we further compared the neighbor and output domains of the corresponding full-length proteins of CBCR GAF domains on all our trees. We found PAS domains mandatory in distantly related canonical phytochromes as the most abundant neighbors, and histidine kinase/HATPase domains as the most prominent output domains in early branching CBCRs on our trees (Fig. 1A–C). The trees presented here, thus, indicate that the LCA of all CBCRs was probably encoded on a phytochrome-like multidomain protein and transduced its signal to a histidine kinase domain.

2.2 Signal for a green/red photocycle in all ancestral CBCR GAF domains

We next determined the photochemical properties of the ancestral CBCR GAF domains. We expressed and purified the three ancestral sequences as recombinant N-terminal His-tagged proteins from *E. coli* harboring a biosynthesis plasmid for the chromophore phycocyanobilin (PCB). The Zn^{2+}-enhanced fluorescence of the purified proteins in an SDS-PAGE gel confirmed the covalent attachment of a bilin chromophore to the apoproteins (Figure S6) [31]. The absorbance spectra of the purified holoproteins showed spectral changes upon illumination with blue ($\lambda_{max} = 448$ nm), green ($\lambda_{max} = 514$ nm), and red light ($\lambda_{max} = 635$ nm). Irradiation with UV ($\lambda_{max} = 355$ nm) and far-red light ($\lambda_{max} = 731$ nm) did not affect the spectra. All ancestral proteins exhibited reversible photoconversion between green (Pg) and red (Pr) absorbing forms (Fig. 2). The bound chromophore species and its configuration were determined using acid denaturation spectra. The acid-denatured red-irradiated state (i.e., Pg) showed a peak at 662 nm and the green-irradiated state (i.e., Pr) at 585 nm, in agreement with *15Z* and *15E* forms of the covalently bound PCB, respectively (Figure S7) [32], indicating that Pg carries *15Z* PCB whereas Pr has *15E* PCB. The *15Z*Pg state showed absorption maxima between 515 nm and 540 nm, and the *15E*Pr state between 600 nm and 656 nm

Fig. 2 Absorption and difference spectra of the purified ancestral proteins. **A–C** Absorption spectra of the ^{15Z}Pg (red line), and of the ^{15E}Pr form (blue and green lines) of Anc1-3. The ^{15Z}Pg form was achieved by irradiation with red, the ^{15E}Pr form by either irradiation with blue or green for one minute. **D** Normalized photochemical difference spectra obtained by subtracting the absorption spectra of the ^{15Z}Pg from those of the ^{15E}Pr of Anc1-3. Difference spectra were normalized to the red photoproduct peak, and are vertically shifted for clarity. (A–C insets) The difference in the color of the ^{15Z}Pg and the ^{15E}Pr forms of Anc1-3 in solution at pH 7.5. All experiments were performed at room temperature
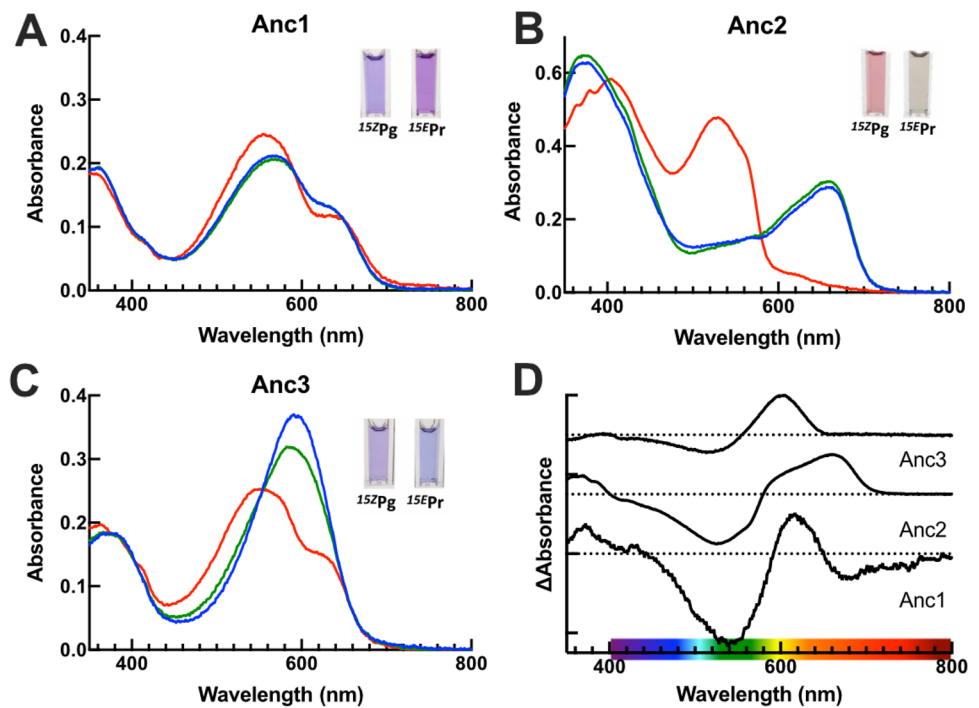
for all the ancestral proteins (Fig. 2, Table 1). For Anc2 and Anc3, irradiation with red ($\lambda_{max}=635$ nm) resulted in almost complete conversion to the ^{15Z}Pg form. For Anc1, we did not yield an apparently homogeneous population of ^{15Z}Pg by red light irradiation, probably due to the significant overlap of the absorption spectra of the two photostates (Fig. 2, Figure S7). The additional incubation of Anc1 overnight in the dark at room temperature allowed a seemingly complete conversion to ^{15Z}Pg (Figure S7). Irradiation with blue ($\lambda_{max}=448$ nm) and green ($\lambda_{max}=514$ nm) rendered almost complete conversion to the ^{15E}Pr state for Anc1 and Anc2. For Anc3, green irradiation resulted in partial conversion. The almost complete conversion was achieved upon blue irradiation, probably due to its good separation from the counteracting

red region (Fig. 2, Figure S7). Although blue light could induce photoconversion, we characterize the ancestral proteins as green-light sensors because the peak wavelengths of the absorption spectra and the difference spectra both fall into the green-light region (Fig. 2D).

