Green light perception paved the way for the diversification of GAF domain photoreceptors

Authors

Nibedita Priyadarshini[a,b,1], Niklas Steube[c,1], Dennis Wiens[c], Rei Narikawa[e,f], Annegret Wilde[a], Georg K. A. Hochberg[c,d,*], and Gen Enomoto[a,*]

Affiliations

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

[b]Spemann Graduate School of Biology and Medicine, Albertstraße-19, 79104 Freiburg, Germany

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

[d]Faculty of Chemistry, University of Marburg, Hans-Meerwein-Straße 4, 35032 Marburg, Germany

[e]Graduate School of Integrated Science and Technology, Shizuoka University, 836 Ohya, Suruga, Shizuoka 422-8529, Japan

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

[1]These authors contributed equally to this work and are listed in alphabetical order.

*To whom correspondence may be addressed. Email: georg.hochberg@mpi-marburg.mpg.de or gen.enomoto@biologie.uni-freiburg.de

Author Contributions: G.K.A.H. and G.E. designed research; N.P., N.S., D.W., and G.E. performed research; all authors analyzed data; all authors contributed to writing the paper.

Competing Interest Statement: The authors declare no competing interest.

Classification: BIOLOGICAL SCIENCES, Biochemistry.

Keywords (3–5): ancestral sequence reconstruction | bilin-based photoreceptor | cyanobacteriochromes | photobiology | GAF domain

This PDF file includes:

Main Text

Figures 1 to 3

Tables 1 to 2

Abstract

Photoreceptors are proteins that sense incident light and then trigger downstream signaling events. Phytochromes are linear tetrapyrrole-binding photoreceptors present in plants, algae, fungi, and various bacteria. Most phytochromes respond to red and far-red light signals. Among the phytochrome superfamily, cyanobacteria-specific cyanobacteriochromes show much more diverse optical properties covering the entire visible region. Both phytochromes and cyanobacteriochromes share the GAF domain scaffold to cradle the chromophore as the light-sensing region. It is unknown what physiological demands drove the evolution of cyanobacteriochromes in cyanobacteria. Here we utilize ancestral sequence reconstruction and report 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 enables the photoreceptor to perceive green light. The ancestral cyanobacteriochromes show modest thermal reversion to the green light-absorbing form, suggesting that they evolved to sense green-rich irradiance rather than red light, which is preferentially utilized for photosynthesis. In contrast to plants and green algae, many cyanobacteria can utilize green light for photosynthesis with their special light-harvesting complexes, phycobilisomes. The evolution of green/red sensing cyanobacteriochromes may therefore have allowed ancient cyanobacteria to acclimate to different light environments by rearranging the absorption capacity of the cyanobacterial antenna complex by chromatic acclimation.

Significance Statement

Light serves as a crucial environmental stimulus affecting the physiology of organisms across all kingdoms of life. Photoreceptors serve as important players of light responses, absorbing light and actuating biological processes. Among a plethora of photoreceptors, cyanobacteriochromes arguably have the wealthiest palette of color sensing, largely contributing to the success of cyanobacteria in various illuminated habitats. Our ancestral sequence reconstruction and the analysis of the resurrected ancestral proteins suggest that the very first cyanobacteriochrome most probably responded to the incident green-to-red light ratio, in contrast to modern red/far-red absorbing plant phytochromes. The deprotonation of the light-absorbing pigment for green light-sensing was a crucial molecular event for the invention of the new class of photoreceptors with their huge color tuning capacity.

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 any light absorption. Thermal reversion is a temperature-dependent process, and therefore the same photoreceptor integrates light and temperature signals (5). 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 (6).

Within the phytochrome superfamily, cyanobacteriochromes (CBCRs) are a distinct class of minimal photoreceptors (7, 8), which only need a single GAF (cGMP phosphodiesterase, adenylyl cyclase, and FhlA) domain to genuinely sense light. This is in contrast to other phytochrome members that strictly require at least another neighboring PHY domain for genuine light perception. The functional light sensing module of canonical phytochromes features a typical PAS-GAF-PHY tridomain architecture, with exception of some members lacking

the PAS domain (PAS-less 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 (6, 9–12), making them a promising scaffold to develop a new generation of optogenetic tools (13–15). Depending on their properties, CBCRs control a diverse range of physiological processes in cyanobacteria (16). Green/red sensing CBCRs with slow reversion kinetics for instance 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 (11, 17, 18). Blue/green sensing CBCRs, on the other hand, are used to detect shading by other cells in cyanobacterial mats (19, 20).

However, the original function of CBCRs which evolved early in evolution remains unknown. We have previously speculated that blue/green perceiving CBCR-mediated cell shade sensing might be the ancestral function of these photoreceptors (19), 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. Further, early-branching cyanobacteria like *Gloeobacter violaceus* PCC 7421 and *Anthocerotibacter panamensis* (21) only possess this kind of blue/green perceiving CBCRs. Here, we used ancestral sequence reconstruction (22) to experimentally test this theory and inferred the photochemistry of the last

common ancestor (LCA) of all extant CBCRs. We show that ancient CBCR proteins most likely sensed the ratio of green to red incident light and that 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 evolved from an ancient CBCR most likely used for chromatic acclimation.

