



HAL
open science

Characterisation of sequence–structure–function space in sensor–effector integrators of phytochrome-regulated diguanylate cyclases

Cornelia Böhm, Geoffrey Gourinchas, Sophie Zweytick, Elvira Hujdur,
Martina Reiter, Sara Trstenjak, Christoph Wilhelm Sensen, Andreas Winkler

► **To cite this version:**

Cornelia Böhm, Geoffrey Gourinchas, Sophie Zweytick, Elvira Hujdur, Martina Reiter, et al..
Characterisation of sequence–structure–function space in sensor–effector integrators of phytochrome-
regulated diguanylate cyclases. *Photochemical & Photobiological Sciences* , 2022, 21 (10), pp.1761-
1779. 10.1007/s43630-022-00255-7 . hal-04217602

HAL Id: hal-04217602

<https://hal.science/hal-04217602>

Submitted on 25 Sep 2023

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

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



Characterisation of sequence–structure–function space in sensor–effector integrators of phytochrome-regulated diguanylate cyclases

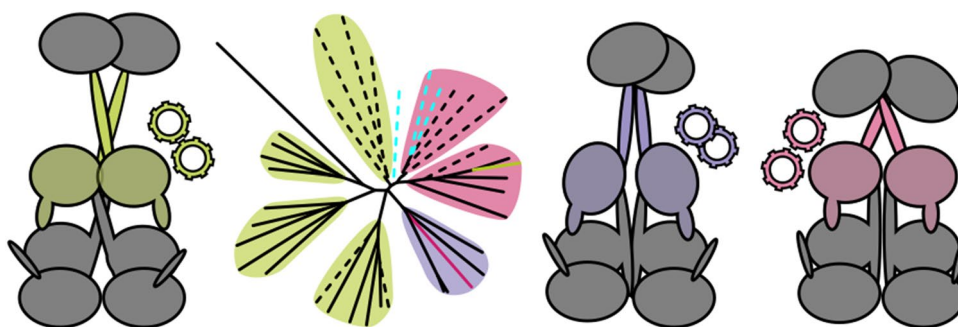
Cornelia Böhm^{1,2} · Geoffrey Gourinchas^{1,3} · Sophie Zweytick¹ · Elvira Hujdur¹ · Martina Reiter¹ · Sara Trstenjak¹ · Christoph Wilhelm Sensen^{2,4} · Andreas Winkler^{1,2} 

Received: 9 February 2022 / Accepted: 8 June 2022 / Published online: 5 July 2022
© The Author(s) 2022

Abstract

Understanding the relationship between protein sequence, structure and function is one of the fundamental challenges in biochemistry. A direct correlation, however, is often not trivial since protein dynamics also play an important functional role—especially in signal transduction processes. In a subfamily of bacterial light sensors, phytochrome-activated diguanylate cyclases (PadCs), a characteristic coiled-coil linker element connects photoreceptor and output module, playing an essential role in signal integration. Combining phylogenetic analyses with biochemical characterisations, we were able to show that length and composition of this linker determine sensor–effector function and as such are under considerable evolutionary pressure. The linker length, together with the upstream PHY-specific domain, influences the dynamic range of effector activation and can even cause light-induced enzyme inhibition. We demonstrate phylogenetic clustering according to linker length, and the development of new linker lengths as well as new protein function within linker families. The biochemical characterisation of PadC homologs revealed that the functional coupling of PHY dimer interface and linker element defines signal integration and regulation of output functionality. A small subfamily of PadCs, characterised by a linker length breaking the coiled-coil pattern, shows a markedly different behaviour from other homologs. The effect of the central helical spine on PadC function highlights its essential role in signal integration as well as direct regulation of diguanylate cyclase activity. Appreciation of sensor–effector linkers as integrator elements and their coevolution with sensory modules is a further step towards the use of functionally diverse homologs as building blocks for rationally designed optogenetic tools.

Graphical abstract



Keywords Bacteriophytochrome · Coiled-coil linker · GGDEF · Photoreceptor · Phylogenetic analysis · Sequence similarity network

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

Extended author information available on the last page of the article

1 Introduction

Protein function depends strongly on protein structure; and the three-dimensional arrangement of amino acids is tightly linked to their sequence. Conservation of protein structure is, however, considerably higher than that of the amino acid sequence [1] and even sequences with no evolutionary relationship can sometimes fold into the same three-dimensional architecture. While similar sequences often imply functional similarities, the situation is further complicated by proteins of different structures performing analogous functions [2]. This complex interplay of sequence, structure and function is one of the fundamental problems in functional annotation of proteins, since the degree of conservation of sequence- and structural features that are essential for different functionalities also provides nature with mechanisms to modulate biological properties and even develop novel activities. While important on a general protein evolution level, this flexibility of sequence–structure–function space is also central to the development of functional sensor–effector systems where domains sensing environmental signals are covalently linked to biological effector units, thereby enabling regulation of enzymatic activities, gene transcription or biological interactions.

One important environmental signal, light, can be sensed by a variety of photoreceptors [3] and frequently these sensors are covalently tethered to their cognate effector domains. In red light sensing bacteriophytochromes, this modularity is quite remarkable [4], but the sequence–function relationship of the corresponding linker elements is generally not well understood. For one bacteriophytochrome subfamily, the linker element has evolved as a central player in modulating bacterial second messenger production. Phytochrome-activated diguanylate cyclases (PadCs) are photoreceptors capable of reacting to red and far red light stimuli, typically resulting in the up- and downregulation of cyclic-dimeric-GMP (c-di-GMP) catalysis by the diguanylate cyclase effector [5–8]. The PadC architecture follows a classical bacteriophytochrome photosensory module of a PAS-GAF-PHY (Period/ARNT/Single-minded—cGMP phosphodiesterase/Adenylate cyclase/FhlA—PHYtochrome-specific) [9] core connected to a GGDEF output domain (Fig. 1a, b). The cofactor binding pocket located in the GAF domain binds the biliverdin IX α chromophore, isomerisation of which from its ZZZssa to ZZEssa configuration induces extensive structural rearrangements in the cofactor environment of bacteriophytochromes [4]. The PHY domain stabilises light-modulated state conformation via its tongue element in an interplay with the N-terminal segment (NTS) [5, 7, 10], which features a strictly conserved biliverdin-binding

cysteine residue. GGDEF domains possess diguanylate cyclase (DGC) activity [11] condensing two molecules of GTP to form the bacterial second messenger c-di-GMP [12, 13]. Several PadCs have been found to C-terminally feature an additional EAL domain, as exemplified by the first characterised member of this family, BphG1 from *R. sphaeroides*, described by Tarutina et al. [14]. The EAL domain provides c-di-GMP phosphodiesterase (EAL-PDE) activity that further hydrolyses the DGC product to the linear dimeric GMP nucleotide 5'-pGpG [15] (Fig. 1d). GGDEF and EAL domains are often found in tandem, linked to various sensory modules, with one of the two effector domains catalytically inactive in many cases [12]; however, bi-functional GGDEF-EAL complexes have been identified as well [16, 17]. The importance of c-di-GMP as a second messenger involved in various biological processes is highlighted by the tight enzymatic control of its formation and degradation [18, 19], affecting bacterial lifestyle (including motility, biofilm-formation and differentiation) on a transcriptional, posttranscriptional and posttranslational level [12]. Even 5'-pGpG has been discussed as a potential second messenger [12, 20], being a member of the nanoRNA class of molecules, which often play a role in the control of gene expression [21].

The importance of c-di-GMP implies the necessity of precise systems to regulate cellular levels of this second messenger in response to various environmental stimuli. Signal transduction in sensory proteins allowing for such regulation is often a dynamic and intricate process [23]. In an attempt to understand these mechanisms of signal integration on a molecular level, we have recently focussed on the characterisation of PadCs with regard to sequence, structure and function.

One PadC member that has been studied in particular detail is *IsPadC*, the homolog from *Idiomarina* sp. A28L [5–7]. The crystal structure of the dark state full-length wildtype *IsPadC* (pdb 5LLW, [5]), as depicted in Fig. 1a, demonstrates the homo-dimeric arrangement which is typical for bacteriophytochromes (BphPs) and DGCs. The active site is formed at the dimer interface of the GGDEF domains [24], and the α -helical sensor–effector linker element features a coiled-coil arrangement. *IsPadC* is considered a prototypical PadC since red light-stimuli lead to pronounced upregulation of DGC activity. Biochemical characterisation of *IsPadC* has revealed a non-canonical spectral behaviour, which features a red light illuminated spectrum that prominently differs from that of several other well-characterised bacteriophytochromes such as *Deinococcus radiodurans* BphP [25] (Fig. 1c) or Agp1 [26]. In the latter cases, both the P_r dark state (phytochrome absorbing in red light) and the P_{fr} conformation (phytochrome absorbing in far red light) formed upon red light activation predominantly contribute to the absorption spectra acquired under the corresponding

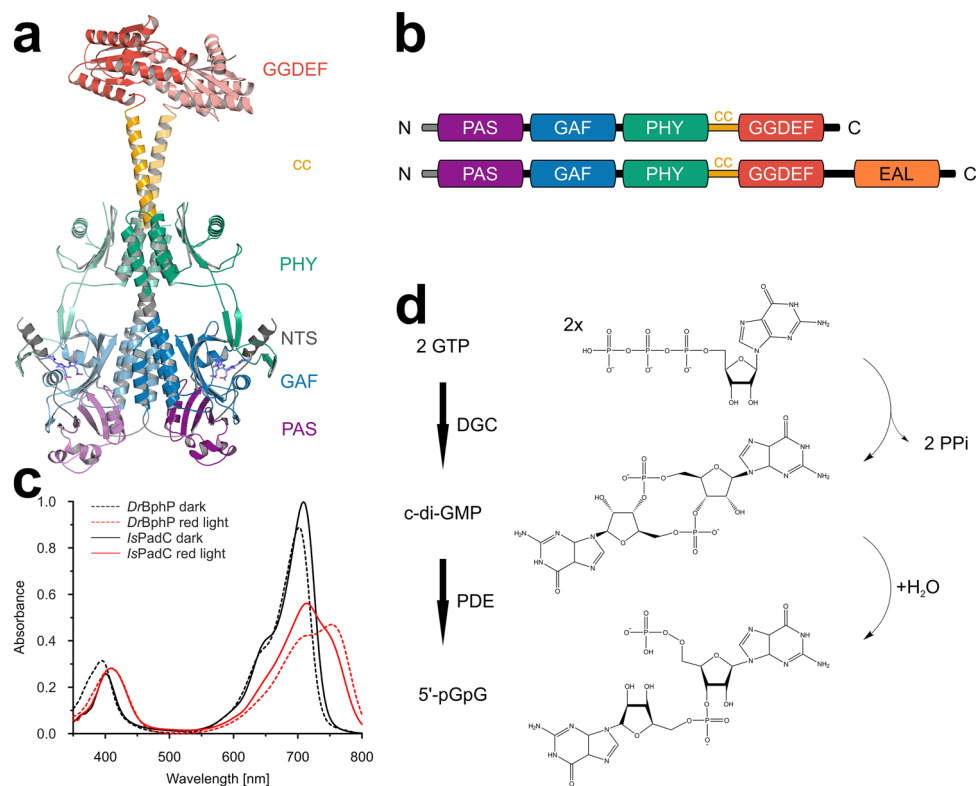


Fig. 1 Phytochrome-activated diguanylate cyclases (PadCs). **a** Dark-state crystal structure of full-length *IsPadC* (pdb 5LLW) represented as cartoon coloured by structural domains [5]. **b** Domain architecture of PadCs and PadC-EALs, coloration of domains as presented in **panel a**. N-terminal segment (NTS)=grey; Period/ARNT/Single-minded (PAS) domain=purple; cGMP phosphodiesterase-Adenylate cyclase-FhlA (GAF) domain=blue; phytochrome-specific (PHY) domain=green; coiled-coil (cc) linker=yellow; GGDEF

domain=red; EAL domain=orange. **c** Comparison of *Deinococcus radiodurans* bacteriophytochrome (*DrBphP*, dashed) and *IsPadC* (continuous) UV/Vis spectra under non-actinic (black) and red light (red) conditions. **d** Catalytic activity of PadCs and PadC-EALs—diguanylate cyclases (DGCs) catalyse turnover of two molecules of GTP to cyclic dimeric GMP (c-di-GMP), which is further hydrolysed to 5'-pGpG [22] by the EAL phosphodiesterase domain (EAL-PDE)

light regimes. *IsPadC*, however, has been shown to not fully occupy a canonical P_{fr} state, even though the chromophore was confirmed to be fully shifted to its 15E isomer. In the crystal structure of a constitutively active variant (pdb 6ET7, *IsPadC* S505V A526V [6]), only one of two PHY tongues is rearranged from β -hairpin to α -helix as observed for both extensions in other bacteriophytochromes [27], leading to the conclusion that the non-canonical light spectrum observed for wild type *IsPadC* might be attributed to differences in the biliverdin environments of the individual protomers. The mechanism of light regulation in *IsPadC* was shown to be based on local rearrangements in the cofactor environment after chromophore isomerisation including the GAF domain binding pocket, as well as the NTS. The interplay of the biliverdin isomerisation state and the altered NTS orientation with the PHY tongue translates into structural and conformational dynamic changes at the PHY domain dimer interface and consequently the central helical spine and the sensor–effector linker helices. The sensor–effector linker undergoes a translational movement corresponding to

reorientation from inhibiting to stimulating coiled-coil register, enabling additional productive encounters of the effector domains and hence increased enzymatic activity [5, 6].