We attempted to characterize further CBCR GAF domains of early branching extant species with a short evolutionary distance to the reconstructed ancestors on our trees, namely *Chlorogloea* sp. CCALA 695 WP_106371463.1, *Oscillatoria* sp. PCC 10802 WP_082218260.1, and *Microcoleus* sp. FACHB1 MBD2125673 WP_190776511.1. As we were not able to heterologously express sufficient amounts of the first two, we characterized the CBCR GAF domain of *Microcoleus* sp. with an evolutionary distance between 0.31 and 0.67

Table 1 Wavelengths of the absorbance peak maxima and the half-lives of thermal reversion of ancestral CBCR proteins at room temperature

Protein	$\lambda_{max, 15Z}$ (nm)	$\lambda_{max, 15E}$ (nm)	$\lambda_{max, 15Z, denatured}$ (nm)	$\lambda_{max, 15E, denatured}$ (nm)	Half-life (min)	References
Anc1 WT	540	610	663	589	233	This work
Anc2 WT	525	656	663	583	180	This work
Anc3 WT	515	600	667	589	310	This work
Anc1 C56V	541	621	n.d.	n.d.	n.d.	This work
Anc1 A54D	535	620	n.d.	n.d.	n.d.	This work
Anc2 E54D	525	660	n.d.	n.d.	n.d.	This work
Anc3 E54D	515	602	n.d.	n.d.	n.d.	This work
Cph1-PCB			669	573		[32]
TePixJ-PVB			600	507		[32]

The peak wavelengths were calculated using the difference spectra upon reversible photoconversion

n.d. not determined

on our trees, and found the same green/red perception as in the ancestral domains (Figure S8).

Taken together, although the spectral shapes are distinct among the three ancestral and the extant CBCR GAF domains, our results show a phylogenetic signal for a green/red photocycle in the LCA of all CBCRs, regardless of the exact branching order of basal CBCRs.

2.3 PCB was the chromophore in ancestral CBCRs

Although most CBCRs incorporate PCB, some CBCRs can bind biliverdin IXa (BV) as the chromophore with variable specificity [33, 34]. To determine the efficiency of BV incorporation by the ancestral proteins, we expressed all of them with a BV biosynthesis plasmid in *E. coli* and purified them (Figure S6). Acid denaturation spectra confirmed the attached chromophore to be BV with the denatured ^{15Z}Pg peaking at around 700 nm (Figure S9) [34]. All ancestral proteins showed slight photoconversion with BV as the chromophore upon irradiation with both green and red light. However, for Anc1 and Anc2, neither lights were sufficient to cause complete photoconversion to either *15E* or *15Z* photostates (Figure S9). Red irradiation caused a complete conversion of Anc3-BV to the *15Z* photostate. However, a complete conversion to the *15E* photostate was not achieved by green irradiation. These data suggest that the ancestral CBCRs may have been able to bind to both PCB and BV, but that photoconversion may have been efficient with PCB. Specificity for BV would then be a derived trait of some crown-group CBCRs [33]. This is consistent with cyanobacterial knotless phytochromes in the outgroup, also being specific for PCB [35, 36]. Besides, PCB is one of the prosthetic groups of the phycobiliproteins of the photosynthetic antenna complex and is much more abundant than BV in cyanobacterial cells [37].

2.4 CBCR GAF reconstructions suggest a function as a sensor of the spectral ratio via a protochromic photocycle

We next asked whether the heterologously expressed ancestral proteins sensed the intensity of green or red light rather than the red/green ratio. To determine this, we measured their rates of thermal reversion. Fast thermal reversion leads to short-lived photoproducts regardless of any counteracting light. Therefore, the population of the photoproduct only depends on the intensity of light that excites the dark state [7–10]. In contrast, slow thermal reversion allows the formation of long-lived photostates and therefore supports sensing of the ratio of two different wavelengths. All three ancestral proteins underwent slow thermal reversion from ^{15E}Pr to ^{15Z}Pg in the dark at room temperature (Figure S10): The half-lives for the thermal reversion in the dark at room

temperature ranged between 180 min and 310 min (Table 1), comparable to the related knotless phytochromes [35]. These half-lives are much longer than those of known intensity-sensing CBCRs, which revert within the range of several seconds [7–10]. Our results, therefore, indicate that the LCA of all CBCRs likely sensed the ratio of green to red incident light rather than the intensity of these wavelengths.

Extant green/red light-sensing CBCRs adopt a protochromic photocycle [15, 38]. The conjugated π system of the bilin chromophore of the green/red CBCRs is deprotonated with a lower p*K*a value in the *15Z* state to absorb green light, whereas it is protonated with a higher p*K*a value in the *15E* state to absorb red light. To assess whether this was also the ancestral photocycle mechanism in CBCR GAF domains, we performed pH titration analysis for the three ancestral proteins.

Anc1–3 showed a decrease in absorption in the red-light region (600–660 nm) and an increase in green-light absorption (520–540 nm) at higher pH conditions (Fig. 3, Figure S11). At lower pH conditions, red-light absorption increased and green-light absorption decreased, except for Anc2 *15Z*, which showed stable green-light absorption under the tested pH conditions. The absorption changes were fitted with one titrating group of the Henderson–Hasselbalch equation to estimate p*K*a [15]. The p*K*a values of the *15Z* chromophore are lower than those of *15E*, indicating that the *15Z* chromophore has a lower affinity to protons (Table 2). The difference in p*K*a values between *15Z* and *15E* was the smallest in Anc1 (Table 2), which may be consistent with its poor spectral shift upon photoconversion under the standard pH condition of 7.5 (Fig. 2). The much lower p*K*a of Anc2 *15Z* may be due to the leucine residue next to the chromophore-binding cysteine, which is important for stabilization of the deprotonation of the chromophore [15, 39]. These results suggest that a photochromic photocycle similar to that of extant green/red CBCRs may have been the ancestral photoswitching mechanism.