Results

Ancestral sequence reconstruction of cyanobacteriochromes

In order to investigate the characteristics of the earliest CBCRs, we first used maximum likelihood phylogenetics and ancestral sequence reconstruction to infer the most likely GAF domain sequence of the LCA of extant CBCRs. We inferred a maximum likelihood phylogeny of 575 CBCR GAF domains with related PAS-less phytochromes' GAF domains as the outgroup. The phylogeny of the CBCR domains was difficult to resolve: we found multiple groups of GAF domains each containing the majority of cyanobacterial diversity. These groups are separated from each other by very short internal branches. In addition, our phylogenetic tree contains one group of CBCR GAF domains, sister to all others, that has a very narrow taxonomic distribution. It is connected by a very long internal branch that attaches to the long branch separating CBCR domains from the outgroup (Figure 1A, Figure S2). To ensure that our estimate of the sequence of the LCA of all

CBCR domains was not biased by this potential long-branch attraction artifact, we inferred two additional phylogenies from pruned alignments: for the first, we removed the sequences belonging to this long branching sister clade to all other CBCR domains from our alignment. In the second, we removed all sequences that had long terminal branches in our original phylogeny or were only poorly aligned. The two resulting topologies have slight rearrangements in branching order within the CBCR domains and represent different, but equally plausible topologies. While the exact branching order remains unclear, the three topologies agree on far-red/orange Ancy2551g3 and green/red SyCcaSg as early branching among the known characterized CBCR GAF domains (Figure 1B,C).

We then used ancestral sequence reconstruction to infer 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 between 0.81 and 0.94 (Figure S3). All ancestral sequences contained the conserved “first” cysteine that binds the bilin chromophore in extant CBCRs but differed at between 37 and 44 out of 142 total residues (Figure 1D, Figure S4).

Although CBCR GAF domains can sense light on their own, they are usually part of multidomain proteins. 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 our trees. We found PAS domains that are mandatory in distantly related canonical phytochromes as the most abundant neighbors, and histidine kinase HATPase domains as the most prominent output domains (Figure 1A-C). This indicates that the LCA of all CBCRs was most parsimoniously encoded on a phytochrome-like multidomain protein, and transduced its signal to a histidine kinase domain.

The ancestral CBCR GAF domain most likely had a green/red photocycle

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-enhanced fluorescence of the purified proteins in an SDS-PAGE gel confirmed the covalent attachment of PCB to the apoproteins (Figure S5) (23). The absorbance spectra of the purified holoproteins showed spectral changes upon illumination with blue, green, and red light. However, irradiation with UV and far-red light did not affect the spectra. All ancestral proteins exhibited reversible photoconversion between green (Pg) and red (Pr) absorbing forms (Figure 2). The bound chromophore and its configuration were determined using acid denaturation spectra with the red-irradiated state (i.e., Pg) peaking at 662 nm and the green-irradiated state (i.e., Pr) peaking at 585 nm, in agreement to *15Z* and *15E* forms of the covalently bound PCB, respectively (Figure S6) (24). The ^{15Z}Pg state showed absorption maxima between 515 nm and 540 nm, and the ^{15E}Pr state between 600 nm and 656 nm for all the ancestral proteins (Figure 2, Table 1). For Anc2 and Anc3, irradiation with red (λ_{max} = 635 nm) resulted in almost complete conversion to the ^{15Z}Pg form. For Anc1, we did not yield a homogeneous population of ^{15Z}Pg, probably due to the significant overlap of the absorption spectra of the two photostates (Figure 2, Figure S6). The additional incubation of Anc1 overnight in the dark at room temperature allowed complete conversion to ^{15Z}Pg (Figure S6). Irradiation with blue (λ_{max} = 448 nm) and green (λ_{max} = 514 nm) rendered complete conversion to the ^{15E}Pr state for Anc1 and Anc2. For Anc3, green irradiation resulted in partial conversion. Almost complete conversion was

achieved upon blue irradiation, probably due to its good separation from the counteracting red region (Figure 2, Figure S6). 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 (Figure 2D). Altogether, these results show that a green/red photocycle most likely existed in the LCA of all CBCRs, regardless of the exact branching order of basal CBCRs.

PCB was the ancestral chromophore in CBCRs

Although most CBCRs incorporate PCB, some CBCRs can bind biliverdin IXa (BV) as the chromophore with variable specificity (25, 26). 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. Acid denaturation spectra confirmed the attached chromophore to be BV with the denatured ^{15Z}Pg peaking at around 700 nm (Figure S7) (26). 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 S7). 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 are able to bind to both PCB and BV, but photoconversion is much more efficient with PCB. Overall, we infer that PCB was the *bona fide* chromophore for the LCA of all CBCR GAF domains. Specificity for BV is therefore a derived trait of some crown group CBCRs (25). This is consistent with canonical phytochromes in the outgroup also being

specific for PCB (27, 28). Besides, PCB is one of the prosthetic groups of the very abundant phycobiliproteins of the photosynthetic antenna complex and is therefore much more abundant in cyanobacterial cells (29).

The ancestral CBCR GAF domain was 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 green/red 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 (6). 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 S8): 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 PAS-less phytochromes (27). These half-lives are much longer than those of known intensity-sensing CBCRs (which revert within the range of several seconds) (6, 30). 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 (11, 31): The chromophore is deprotonated with a lower pK_a value in the *15Z* state to absorb green light, whereas it is protonated with a higher pK_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 of red-light absorption and an increase of green-light absorption at higher pH conditions (Figure 3, Figure S9). 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 pK_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 of pK_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 (Figure 2). Our analyses therefore suggest that a photochromic photocycle similar to that of extant green/red CBCRs was also the ancestral photo-switching mechanism.