A second homolog that attracted attention due to its fundamental differences to *IsPadC* is *MpPadC* (found in the organism *Marinobacter persicus*). *MpPadC* shows barely any red light-induced upregulation of enzymatic activity and it features a canonical bacteriophytochrome light state fully populating P_{fr} . On a sequence level a major discrepancy in linker length, i.e. the distance between PHY and GGDEF domains, was detected. For most PadCs, differences in linker length are multiples of seven amino acids, which is indicative of heptad repeats essential for coiled-coil formation. Indeed, in *IsPadC* the linker sequence follows a classical h-p-p-h-p-p-p (hydrophobic; polar) pattern, conventionally denoted as *a b c d e f g* [28]. Importantly, two overlapping heptad repeat patterns can formally be defined for many PadC homologs, which presumably allows for the register-switching necessary for profound upregulation of DGC activity [6]. *MpPadC*,

in contrast, features a linker length that is five residues shorter in comparison, suggesting that continuous coiled-coil interactions are impeded in this protein [8]. A series of *IsMpPadC* chimaeras revealed that the sequence of *MpPadC* retains the ability for red light-based regulation of enzymatic activity if the missing residues are inserted, and that the PHY dimer interface plays an essential role in signal integration, which even overrides the effect of a functional linker length [8].

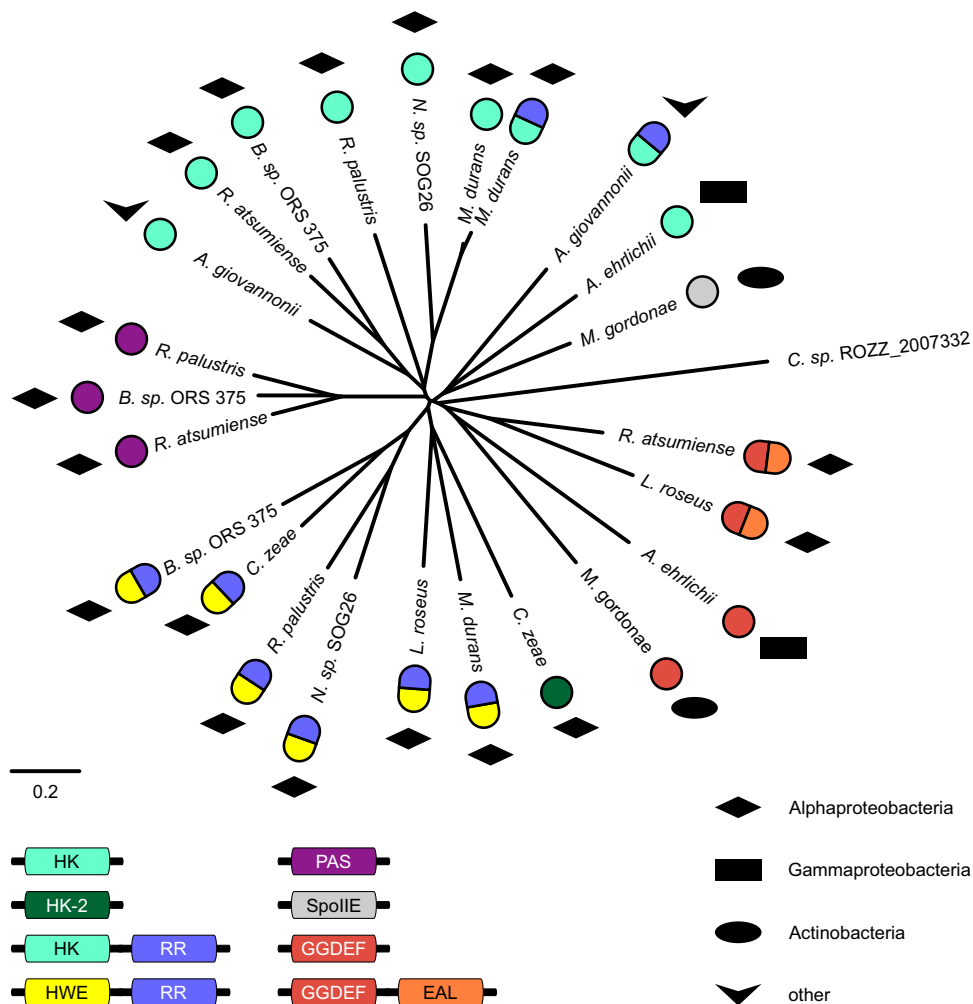
To understand how these drastically different functional properties are defined on a molecular level, we characterised these phytochrome-based DGCs more extensively. We performed a phylogenetic analysis of PadCs and PadC-EALs, as well as the biochemical characterisation of representative PadC homologs, which provided novel insights into the evolution as well as the importance of the central signal integration unit composed of the linker element and the PHY dimer interface. In addition, we have addressed the evolutionary emergence of different bacteriophytochrome-effector combinations to better understand the mechanisms underlying PadC sequence conservation and to get closer to answering

the question of how different structural features affect overall protein function in these sensor–effector systems.

2 Results

The importance of reacting to the presence or absence of (red) light is reflected in the diversity of systems sensing different qualities of light and integrating this information in various signalling pathways. Bacteriophytochromes can be linked to different effector domains, which permit not only the regulation of c-di-GMP levels but also control of auto-phosphorylation or phosphatase activities, amongst other reactions. Interestingly, several bacterial organisms feature proteins with more than one effector domain linked to a phytochrome-based sensory module. Phylogenetic relationships between representatives of such PAS-GAF-PHY architectures with different effectors present in one organism are represented in Fig. 2. For *Rhodospseudomonas palustris*, for example, database entries exist for bacteriophytochromes linked to a PAS domain, a histidine kinase, a HWE histidine

Fig. 2 Distance tree of representative bacteriophytochromes with more than one PAS-GAF-PHY-linked output domain per organism, the cryptophyte phytochrome from *Chroomonas* sp. ROZZ_2007332 serves as outgroup. Homologous sequences tend to cluster more according to effector (coloured beads) than to evolutionary lineage (filled symbols). Histidine kinases (HK and HK-2) = turquoise and green, respectively; response regulator (RR) = blue; HWE histidine kinase = yellow; GGDEF/diguanylate cyclases = red; EAL/phosphodiesterase = orange; PAS = purple; *Bacillus subtilis* stage II sporulation protein E (SpoIIE)/PPM-type phosphatase domain = grey. Black shapes represent taxonomy class



kinase with an attached response regulator, and even the sensory module without any output domain (not depicted in the distance tree).

According to PHYLIP-predicted distances, bacteriophytochrome sequences sharing the same effector, but occurring in different species, have closer relationships than bacteriophytochrome sequences linked to different effectors within a single species. For the majority of effectors, this is corroborated by sequence identities of PAS-GAF-PHY core modules. Based on this observation, we conclude that the occurrence of multiple phytochrome-linked effector combinations is predominantly based on horizontal gene transfer, a well-established major player in microbial evolution [29], rather than recurring internal recombination events. This is further supported by the fact that relatively closely related sequences also occur in different phyla and classes of the bacterial kingdom. Nevertheless, there is also no strict clustering according to effector domains, suggesting that comparable sensor–effector recombination events have happened multiple times, with different fusions improving species fitness.

For the selected organisms containing GGDEF or GGDEF-EAL outputs along other bacteriophytochrome-linked effectors, frequently a closer relationship is observed between identical sensor–effector combinations rather than the bacteriophytochrome sequences within the respective organism. While this shows that functional systems are present in many organisms, it also indicates that successful recombination events which lead to novel red light-regulated systems have been rather rare.

2.1 A defined set of linker lengths is found in naturally occurring PadCs

In this study, we focussed on a group of bacteriophytochromes for which the linker connecting sensor and effector domain has emerged as an important structural and functional element [5–8], i.e. the family of phytochrome-activated diguanylate cyclases. PadCs are defined by the presence of a set of essential motifs and residues: A highly conserved cysteine in the NTS, which covalently binds to the chromophore [30, 31] (C17 in *IsPadC*), the DIP motif in the GAF domain cofactor binding pocket [32], the PRXSF motif in the PHY tongue [10] and the DGC activity defining GGDEF/GGEF [33] motif. PadC-EALs are required to additionally feature the EAL [15] or EXL motif [34], which is part of the EAL-PDE active site [35], where X can be any hydrophobic, aliphatic residue. The number of sequences obtained when searching the NCBI protein database [36] using the iterative PSI-blast algorithm [37] under these constraints ranged around 200 for PadCs (~140) and PadC-EALs (~60) combined (consult Supp. Tables 1–5 for additional information).

In order to further investigate the evolutionary relationship amongst PadC sequences, we subjected a structure-oriented global alignment (complete alignment in Supp. Figure 1), of which a subset was previously used to motivate the characterisation of *MpPadC* [8], to a PHYLIP distance analysis with 1000 delete-half Jackknife sub-alignments [38]. The resulting phylogenetic tree (Fig. 3, see Supp. Tables 1 and 2 for NCBI accession numbers and additional information) constructed from representative PadC and PadC-EAL sequences uses the BphP-like fungal phytochrome FphA from *Aspergillus nidulans* [39, 40] as an outgroup. This bootstrap tree clearly highlights the importance of the coiled-coil linker element and we therefore include the corresponding linker length information in all homolog descriptions as superscript from now on—e.g. *IsPadC*⁺⁷, corresponding to a linker length of +7 residues relative to the reference system *TsPadC*⁰ (*Thioalkalivibrio* sp. ALMg3, with a 28 residues long linker). The latter is defined as the number of amino acid residues between a conserved hydrophobic residue in the terminal PHY-helix (I495 in *IsPadC*⁺⁷) and the DXLT motif at the beginning of the effector domain. In our phylogenetic analysis, three homologs were included despite not corresponding to the definition of PadCs and PadC-EALs as described above—*HmPadC*⁺¹⁴ (*Halomonas meridiana*), a homolog missing all essential motifs and residues that regardlessly clusters tightly with *IsPadC*⁺⁷, as well as *BbPadC-EAL*⁻⁷ (*Betaproteobacteria bacterium* SCN2) and *NsPadC-EAL*⁻⁷ (*Nevskia soli*), which feature an EAL motif mutated to EAF, most likely resulting in no EAL-PDE activity in those systems.