2.5 The amino acids aligned at the conserved CBCR hallmark residues do not control the green/red photocycle

Lastly, we sought to gain insights into the molecular mechanisms of color tuning of the reconstructed CBCR proteins relative to canonical red/far-red phytochromes. We first focused on what allows deprotonation of the chromophore. In canonical phytochromes, the chromophore is protonated in both photostates [35, 38, 40, 41]. The protonated state is stabilized by a conserved aspartate (Asp) residue at position 54 [the numbering of the amino acid is based on the multiple sequence alignment (Figure S4)] that forms a hydrogen bond network with the nitrogen atoms of the B and C pyrrole rings of the chromophore [42–45]. The resurrected CBCR

Fig. 3 Protochromic absorption spectra changes of the ancestral proteins. **A**–**F** pH-dependent absorbance spectra of Anc1-3 with the configuration of *15Z* (**A**, **C**, **E**) or *15E* (**B**, **D**, **F**) measured in buffers with pH between 5.0 (dark red) and 11.0 (dark purple) in 0.5 pH steps. Increased scattering was observed at lower pH of 5.0 and 5.5, probably due to partial protein aggregation. For the analysis, samples were irradiated to obtain homogenous *15Z* and *15E* photostates, followed by mixing with 1 M buffers of different pH in 1:4 ratio and immediate measurement of absorption spectra. Note that the homogenous *15Z* of Anc1 was prepared by overnight incubation of the protein in the dark
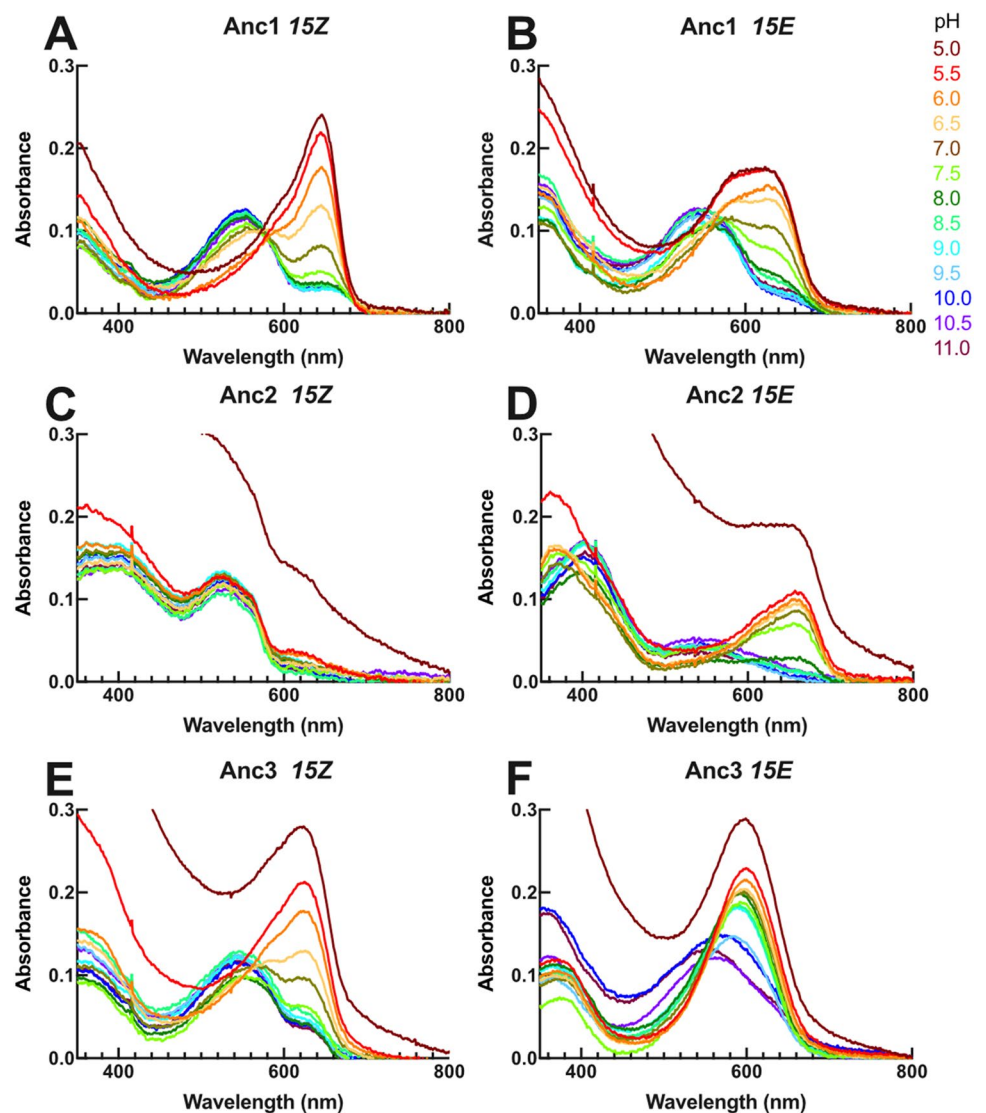
ancestral proteins feature either alanine or glutamate residue at this position, suggesting that the substitution of Asp to a different amino acid might have allowed the deprotonation of the chromophore. To test this hypothesis, we mutated this site to Asp in all three ancestral proteins, mimicking the situation in canonical phytochromes and most CBCRs. We then determined whether the deprotonation of the chromophore was affected. Surprisingly, green-light absorption and deprotonation were unaffected in all three mutants (Tables 1, 2, Figure S12–13). This suggests that the loss of the protonation-stabilizing Asp was neither essential for the evolution of a deprotonated chromophore in the *15Z* photostate nor for green-light absorption.

Finally, we investigated the influence of another site—the so-called 'second cysteine' at position 56 that is known to influence spectral tuning in extant CBCRs. CBCRs containing this Cys residue form a thioether linkage with the C10 position of the bilin chromophore [46]. The disruption of the π-conjugated system at the C10 position leads to absorption in the UV-to-blue region [14, 47]. The covalent bond formation between the chromophore and the second Cys can be reversibly induced by the light-induced conformational change of the chromophore and the protein. Some 2nd-Cys-containing proteins retain the covalent bond in both *15Z* and *15E* states. The evolution of this second Cys could have contributed to the spectral properties that distinguish CBCRs from canonical phytochromes. However, the predicted ancestral sequences are in disagreement with the presence of the second Cys in the LCA of all CBCRs: only Anc1 harbors the second Cys residue, whereas Anc2 and Anc3 have a valine at this position (Fig. 1D, Figure S4). Although all three proteins have a green/red photocycle, this introduces ambiguity about whether the second Cys played an essential role in the evolution of the green/red photocycle. The function of this cysteine may depend on the specific context of the protein, such as the neighboring amino acid residues,