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 ancestral 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 (27, 31–33). The protonated state is stabilized by a conserved aspartate (Asp) residue that forms a hydrogen bond network with the nitrogen atoms of the B and C pyrrole rings of the chromophore (34–37). The resurrected CBCR ancestors feature either an alanine or glutamate residue at the position 54, suggesting that the substitution of Asp to a different amino acid 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 CBCRs. We then determined whether the deprotonation of the chromophore was affected, and found that green-light absorption and deprotonation were both unaffected in all three mutants (Table 1–2, Figure S10–11). 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 form a thioether linkage with the C10 position of the bilin chromophore (38). The disruption of the π-conjugated system at the C10 position leads to absorption in the UV-to-blue light region (10, 39). Therefore, 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 (Figure 1D, Figure S4), whereas Anc2 and Anc3 have a valine at this position. Although all three proteins have a green/red photocycle, this introduces ambiguity about whether the second Cys played an important role in the evolution of the green/red photocycle: It is possible that the function of this cysteine depends on the specific context of the protein such as the neighboring amino acid residues. 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 (Table 1–2, Figure S10–11). This confirms that a green/red photocycle

was present in the LCA of all CBCR GAF domains, regardless of the presence of the second cysteine in the ancestral protein.

Discussion

The first CBCRs most probably functioned in chromatic acclimation

Our results strongly suggest that the LCA of extant CBCRs functioned as a green/red light sensor with slow thermal reversion that used a photochromic photocycle similar to that of extant green/red sensing CBCRs. What physiological function did this receptor fulfil? One plausible answer upon comparison with extant CBCRs that have similar photocycles suggests their involvement in regulating the relative amounts of red-absorbing phycocyanin and green-absorbing phycoerythrin in phycobilisomes during chromatic acclimation (18). This hypothesis is further supported by the analysis of neighboring domains as the extant chromatic acclimation regulators also harbor an additional PAS domain and a histidine kinase as the output domain (17). This implies that the LCA of all extant cyanobacteria, in which the here identified ancestral GAF domain would have existed, already possessed phycoerythrin. Members of the Gloeobacterales (the earliest diverging clade of cyanobacteria) usually possess phycoerythrin, suggesting that the pigment has an ancient origin (21, 29, 40). This implies that phycoerythrin and the ability for chromatic acclimation already existed in the earliest cyanobacteria. It is of note that extant green/red CBCRs are also regulators of the different types of chromatic acclimation to control the relative amounts of the yellow–green-absorbing phycoerythrocyanin protein or a rod-membrane linker CpcL protein, which assembles the photosystem I-specific

phycobilisome (17, 41). 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 (42). In these environments, 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 (19).

Tuning of the chromophore towards green/red sensing

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

Our results show that the *15Z* state was deprotonated in the ancestral proteins. This is different from phytochromes, in which the bilin chromophore is protonated in both photostates (27, 31–33), suggesting that deprotonation of the chromophore was an important requirement for evolving green-light absorption. The ancestral proteins all lack the conserved Asp, important for stabilization of the protonated state in phytochromes (36, 37), suggesting that this substitution may have allowed for deprotonation. 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 not sufficient to yield green light absorption: the

337 cyanobacterial canonical phytochrome Cph1 exhibits a pK_a of ~9.0 in the *15Z* and
338 *15E* photostates to stabilize the protonated chromophore. Increasing the solvent
339 pH induces a decrease in red-light absorption by Cph1 but does not cause an
340 increase in green-light absorption (43). The red/green CBCR AnPixJg2 retains
341 the protonated chromophore even at the green-absorbing state, and artificial
342 deprotonation does not affect the green absorption (44). This suggests that green
343 absorption requires additional amino acid substitutions affecting the light
344 wavelength absorbed by the deprotonated chromophore.
345 The *15E* state is also hypsochromically shifted from far-red to red
346 absorption. This could have occurred through the loss of the adjacent PHY
347 domain from an ancestral phytochrome-like precursor. Such truncations led to a
348 blue shift of the far-red absorbing state of extant phytochromes (28, 34, 45).
349 Another suggested tuning mechanism is the “second” Cys, that is found near the
350 chromophore and is known to influence the absorption properties of proteins from
351 various lineages of CBCR GAF domains (10, 39, 46, 47). However, the
352 reconstructed ancestral proteins vary in the amino acid at that position; Anc1 has
353 a cysteine, whereas Anc2 and Anc3 both have valine. Mutating this cysteine in
354 Anc1 has no effect on optical properties, suggesting that in the LCA of all CBCR
355 GAF domains this site was not essential for color tuning. Further exploration
356 would throw light upon the exact genetic mechanism that transformed a likely
357 red/far-red sensing phytochrome into a green/red sensing CBCR.

The genetic basis of CBCRs 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 all possess the ability to also sense blue light, which was likely 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.

In addition, 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 pK_a of the chromophore. Our characterization of sequences representative of the first CBCR sets the stage to elucidate exactly how this tinkering occurred in the colorful history of CBCR proteins.