Interestingly, while only a limited set of linker lengths is found within the compiled set of PadCs, the group of PadC-EALs additionally features the unique linker lengths of -5, -2 and +24 (*ApPadC-EAL*⁻⁵, *Acidisarcina polymorpha*; *RnPadC-EAL*⁻², *Roseomonas nepalensis*; *AtPadC-EAL*⁺²⁴, *Aquabacterium tepidophilum*). It should also be noted that no single organism features more than one PadC homolog with different linker lengths. Figure 4 depicts the complete set of PadCs in the form of a sequence similarity network (SSN), with BLAST-generated [41, 42] bitscores (S) above a threshold of 380 serving as edges and nodes representing homologs coloured according to linker length. The majority of PadCs (54%) feature linker length differences divisible by seven, which is indicative of the heptad repeat patterns necessary for coiled-coil formation, as has also been confirmed experimentally [5, 6]. Interestingly, with 38% the number of identified homologs with a seemingly non-conforming linker length of -4 (*PadCs*⁻⁴) is surprisingly high and, in addition, a small number of PadCs was found to have a linker length of +2 (*PadCs*⁺²) [8]. The relative abundance of all linker lengths is represented in Fig. 4a; and Fig. 5 depicts close-ups of all linker area sequences as well as the residues involved in stabilising the predicted inhibiting and

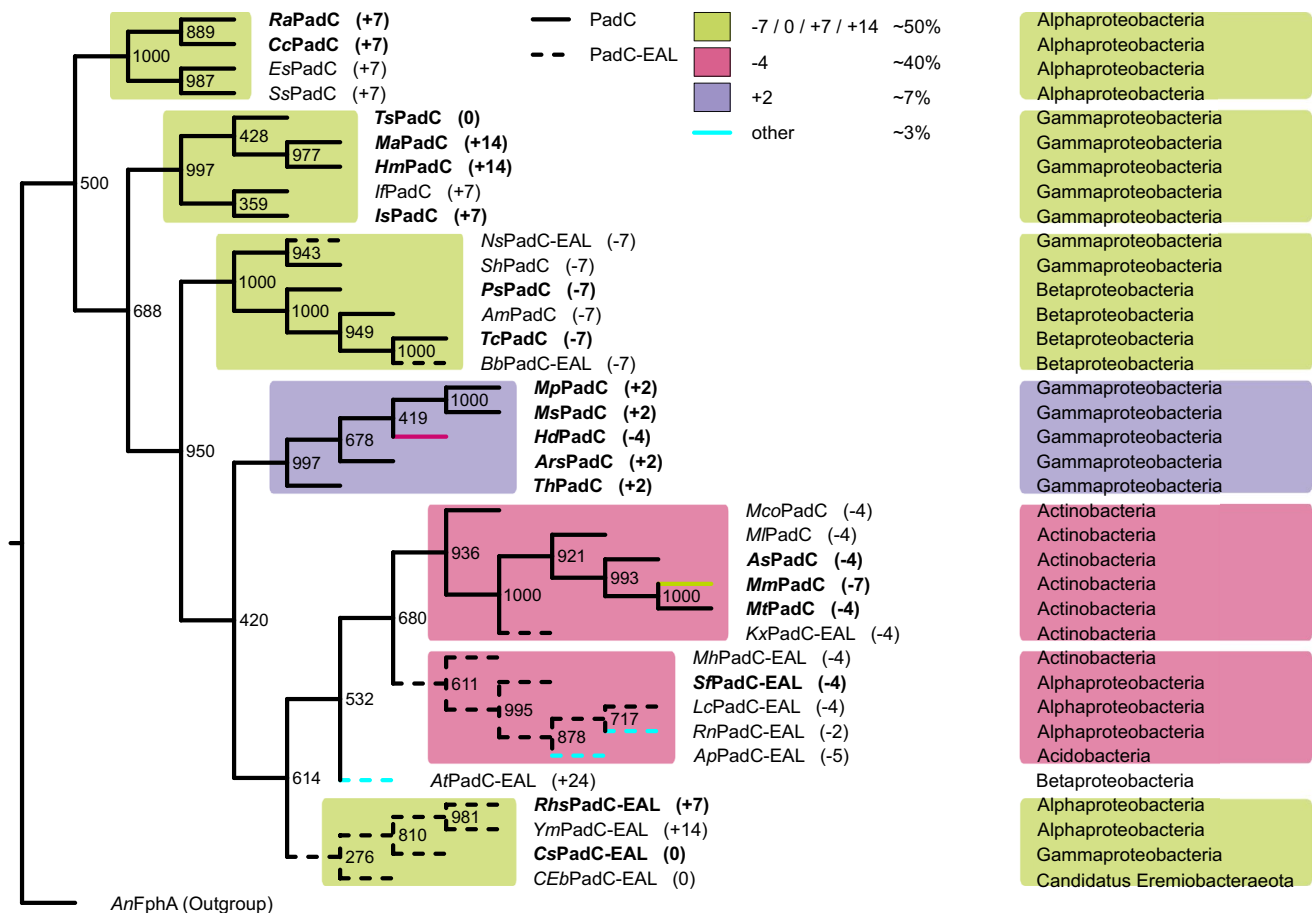


Fig. 3 Bootstrap tree of representative PadC (continuous) and PadC-EAL (dashed) homologs, relative linker length prevalence is indicated for PadCs and PadC-EALs combined. Bold entries represent biochemically characterised homologs. PadCs cluster according to linker

length and linker lengths differing by multiples of seven are represented in green, -4 in pink and $+2$ in purple; all others are highlighted in turquoise. The corresponding taxonomy class is provided on the right for every homolog

stimulating registers. Whereas PadCs cluster according to linker length in the SSN, at the chosen bitscore cut-off, all PadC-EALs group together regardless of linker length. At increasing cut-offs, PadC-EALs also separate into linker length-specific clusters as seen in the bootstrap tree; however, individual PadC subfamilies are no longer linked to the SSN at that point.

2.2 Different linker length families are most likely based on parallel evolution

The phylogenetic tree (Fig. 3) features strong clustering according to linker length, with one major branch of mixed linker lengths divisible by seven including both *IsPadC*⁺⁷ and *TsPadC*⁰ [5], as corroborated by bootstrap values. High bootstrap values for each individual branch indicate that they are significant as separate branches (> 950, i.e. a 95% confidence interval), whereas the relatively low bootstrap values defining relation of the branches to each

other point most likely towards parallel evolution of individual linker length families. The exception to this trend of parallel evolution is the extensive cluster of PadCs⁻⁴ and PadC-EALs⁻⁴, which are closely related to each other, though not significant as an individual branch. *KxPadC-EAL*⁻⁴ (*Kineococcus xinjiangensis*), interestingly, clusters with PadCs⁻⁴ rather than PadC-EALs⁻⁴, as confirmed by a bootstrap value of 1000. All PadCs⁻⁴, as well as roughly 50% of the PadC-EAL⁻⁴ homologs, belong to the class of Actinobacteria (as visualised in the SSN depicted in Fig. 4), indicating that PadCs with a linker length of -4 evolved from PadC-EALs⁻⁴ through loss of the EAL domain. The odd PadC⁻⁴ family member *MmPadC*⁻⁷ (*Microbacterium mangrove*—see Fig. 3 and Supp. Fig. 2) is present in an Actinobacterium, like its PadC⁻⁴ branch members. In contrast, *HdPadC*⁻⁴ (*Halospina denitrificans*), which clusters in the PadC⁺² group, is found in a Gammaproteobacterium, further reinforcing the PHYLLIP-predicted relationships of both homologs in spite of their

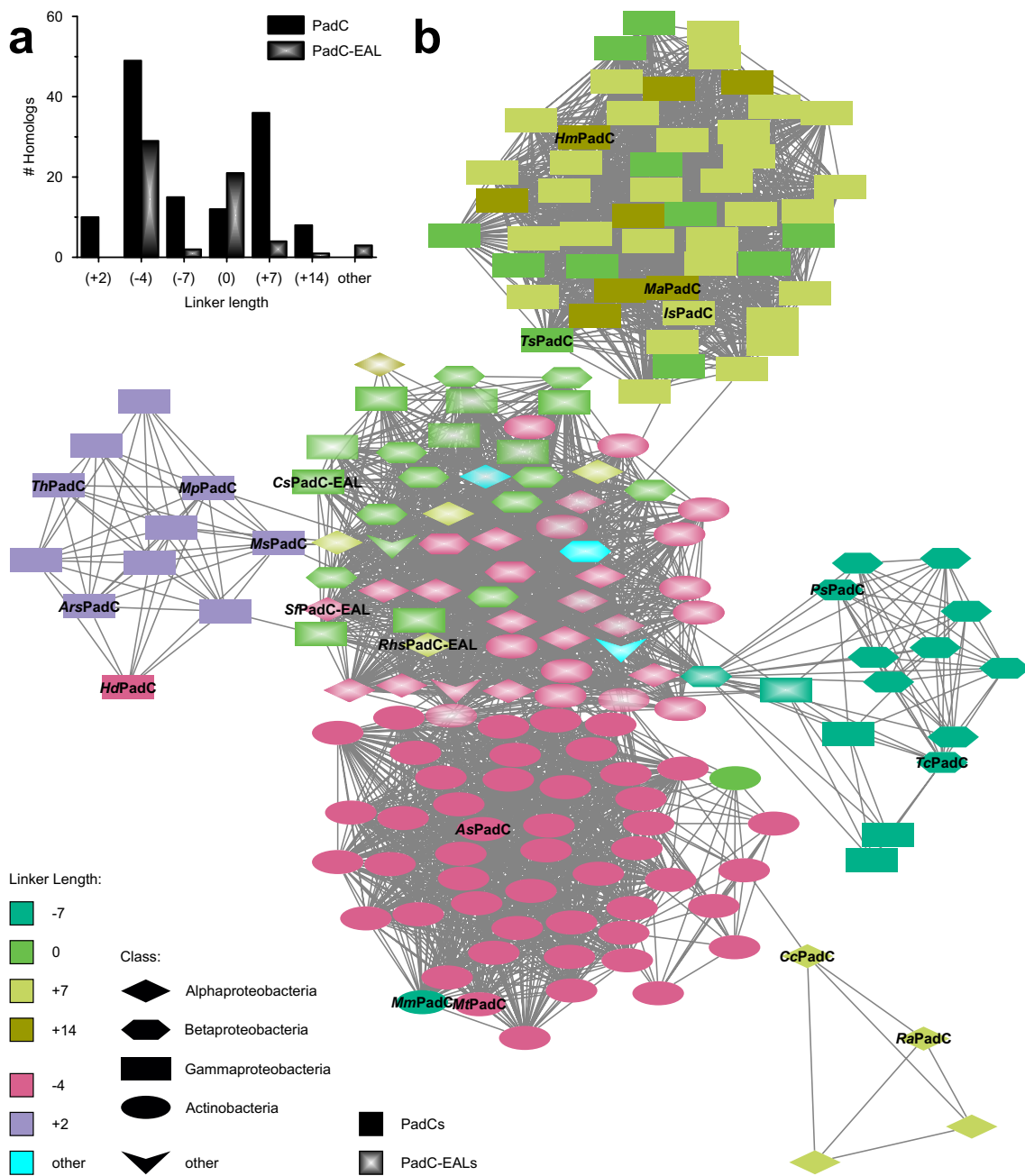


Fig. 4 Evolutionary distribution of various PadC linker lengths. **a** Relative frequencies of naturally occurring linker lengths in PadCs (black) and PadC-EALs (opaque). **b** Sequence similarity network of PadCs and PadC-EALs. Nodes correspond to different homologs; edges represent BLAST-generated bitscores (S) above a minimal threshold of 380. Linker lengths divisible by seven are shown in

shades of green, PadCs⁺² in purple and PadCs⁻⁴ as well as PadC-EALs⁻⁴ in pink; turquoise denotes other rare linker lengths. The different node shapes refer to the phylogenetic class. All expressed and/or biochemically characterised homologs are indicated by name as referred to in previous publications [5, 7, 8] or this manuscript

divergent linker lengths. Annotated linker lengths of both homologs were confirmed as correct through re-sequencing of DSMZ DNA. That the few observed outliers also feature linker lengths observed in other branches further shows the high evolutionary pressure restricting linkers to a confined length and to a lesser degree sequence space.