Table 2 The estimated pKa values of the ancestral CBCR proteins

Protein/configuration	pKa	Absorption peaks for fitting (nm)	R^2
Anc1 WT/*15Z*	6.54	650	0.9994
Anc1 WT/*15E*	7.22	635	0.9936
Anc2 WT/*15Z*	<5.0	n.a.	n.a.
Anc2 WT/*15E*	7.57	670	0.9613
Anc3 WT/*15Z*	6.59	630	0.9859
Anc3 WT/*15E*	9.35	610	0.9501
Anc1 C56V/*15Z*	6.77	650	0.9959
Anc1 C56V/*15E*	7.46	635	0.9887
Anc1 A54D/*15Z*	6.57	650	0.9941
Anc1 A54D/*15E*	7.61	635	0.9913
Anc2 E54D/*15Z*	<5.0	n.a.	n.a.
Anc2 E54D/*15E*	7.75	670	0.9955
Anc3 E54D/*15Z*	6.58	630	0.9653
Anc3 E54D/*15E*	9.48	610	0.9697

The pKa values were calculated using the data of the absorption changes in the pH titration experiments in Fig. 3

n.a. not applicable

although the second Cys is functional in many proteins from different lineages within CBCRs [47]. To address this issue, we mutated the Cys at position 56 of Anc1 to valine (identical to the state in Anc2 and Anc3) and tested for differences in spectral properties. The mutation only slightly elevated the absorbance in the red region compared to the green one of both *15E* and *15Z* photostates, but without affecting the absorption maxima (Tables 1, 2, Figure S12–13). This confirms that a green/red photocycle was likely present in the LCA of all CBCR GAF domains, regardless of the presence of the second cysteine in the ancestral protein.

3 Discussion

3.1 The first CBCRs could have functioned in chromatic acclimation

Our results suggest that the LCA of extant CBCRs may have functioned as a green/red light sensor with slow thermal reversion that used a protochromic photocycle similar to that of extant green/red sensing CBCRs. However, we caution that this inference is based on trees with unresolved and presumably incorrect topologies within the CBCRs. The fact that we observed similar properties on three different topologies is encouraging, suggesting that the signal for a green/red photocycle may persist independently of the exact topology. However, biases in the data systematically could favor incorrect topologies that then lead to ancestors with misleading biochemical properties [48]. A green/red photocycle

might be the genetically simplest one, and we may observe it because our reconstructions fail to correctly incorporate all states necessary to produce any other kind of photochemistry. In light of these caveats, we do not exclude that the LCA of all CBCR GAF domains had different characteristics.

It is unlikely that more CBCR GAF sequences would improve our inference in the future. Fundamentally, we are limited by the small size (~ 140 a.a.) and fast evolution of CBCR GAF domains. The complex architecture of CBCR GAF domain-containing proteins further complicates the phylogeny of these proteins. Our trees must contain gene duplications of entire CBCR GAF domain-containing proteins, internal duplications that produce proteins containing two or more CBCR GAF domains, and possibly horizontal transfers, domain fusions, and gene conversion events between adjacent CBCR GAF domains. This makes the gene trees of these domains extremely difficult to interpret. Solving this problem will likely require inferring the histories of other domains in CBCR GAF domain-containing proteins and using reconciliation approaches to infer a global history of how CBCR GAF domains were added and removed from different proteins.

An ancestral green/red photocycle is, however, also likely in the light of ecological relevance. What might have been the physiological function of green/red sensing ancestral CBCRs? The first discovered CBCR, RcaE, is a green/red sensing protein as the regulator of chromatic acclimation [15, 22]. One plausible answer upon comparison with such extant CBCRs with similar photocycle suggests their involvement in regulating the relative amounts of red-absorbing phycocyanin and green-absorbing phycoerythrin in phycobilisomes during chromatic acclimation [21]. This implies that the LCA of all extant cyanobacteria, in which the here identified ancestral GAF domain would have existed, already possessed phycoerythrin. The Gloeobacterales (the earliest diverging clade of cyanobacteria) usually possess phycoerythrin, suggesting that the pigment has an ancient origin [26, 37, 49] and that the ability for chromatic acclimation already existed in the earliest cyanobacteria. The analysis of neighboring domains further supports this hypothesis as the extant known chromatic acclimation regulators harbor an additional PAS domain and a histidine kinase as the output domain [20]. It is of note that extant green/red CBCRs also regulate other types of chromatic acclimation, such as controlling the relative amounts of the yellow-green-absorbing phycoerythrocyanin protein or a rod-membrane linker CpcL protein, which assembles a photosystem I-specific phycobilisome only in green light [20, 50]. Thus, green/red light sensing could be crucial even for cyanobacterial strains lacking green-absorbing phycoerythrin.

Chromatic acclimation was likely important to early cyanobacteria, as a current analysis points to them having lived in sessile microbial mats [51]. In these environments,

the availability of different wavelengths of light can change dramatically and rapidly across minute distances, depending on the depth of the cell in the mat or the composition of the overlying cells [23].

3.2 Tuning of the chromophore towards green/red sensing

Based on our current work, we can speculate about the genetic mechanism responsible for the evolution of the CBCR's green/red light sensitivity from red/far-red sensing canonical phytochromes. If the green/red photocycle was ancestral to all CBCRs, two changes must have occurred relative to canonical phytochromes: the shift of the *15Z* state from red to green, and that of the *15E* state from far-red to red-light absorption.

In the resurrected CBCR ancestral proteins, the *15Z* state is deprotonated. This is different from phytochromes, in which the bilin chromophore is protonated in both photostates [35, 38, 40, 41], implying that deprotonation of the chromophore is important for green-light absorption. The ancestral proteins all lack the conserved Asp, which is allegedly important for the stabilization of the protonated state in phytochromes and CBCRs [44, 45], suggesting that this substitution may have allowed for deprotonation. The side chain of the Asp residue is involved in the hydrogen bond network with the bilin chromophore in CBCRs [43, 52], whereas it is generally oriented toward the outside of the chromophore-binding pocket in phytochromes [53, 54]. The Alphafold2 prediction of the structures of the Anc proteins suggests that the amino acids at the hallmark Asp position could form the hydrogen bond network with the chromophore (Figure S14). However, introducing the Asp back into the ancestral photoreceptors does not abolish deprotonation, implying the involvement of other factors for deprotonation of the chromophore.