Methods

Phylogenetics and ancestral sequence reconstruction

Amino acid sequences of cyanobacterial proteins containing GAF domains were gathered using protein-protein BLAST (nr database) and a CBCR protein as a query (48). Models (XM/XP) and uncultured/environmental sample sequences were excluded from the search. Protein sequences were selected to represent the whole cyanobacterial species phylogeny based on recently published data (49). 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 (50). CBCR GAF domain sequences were aligned with MUSCLE (51), and the alignment was manually cropped to remove gaps by deleting lineage-specific and non-parsimonious inserts (52). The cropped alignment was used to infer an initial ML phylogeny using RAxML (53) in the PROTGAMMAAUTO mode resulting in the LG likelihood model with fixed base frequencies. The resulting tree was rooted using GAF domain sequences of proteins lacking the PAS domain but containing the C-terminal PHY domain as an outgroup (cyanobacterial PAS-less phytochromes) (54). The last common ancestor of all CBCR GAF domains (Anc1) was reconstructed at the internal node indicated in Figure 1A on Tree A using the CodeML package of PAML (55) with the LG substitution model and 16 gamma categories. Due to the suspicious long branch of the first branching sequences, an alternative tree (Tree B) was inferred by deletion of these sequences from the multiple sequence 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 (56). Additionally, approximate likelihood ratios were calculated with PhyML (57). The consensus neighbor and output domains of each group on the trees were determined manually and mapped next to the topologies (Figure 1).

Plasmid construction

Codon-optimized sequences for *E. coli* encoding the ancestral CBCR GAF domains of Anc1, Anc2, and Anc3 (Table S1) were obtained from Twist Bioscience (San Francisco, California, USA) on the pTwist Amp vector. 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) (58). 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) (59). 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.

Protein expression and purification

The *E.coli* strain BL21(DE3) was co-transformed with one of the pET28V plasmids harboring the gene for Anc1, Anc2 or Anc3, respectively, and either the PCB producing pTDho1pcyA-HisTag plasmid or the BV producing pTDho1 plasmid. The cultures were induced with 0.1 M isopropyl-β-D-thiogalactopyranosid and grown overnight at 25°C in LB medium with appropriate antibiotics. The cells were harvested and disrupted three times using a french press at 20000 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.). Elution was carried out at a flow rate of 1 ml/min with all solutions maintained at 4 °C and a linear imidazole concentration gradient from 30 mM to 530 mM.

SDS-PAGE and fluorescence detection of PCB

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 (60). 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 a F595 Y3 filter. The gel was further stained with Coomassie G-250.

Light sources

To irradiate purified proteins, LEDs illuminating at 355 nm for UV light, 448 nm for blue light, 514 nm for green light, 635 nm for red light, and 731 nm for far-red light were used (Figure S12).

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 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 dark overnight at room temperature. To acquire the absorption spectra of the acid denatured proteins, 140 µl of the 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 5ml; Cytiva) and was diluted with the same buffer in a 1:1 ratio. 560µl of the diluted 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 pK_a values of the

462 chromophore were determined using Henderson-Hasselbalch equations of a

463 single titrating group (11, 36).

Acknowledgments

We thank Prof. Nicole Frankenberg-Dinkel for the kind gift of the chromophore biosynthesis plasmid, and Prof. J. Clark Lagarias and Dr. Nathan C. Rockwell for providing the alignment of GAF domains and fruitful discussions. G.E. is supported by the EMBO postdoctoral fellowship (ALTF 274-2017) and JSPS Overseas Research Fellowships. N.P. was supported in part by the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School) and in part by the Ministry for Science, Research and Arts of the State of Baden-Wuerttemberg. N.S., D.W., and G.K.A.H. are supported by the Max Planck Society.

References

1. A. Möglich, X. Yang, R. A. Ayers, K. Moffat, Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol.* **61**, 21–47 (2010).
2. N. C. Rockwell, J. C. Lagarias, A brief history of phytochromes. *Chemphyschem* **11**, 1172–80 (2010).
3. K. Anders, L. O. Essen, The family of phytochrome-like photoreceptors: diverse, complex and multi-colored, but very useful. *Curr. Opin. Struct. Biol.* **35**, 7–16 (2015).
4. C. Song, *et al.*, The D-ring, not the A-ring, rotates in *Synechococcus* OS-B ' phytochrome. *J. Biol. Chem.* **289**, 2552–62 (2014).
5. C. Klose, F. Nagy, E. Schäfer, Thermal Reversion of Plant Phytochromes. *Mol. Plant* **13**, 386–397 (2020).
6. N. C. Rockwell, S. S. Martin, J. C. Lagarias, Red/green cyanobacteriochromes: sensors of color and power. *Biochemistry* **51**, 9667–9677 (2012).
7. M. Ikeuchi, T. Ishizuka, Cyanobacteriochromes: a new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. *Photochem Photobiol Sci* **7**, 1159–67 (2008).
8. N. C. Rockwell, J. C. Lagarias, Phytochrome evolution in 3D: deletion, duplication, and diversification. *New Phytol.* **225**, 2283–2300 (2020).
9. K. Fushimi, R. Narikawa, Cyanobacteriochromes: photoreceptors covering the entire UV-to-visible spectrum. *Curr. Opin. Struct. Biol.* **57**, 39–46 (2019).
10. N. C. Rockwell, S. S. Martin, K. Feoktistova, J. C. Lagarias, Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. *Proc Natl Acad Sci U S A* **108**, 11854–9 (2011).
11. Y. Hirose, *et al.*, Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. *Proc Natl Acad Sci U S A* **110**, 4974–9 (2013).
12. N. C. Rockwell, S. S. Martin, A. G. Gulevich, J. C. Lagarias, Conserved phenylalanine residues are required for blue-shifting of cyanobacteriochrome photoproducts. *Biochemistry* **53**, 3118–30 (2014).
13. O. S. Oliinyk, K. G. Chernov, V. V. Verkhusha, Bacterial phytochromes, cyanobacteriochromes and allophycocyanins as a source of near-infrared fluorescent probes. *Int J Mol Sci* **18** (2017).
14. M. Blain-Hartung, N. C. Rockwell, J. C. Lagarias, Light-Regulated Synthesis of Cyclic-di-GMP by a Bidomain Construct of the Cyanobacteriochrome Tlr0924 (SesA) without Stable Dimerization. *Biochemistry* **56**, 6145–6154 (2017).
15. K. Fushimi, G. Enomoto, M. Ikeuchi, R. Narikawa, Distinctive properties of dark reversion kinetics between two red/green-type cyanobacteriochromes and their application in the photoregulation of cAMP synthesis. *Photochem. Photobiol.* **93**, 681–691 (2017).
16. L. B. Wiltbank, D. M. Kehoe, Diverse light responses of cyanobacteria mediated by phytochrome superfamily photoreceptors. *Nat. Rev. Microbiol.* **17**, 37–50 (2019).
17. Y. Hirose, *et al.*, Diverse Chromatic Acclimation Processes Regulating Phycoerythrocyanin and Rod-Shaped Phycobilisome in Cyanobacteria. *Mol. Plant* **12**, 715–725 (2019).
18. J. E. Sanfilippo, L. Garczarek, F. Partensky, D. M. Kehoe, Chromatic