In general, PadC-EALs—though not equally robust in their linker length-related clustering as PadCs—firmly group together. Moreover, in a bootstrap tree of the EAL domains only (Supp. Figure 3), no individual branches are statistically significant, suggesting a close evolutionary relationship of the EAL domains. The strong conservation of

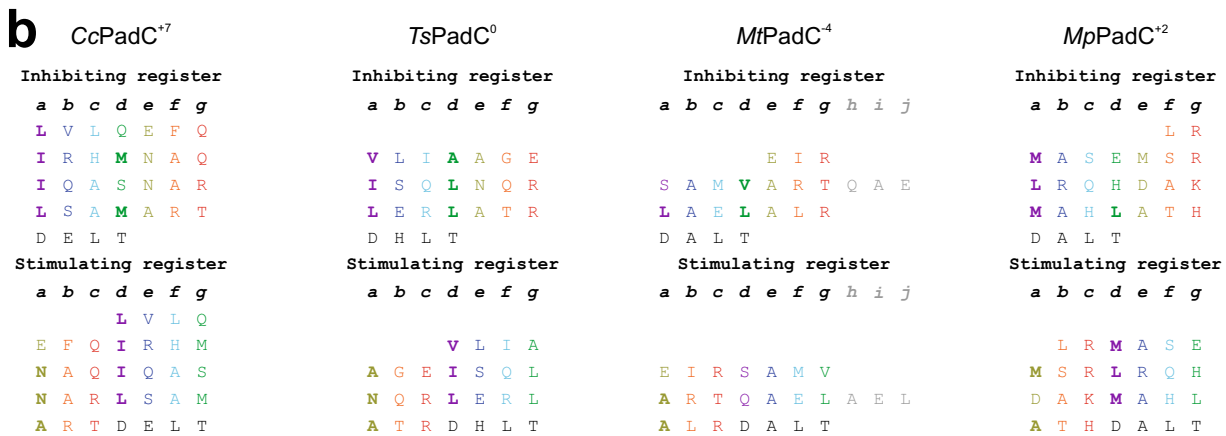
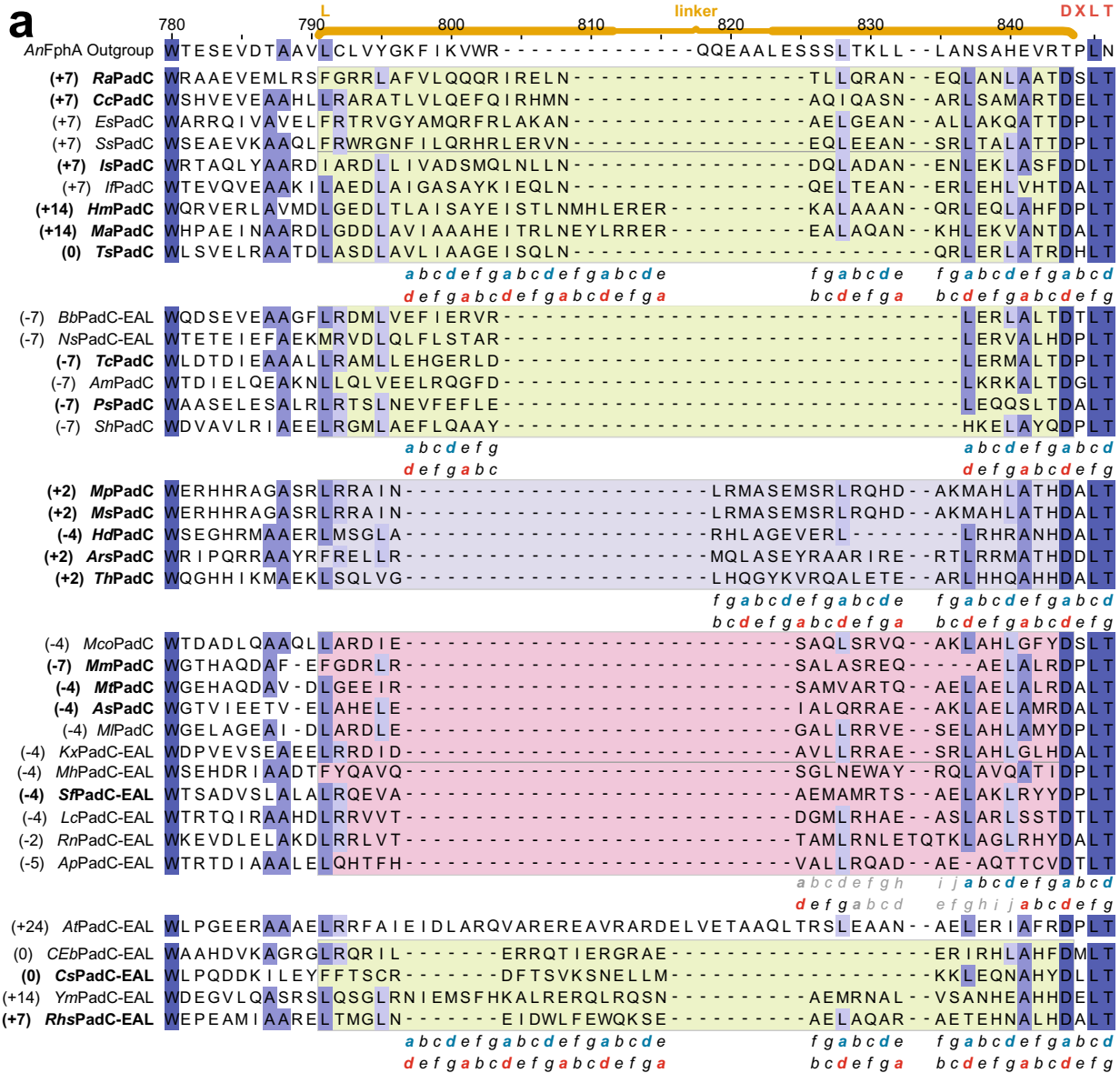


Fig. 5 Sequences and register predictions of the coiled-coil linker elements. **a** Alignment of the coiled-coil linker, with linker lengths represented by colour (green=divisible by seven, pink=-4, purple=+2). Highlighting of residues corresponds to the degree of conservation (hues of dark blue to the respective background colour= $\geq 84\%$, $\geq 68\% \geq 40\%$, $< 40\%$). Heptad repeat patterns typical for coiled-coils are listed for inhibiting (blue) and stimulating (red) registers, grey residues correspond to the overwinding postulated for PadCs⁻⁴ [43]. Bold entries represent investigated homologs. **b** Heptad repeat definitions in both inhibiting and stimulating registers for CcPadC⁺⁷, TsPadC⁰, MtPadC⁻⁴ and MpPadC⁺²; see [5] for a representation of IsPadC⁺⁷. Coiled-coil stabilising residues (hydrophobics and Asn) are bold

length observed for the linker element between GGDEF and EAL domains in addition to numerous conserved residues (Supp. Fig. 4) further supports the hypothesis of a common GGDEF-EAL ancestor. This conservation of the EAL-linker is seen even in the two exceptions to the cluster of PadC-EALs. BbPadC-EAL⁻⁷ and NsPadC-EAL⁻⁷, both featuring a linker length of -7, are more closely related to PadCs⁻⁷ than to other PadC-EAL homologs. Interestingly, when considering relationships of the EAL domains, PadC-EALs⁻⁷ are not significant as a branch despite the loss of residues necessary for catalytic competence.

2.3 Homologs with a functional coiled-coil linker permit efficient red light-induced regulation of enzymatic activity

As previously published [8] and outlined in the introduction, two PadCs of different linker lengths that were characterised in greater detail—IsPadC⁺⁷ and MpPadC⁺²—distinctly differ in their biochemical behaviour. In order to examine how far these observed disparities in light-controlled regulation of enzymatic activity, as well as spectral behaviour, correlate with linker length, we have biochemically characterised additional representative PadCs of each branch. The different homologs were selected based on phylogenetic diversity in an attempt to address how specific characteristics are preserved within branches and linker length groups. In addition to archetypic representatives of the individual groups, HdPadC⁻⁴ and MmPadC⁻⁷ were investigated, to see whether their behaviour is influenced more strongly by linker length or overall sequence similarity. Three representatives of the PadC-EAL subgroup were included as well—SfPadC-EAL⁻⁴ (*Salinibacterium flavidus*), RhsPadC-EAL⁺⁷ (*Rhodobacter* sp. JA431) and CsPadC-EAL⁰ (*Catenovulum sediminis*). The three latter homologs showed different degrees of oligomerisation and degradation, are regulated by red light and all three feature DGC activity. RhsPadC-EAL⁺⁷ is the only homolog investigated for which EAL-PDE activity was measured as well, and its partial cleavage pattern corresponds to what has previously been reported for the *Rhodobacter sphaeroides* PadC-EAL homolog [14], a member of the PadC-EAL⁻⁴ family. Due to low overall yields

and the observed mixtures of oligomeric species of full-length and partially truncated forms of the proteins, a quantitative description of kinetic and spectral properties would go beyond the scope of this manuscript.

2.3.1 Homologs with linker lengths divisible by 7

From the three branches containing linker lengths divisible by seven, a total of six homologs were characterised in detail (see Supp. Tables 1 and 2 for NCBI accession numbers and taxonomy). CcPadC⁺⁷ (*Cereibacter changlensis*) and RaPadC⁺⁷ (*Rhodoligotrophos appendicifer*) represent a branch exclusively comprised of homologs with a linker length of +7. Of the PadC⁻⁷ branch, only PsPadC⁻⁷ (*Pusillimonas maritima*) could be characterised, as expression levels of TcPadC⁻⁷ (*Thauera chlorobenzoica*) were too low to allow efficient purification. The same held true for HmPadC⁺¹⁴, the homolog lacking all essential motifs and residues. From the central branch of homologs with a canonical linker length, MaPadC⁺¹⁴ (*Marinimicrobium agarilyticum*), featuring a linker length of +14, was previously characterised in addition to IsPadC⁺⁷ and TsPadC⁰ [5].

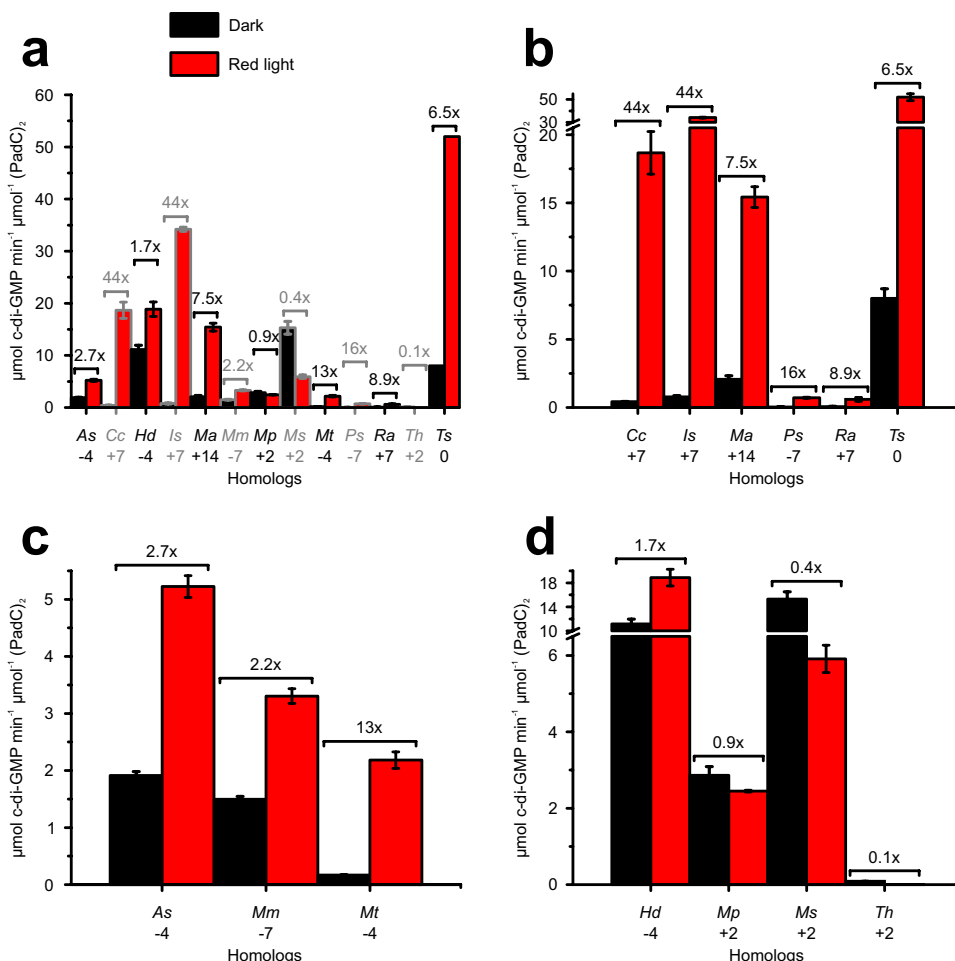
All six homologs listed above show strong red light-dependent upregulation of enzymatic activity, as visualised in Fig. 6a, b and detailed in Supp. Table 6. Dynamic ranges reach from 6.5-fold upregulation, as measured for TsPadC⁰, to a roughly 40-fold increase seen in IsPadC⁺⁷ or CcPadC⁺⁷. Neither are these differences in light regulation homogenous within individual branches of the phylogenetic tree, nor are absolute turnover rates under both red light and non-actinic conditions. However, a significant increase in enzymatic activity upon illumination is observed for all investigated PadC homologs with a linker length divisible by seven.