In addition, observations from extant CBCRs and phytochromes suggest that deprotonation alone is likely insufficient to yield green-light absorption: the cyanobacterial canonical phytochrome Cph1 exhibits a pKa of ~9.0 in the *15Z* and *15E* photostates to stabilize the protonated chromophore. Increasing the solvent pH induces a decrease in red-light absorption by Cph1 but does not cause an increase in green-light absorption [55]. The red/green CBCR AnPixJg2 retains the protonated chromophore even at the green-absorbing state, and artificial deprotonation does not affect the green absorption [56]. This suggests that green absorption requires additional amino acid substitutions affecting the light wavelength absorbed by the deprotonated chromophore.

The *15E* state is also hypsochromically shifted from far-red to red absorption. This could have occurred through the loss of the adjacent PHY domain from an ancestral phytochrome-like precursor. Such truncations led to a blue shift of the far-red absorbing state of extant phytochromes [36, 42, 57]. Another suggested tuning mechanism is the "second" Cys, which is found near the chromophore and is known to influence the absorption properties of proteins from various lineages of CBCR GAF domains [14, 47, 58, 59]. However, the reconstructed ancestral proteins vary in the amino acid at that position; Anc1 has a Cys, whereas Anc2 and Anc3 both have valine. Although the Alphafold2 prediction locates the second Cys near the C10 of the chromophore (Figure S14), mutating this cysteine in Anc1 has essentially no effect on optical properties, suggesting that in the LCA of all CBCR GAF domains this site is likely not involved in color tuning. Further exploration would be necessary to shed light on the exact genetic mechanism that transformed a likely red/far-red sensing phytochrome into a green/red sensing CBCR.

3.3 The genetic basis of CBCRs may have diversified from an ancestral green/red light sensor

Our results hint at how the remarkable diversity of colors found in extant CBCRs may have evolved from a green/red sensing ancestor. The ancestral proteins reconstructed in this work possess the ability also to sense blue light, which was perhaps later exploited in CBCRs with blue-light photocycles. Additionally, the ancestral photoreceptors most likely already had the ability to bind BV, which could have enabled the evolution of several extant CBCR groups that utilize BV in their photocycle and are hence able to perceive different wavelengths. The evolution of two-color sensing in the LCA of CBCR GAF domains probably made it easier to further tinker with the exact wavelengths of the *15Z* and *15E* photostates through changes affecting the local environment and pKa of the chromophore. Our characterization of sequences representative of the first CBCR is a first step in understanding how this tinkering occurred in the colorful history of CBCR proteins.

4 Methods

4.1 Phylogenetics and ancestral sequence reconstruction

Amino acid sequences of cyanobacterial proteins containing GAF domains were gathered using protein–protein BLAST (non-redundant protein sequences (nr) database) and a CBCR protein as a query [60]. Models (XM/XP) and uncultured/environmental sample sequences were excluded from the search. Protein sequences were manually selected to represent all large groups of the whole known cyanobacterial species phylogeny based on recently published data [61].

Sequences that were annotated to multiple species as well as incomplete sequences were excluded. Conserved domains of each sequence were identified with the HMMER web server using the Pfam database [62]. GAF domain sequences were aligned with MUSCLE 3.8 [63], and the alignment was manually cropped to remove gaps by deleting lineage-specific inserts [64]. The cropped alignment was used to infer an initial ML phylogeny using RAxML [65] in the PROTGAMMAAUTO mode resulting in the LG likelihood model with fixed base frequencies. The resulting tree was rooted using GAF domain sequences of cyanobacterial proteins lacking the PAS domain but containing a PHY domain as an outgroup (cyanobacterial knotless phytochromes) [66]. The last common ancestor of all CBCR GAF domains (Anc1) was reconstructed at the internal node indicated in Fig. 1A on Tree A using the CodeML package of PAML [67] with the LG substitution model and 16 gamma categories. Due to the suspicious long branch of the 19 first branching sequences, an alternative tree (Tree B) was inferred by the deletion of these sequences from the corresponding alignment. An alternative ancestor (Anc2) was equivalently reconstructed on Tree B. For the third ancestral sequence (Anc3), Tree C was inferred after deleting all domains with particular long branches or poorly aligned sequences from the alignment. The robustness of each topology was tested by running 100 nonparametric bootstraps and calculating the transfer bootstrap estimates (TBE) for internal nodes using the BOOSTER web tool [68]. Additionally, approximate likelihood ratios were calculated with PhyML [69]. The consensus neighbor and output domains of each group on the trees were determined manually and mapped next to the topologies (Fig. 1).

4.2 Plasmid construction

Codon-optimized sequences for *E. coli* encoding the ancestral CBCR GAF domains of Anc1, Anc2, and Anc3, and *Microcoleus* sp. FACHB1 MBD2125673 WP_190776511.1 (Table S1) were obtained from Twist Bioscience (San Francisco, California, USA) or Eurofins Genomics (Ebersberg, Germany). The synthesized gene fragments were amplified by PCR and subcloned into a pET28V vector containing an N-terminal, TEV-cleavable 6×His tag via assembly cloning (AQUA cloning) [70]. Utilized oligonucleotides are provided in Table S2. Sequences of the constructs were confirmed by Sanger sequencing.

The PCB chromophore biosynthesis plasmid pTDho1pcyA was a kind gift from Prof. Nicole Frankenberg-Dinkel (University of Kaiserslautern) [71]. The N-terminal 6xHis tag of PcyA was removed via AQUA cloning using the primers pTDho1pcyA-1F/-2R to obtain pTDho1pcyA-HisTag. For the construction of the BV-producing plasmid, the *pcyA*

gene was deleted via AQUA cloning using the primers pTDho1pcyA-3bF/-4bR to obtain the pTDho1 plasmid.