Acclimation in Cyanobacteria: A Diverse and Widespread Process for Optimizing Photosynthesis. *Annu. Rev. Microbiol.* **73**, 407–433 (2019).

19. G. Enomoto, M. Ikeuchi, Blue-/Green-Light-Responsive Cyanobacteriochromes Are Cell Shade Sensors in Red-Light Replete Niches. *iScience* **23**, 100936 (2020).
20. F. D. Conradi, C. W. Mullineaux, A. Wilde, The Role of the Cyanobacterial Type IV Pilus Machinery in Finding and Maintaining a Favourable Environment. *Life* **10**, 252 (2020).
21. N. Rahmatpour, *et al.*, A novel thylakoid-less isolate fills a billion-year gap in the evolution of Cyanobacteria. *Curr. Biol.* **31**, 2857-2867.e4 (2021).
22. G. K. A. Hochberg, J. W. Thornton, Reconstructing Ancient Proteins to Understand the Causes of Structure and Function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).
23. T. R. Berkelman, J. C. Lagarias, Visualization of bilin-linked peptides and proteins in polyacrylamide gels. *Anal. Biochem.* **156**, 194–201 (1986).
24. T. Ishizuka, R. Narikawa, T. Kohchi, M. Katayama, M. Ikeuchi, Cyanobacteriochrome TePixJ of *Thermosynechococcus elongatus* harbors phycoviolobilin as a chromophore. *Plant Cell Physiol.* **48**, 1385–90 (2007).
25. M. V. Moreno, N. C. Rockwell, M. Mora, A. J. Fisher, J. C. Lagarias, A far-red cyanobacteriochrome lineage specific for verdins. *Proc Natl Acad Sci U S A* **117**, 27962–27970 (2020).
26. R. Narikawa, *et al.*, A biliverdin-binding cyanobacteriochrome from the chlorophyll d-bearing cyanobacterium *Acaryochloris marina*. *Sci Rep* **5**, 7950 (2015).
27. K. Anders, *et al.*, Spectroscopic and photochemical characterization of the red-light sensitive photosensory module of Cph2 from *Synechocystis* PCC 6803. *Photochem. Photobiol.* **87**, 160–73 (2011).
28. S. H. Wu, J. C. Lagarias, Defining the bilin lyase domain: lessons from the extended phytochrome superfamily. *Biochemistry* **39**, 13487–95 (2000).
29. M. Watanabe, M. Ikeuchi, Phycobilisome: architecture of a light-harvesting supercomplex. *Photosynth. Res.* **116**, 265–76 (2013).
30. M. Hasegawa, *et al.*, Molecular characterization of DXCF cyanobacteriochromes from the cyanobacterium *Acaryochloris marina* identifies a blue-light power sensor. *J. Biol. Chem.* **293**, 1713–1727 (2018).
31. S. Osoegawa, *et al.*, Identification of the Deprotonated Pyrrole Nitrogen of the Bilin-Based Photoreceptor by Raman Spectroscopy with an Advanced Computational Analysis. *J. Phys. Chem. B* **123**, 3242–3247 (2019).
32. E. S. Burgie, R. D. Vierstra, Phytochromes: An Atomic Perspective on Photoactivation and Signaling. *Plant Cell* **26**, 4568–4583 (2014).
33. Q.-Z. Xu, *et al.*, MAS NMR on a Red/Far-Red Photochromic Cyanobacteriochrome All2699 from *Nostoc*. *Int. J. Mol. Sci.* **20**, 3656 (2019).
34. J. R. Wagner, J. S. Brunzelle, K. T. Forest, R. D. Vierstra, A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**, 325–31 (2005).
35. R. Narikawa, *et al.*, Structures of cyanobacteriochromes from phototaxis regulators AnPixJ and TePixJ reveal general and specific photoconversion mechanism. *Proc Natl Acad Sci U S A* **110**, 918–23 (2013).
36. T. Sato, *et al.*, Protochromic absorption changes in the two-cysteine photocycle of a blue/orange cyanobacteriochrome. *J. Biol. Chem.* **294**, 18909–18922 (2019).

37. D. von Stetten, *et al.*, Highly conserved residues Asp-197 and His-250 in Agp1 phytochrome control the proton affinity of the chromophore and Pfr formation. *J. Biol. Chem.* **282**, 2116–23 (2007).
38. E. S. Burgie, J. M. Walker, G. N. Phillips, R. D. Vierstra, A photo-labile thioether linkage to phycoviolobilin provides the foundation for the blue/green photocycles in DXCF-cyanobacteriochromes. *Structure* **21**, 88–97 (2013).
39. N. C. Rockwell, S. S. Martin, J. C. Lagarias, There and Back Again: Loss and Reacquisition of Two-Cys Photocycles in Cyanobacteriochromes. *Photochem. Photobiol.* **93**, 741–754 (2017).
40. C. L. Grettenberger, *et al.*, A phylogenetically novel cyanobacterium most closely related to *Gloeobacter*. *ISME J.* **14**, 2142–2152 (2020).
41. M. Watanabe, *et al.*, Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria. *Proc Natl Acad Sci U S A* **111**, 2512–7 (2014).
42. K. Hammerschmidt, G. Landan, F. Domingues Kümmel Tria, J. Alcorta, T. Dagan, The Order of Trait Emergence in the Evolution of Cyanobacterial Multicellularity. *Genome Biol. Evol.* **13**, evaa249 (2021).
43. F. Velazquez Escobar, *et al.*, Protonation-Dependent Structural Heterogeneity in the Chromophore Binding Site of Cyanobacterial Phytochrome Cph1. *J. Phys. Chem. B* **121**, 47–57 (2017).
44. C. Song, *et al.*, A Red/Green Cyanobacteriochrome Sustains Its Color Despite a Change in the Bilin Chromophore's Protonation State. *Biochemistry* **54**, 5839–5848 (2015).
45. T. Fischer, *et al.*, Effect of the PHY Domain on the Photoisomerization Step of the Forward $P_r \rightarrow P_{fr}$ Conversion of a Knotless Phytochrome. *Chem. – Eur. J.* **26**, 17261–17266 (2020).
46. R. Narikawa, G. Enomoto, W. Ni Ni, K. Fushimi, M. Ikeuchi, 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**, 5051–9 (2014).
47. M. Blain-Hartung, N. C. Rockwell, J. C. Lagarias, Natural diversity provides a broadspectrum of cyanobacteriochrome-based diguanylate cyclases. *Plant Physiol.*, kiab240 (2021).
48. S. Altschul, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
49. K. R. Moore, *et al.*, An Expanded Ribosomal Phylogeny of Cyanobacteria Supports a Deep Placement of Plastids. *Front. Microbiol.* **10**, 1612 (2019).
50. S. C. Potter, *et al.*, HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).
51. F. Madeira, *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
52. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–7 (2004).
53. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
54. A. T. Ulijasz, *et al.*, Characterization of Two Thermostable Cyanobacterial Phytochromes Reveals Global Movements in the Chromophore-binding Domain during Photoconversion. *J. Biol. Chem.* **283**, 21251–21266 (2008).
55. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
56. F. Lemoine, *et al.*, Renewing Felsenstein's phylogenetic bootstrap in the era of

big data. *Nature* **556**, 452–456 (2018).
57. S. Guindon, *et al.*, New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
58. H. M. Beyer, *et al.*, AQUA cloning: a versatile and simple enzyme-free cloning approach. *PLoS One* **10**, e0137652 (2015).
59. T. Dammeyer, S. C. Bagby, M. B. Sullivan, S. W. Chisholm, N. Frankenberg-Dinkel, Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr. Biol.* **18**, 442–8 (2008).
60. H. Schägger, Tricine–SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).

Figures and Tables

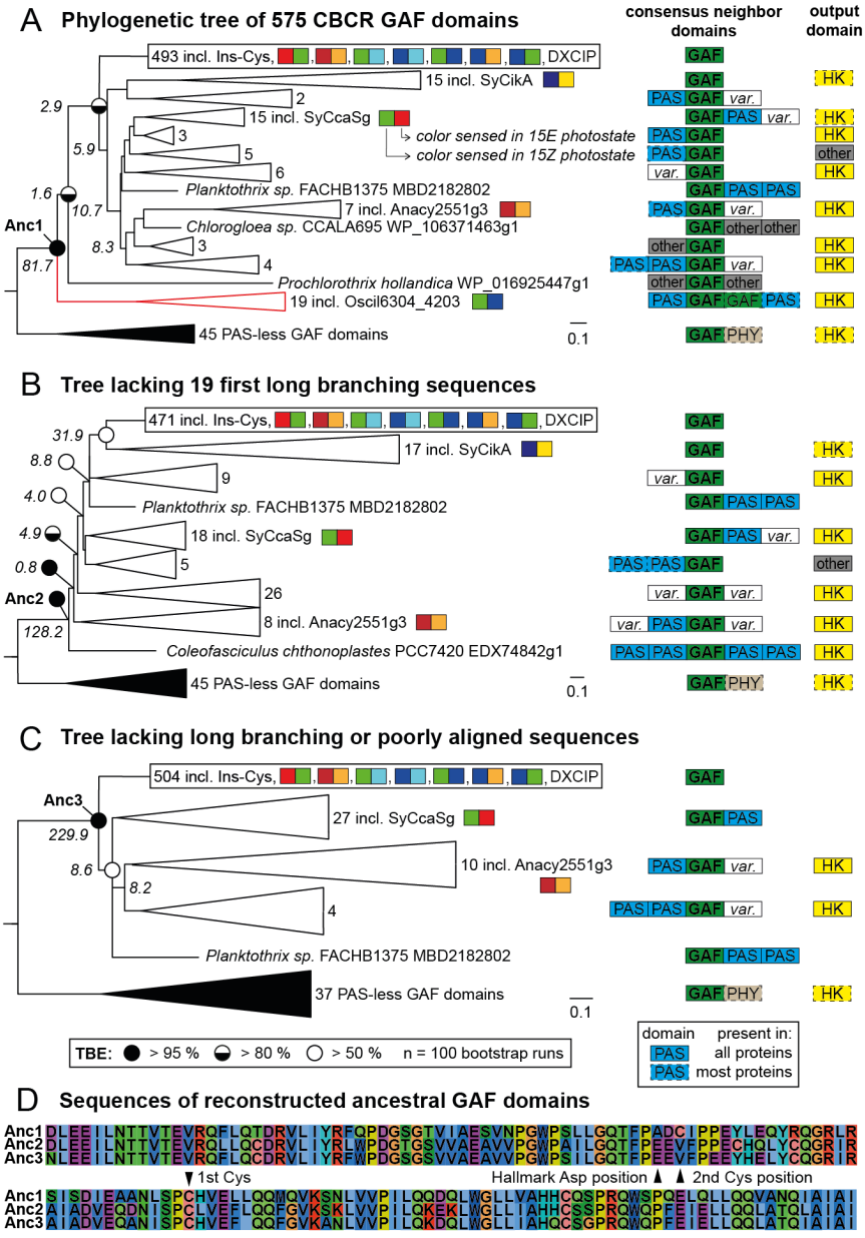


Figure 1. Ancestral CBCR GAF domain reconstruction on ML phylogenies.

(A-C) Maximum Likelihood (ML) phylogenetic trees of cyanobacterial GAF domains used for ancestral sequence reconstruction. Numbers labelling clades denote quantity of taxa. Colored squares highlight biochemically characterized domains and the colors they sense. “Ins-Cys” and “DXCIP” denote families sensing various colors. Node

support is shown as approximate likelihood test statistics (italic numbers) and transfer bootstrap expectations (TBE, 100 replicates, circles). Scale bar: 0.1 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, PHY (Phytochrome-specific domain) or HK (histidine kinase/HATPase). (D) Amino acid sequences of the reconstructed ancestral GAF domains. Arrows point positions important for color sensing in extant CBCRs.

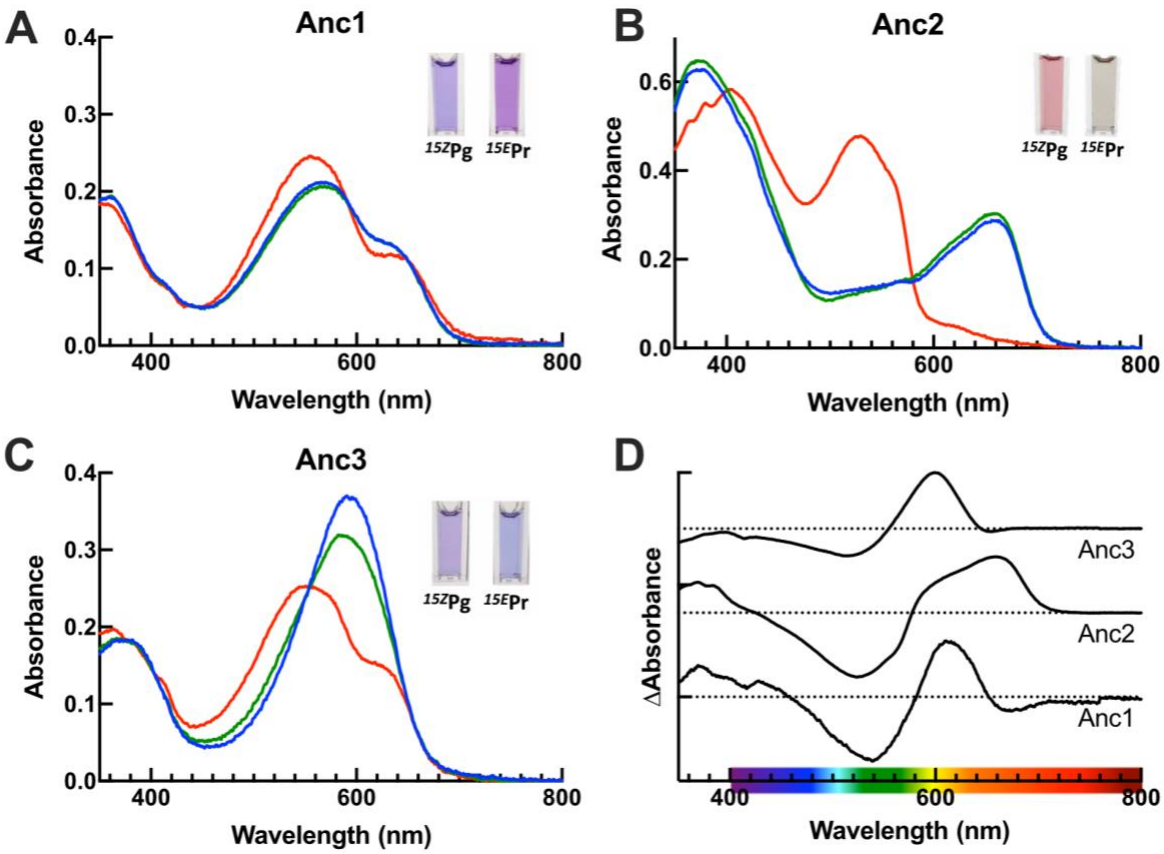
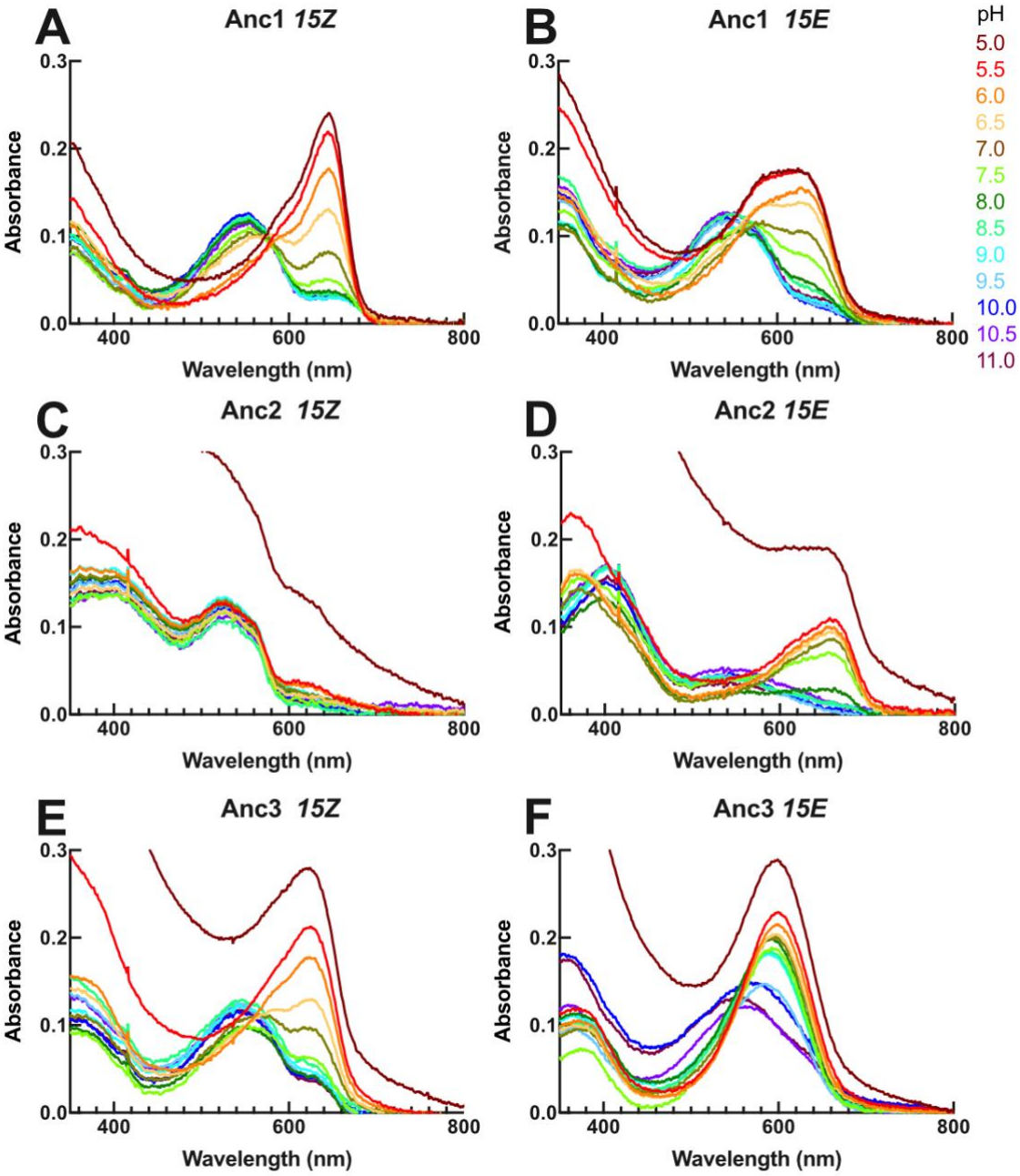


Figure 2. Absorption and difference spectra of the purified ancestral proteins.

(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 form of Anc1-3. Difference spectra were normalized on 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.



657

658 **Figure 3. Protochromic absorption spectra changes of the ancestral**

659 **proteins.**

660 (A-F) pH-dependent absorbance spectra of Anc1-3 with the configuration of *15Z*

661 (A, C, E) or *15E* (B, D, F) measured in buffers with pH between 5.0 (dark red)

662 and 11.0 (dark purple) in 0.5 pH steps. Increased scattering was observed at

663 lower pH of 5.0 and 5.5, probably due to partial protein aggregation. For the

664 analysis, samples were irradiated or dark-incubated to obtain homogenous *15Z*

665 and *15E* photostates, followed by mixing with 1 M buffers of different pH in 1:4

666 ratio and immediate measurement of absorption spectra.

667 **Table 1. Wavelengths of the absorbance peak maxima and the half-lives of**

668 **thermal reversion of ancestral CBCR proteins at room temperature.**

Protein	$\lambda_{max,\ 15Z}$	$\lambda_{max,\ 15E}$	Half-life
Anc1 WT	540 nm	610 nm	233 min
Anc2 WT	525 nm	656 nm	180 min
Anc3 WT	515 nm	600 nm	310 min
Anc1 C56V	541 nm	621 nm	n.d.
Anc1 A54D	535 nm	620 nm	n.d.
Anc2 E54D	525 nm	660 nm	n.d.
Anc3 E54D	515 nm	602 nm	n.d.

669 The peak wavelengths were calculated using the difference spectra upon reversible
670 photoconversion. n.d., not determined
671

672 **Table 2. The estimated pK_a values of the ancestral CBCR proteins.**

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

673 The pK_a values were calculated using the data of the absorption changes in the pH titration
674 experiments in Figure 3. n.a., not applicable