Similarly, all homologs of this group feature spectra with a non-canonical light state (see Fig. 7), as previously published for IsPadC⁺⁷ [5, 6]. Interestingly, the relative contributions of P_{fr} and P_r-resembling [44] Q-bands to the light-activated spectra differ substantially amongst homologs, indicating different contributions of the canonical P_{fr} state to the individual light state spectra [8]. Wavelengths of both P_r and P_{fr} Q-band maxima are relatively similar amongst PadCs with linker lengths divisible by seven; with P_r maxima ranging from 700 to 708 nm and P_{fr} Q-band maxima from 737 to 751 nm. Within this group, P_{fr} state stabilities vary considerably, with mean lifetimes averaging from 1 to 2 min (CcPadC⁺⁷, IsPadC⁺⁷ [5]) to approximately 1 h (MaPadC⁺¹⁴, TsPadC⁰ [5]).

2.3.2 Homologs belonging to the branch featuring predominantly linker lengths of -4

P_r and P_{fr} Q-band absorption maxima of PadCs⁻⁴—AsPadC⁻⁴ (*Agromyces* sp. Leaf222) and

Fig. 6 Specific DGC activities for representative PadCs. All constructs are compared at 200 μM GTP and normalised to dimer concentrations of protein, *IsPadC*⁺⁷, *MaPadC*⁺¹⁴ and *TsPadC*⁰ specific activities correspond to those previously published [5, 7]. Turnover rates, measured under non-actinic (black) and red light (red) conditions, were calculated from four timepoints each. Error bars represent the SE of the estimate from the linear regression. Fold differences between dark and red light measurements are indicated by brackets above the bars, numbers < 1 denote downregulation of DGC activity. **a** Overall comparison of specific activities measured for every investigated homolog. **b–d** Specific activities of PadCs with linker lengths divisible by seven, PadCs⁻⁴ (including *MmPadC*⁻⁷) and PadCs⁺² (including *HdPadC*⁻⁴), respectively



MtPadC⁻⁴ (*Microbacterium trichothecenolyticum*)—as well as the closely related *MmPadC*⁻⁷ are found to be in a similar range; 698 / 699 nm and between 740 and 745 nm, respectively. Interestingly, the extent of P_{fr} contributions to the steady-state spectrum during red light illumination (680 nm) in this group is more homogenous than within the branches with linker lengths divisible by 7; and all three homologs feature non-canonical light state spectra. Mean P_{fr} state life times were measured to be in the range of one hour.

The dynamic ranges of light activation measured for this group are generally noticeably smaller (ranging from 2.2 to 13 fold upregulation) than observed for linker families with complete heptad repeats, but specific turnover numbers are considerably more uniform (see Fig. 6 and Supp. Table 6). *MmPadC*⁻⁷ shows behaviour very similar to that of *AsPadC*⁻⁴; red light illumination triggers upregulation of enzymatic activity in both homologs, but on a smaller scale than observed for *MtPadC*⁻⁴ or PadCs belonging to the sub-families with linker lengths divisible by seven.

2.3.3 Homologs belonging to the branch featuring predominantly linker lengths of +2

In contrast, homologs with a linker length of +2 differ substantially in their behaviour from all other PadCs, as recently described in detail for *MpPadC*⁺² [8]. In addition to this representative, four additional homologs from this branch of the phylogenetic tree were characterised—*ArsPadC*⁺² (*Arhodomonas* sp. KWT), *MsPadC*⁺² (*Marinobacter* sp. T13-3), *ThPadC*⁺² (*Thiohalospira halophila*) as well as *HdPadC*⁻⁴. Similar to members of the PadC⁻⁴ branch, *HdPadC*⁻⁴ shows some upregulation of enzymatic activity upon illumination (a factor of 1.7). *MpPadC*⁺², *MsPadC*⁺² and *ThPadC*⁺², in contrast, feature either almost no effect of illumination (*MpPadC*⁺²) or even red light-induced downregulation of enzymatic turnover (*MsPadC*⁺², *ThPadC*⁺²) up to only a tenth of the activity measured under non-actinic conditions (Fig. 6, Supp. Table 6). Contrary to PadCs with a linker length divisible by seven, which feature highly efficient inhibition in the absence of red light, a relatively high basal activity was observed for *MpPadC*⁺² and *MsPadC*⁺² as well

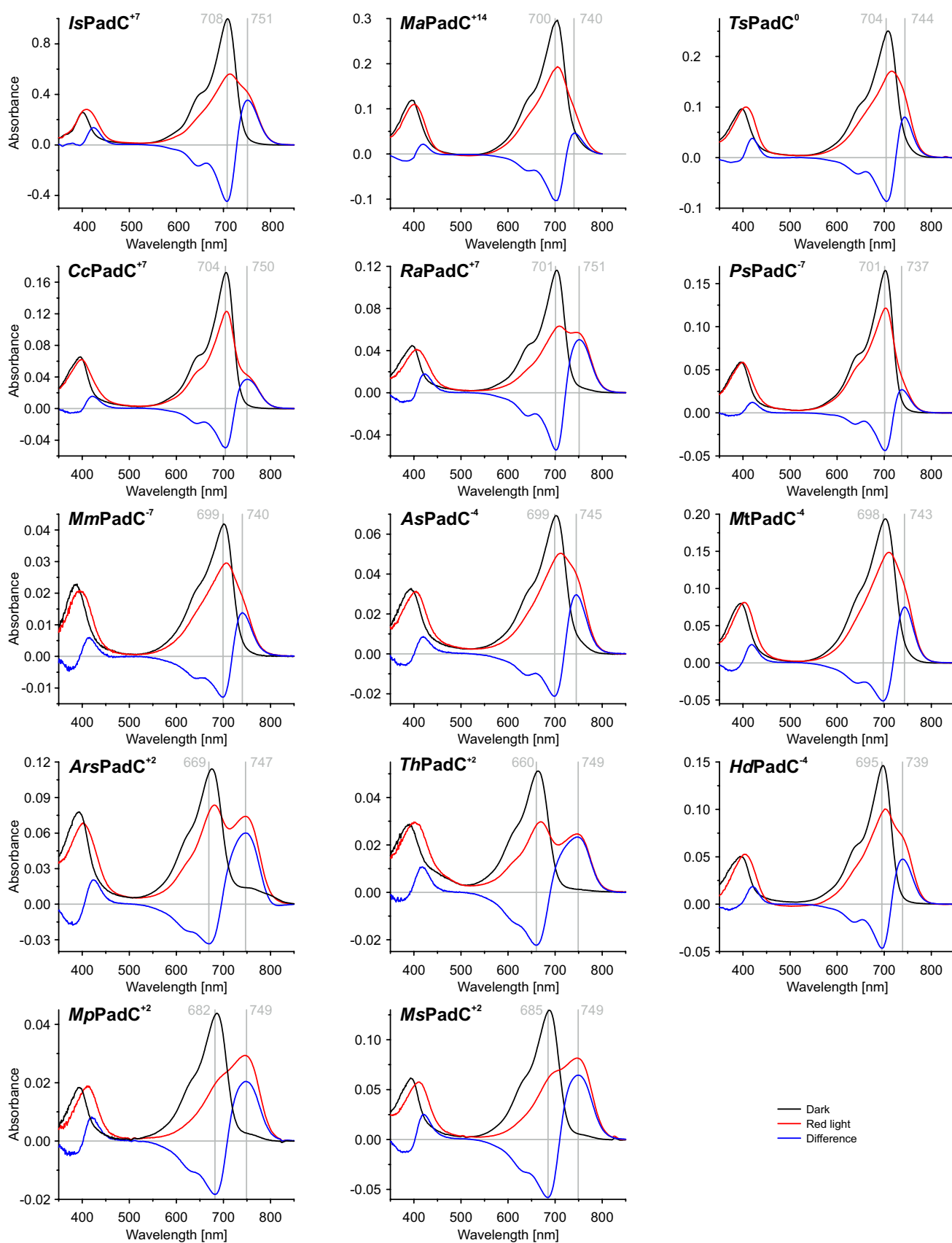


Fig. 7 UV/Vis absorption spectra of all characterised homologs. Dark state (black), red light (red) and difference (light—dark; blue) in-solution spectra obtained at 20 °C and with identical illumination protocols

as *HdPadC*⁻⁴. Interestingly, the HPLC-based assay used to quantify enzymatic activity also exposed a secondary product peak—considerably smaller than that measured for *c*-di-GMP—in *ThPadC*⁺², *ArsPadC*⁺², to our surprise, produced only this additional, as-of-yet unidentified compound, and no measurable quantities of *c*-di-GMP.

In-solution spectra further emphasise the differences between the *PadC*⁺² family and homologs with other linker lengths. While P_{fr} Q-band absorption maxima fall within the scope observed for the previously introduced groups (ranging from 739 to 750 nm), P_r Q-maxima show considerable variations (660–695 nm). The most extreme value was measured for *ThPadC*⁺². Only one additionally investigated member of the *PadC*⁺² family, *MsPadC*⁺², features a symmetric, canonical light state as has been reported for *MpPadC*⁺² [8]; in contrast, light state spectra measured for *ArsPadC*⁺², *ThPadC*⁺² and *HdPadC*⁻⁴ rather feature non-canonical spectra as observed for *IsPadC*⁺⁷, albeit to a lesser degree. P_r state recovery kinetics feature mean lifetimes of approximately one hour for *HdPadC*⁻⁴, and range up to about 30 h measured for *MpPadC*⁺² [8], showing considerably higher P_{fr} state stability than all other linker length families.

2.4 The central helical spine strongly influences effector domain positioning

The ability of a *PadC* dimer to form a functional coiled-coil in its sensor–effector linker has a profound effect on protein structure as well as dynamics. In the absence of experimentally determined structures, we calculated AlphaFold2 [45] predictions for all representative *PadC* homologs (see Supp. Figure 5 and 6). 5LLW, the published dark state structure of *IsPadC*⁺⁷ [5], was part of the array used to train the AlphaFold2 AI [45], eliminating a potential benchmark. According to pLDDT and PAE scores (see Supp. Figs. 7 and 8, respectively), confidence in predictions of the PAS-GAF core and its dimer interface is high. PHY dimerisation and coiled-coil linker architecture are expectedly less confident, and the prediction of the relative GGDEF effector module orientations is not well defined. The latter was expected due to the high degree of freedom observed for the DGC effector in SAXS measurements of *IsPadC*⁺⁷ [5]. Eventually, the extent of this conformational freedom depends on the degrees of freedom in the PHY domain dimer interface as well as its compatibility with a coiled-coil linker architecture. AlphaFold2 predictions, while not as robust as an experimentally determined structure, still serve to illustrate the intrusive effect of very short linkers (*PsPadC*⁻⁷, Supp. Fig. 5b), or the lack of functional coiled-coils (Supp. Fig. 5c). Strong conservation of coiled-coil contacts in the sensor–effector linker reaching all the way to the characteristic wide turn corresponding to the DXLT motif in the DGC effector domain confirmed by coevolution analysis of *IsPadC*⁺⁷ using the GREMLIN

webserver (Supp. Fig. 9) further highlights the importance of the central helical spine in *PadC* signal integration. Several residues at the C-terminal tip of the GGDEF α_0 helix and in the GTP-binding site also coevolved with coiled-coil stabilising residues in the linker, highlighting not only direct contacts but also the allosteric regulatory networks at play. Interestingly, no coevolution is observed for the PHY domain dimer interface according to GREMLIN. This is in line with the low degree of conservation of residues in the PHY domain (excluding the PHY tongue) and most likely also reflects the functional divergence of the PHY domain as integrator unit in the different subfamilies of *PadCs*.

3 Discussion

In *PadCs* specifically [6, 8], and bacteriophytochromes in general [4, 46], the architecture of the central helical spine at the dimer interface greatly affects the degrees of freedom in effector domain orientation. Interestingly, the relatively loose dimer interface suggested by AlphaFold2 predictions of *MpPadC*⁺², *MsPadC*⁺² and *ThPadC*⁺² structures correlates with the comparatively open PHY interface conformation observed for the *Deinococcus radiodurans* phytochrome [27, 47, 48]. In the absence of a functional coiled-coil, due to the incomplete heptad repeats of a +2 linker, the central helical spine underlies fewer constraints and can be considerably more dynamic than observed for *IsPadC*⁺⁷ and its linker length family members. Additionally, it has been observed that even a functional linker length does not suffice to override the effect of the *MpPadC*⁺² PHY dimer interface [8]. In the context of the well-established light-induced structural rearrangements of the PHY tongue [10, 27], this could imply different signal integration mechanisms depending on relative PHY dimer interactions.

Even though PHY domain dimerisation was shown to be a key player in signal integration of *PadCs* [8], the PHY domain itself—excluding the tongue region—is subject to relatively little evolutionary pressure according to the low degree of sequence conservation (Supp. Fig. 1). Based on its importance for *PadC* functionality we conclude that the PHY dimer interface provides the evolutionary playground for adapting to various effectors. Whereas the PAS-GAF core is rigidly conserved, orientation of the PHY dimer interface and the central helical spine remains malleable. Potentially, this could also be one explanation for why *PadCs* tend to allow higher dynamic ranges than many PHY-less cyanobacteriochrome-based diguanylate cyclases [49]. However, tight interaction of the central helical spine is not necessarily a functional prerequisite for other bacteriophytochromes, as observed for the phytochrome-linked histidine kinase/phosphatase of *Deinococcus radiodurans* with its relatively loose PHY domain interaction [50] or the *Xanthomonas*

campestris BphP where breaks of the helical spine affect the quaternary assembly [46]. Nonetheless, neither regulation mechanism appears applicable to PadCs and the high degree of conformational flexibility observed for their GGDEF domains [5]. Modulation of DGC activity is known to be regulated by either dimerisation events or reorganisation of input domains enabling productive encounter [33].

Still, light-controlled regulation of enzymatic activity in PadCs depends strongly on PHY dimer interface interactions and coiled-coil register switching [5, 8]. Interestingly, the importance of register switching in coiled-coils influencing DGC activity has recently also been shown for a Rec-GGDEF system [51]. In PadCs, a balance between stabilisation of inhibiting and stimulating registers appears to provide the biophysical framework for efficient register switching, as observed for *IsPadC*⁺⁷ or *CcPadC*⁺⁷ (Fig. 5b). In *IsPadC*⁺⁷, variants with additional hydrophobic residues in *a* or *d* position of the inhibiting or stimulating register, respectively, suffice to trap the protein in either an inhibited or constitutively active conformation, respectively [6]. However, as seen in the linker regions of the PadC⁺² family (Fig. 5), a loss of the heptad repeat pattern as well as low conservation of hydrophobic residues in *a* and *d* positions impede the efficient stabilisation of different coiled-coil arrangements, thus implying a different mechanism for the light inhibition of DGC activity observed for some of its family members. For PadCs⁻⁴, on the other hand, the interference with coiled-coil formation is less pronounced than for PadCs⁺², as the loss of four residues corresponds to a deletion of roughly one helical turn. In this case, the -4 deletion is compensated by overwinding of the coiled-coil resulting in a 10 amino acid stretch for 3 turns as shown in Fig. 5 [43, 52]. In analogy to the inhibiting and stimulating registers observed for the families with classical heptad repeats, similar alternative register architectures can be identified in the PadC⁻⁴ subfamily (Fig. 5b). In this respect, it is not surprising that the biochemical behaviour of this linker length family more closely resembles that of the branch including *IsPadC*⁺⁷, albeit on average with lower dynamic ranges of enzymatic light activation and in line with the roughly fourfold upregulation observed for the *R. sphaeroides* BphG1 EAL-less construct [14].

The evolutionary behaviour of bacteriophytochromes in general complements what is observed on a smaller scale for linker length constraints in PadCs. Across the range of naturally occurring bacteriophytochromes, more than one functionality can be linked to a biliverdin-based sensory module even within individual organisms, as well-established for Agp1 and Agp2 in *Agrobacterium tumefaciens* [53] or the multiple BphPs identified in *Rhodospseudomonas palustris* [54], and representatives tend to cluster in distance trees according to their output domain. Within an effector family, amino acid substitutions with varying functional

consequences can happen occasionally, however, the resulting sequences are only retained if they either improve overall fitness, or generate no additional metabolic burden. We believe that of such random deletions or insertions in PadC linkers, only few survived, which were the origins of *MmPadC*⁻⁷ in the PadC⁻⁴ branch and *HdPadC*⁻⁴ in the +2 branch. The considerable evolutionary pressure on PadC linker length is further corroborated by the very distinct groups of linker lengths that are observed. With the exception of PadC-EALs, no naturally occurring linker lengths other than -4 and +2 consisting of incomplete heptad repeats were identified. A similar conservation of linker length is seen in LOV-regulated effector domains and especially pronounced for LOV-regulated DGCs [55]. We have previously suggested that PadCs with a linker length of +2 might not function as light-sensing signalling systems based on the comparison of apo- to holoprotein DGC activity as well as P_{fr} state stability, thus the loss in linker length may rather be a symptom of a successful functional adaptation than the cause [8]. A functional importance of apoforms has also been shown for other bacteriophytochromes in low light responses [56]. The unidentified enzymatic turnover of a side-product observed for *ArsPadC*⁺² as well as—to a lesser extent—*ThPadC*⁺² further suggests that this particular linker family serves a different function on a cellular signalling level than other PadCs, negating the evolutionary pressure on coiled-coil linker architecture. From an evolutionary perspective, protein function can change for particular homologs of a family as they are recruited for a different purpose without requiring extensive alterations to the sequence [57]. Moreover, Mantoni et al. [58] have recently shown that if productive active site formation at the GGDEF dimer is prevented, these domains are capable of slowly converting GTP to GMP. HPLC–MS analysis of the unknown product produced by *ArsPadC*⁺² as well as *ThPadC*⁺² does, however, not correspond to GMP, nor to GDP, 5'-pGpG or pppGpG (the intermediate frequently observed for DGC reactions [59]). Basal DGC activity levels measured for *ThPadC*⁺² are notably low and non-existent in *ArsPadC*⁺², suggesting that in some PadCs with a non-prototypical linker length the productive encounter of the effector domains is highly inefficient, allowing for production of various side-products.

The hypothesis of parallel development of different PadC linker length families based on low bootstrap values between individual branches (Fig. 3) is further supported by the relatively strong conservation of taxonomic classes, as also visualised in the SSN (Fig. 4). Within the family of PadCs, many significant linker length branches correspond exclusively to Alpha- (PadCs⁺⁷) or Gammaproteobacteria (PadCs⁺², branch containing *IsPadC*⁺⁷); PadCs⁻⁷ contain both Alpha- and Betaproteobacteria, whereas the branch of PadCs⁻⁴ contains Actinobacteria only. In contrast, PadC-EALs feature Acidobacteria as well as Alpha-, Beta-,

Gammaproteobacteria and Actinobacteria. However, phylogenetic clustering by linker length has been shown to be less robust in PadC-EALs, as corroborated by the bootstrap values. PadC-EALs appear to be more flexible overall when it comes to linker length, featuring three unique linker lengths of -5 , -2 and $+24$, which could be an indicator that their EAL-PDE activity is of higher relevance to cellular signaling than that of the DGC output module.

On an evolutionary level, we believe that fusion events of GGDEF-EAL couples to bacteriophytochrome cores were rare during evolution and that the EAL domain was then repeatedly lost, yielding the groups of PadCs⁻⁴ and PadCs⁻⁷. This conclusion is based solidly on the relationship of PadC-EAL⁻⁷ EAL domains to other homologs' as well as the degree of conservation observed for the EAL-linker length and the numerous conserved residues between DGC and EAL domains (see Supp. Fig. 4), which is highly unlikely to have developed repeatedly and independently by coincidence. Moreover, the structural element preceding the EAL domain has been demonstrated to be a structurally well-defined helical linker in GGDEF-EAL couples [60] and also an important EAL-regulatory element in BLUF-EAL systems [61].

NsPadC-EAL⁻⁷ and *BbPadC-EAL⁻⁷* further attract attention by clustering with the significant branch of PadCs⁻⁷. Despite the loss of a highly conserved NTS motif, biliverdin binding and light regulation of DGC activity do not appear to be impeded in this group. In contrast, the EAL motif itself features disruptive mutations to EAF, in addition to mutations in many motifs listed as essential for diguanylate phosphodiesterase activity (e.g. the DDFGTG motif) [34], and we expect neither homolog to feature any EAL-PDE activity. In fact, it has been shown that in proteins featuring a tandem GGDEF-EAL effector module, often one of the two output domains is inactive [22, 34]; however, the catalytically incompetent domain frequently still affects enzymatic activity of the functional output module. Since EAL domains usually require dimerisation for efficient catalysis [34, 62], oligomerisation events of the secondary output domain might influence the conformational sampling of the GGDEF domains in PadC-EALs⁻⁷.

The majority of phylogenetic relationships observed for PHYLP-generated distance trees of PadCs is also reflected in sequence similarity-based networks, including the assignment of atypical linker lengths within a branch (*MmPadC⁻⁷*, *HdPadC⁻⁴*). Use of bitscores (S) rather than E-values means query length does not affect the output value, which allows for comparison of PadCs and PadC-EALs in the same network. Relationships visualised in SSNs have been shown to correlate well with known functional relationships [63], and we confirm that the linker length representation in a sequence similarity network of PadCs and PadC-EALs largely concurs with phylogenetic as well as biochemical

observations. Sequence similarity may not permit unambiguous statements about structure or function, however, the visualised relationships detailed in this paper correlate well with known functional relationships between PadC linker length and enzymatic behaviour. Here, we demonstrated that biochemical behaviour of PadCs cannot be transferred par for par to phylogenetic relationships; however, a strong correlation—especially with regard to linker length and the presumably coevolved PHY interface—is clearly observable.

In summary, we were able to show that the high level of conservation and strongly biased evolutionary selection observed for the linker length in phytochrome-activated diguanylate cyclases corresponds prominently to functional characteristics. The rare occurrence of linker length variation within PadC subfamilies only to lengths observed in other subfamilies further highlights the importance of the coiled-coil linker conformation for biochemical function. Successful coiled-coil formation and its effect on orientational sampling of the effector domains affects not only efficiency of signal integration but also spectral properties of the photoreceptor and ultimately regulation of the cellular levels of an important bacterial second messenger. This importance of linker length and sequence on signal integration has also been observed in the past for many naturally occurring systems as well as artificial designs based on bacteriophytochromes for the generation of novel optogenetic tools [4, 64]. Purely sequence-based functional annotation of PadCs, however, remains a challenge. Though doubtlessly the most challenging aspect to automate, the importance of biochemical characterisation in understanding sequence–structure–function space cannot be overstated.

4 Materials and methods

4.1 Protein preparation, expression and purification

The coding sequences for all homologs (corresponding NCBI accession numbers are listed in Supp. Table 1) were subjected to *Escherichia coli* codon optimisation and synthesised by GeneArt (Life Technologies). The ordered gene strings were cloned into pETM-11 vectors using the NEBuilder[®] HiFi DNA Assembly Cloning Kit.

Homologs were expressed and purified as described previously [5]. Briefly, *E. coli* BL21 (DE3) cells containing an additional pT7-ho1 helper plasmid coding for heme oxygenase (HO-1) were used to permit biliverdin synthesis for holo-protein expression. After transformation with the respective PadC-(EAL) plasmids, cells were cultivated in LB medium at 37 °C until OD₆₀₀ 0.45, then cooled to 16 °C. 10 mg L⁻¹ δ-aminolevulinic acid were supplied, and expression induced with 0.25 mM isopropyl-β-D-thiogalactopyranoside

(IPTG) at an OD_{600} of ~ 0.7 . Harvested cells were treated with $100 \mu\text{g mL}^{-1}$ lysozyme and protease inhibitor cocktail (Roche cOmplete ULTRA tablets) before sonication (5×5 min, 70%, pulsed mode; Labsonic L, Sartorius). The soluble fraction was separated by ultracentrifugation (206,000 rcf, 45 min).

Using a Ni^{2+} -Sepharose matrix (Ni Sepharose 6 Fast Flow, GE Healthcare), soluble proteins were purified by affinity chromatography. Proteins were eluted (lysis buffer with 250 mM imidazole) following a washing step (lysis buffer with 30 mM imidazole). During overnight dialysis at 4°C in dialysis buffer (50 mM HEPES pH 7.0, 500 mM NaCl, 2 mM MgCl_2 , 20 mM imidazole) His-tags were cleaved with histidine-tagged tobacco etch virus (TEV) protease ($\sim 1:16$ TEV:substrate). The TEV protease as well as the cleaved tag were removed during another step of affinity chromatography, where the proteins in the flowthrough were concentrated (using ultra centrifugal filters—Amicon Ultra-15; Merck Millipore) and subsequently loaded onto a size exclusion chromatography column (HiLoad 16/60 Superdex 200 prep grade; GE Healthcare) equilibrated in storage buffer (10 mM Hepes pH 7.0, 500 mM NaCl, 2 mM MgCl_2). Fractions were further concentrated and aliquots flash-frozen in liquid nitrogen to be stored at -80°C .

4.2 Spectroscopic characterisation

UV–Vis absorption spectra of $\sim 2 \mu\text{M}$ samples (diluted in storage buffer) were acquired using a Specord 200 Plus spectrophotometer (Analytic Jena) at 20°C . Dark-state P_r spectra were measured under non-actinic conditions, minimising contact with measuring light. Light state spectra were recorded after 1 min of red light illumination (660 nm, $20.9 \mu\text{W}/\text{mm}^2$, Thorlabs).

P_r state recovery kinetics were measured for wavelength maxima and minima of light–dark difference spectra obtained after 1 min of red light illumination (Specord 200 Plus spectrophotometer (Analytic Jena) with 10 ms integration time). As measuring light can affect P_r/P_{fr} steady states, estimated recovery times should not be over-interpreted quantitatively.

4.3 Kinetic analysis

High-performance liquid chromatography (HPLC) was used to monitor turnover of GTP to c-di-GMP as described previously ([8], adapted from [5]). $400 \mu\text{M}$ GTP were added to purified protein diluted in reaction buffer (10 mM Hepes pH 7.0, 500 mM NaCl, 50 mM MgCl_2), yielding a final GTP concentration of $200 \mu\text{M}$, after a 1 min incubation at 20°C under non-actinic light or constant red light illumination (660 nm, $20.9 \mu\text{W}/\text{mm}^2$, Thorlabs). After incubation at four different time points under non-actinic or constant red light

conditions, samples were thermally inactivated by 1 min incubation at $\sim 90^\circ\text{C}$. Following a centrifugation step, nucleotides were separated by reversed phase HPLC (ProntoSil C18 ace-EPS column, Bischoff) using an aqueous mobile phase (10 mM K_2HPO_4 pH 8.0, 5% MeOH, 5 min) under isocratic conditions at 35°C . All activities were normalised to the dimer concentration of the respective constructs.

4.4 Phylogenetic analysis

Following an iterative PSI-blast search [37], sequences matching the requirements outlined in the results section, paragraph 2.1, were collected from the NCBI website [65] and compiled according to their official status on 26th January 2021 (accession numbers are listed in Supp. Table 1). Alignments were performed using the T-COFFEE Multiple Alignment Sequence Server (version 11.00) [66, 67], including structural information with the dark state structure for *IsPadC* (advanced: *t_coffee -in = file.fa -mode = expresso -blast = LOCAL -pdb = 5llw.pdb -evaluate_mode = t_coffee_slow -output = score_html clustalw_aln fasta_aln score_ascii phylip -maxnseq = 150 -maxlen = 2500 -case = upper -seqnos = off -outorder = input -run_name = result -multi_core = 4 -pdb_db = /db/pdb/derived_data_format/blast/latest/pdb_seqs.fa -protein_db = /db/ncbi/201511/blast/db/nr.fa -quiet = stdout*). The alignment was manually curated in Jalview [68] and reduced for the creation of the final tree, which is included in the paper, to four representatives per branch, excluding outliers; with sequences annotated as “hypothetical” excluded. Individual sequences were chosen based on maximum diversity and distance. Distance as well as bootstrap trees of the curated alignment were created using the PHYLIP 3.6 package (SEQBOOT, PROTDIST, FITCH and CONSENSE, [38]). 1000 Jackknife sub-alignments were created using SEQBOOT and subjected to a bootstrapped protein distance analysis. The fungal phytochrome FphA from *Aspergillus nidulans* was used as an outgroup and trees were visualised using Figtree [69]. After re-evaluation of the compiled sequences on 15th November 2021, the entries listed in Supp. Table 5 were removed from the complete collection of PadCs, since they had been re-annotated as obsolete in the meantime (i.e. “This protein record was suppressed because it is no longer annotated on any genome”). For phylogenetic work as well as sequence similarity networks, the sequence of WP_114419467 (*PsPadC*) was substituted with the current version WP_119515934; however, the sequence of the biochemically characterised construct correlates with the obsolete entry. KQM84620 (*AsPadC*), now listed as “hypothetical”, was left in the alignment as no surrogate of the same length was found.

For comparison of bacteriophytochromes with different effectors, InterPro was searched for entries with a PAS

(IPR013654)—GAF (IPR003018)—PHY (IPR013515) architecture. From all domain structures with ≥ 5 hits on 7th April 2021, those with organisms present in more than one architecture family were chosen. A list of representative sequences as depicted in Fig. 2 can be found in Supp. Table 6; alignments were generated with standard T-COFFEE settings and distance trees generated using the PHYLIP package.

4.5 Sequence similarity networks

From the complete set of PadC and PadC-EAL sequences compiled as described above (see Supp. Table 3), databases were created using the *makeblastdb* option of NCBI BLAST+ [65], followed by an all-vs-all BLAST. Networks visualised in Cytoscape 3.8.2 [70] represent BLAST-generated bitscores (S) as edges, which were restricted to values higher than 380. Of a given pair of bitscores (S) for two sequences, only the higher one was kept, as described in [63]

4.6 AlphaFold2

Structure predictions were computed using AlphaFold v2.1.0 [45] in a Jupyter notebook as provided by [71]; with default settings, and sequences and dimeric configuration provided as input. The results with the best overall PAE score (Supp. Fig. 8) are presented in this paper, pLDDT values for which are depicted in Supp. Fig. 7.

4.7 Analysis of coevolution

Coevolution of PadCs was addressed using the Gremlin server of the Baker lab (<http://gremlin.bakerlab.org/>) [72]. An alignment of all ~200 PadC and PadC-EAL sequences (provided in Supp. Table 3) featuring the PSM-GGDEF parts and *IsPadC* as reference was not sufficient for identification of critical coevolving residues. Therefore, the alignment was extended by the HHblits algorithm using 8 iterations and an *E*-value cut-off of 10^{-4} . While still somewhat noisy, the coevolution pattern clearly supports the functional role of previously known phytochrome and GGDEF elements.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43630-022-00255-7>.

Acknowledgements We are grateful to M. Totaro for local computational expertise and support, to E. Zenzmaier for technical support and to N. Galler for help with protein expression and purification. We further wish to thank P. Macheroux and G. Oberdorfer (Institute of Biochemistry, Graz University of Technology) for stimulating discussions. The authors gratefully acknowledge support from NAWI Graz.

This research was supported by the Austrian Science Fund (FWF): P32022 (AW).

Funding Open access funding provided by Austrian Science Fund (FWF).

Declarations

Conflict of interest There are no competing interests related to this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References


- Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO Journal*, 5, 823–826.
- Whisstock, J. C., & Lesk, A. M. (2003). Prediction of protein function from protein sequence and structure. *Quarterly Reviews of Biophysics*, 36, 307–340. <https://doi.org/10.1017/S0033583503003901>
- Fraikin, G. Y., Strakhovskaya, M. G., & Rubin, A. B. (2013). Biological photoreceptors of light-dependent regulatory processes. *Biochemistry (Moscow)*, 78, 1238–1253. <https://doi.org/10.1134/S0006297913110047>
- Gourinchas, G., Ettl, S., & Winkler, A. (2019). Bacteriophytochromes—From informative model systems of phytochrome function to powerful tools in cell biology. *Current Opinion in Structural Biology*, 57, 72–83. <https://doi.org/10.1016/j.sbi.2019.02.005>
- Gourinchas, G., Ettl, S., Göbl, C., Vide, U., Madl, T., & Winkler, A. (2017). Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Science Advances*, 3, e1602498. <https://doi.org/10.1126/sciadv.1602498>
- Gourinchas, G., Heintz, U., & Winkler, A. (2018). Asymmetric activation mechanism of a homodimeric red light-regulated photoreceptor. *eLife*, 7, e34815. <https://doi.org/10.7554/eLife.34815>
- Gourinchas, G., Vide, U., & Winkler, A. (2019). Influence of the N-terminal segment and the PHY-tongue element on light-regulation in bacteriophytochromes. *Journal of Biological Chemistry*, 294, 4498–4510. <https://doi.org/10.1074/jbc.RA118.007260>
- Böhm, C., Todorović, N., Balasso, M., Gourinchas, G., & Winkler, A. (2021). The PHY domain dimer interface of bacteriophytochromes mediates cross-talk between photosensory modules and output domains. *Journal of Molecular Biology*, 433, 167092. <https://doi.org/10.1016/j.jmb.2021.167092>
- Rockwell, N. C., Su, Y. S., & Lagarias, J. C. (2006). Phytochrome structure and signaling mechanisms. *Annual Review of Plant Biology*, 57, 837–858. <https://doi.org/10.1146/annurev.arplant.56.032604.144208>

10. 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 of the United States of America*, 105, 14709–14714. <https://doi.org/10.1073/pnas.0806477105>
11. Ryjenkov, D. A., Tarutina, M., Moskvina, O. V., & Gomelsky, M. (2005). Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. *Journal of Bacteriology*, 187, 1792–1798. <https://doi.org/10.1128/JB.187.5.1792-1798.2005>
12. Römling, U., Galperin, M. Y., & Gomelsky, M. (2013). Cyclic di-GMP: The first 25 years of a universal bacterial second messenger. *Microbiology and Molecular Biology Reviews*, 77, 1–52. <https://doi.org/10.1128/mmr.00043-12>
13. Jenal, U., Reinders, A., & Lori, C. (2017). Cyclic di-GMP: Second messenger extraordinaire. *Nature Reviews Microbiology*, 15, 271–284. <https://doi.org/10.1038/nrmicro.2016.190>
14. Tarutina, M., Ryjenkov, D. A., & Gomelsky, M. (2006). An unorthodox bacteriophytochrome from *Rhodospirillum rubrum* involved in turnover of the second messenger c-di-GMP. *Journal of Biological Chemistry*, 281, 34751–34758. <https://doi.org/10.1074/jbc.M604819200>
15. Tal, R., Wong, H. C., Calhoun, R., Gelfand, D., Fear, A. L., Volman, G., et al. (1998). Three cdg operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: Genetic organization and occurrence of conserved domains in isoenzymes. *Journal of Bacteriology*, 180, 4416–4425. <https://doi.org/10.1128/jb.180.17.4416-4425.1998>
16. Levet-Paulo, M., Lazzaroni, J. C., Gilbert, C., Atlan, D., Doublet, P., & Vianney, A. (2011). The atypical two-component sensor kinase Lpl0330 from *Legionella pneumophila* controls the bifunctional diguanylate cyclase-phosphodiesterase Lpl0329 to modulate Bis-(3'-5')-cyclic dimeric GMP synthesis. *Journal of Biological Chemistry*, 286, 31136–31144. <https://doi.org/10.1074/jbc.M111.231340>
17. Bharati, B. K., Sharma, I. M., Kasetty, S., Kumar, M., Mukherjee, R., & Chatterji, D. (2012). A full-length bifunctional protein involved in c-di-GMP turnover is required for long-term survival under nutrient starvation in *Mycobacterium smegmatis*. *Microbiology (United Kingdom)*, 158, 1415–1427. <https://doi.org/10.1099/mic.0.053892-0>
18. Simm, R., Morr, M., Kader, A., Nimtz, M., & Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Molecular Microbiology*, 53, 1123–1134. <https://doi.org/10.1111/j.1365-2958.2004.04206.x>
19. Petchiappan, A., Naik, S. Y., & Chatterji, D. (2020). Tracking the homeostasis of second messenger cyclic-di-GMP in bacteria. *Biophysical Reviews*, 12, 719–730. <https://doi.org/10.1007/s12551-020-00636-1>
20. Römling, U., Gomelsky, M., & Galperin, M. Y. (2005). C-di-GMP: The dawning of a novel bacterial signalling system. *Molecular Microbiology*, 57, 629–639. <https://doi.org/10.1111/j.1365-2958.2005.04697.x>
21. Goldman, S. R., Sharp, J. S., Vvedenskaya, I. O., Livny, J., Dove, S. L., & Nickels, B. E. (2011). NanoRNAs prime transcription initiation in vivo. *Molecular Cell*, 42, 817–825. <https://doi.org/10.1016/j.molcel.2011.06.005>
22. Christen, M., Christen, B., Folcher, M., Schauerte, A., & Jenal, U. (2005). Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *Journal of Biological Chemistry*, 280, 30829–30837. <https://doi.org/10.1074/jbc.M504429200>
23. Ahuja, L. G., Taylor, S. S., & Kornev, A. P. (2019). Tuning the “Violin” of protein kinases: The role of dynamics-based allostery. *IUBMB Life*, 71, 685–696. <https://doi.org/10.1002/iub.2057>
24. Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H., & Jenal, U. (2007). Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *Journal of Biological Chemistry*, 282, 29170–29177. <https://doi.org/10.1074/jbc.M704702200>
25. Wagner, J. R., Zhang, J., Brunzelle, J. S., Vierstra, R. D., & Forest, K. T. (2007). High resolution structure of *Deinococcus bacteriophytochrome* yields new insights into phytochrome architecture and evolution. *Journal of Biological Chemistry*, 282, 12298–12309. <https://doi.org/10.1074/jbc.M611824200>
26. Lamparter, T., Michael, N., Caspani, O., Miyata, T., Shirai, K., & Inomata, K. (2003). Biliverdin binds covalently to *Agrobacterium phytochrome* Agp1 via its ring a vinyl side chain. *Journal of Biological Chemistry*, 278, 33786–33792. <https://doi.org/10.1074/jbc.M305563200>
27. Takala, H., Björling, A., Berntsson, O., Lehtivuori, H., Niebling, S., Hoernke, M., et al. (2014). Signal amplification and transduction in phytochrome photosensors. *Nature*, 509, 245–248. <https://doi.org/10.1038/nature13310>
28. Moutevelis, E., & Woolfson, D. N. (2009). A periodic table of coiled-coil protein structures. *Journal of Molecular Biology*, 385, 726–732. <https://doi.org/10.1016/j.jmb.2008.11.028>
29. Smets, B. F., & Barkay, T. (2005). Horizontal gene transfer: Perspectives at a crossroads of scientific disciplines. *Nature Reviews Microbiology*, 3, 675–678. <https://doi.org/10.1038/nrmicro1253>
30. 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. <https://doi.org/10.1038/nature04118>
31. Lamparter, T., Carrascal, M., Michael, N., Martinez, E., Rottwinkel, G., & Abian, J. (2004). The biliverdin chromophore binds covalently to a conserved cysteine residue in the N-terminus of *Agrobacterium phytochrome* Agp1. *Biochemistry*, 43, 3659–3669. <https://doi.org/10.1021/bi035693i>
32. Burgie, E. S., & Vierstra, R. D. (2014). Phytochromes: An atomic perspective on photoactivation and signaling. *The Plant Cell*, 26, 4568–4583. <https://doi.org/10.1105/tpc.114.131623>
33. Schirmer, T. (2016). C-di-GMP synthesis: Structural aspects of evolution, catalysis and regulation. *Journal of Molecular Biology*, 428, 3683–3701. <https://doi.org/10.1016/j.jmb.2016.07.023>
34. Schmidt, A. J., Ryjenkov, D. A., & Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *Journal of Bacteriology*, 187, 4774–4781. <https://doi.org/10.1128/JB.187.14.4774-4781.2005>
35. Minasov, G., Padavattan, S., Shuvalova, L., Brunzelle, J. S., Miller, D. J., Baslé, A., et al. (2009). Crystal structures of Ykull and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. *Journal of Biological Chemistry*, 284, 13174–13184. <https://doi.org/10.1074/jbc.M808221200>
36. (2004) NCBI Protein [Internet]. Bethesda Natl. Libr. Med. (US), Natl. Cent. Biotechnol. Inf.
37. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
38. Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) version 3.6*. Distributed by the author. University of Washington, Department of Genome Sciences.
39. Blumenstein, A., Vienken, K., Tasler, R., Purschwitz, J., Veith, D., Frankenberg-Dinkel, N., et al. (2005). The *Aspergillus nidulans*

- phytochrome FphA represses sexual development in red light. *Current Biology*, 15, 1833–1838. <https://doi.org/10.1016/j.cub.2005.08.061>
40. Brandt, S., Von Stetten, D., Günther, M., Hildebrandt, P., & Frankenberg-Dinkel, N. (2008). The fungal phytochrome FphA from *Aspergillus nidulans*. *Journal of Biological Chemistry*, 283, 34605–34614. <https://doi.org/10.1074/jbc.M805506200>
 41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
 42. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 1–9. <https://doi.org/10.1186/1471-2105-10-421>
 43. Hartmann, M. D., Mendler, C. T., Bassler, J., Karamichali, I., Ridderbusch, O., Lupas, A. N., et al. (2016). A/B coiled coils. *eLife*. <https://doi.org/10.7554/eLife.11861>
 44. Buhrke, D., Gourinchas, G., Müller, M., Michael, N., Hildebrandt, P., & Winkler, A. (2020). Distinct chromophore–protein environments enable asymmetric activation of a bacteriophytochrome-activated diguanylate cyclase. *Journal of Biological Chemistry*, 295, 539–551. <https://doi.org/10.1074/jbc.RA119.011915>
 45. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
 46. Otero, L. H., Foscardi, S., Antelo, G. T., Rosano, G. L., Sirigu, S., Klinke, S., et al. (2021). Structural basis for the Pr-Pfr long-range signaling mechanism of a full-length bacterial phytochrome at the atomic level. *Science Advances*. <https://doi.org/10.1126/sciadv.abh1097>
 47. Burgie, E. S., Wang, T., Bussell, A. N., Walker, J. M., Li, H., & Vierstra, R. D. (2014). Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion. *Journal of Biological Chemistry*, 289, 24573–24587. <https://doi.org/10.1074/jbc.M114.571661>
 48. Björling, A., Berntsson, O., Lehtivuori, H., Takala, H., Hughes, A. J., Panman, M., et al. (2016). Structural photoactivation of a full-length bacterial phytochrome. *Science Advances*, 2, e1600920. <https://doi.org/10.1126/sciadv.1600920>
 49. Blain-Hartung, M., Rockwell, N. C., & Lagarias, J. C. (2021). Natural diversity provides a broad spectrum of cyanobacteriochrome-based diguanylate cyclases. *Plant Physiology*, 187, 632–645. <https://doi.org/10.1093/plphys/kiab240>
 50. Multamäki, E., Nanekar, R., Morozov, D., Lievonen, T., Golonka, D., Wahlgren, W. Y., et al. (2021). Comparative analysis of two paradigm bacteriophytochromes reveals opposite functionalities in two-component signaling. *Nature Communications*, 12, 4394. <https://doi.org/10.1038/s41467-021-24676-7>
 51. Teixeira, R. D., Holzschuh, F., & Schirmer, T. (2021). Activation mechanism of a small prototypic Rec-GGDEF diguanylate cyclase. *Nature Communications*, 12, 1–15. <https://doi.org/10.1038/s41467-021-22492-7>
 52. Brown, J. H., Cohen, C., & Parry, D. A. D. (1996). Heptad breaks in α -helical coiled coils: Stutters and stammers. *Proteins: Structure Function and Genetics*, 26, 134–145. [https://doi.org/10.1002/\(SICI\)1097-0134\(199610\)26:2%3c134::AID-PROT3%3e3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0134(199610)26:2%3c134::AID-PROT3%3e3.0.CO;2-G)
 53. Lamparter, T., Krauß, N., & Scheerer, P. (2017). Phytochromes from *Agrobacterium fabrum*. *Photochemistry and Photobiology*, 93, 642–655. <https://doi.org/10.1111/php.12761>
 54. Giraud, E., Zappa, S., Vuillet, L., Adriano, J. M., Hannibal, L., Fardoux, J., et al. (2005). A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodospseudomonas palustris*. *Journal of Biological Chemistry*, 280, 32389–32397. <https://doi.org/10.1074/jbc.M506890200>
 55. Glantz, S. T., Carpenter, E. J., Melkonian, M., Gardner, K. H., Boyden, E. S., Wong, G. K. S., et al. (2016). Functional and topological diversity of LOV domain photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E1442–E1451. <https://doi.org/10.1073/pnas.1509428113>
 56. Fixen, K. R., Baker, A. W., Stojkovic, E. A., Beatty, J. T., & Harwood, C. S. (2014). Apo-bacteriophytochromes modulate bacterial photosynthesis in response to low light. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1322410111>
 57. Ganfornina, M. D., & Sánchez, D. (1999). Generation of evolutionary novelty by functional shift. *BioEssays*, 21, 432–439. [https://doi.org/10.1002/\(SICI\)1521-1878\(199905\)21:5%3c432::AID-BIES10%3e3.0.CO;2-T](https://doi.org/10.1002/(SICI)1521-1878(199905)21:5%3c432::AID-BIES10%3e3.0.CO;2-T)
 58. Mantoni, F., Rossi, C. S., Paiardini, A., Di Matteo, A., Cappellacci, L., Petrelli, R., et al. (2021). Studying ggdef domain in the act: Minimize conformational frustration to prevent artefacts. *Life*, 11, 1–13. <https://doi.org/10.3390/life11010031>
 59. Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., et al. (1987). Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature*, 325, 279–281. <https://doi.org/10.1038/325279a0>
 60. Liu, C., Liew, C. W., Wong, Y. H., Tan, S. T., Poh, W. H., Manimekalai, M. S. S., et al. (2018). Insights into biofilm dispersal regulation from the crystal structure of the PAS-GGDEF-EAL region of RbdA from *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 200, 1–19. <https://doi.org/10.1128/JB.00515-17>
 61. Winkler, A., Udvarhelyi, A., Hartmann, E., Reinstein, J., Menzel, A., Shoeman, R. L., et al. (2014). Characterization of elements involved in allosteric light regulation of phosphodiesterase activity by comparison of different functional BlrP1 states. *Journal of Molecular Biology*, 426, 853–868. <https://doi.org/10.1016/j.jmb.2013.11.018>
 62. Sundriyal, A., Massa, C., Samoray, D., Zehender, F., Sharpe, T., Jenal, U., et al. (2014). Inherent regulation of EAL domain-catalyzed hydrolysis of second messenger cyclic di-GMP. *Journal of Biological Chemistry*, 289, 6978–6990. <https://doi.org/10.1074/jbc.M113.516195>
 63. Atkinson, H. J., Morris, J. H., Ferrin, T. E., & Babbitt, P. C. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE*, 4, e4345. <https://doi.org/10.1371/journal.pone.0004345>
 64. Takala, H., Edlund, P., Ihalainen, J. A., & Westenhoff, S. (2020). Tips and turns of bacteriophytochrome photoactivation. *Photochemical and Photobiological Sciences*, 19, 1488–1510. <https://doi.org/10.1039/d0pp00117a>
 65. Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., et al. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 46, D8–13. <https://doi.org/10.1093/nar/gkx1095>
 66. Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302, 205–217. <https://doi.org/10.1006/jmbi.2000.4042>
 67. Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J. M., et al. (2011). T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39, 13–17. <https://doi.org/10.1093/nar/gkr245>
 68. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>

69. Rambaut, A. (2018) FigTree v1.4.4. <http://tree.bio.ed.ac.uk/software/figtree/>
70. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498–2504. <https://doi.org/10.1101/gr.1239303.metabolite>
71. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
72. Ovchinnikov, S., Kamisetty, H., & Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife*, 2014, 1–21. <https://doi.org/10.7554/eLife.02030>

Authors and Affiliations

Cornelia Böhm^{1,2} · Geoffrey Gourinchas^{1,3} · Sophie Zweytick¹ · Elvira Hujdur¹ · Martina Reiter¹ · Sara Trstenjak¹ · Christoph Wilhelm Sensen^{2,4} · Andreas Winkler^{1,2} 

✉ Andreas Winkler
Andreas.Winkler@TUGraz.at

¹ Institute of Biochemistry, Graz University of Technology,
8010 Graz, Austria

² BioTechMed-Graz, 8010 Graz, Austria

³ Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC),
67404 Illkirch, France

⁴ Hungarian Centre of Excellence for Molecular Medicine,
Római körút 21, 6723 Szeged, Hungary