4.3 Protein expression and purification

The *E. coli* strain BL21(DE3) was co-transformed with one of the pET28V plasmids harboring the gene for the target CBCR GAF domains, and either the PCB-producing pTDho1pcyA-HisTag plasmid or the BV-producing pTDho1 plasmid. The cultures were induced with 0.1 mM isopropyl-β-D-thiogalactopyranosid and grown overnight at 18 °C in LB medium with appropriate antibiotics. The cells were harvested and disrupted three times using a French cell press (50 ml, Aminco French Pressure Cell Press) at 20,000 psi in 50 mM HEPES·NaOH, pH 7.5; 300 mM NaCl, 10% (w/v) glycerol, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), and 30 mM imidazole. The His-tagged proteins were purified by affinity chromatography with nickel affinity columns (HisTrap 1 ml; Cytiva) using the Äkta pure system (GE Healthcare UK Ltd.) from approximately 35 ml of extract. The column was washed with 10 ml of 50 mM HEPES·NaOH, pH 7.5; 300 mM NaCl, 10% (w/v) glycerol, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), and 30 mM imidazole at a flow rate of 1 ml/min after application of the sample. Elution was carried out at a flow rate of 1 ml/min with all solutions maintained at 4 °C at a linear imidazole concentration gradient from 30 to 530 mM.

4.4 SDS-PAGE and fluorescence detection of the bound bilin chromophore

To check the purity of the protein samples, they were first denatured using 62.5 mM Tris-HCl, pH 6.8; 11.25% (w/v) glycerol, 4% SDS, 10 mM DTT, and 0.0125% (w/v) bromophenol blue and incubated at 95 °C for 5 min. They were separated by SDS polyacrylamide gel electrophoresis using a 16% Tris-Tricine acrylamide gel [72]. The gel was then incubated in 2 mM zinc acetate solution for 15 min and fluorescence signals were imaged using a Fusion SL (Peqlab) with an F595 Y3 filter. The gel was further stained with Coomassie G-250.

4.5 Light sources

To irradiate purified proteins, LEDs illuminating at 355 nm for UV light (0.45 µmol photons m^{-2} s^{-1}), 448 nm for blue light (516 µmol photons m^{-2} s^{-1}), 514 nm for green light (540 µmol photons m^{-2} s^{-1}), 635 nm for red light (600 µmol photons m^{-2} s^{-1}), and 731 nm for far-red light (241 µmol photons m^{-2} s^{-1}) were used (Figure S15).

4.6 Spectroscopy and pH titration analysis

To measure the absorption spectra, the purified proteins were dialyzed in 50 mM HEPES·NaOH, pH 7.5; 300 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP using desalting columns (HiTrap 5 ml; Cytiva), followed by irradiation with a specific wavelength for around one minute each at room temperature. The absorption spectra were acquired using a UV-2450 spectrophotometer (Shimadzu) in the dark. Thermal reversion was achieved by incubating the samples in the dark overnight at room temperature. To acquire the absorption spectra of the acid-denatured proteins, 140 µl of the protein sample was mixed with 560 µl of 10 M urea (pH 2.0) by pipetting, followed by immediate measurement of absorbance spectra.

For pH titration, the purified protein was dialyzed in 10 mM HEPES·NaOH, pH 7.5; 300 mM NaCl, 0.5 mM TCEP using desalting columns (HiTrap 5 ml; Cytiva). 560 µl of the protein was converted to either *15E* or *15Z* photostate by irradiation of either blue, green, or red light for one minute or incubation in the dark overnight, followed by the addition of 140 µl of the following buffers in the dark (each 1M): MES-NaOH for pH 5.0–6.5; HEPES-NaOH for pH 7.0–8.5; or glycine-NaOH for pH 9.0–11.0. The pH titration data were analyzed by fitting the absorbance value at a particular wavelength using nonlinear regression in Prism software. The pKa values of the chromophore were determined using Henderson–Hasselbalch equations of a single titrating group [15, 44].

4.7 AlphaFold2 structure predictions

AlphaFold2 structural predictions of the ancestral CBCR GAF domains (Anc1-3) were generated utilizing the ColabFold server on 10/18/2022 with default settings [73]. Structures were aligned to the crystal structure of the chromophore-bound NpR6012g4 (PDB ID: 6BHN) and TePixJg (PDB ID: 4GLQ) [46, 74]. Data were visualized with the Pymol Molecular Graphics System v2.4.0 (Schrödinger, LLC; New York, NY). Hallmark residues for the interaction with the chromophore in Anc1-3 were displayed in Fig. S14.

Declarations

References

1. Möglich, A., Yang, X., Ayers, R. A., & Moffat, K. (2010). Structure and function of plant photoreceptors. *Annual Review of Plant Biology, 61*, 21–47.
2. Rockwell, N. C., & Lagarias, J. C. (2010). A brief history of phytochromes. *ChemPhysChem, 11*, 1172–1180.
3. Anders, K., & Essen, L. O. (2015). The family of phytochrome-like photoreceptors: Diverse, complex and multi-colored, but very useful. *Current Opinion in Structural Biology, 35*, 7–16.
4. Song, C., et al. (2014). The D-ring, not the A-ring, rotates in *Synechococcus* OS-B' phytochrome. *Journal of Biological Chemistry, 289*, 2552–2562.
5. Klose, C., Nagy, F., & Schäfer, E. (2020). Thermal reversion of plant phytochromes. *Molecular Plant, 13*, 386–397.
6. Jung, J. H., et al. (2016). Phytochromes function as thermosensors in Arabidopsis. *Science, 354*, 886–889.
7. Rockwell, N. C., Martin, S. S., & Lagarias, J. C. (2012). Red/green cyanobacteriochromes: Sensors of color and power. *Biochemistry, 51*, 9667–9677.
8. Hasegawa, M., et al. (2018). Molecular characterization of DXCF cyanobacteriochromes from the cyanobacterium *Acaryochloris marina* identifies a blue-light power sensor. *Journal of Biological Chemistry, 293*, 1713–1727.
9. Fushimi, K., et al. (2016). Cyanobacteriochrome photoreceptors lacking the canonical Cys residue. *Biochemistry, 55*, 6981–6995.
10. Fushimi, K., Enomoto, G., Ikeuchi, M., & Narikawa, R. (2017). Distinctive properties of dark reversion kinetics between two red/green-type cyanobacteriochromes and their application in the photoregulation of cAMP synthesis. *Photochemistry and Photobiology, 93*, 681–691.
11. Ikeuchi, M., & Ishizuka, T. (2008). Cyanobacteriochromes: A new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochemical & Photobiological Sciences, 7*, 1159–1167.

12. Rockwell, N. C., & Lagarias, J. C. (2020). Phytochrome evolution in 3D: Deletion, duplication, and diversification. *New Phytologist, 225*, 2283–2300.

13. Fushimi, K., & Narikawa, R. (2019). Cyanobacteriochromes: Photoreceptors covering the entire UV-to-visible spectrum. *Current Opinion in Structural Biology, 57*, 39–46.

14. Rockwell, N. C., Martin, S. S., Feoktistova, K., & Lagarias, J. C. (2011). Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. *Proceedings of the National Academy of Sciences USA, 108*, 11854–11859.

15. Hirose, Y., et al. (2013). Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. *Proceedings of the National Academy of Sciences USA, 110*, 4974–4979.

16. Rockwell, N. C., Martin, S. S., Gulevich, A. G., & Lagarias, J. C. (2014). Conserved phenylalanine residues are required for blue-shifting of cyanobacteriochrome photoproducts. *Biochemistry, 53*, 3118–3130.

17. Oliinyk, O. S., Chernov, K. G., & Verkhusha, V. V. (2017). Bacterial phytochromes, cyanobacteriochromes and allophyco-cyanins as a source of near-infrared fluorescent probes. *International Journal of Molecular Sciences, 18*, 1691.

18. Blain-Hartung, M., Rockwell, N. C., & Lagarias, J. C. (2017). Light-regulated synthesis of cyclic-di-GMP by a bidomain construct of the cyanobacteriochrome Tlr0924 (SesA) without stable dimerization. *Biochemistry, 56*, 6145–6154.

19. Wiltbank, L. B., & Kehoe, D. M. (2019). Diverse light responses of cyanobacteria mediated by phytochrome superfamily photoreceptors. *Nature Reviews Microbiology, 17*, 37–50.

20. Hirose, Y., et al. (2019). Diverse chromatic acclimation processes regulating phycoerythrocyanin and rod-shaped phycobilisome in cyanobacteria. *Molecular Plant, 12*, 715–725.

21. Sanfilippo, J. E., Garczarek, L., Partensky, F., & Kehoe, D. M. (2019). Chromatic acclimation in cyanobacteria: A diverse and widespread process for optimizing photosynthesis. *Annual Review of Microbiology, 73*, 407–433.

22. Kehoe, D. M., & Grossman, A. R. (1996). Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science, 273*, 1409–1412.

23. Enomoto, G., & Ikeuchi, M. (2020). Blue-/green-light-responsive cyanobacteriochromes are cell shade sensors in red-light replete niches. *iScience, 23*, 100936.

24. Conradi, F. D., Mullineaux, C. W., & Wilde, A. (2020). The role of the cyanobacterial type IV pilus machinery in finding and maintaining a favourable environment. *Life, 10*, 252.

25. Ohkubo, S., & Miyashita, H. (2017). A niche for cyanobacteria producing chlorophyll *f* within a microbial mat. *ISME Journal, 11*, 2368–2378.

26. Rahmatpour, N., et al. (2021). A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. *Current Biology, 31*, 2857-2867.e4.

27. Rockwell, N. C., Martin, S. S., & Lagarias, J. C. (2015). Identification of DXCF cyanobacteriochrome lineages with predictable photocycles. *Photochemical & Photobiological Sciences, 14*, 929–941.

28. Hochberg, G. K. A., & Thornton, J. W. (2017). Reconstructing ancient proteins to understand the causes of structure and function. *Annual Review of Biophysics, 46*, 247–269.

29. Ulijasz, A. T., & Vierstra, R. D. (2011). Phytochrome structure and photochemistry: Recent advances toward a complete molecular picture. *Current Opinion in Plant Biology, 14*, 498–506.

30. Hanson-Smith, V., Kolaczkowski, B., & Thornton, J. W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular Biology and Evolution, 27*, 1988–1999.

31. Berkelman, T. R., & Lagarias, J. C. (1986). Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Analytical Biochemistry, 156*, 194–201.

32. Ishizuka, T., Narikawa, R., Kohchi, T., Katayama, M., & Ikeuchi, M. (2007). Cyanobacteriochrome TePixJ of *Thermosynechococcus elongatus* harbors phycoviolobilin as a chromophore. *Plant and Cell Physiology, 48*, 1385–1390.

33. Moreno, M. V., Rockwell, N. C., Mora, M., Fisher, A. J., & Lagarias, J. C. (2020). A far-red cyanobacteriochrome lineage specific for verdins. *Proceedings of the National Academy of Sciences, 117*, 27962–27970.

34. Narikawa, R., et al. (2015). A biliverdin-binding cyanobacteriochrome from the chlorophyll d-bearing cyanobacterium *Acaryochloris marina*. *Science and Reports, 5*, 7950.

35. Anders, K., et al. (2011). Spectroscopic and photochemical characterization of the red-light sensitive photosensory module of Cph2 from *Synechocystis* PCC 6803. *Photochemistry and Photobiology, 87*, 160–173.

36. Wu, S. H., & Lagarias, J. C. (2000). Defining the bilin lyase domain: Lessons from the extended phytochrome superfamily. *Biochemistry, 39*, 13487–13495.

37. Watanabe, M., & Ikeuchi, M. (2013). Phycobilisome: Architecture of a light-harvesting supercomplex. *Photosynthesis Research, 116*, 265–276.

38. Osoegawa, S., et al. (2019). Identification of the deprotonated pyrrole nitrogen of the bilin-based photoreceptor by Raman spectroscopy with an advanced computational analysis. *The Journal of Physical Chemistry B, 123*, 3242–3247.

39. Nagae, T., et al. (2021). Structural basis of the protochromic green/red photocycle of the chromatic acclimation sensor RcaE. *Proceedings of the National Academy of Sciences, 118*, e2024583118.

40. Burgie, E. S., & Vierstra, R. D. (2014). Phytochromes: An atomic perspective on photoactivation and signaling. *The Plant Cell, 26*, 4568–4583.

41. Xu, Q.-Z., et al. (2019). MAS NMR on a Red/Far-Red Photochromic Cyanobacteriochrome All2699 from *Nostoc*. *International Journal of Molecular Sciences, 20*, 3656.

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

43. Narikawa, R., et al. (2013). Structures of cyanobacteriochromes from phototaxis regulators AnPixJ and TePixJ reveal general and specific photoconversion mechanism. *Proceedings of the National Academy of Sciences USA, 110*, 918–923.

44. Sato, T., et al. (2019). Protochromic absorption changes in the two-cysteine photocycle of a blue/orange cyanobacteriochrome. *Journal of Biological Chemistry, 294*, 18909–18922.

45. von Stetten, D., et al. (2007). Highly conserved residues Asp-197 and His-250 in Agp1 phytochrome control the proton affinity of the chromophore and Pfr formation. *Journal of Biological Chemistry, 282*, 2116–2123.

46. Burgie, E. S., Walker, J. M., Phillips, G. N., & Vierstra, R. D. (2013). A photo-labile thioether linkage to phycoviolobilin provides the foundation for the blue/green photocycles in DXCF-cyanobacteriochromes. *Structure, 21*, 88–97.

47. Rockwell, N. C., Martin, S. S., & Lagarias, J. C. (2017). There and back again: Loss and reacquisition of two-Cys photocycles in cyanobacteriochromes. *Photochemistry and Photobiology, 93*, 741–754.

48. Park, Y., Patton, J. E. J., Hochberg, G. K. A., & Thornton, J. W. (2020). Comment on "Ancient origins of allosteric activation in a Ser-Thr kinase." *Science, 370*, eabc8301.

49. Grettenberger, C. L., et al. (2020). A phylogenetically novel cyanobacterium most closely related to Gloeobacter. *ISME Journal, 14*, 2142–2152.

50. Watanabe, M., et al. (2014). Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proceedings of the National Academy of Sciences USA, 111*, 2512–2517.

51. Hammerschmidt, K., Landan, G., Domingues Kümmel Tria, F., Alcorta, J., & Dagan, T. (2021). The order of trait emergence in the evolution of cyanobacterial multicellularity. *Genome Biology and Evolution, 13*, 249.

52. Xu, X., et al. (2020). Structural elements regulating the photochromicity in a cyanobacteriochrome. *Proceedings of the National Academy of Sciences, 117*, 2432–2440.

53. Essen, L. O., Mailliet, J., & Hughes, J. (2008). The structure of a complete phytochrome sensory module in the Pr ground state. *Proceedings of the National Academy of Sciences USA, 105*, 14709–14714.

54. Anders, K., Daminelli-Widany, G., Mroginski, M. A., von Stetten, D., & Essen, L. O. (2013). Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. *Journal of Biological Chemistry, 288*, 35714–35725.

55. Velazquez Escobar, F., et al. (2017). Protonation-dependent structural heterogeneity in the chromophore binding site of cyanobacterial phytochrome Cph1. *The Journal of Physical Chemistry B, 121*, 47–57.

56. Song, C., et al. (2015). A red/green cyanobacteriochrome sustains its color despite a change in the bilin chromophore's protonation state. *Biochemistry, 54*, 5839–5848.

57. Fischer, T., et al. (2020). Effect of the PHY domain on the photoisomerization step of the forward $P_r \rightarrow P_{fr}$ conversion of a knotless phytochrome. *Chemistry – European Journal, 26*, 17261–17266.

58. Narikawa, R., Enomoto, G., Ni Ni, W., Fushimi, K., & Ikeuchi, M. (2014). A new type of dual-Cys cyanobacteriochrome GAF domain found in cyanobacterium *Acaryochloris marina*, which has an unusual red/blue reversible photoconversion cycle. *Biochemistry, 53*, 5051–5059.

59. Blain-Hartung, M., Rockwell, N. C., & Lagarias, J. C. (2021). Natural diversity provides a broadspectrum of cyanobacteriochrome-based diguanylate cyclases. *Plant Physiology, 187*, kiab240.

60. Altschul, S. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389–3402.

61. Moore, K. R., et al. (2019). An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. *Frontiers in Microbiology, 10*, 1612.

62. Potter, S. C., et al. (2018). HMMER web server: 2018 update. *Nucleic Acids Research, 46*, W200–W204.

63. Madeira, F., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research, 47*, W636–W641.

64. Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*, 1792–1797.

65. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*, 1312–1313.

66. Ulijasz, A. T., et al. (2008). Characterization of two thermostable cyanobacterial phytochromes reveals global movements in the chromophore-binding domain during photoconversion. *Journal of Biological Chemistry, 283*, 21251–21266.

67. Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*, 1586–1591.

68. Lemoine, F., et al. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature, 556*, 452–456.

69. Guindon, S., et al. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology, 59*, 307–321.

70. Beyer, H. M., et al. (2015). AQUA cloning: A versatile and simple enzyme-free cloning approach. *PLoS ONE, 10*, e0137652.

71. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W., & Frankenberg-Dinkel, N. (2008). Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology, 18*, 442–448.

72. Schägger, H. (2006). Tricine–SDS-PAGE. *Nature Protocols, 1*, 16–22.

73. Mirdita, M., et al. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods, 19*, 679–682.

74. Lim, S., et al. (2018). Correlating structural and photochemical heterogeneity in cyanobacteriochrome NpR6012g4. *Proceedings of the National Academy of Sciences, 115*, 4387–4392.

Authors and Affiliations

Nibedita Priyadarshini[1,2] · **Niklas Steube**[3] · **Dennis Wiens**[3] · **Rei Narikawa**[4] · **Annegret Wilde**[1] · **Georg K. A. Hochberg**[3,5,6] · **Gen Enomoto**[1,7]

✉ Georg K. A. Hochberg
georg.hochberg@mpi-marburg.mpg.de

✉ Gen Enomoto
genenomoto@gmail.com

1 Faculty of Biology, Institute of Biology III, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany

2 Spemann Graduate School of Biology and Medicine, Albertstr. 19, 79104 Freiburg, Germany

3 Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Str. 10, 35043 Marburg, Germany

4 Graduate School of Biological Sciences, Faculty of Science, Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji, Tokyo 192-0397, Japan

5 Faculty of Chemistry, University of Marburg, Hans-Meerwein-Str. 4, 35032 Marburg, Germany

6 Center for Synthetic Microbiology (SYNMIKRO), Karl-von-Frisch-Str. 14, 35032 Marburg, Germany

7 Present Address: Department of Engineering Science, Graduate School of Informatics and Engineering, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan