B_{12}-based photoreceptors: from structure and function to applications in optogenetics and synthetic biology

S. Padmanabhan[a], Ricardo Pérez-Castaño[b], and Montserrat Elías-Arnanz[b]*

[a]Instituto de Química Física "Rocasolano", Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain; email: padhu@iqfr.csic.es

[b]Departamento de Genética y Microbiología, Área de Genética (Unidad Asociada al Instituto de Química Física "Rocasolano", CSIC), Facultad de Biología, Universidad de Murcia, Murcia 30100, Spain; email: ricardo.perez@um.es; melias@um.es

Shortened Running Title: B_{12}-based Photoreceptors

Key Words: CarH; optogenetics; cobalamin; photochemistry; synthetic biology; photoregulation

Corresponding Author Contact Information:

Montserrat Elías-Arnanz

Departamento de Genética y Microbiología, Área de Genética (Unidad Asociada al Instituto de Química Física "Rocasolano", CSIC),

Facultad de Biología,

Universidad de Murcia, Murcia 30100, Spain

E-mail: melias@um.es

Telephone: (34) 868887134; FAX: (34) 868883693

Abstract

Vitamin B_{12}-based photoreceptor proteins sense ultraviolet (UV), blue or green light using 5′-deoxyadenosylcobalamin (AdoCbl). The prototype of this widespread bacterial photoreceptor family, CarH, controls light-dependent gene expression in photoprotective cellular responses. It represses transcription in the dark by binding to operator DNA as an AdoCbl-bound tetramer, whose disruption by light relieves operator binding to allow transcription. Structures of the "dark" (free and DNA-bound) and "light" CarH states and studies on the unusual AdoCbl photochemistry have provided fundamental insights into these photoreceptors. We highlight these, the plasticity within a conserved mode of action among CarH homologs, their distribution, and their promising applications in optogenetics and synthetic biology.

Introduction

Light, a crucial energy source and ubiquitous environmental factor, determines various biological processes ranging from photosynthesis, vision and circadian rhythms to mechanisms limiting photooxidative cellular damage [1-8]. Specialized proteins called photoreceptors enable light sensing to evoke the appropriate cellular response. Those sensing visible light (380-760 nm range of the electromagnetic spectrum) use covalently or noncovalently bound components called chromophores, based on which there are about ten photoreceptor classes [6,9-14]. Generally, the "dark" state chromophore absorbs light of given wavelength(s) and undergoes changes in conformation and chemistry to a "light" state. These changes are transmitted to the photosensory "input" protein domain housing the chromophore, and then on to an effector or "output" module in the same or a different protein molecule, whose defined action (DNA binding, enzyme activity, etc.) determines the final output. A recurrent feature of photoreceptors is a modular architecture, which allows combinatorial domain mixing for the evolution of signaling and regulation complexity [15]. It also enables use of the photosensory module in optogenetics and synthetic biology. We survey these aspects of CarH photoreceptors, whose light-sensing chromophore is the vitamin B_{12} form AdoCbl, and highlight recent advances in our knowledge of these proteins, their distribution, functions, mode of action, and applications.

CarH photoreceptors: functions and mechanism of action

CarH photoreceptors are widely distributed cytoplasmic bacterial transcription factors that sense UV (UV-A, 315-400 nm; and likely UV-B, 280-315 nm), blue (400-500 nm) and green (500-570 nm) but not red (>600 nm) light. The discovery that CarH senses light using AdoCbl as the chromophore unveiled both a new photoreceptor family and a novel functional facet of this vitamin [13•,16••,17••]. CarH regulates light-dependent

expression usually of its own gene and of those for the synthesis of carotenoids, which protect cells against photooxidative damage by quenching singlet oxygen and other reactive oxygen species [8,13•,17••-21]. In some bacteria, CarH also controls expression of photolyase, which repairs UV-induced pyrimidine dimers in DNA, or of a general redox response regulator [17••,19,20,22••]. These functions were identified by studying CarH from the Gram-negative bacteria *Myxococcus xanthus* and *Thermus thermophilus*, and from the Gram-positive *Bacillus megaterium*. Studies of CarH homologs in other bacteria may conceivably uncover new functions.

The AdoCbl CarH chromophore is a complex organometallic molecule. Its main feature is a corrin ring (a cyclic tetrapyrrole with a central hexacoordinated cobalt atom in the Co^{3+} oxidation state) covalently ligated to additional groups, including a 5,6-dimethylbenzimidazole (DMB) moiety (Figure 1a). The lower (α) axial Co^{3+} ligand is DMB, and the upper (β) one, linked via a highly photosensitive Co–C bond, is a 5′-deoxyadenosyl (Ado) moiety [13•,23-26]. In many B_{12}-binding proteins including CarH, the imidazole sidechain of a His in a D/ExHx_2G motif (x=any amino acid) displaces DMB as the lower Co^{3+} ligand, in the so-called "base-off/His-on" B_{12}-binding mode [16••,27-32]. AdoCbl biosynthesis is complex and confined to some microorganisms; other organisms (*e.g.* human) depend on exogenous B_{12} sources, while plants, fungi and many prokaryotes do without it entirely [33]. In enzymes, the AdoCbl Co–C bond undergoes reversible, homolytic cleavage to yield a transient, highly reactive Ado radical that can initiate diverse radical-based rearrangement reactions [26-29,31,32]. However, CarH, which lacks enzymatic activity, exploits AdoCbl Co–C bond photosensitivity to sense light and regulate transcription [13•,16••,17••].

The molecular mechanism underlying light-dependent CarH activity is, essentially, the interplay between AdoCbl and light to modulate CarH oligomerization and thereby

operator DNA binding [17••]. In the dark, an AdoCbl-bound CarH tetramer binds to an operator that overlaps with the target gene promoter thwarting access to RNA polymerase (RNAP) and transcription. Light cleaves the Co–C bond and frees the Ado group, which provokes tetramer disassembly and loss of operator binding thus allowing RNAP to access the promoter and initiate transcription (Figure 1b; [13•,16••-18,20,22,34,35]). The repressor activity of CarH homologs studied thus far requires their binding to operator DNA as tetramers, whose protomers are each bound to an intact AdoCbl molecule. The cofactor-free apo and light-exposed holo forms do not bind operator, and their oligomeric states can vary among homologs (Figure 1b). For instance, both forms of *T. thermophilus* or *M. xanthus* CarH are monomeric [16••,17••,34], whereas *B. megaterium* apoCarH is a loosely-folded, molten globule tetramer and its light-exposed holo form is a dimer (Figure 1b; [20,22••]). The molecular basis for this plasticity in oligomerization remains to be resolved.

Besides oligomerization, plasticity also characterizes CarH operator design [22••]. Known CarH operators span three or four tandem 11-bp direct repeats (DRs) (Figure 1b; [13•,16••,22••]), differing notably from MerR proteins whose similar DNA-binding domains (DBDs) recognize smaller, inverted-repeat operators [36]. CarH operators overlap with the -35 and/or -10 promoter elements recognized by RNAP associated with the primary σ factor [13•,16••,20,22••,35]. Consequently, CarH can repress transcription by sequestering either or both of these promoter elements.

AdoCbl photolysis releases the Ado group and leaves the remaining cobalamin (Cbl) molecule bound to CarH. The chromophore is therefore irreversibly altered in CarH, unlike in most other photoreceptors where the light-induced molecular changes in the chromophore and transitions between the "dark" and “light” states (the photocycle) are reversible. Recovery of the “dark” CarH repressor therefore requires exchange of Cbl with

intact AdoCbl (if Cbl binding is labile), or *in situ* regeneration of AdoCbl through unknown cellular mechanisms (see below). Also, because light usually relieves *carH* autorepression, the freshly generated holoprotein can replace photolyzed CarH that turns over.

CarH photoreceptor structures and photochemistry

Fundamental insights into CarH modular architecture, oligomerization, DNA binding and light-dependent conformational changes emerged from the crystal structures of the *T. thermophilus* CarH tetramer, free or bound to DNA in the dark, and of its light-inactivated form [16••]. Each CarH protomer has two structured modules, an N-terminal "effector/ output" DBD and a C-terminal, AdoCbl-binding "photosensor/input" domain (CBD), connected by a flexible, disordered linker (Figure 2a). The DBD, similar to the winged-helix domain of bacterial MerR-type transcription factors as noted before [16••], contains many highly conserved residues among CarH homologs, suggesting a common DNA recognition mode (Figure 2a; [13•]). AdoCbl is sandwiched between a four-helix bundle and a Rossmann fold subdomain in the photosensory CBD (Figure 2a; [16••]). This CBD structurally mirrors the methionine synthase (MetH) domain for binding methylcobalamin, whose upper axial methyl Co^{3+} ligand is, however, much smaller than Ado [30]. Highly conserved residues occur in both CarH CBD subdomains. A His (H177, Figure 2a) in an $ExHx_2G$ motif in the Rossmann fold supplies the "base-off/His-on" lower axial Co^{3+} ligand (as in MetH); but the specific AdoCbl binding determinant, which is absent in MetH, is a Wx_9EH motif in the four-helix bundle that caps the upper axial Ado (Figure 2a; [16••]). The Trp of the Wx_9EH motif packs against the Ado group resulting in an upright protomer conformation that favors formation of head-to-tail dimers, which readily assemble to a tetramer with the four DBDs splayed on the surface (Figure 2b,c; [16••]). The

consequence is an unusual DNA-binding mode where only three out of the four DBDs contact three tandem 11-bp DRs on one side of operator DNA (Figure 2d; [16••]). The consensus DR sequence is nAnn**T**nn**ACA**n, whose most conserved bases (boldface and underlined; n is any base; Figure 2d) make crucial contacts with the DBD recognition helix, although other less-conserved bases also determine DNA binding. Interestingly, the *B. megaterium* CarH tetramer uses the same DNA binding mode to contact four DRs, presumably using all four DBDs, but can also bind to three DRs revealing a plasticity in operator design and DNA binding [22••]. High-resolution structures are unavailable for this homolog, which conserves nearly all of the critical AdoCbl- and DNA-binding residues of its *T. thermophilus* counterpart and has a very similar secondary structure. Binding to the fourth DR may be facilitated by a longer linker between the photosensory CBD and the effector [22••]. This linker, which could conceivably channel sensor-to-effector allosteric changes, varies in size (4-120 residues) and composition among CarH homologs [13•]. Recent data indicate profound, not fully understood effects of the nature of the linker and the protein context in which it occurs on CarH properties [22••], which have to be further explored.

The light-exposed *T. thermophilus* CarH structure revealed a large shift (>8 Å) of the helix bundle relative to the Rossmann fold due to liberation of Ado from photolyzed AdoCbl (Figure 2e; [16••]). The altered dimer interface provokes tetramer collapse and loss of DNA binding. The "light" monomer retains Cbl tightly in a bis-His linkage to Co^{3+} where the upper axial ligand is a His (H132) adjacent to the Trp of the Wx_9EH motif [16••]. This prevents exchange between the bound Cbl and intact AdoCbl. Many CarH homologs, however, lack the upper axial His, and Cbl can be readily displaced by AdoCbl [13•,16••,22••]. Ultrafast time-resolved spectroscopy [37••] and analysis of photolysis products [38••] indicated that the CarH-bound AdoCbl photochemistry differs from that

established for free or enzyme-bound AdoCbl, where Co–C bond homolysis generates a cob(II)alamin/Ado radical pair [27-29,31,32]. For CarH-bound AdoCbl, formation of the highly reactive Ado radical was proposed to be circumvented by an unprecedented heterolytic Co–C cleavage that is coupled to CarH structural changes occurring over a ns-ms timescale [26,37••]. Moreover, AdoCbl photolysis in CarH generates 4′,5′-anhydro-adenosine, an innocuous product hardly observed on photolysis of enzyme-bound or free AdoCbl [38••]. CarH thus appears to alter AdoCbl photochemistry to use it safely as a chromophore. The molecular mechanism for this remains unresolved but it has been speculated that molecular oxygen and residues around the Ado group, notably Trp, Glu and His of the Wx_9EH motif, may be important [13•,26,37••,38••].

Distribution of CarH photosensory modules

Genome sequences reveal hundreds of CarH homologs widely distributed among diverse bacterial species [13•,17••]. An unrooted phylogenetic tree for a large set of CarH homologs [13•] indicates clustering of those from specific phyla as monophyletic groups, the largest corresponding to Actinobacteria, Bacteroidetes, Betaproteobacteria and Firmicutes, with Deltaproteobacteria, Thermus-Deinococcus and Chloroflexi also well represented (Figure 3a). These phyla also include members lacking CarH, perhaps reflecting differences in environmental niche and lifestyle. Like many other types of photosensory modules, CarH CBD is an autonomous unit [17••] that can occur as a standalone protein without an effector domain, or fused to a sensor domain other than a DBD such as a histidine kinase or a GAF domain (present in cGMP-specific phosphodiesterases, adenylyl cyclases, and the transcriptipon factor FhlA) whose functions and modes of action await elucidation [13•,17••]. Intriguingly, these standalone proteins often coexist with CarH in many species, raising the possibility of

crosstalk between their cellular functions. In an unrooted phylogenetic tree for the standalone proteins, the largest monophyletic groups correspond to Firmicutes and Actinobacteria, and a smaller one to Deltaproteobacteria (Figure 3b). Lack of an effector module suggests that such standalone proteins may exhibit AdoCbl/light-dependent activities based on heterotypic interactions with discrete effector proteins. Interestingly, known CarH interactions are homotypic, although heterotypic ones may very well surface in future studies. Identifying such interactions and elucidating their molecular basis is of functional relevance, and would expand the optogenetics/synthetic biology toolbox based on this photoreceptor module.

CarH applications in optogenetics and synthetic biology

CarH has now been applied as a versatile tool in optogenetics and synthetic biology [39••-42•]. Its CBD can be swapped and fused to a variety of effector domains for light-driven modulation of protein oligomerization and function, and a major asset is its green-light responsiveness, which is underrepresented compared to the available blue and red light repertoire in optogenetics [14,39••,40••,43,44]. Existence of CBD homologs with divergent properties also expands the breadth of potential applications. CarH was first exploited for B_{12}/light-controlled gene expression in *M. xanthus* to assess gene essentiality [45] in a manner that, in principle, can be adapted to other bacteria. It is based on a single plasmid construct that expresses CarH constitutively and the gene of interest (*X*) under the control of the CarH operator-promoter such that, in the presence of AdoCbl, expression of *X* occurs only when bacterial cells are grown in the light (Figure 4a). A system using CarH was recently developed for targeted control of gene expression in eukaryotes, specifically in mammalian and plant cells [39••]. Reportedly, it is the first operative green-light-responsive gene switch for such cells and its design illustrates how CarH can be

repurposed for activation (Figure 4b). The system employs an activator plasmid to constitutively express CarH fused to a transactivator domain (VP16) and a reporter plasmid with multiple CarH operators (eight yielded maximum response) upstream of a minimal eukaryotic promoter to control gene *X* expression. On providing AdoCbl, the CarH-VP16 fusion protein forms tetramers in the dark that bind to the multimeric CarH operator. This enables VP16-mediated recruitment of the transcription initiation complex for RNAP II-dependent expression of *X*, which is abrogated upon exposure to green light. The system exhibits high induction (~350-fold), low leakiness, good reversibility and works in various mammalian cell lines and in *Arabidopsis thaliana* protoplasts. This CarH-based synthetic switch has added green light to the existing set of UV-B, blue, red, and far-red light sensitive systems for mammalian cells, and expanded the plant optogenetic toolbox beyond the red light-inducible switch [39••].

The isolated CarH CBD module has also been used successfully. In one application [40••], it was fused to the intracellular domain of a receptor tyrosine kinase (RTK), murine fibroblast growth factor receptor 1, whose N-terminal ligand binding and transmembrane domains were replaced by a myristoylation membrane anchor (Figure 4c). In cultured human cells supplied with AdoCbl, fusion protein oligomerization in the dark induces RTK-regulated cell signaling, while green light (also blue, but not red) abolishes oligomerization and hence signaling. The optically controlled oligomerization was reversible if, after illumination, CBD-Cbl binding was labile (no upper axial His for bis-His Co^{3+} ligation) and irreversible if tight (upper axial His present). Moreover, this CarH-based system was transplantable to a whole organism: the fusion receptor expressed in zebrafish embryos conferred robust temporal green-light control of signaling during development (Figure 4c; [40••]).

Two recent applications harnessed CarH CBD to SpyTag/SpyCatcher technology [46-48] to produce AdoCbl/light-sensitive protein hydrogels [41••] and layer-by-layer nanofilms [42•]. These water-swollen viscoelastic, 3-D cross-linked polymer networks have the biomimetic and facile release properties desired for biomedical applications such as drug delivery, stem cell therapy and tissue engineering [49,50]. The strategy used (Figure 4d) was to engineer SpyTag and SpyCatcher separately into the termini of two ELP (elastin-like pentapeptide repeat) units bridged by CarH CBD [41••]. Mixing the two purified constructs covalently stitches them into protein polymers (through isopeptide linkage between the SpyTag and its SpyCatcher partner). Interchain interactions mediated by CBD tetramerization in the dark when AdoCbl is present generates hydrogels that can undergo rapid gel-sol transition upon exposure to light. These hydrogels allowed light-controlled encapsulation and release of proteins or cells (fibroblasts and mesenchymal stem cells) [41••]. The procedure has been extended to build layer-by-layer protein nanofilms on gold surfaces [42•] for optically controlled release of protein cargos, with potential for its therapeutic delivery.

Conclusions

Discovery of the large AdoCbl-dependent CarH photoreceptor family and the mechanism of action was followed by high-resolution structure determination of their functionally relevant states and the finding of an unusual photochemistry. CarH has now emerged as an optogenetics and synthetic biology tool for light-regulated gene expression and receptor signaling, and for fabrication of photoresponsive protein hydrogels and nanofilms with potential in drug/cell delivery and tissue engineering. The mode of action of CarH is conserved and yet, even among the few homologs studied, a plasticity is observed whose molecular basis is unresolved. The mechanism behind the unusual photochemistry also

remains enigmatic. The large number of standalone proteins of unknown functions corresponding to only the CarH photosensor domain very likely act via heterotypic interactions with discrete effector proteins, and their study will surely uncover new variations on the theme of B_{12}/light-dependent regulation. Future work can be expected to shed more light on the biology and chemistry of these photoreceptors and to enlarge the toolkit for optogenetics and synthetic biology.

Conflict of interest statement

No conflict of interest.

Acknowledgements

We thank all members of our group and collaborators in the work on B_{12}-based photoreceptors. The authors were funded by grants BFU2015-67968-C2-1-P (to ME-A) and BFU2015-67968-C2-2-P (to SP) from the Ministerio de Economía and Competitividad, Spain and cofinanced by the Fondo Europeo de Desarrollo Regional (FEDER).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hohmann-Marriott MF, Blankenship RE: **Evolution of photosynthesis**. *Annu Rev Plant Biol* 2011, **62**:515-548.

2. Nelson N, Junge W: **Structure and energy transfer in photosystems of oxygenic photosynthesis**. *Annu Rev Biochem* 2015, **84**:659-683.

3. Johnson CH, Zhao C, Xu Y, Mori T: **Timing the day: what makes bacterial clocks tick?** *Nat Rev Microbiol* 2017, **15**:232-242.

4. Palczewski K: **Chemistry and biology of vision**. *J Biol Chem* 2012, **287**:1612-1619.

5. Reddy AB, Rey G: **Metabolic and nontranscriptional circadian clocks: eukaryotes**. *Annu Rev Biochem* 2014, **83**:165-189.

6. Purcell EB, Crosson S: **Photoregulation in prokaryotes**. *Curr Opin Microbiol* 2008, **11**:168-178.

7. Ziegelhoffer EC, Donohue TJ: **Bacterial responses to photo-oxidative stress**. *Nat Rev Microbiol* 2009, **7**:856-863.

8. Elías-Arnanz M, Padmanabhan S, Murillo FJ: **Light-dependent gene regulation in nonphototrophic bacteria**. *Curr Opin Microbiol* 2011, **14**:128-135.

9. Anders K, Essen LO: **The family of phytochrome-like photoreceptors: diverse, complex and multi-colored, but very useful**. *Curr Opin Struct Biol* 2015, **35**:7-16.

10. Govorunova EG, Sineshchekov OA, Li H, Spudich JL: **Microbial Rhodopsins: diversity, mechanisms, and optogenetic applications**. *Annu Rev Biochem* 2017, **86**:845-872.

11. Losi A, Gartner W: **The evolution of flavin-binding photoreceptors: an ancient chromophore serving trendy blue-light sensors**. *Annu Rev Plant Biol* 2012, **63**:49-72.

12. Moglich A, Yang X, Ayers RA, Moffat K: **Structure and function of plant photoreceptors**. *Annu Rev Plant Biol* 2010, **61**:21-47.

13. Padmanabhan S, Jost M, Drennan CL, Elías-Arnanz M: **A new facet of vitamin B_{12}: gene regulation by cobalamin-based photoreceptors**. *Annu Rev Biochem* 2017, **86**:485-514.

• A comprehensive recent review on B_{12}-based photoreceptors, their discovery, functions, mode of action, structures, photochemistry and extensive distribution in bacteria.

14. Losi A, Gardner KH, Moglich A: **Blue-light receptors for optogenetics**. *Chem Rev* 2018.

15. Gordley RM, Bugaj LJ, Lim WA: **Modular engineering of cellular signaling proteins and networks**. *Curr Opin Struct Biol* 2016, **39**:106-114.

16. Jost M, Fernández-Zapata J, Polanco MC, Ortiz-Guerrero JM, Chen PY, Kang G, Padmanabhan S, Elías-Arnanz M, Drennan CL: **Structural basis for gene regulation by a B_{12}-dependent photoreceptor**. *Nature* 2015, **526**:536-541.

•• This paper revealed the structural-molecular basis for CarH light-dependent transcriptional regulation by providing transformative crystal structures of all of the functionally relevant CarH forms, the "dark" free and DNA-bound tetramer form and the "light" monomer form, together with corroborative experimental data.

17. Ortiz-Guerrero JM, Polanco MC, Murillo FJ, Padmanabhan S, Elías-Arnanz M: **Light-dependent gene regulation by a coenzyme B_{12}-based photoreceptor**. *Proc. Natl. Acad. Sci. U.S.A.* 2011, **108**:7565-7570.

•• This paper laid the foundation for all subsequent work on CarH. It reported studies demonstrating that CarH is an AdoCbl-based photoreceptor and uncovered its mode of action.

18. Pérez-Marín MC, Padmanabhan S, Polanco MC, Murillo FJ, Elías-Arnanz M: **Vitamin B_{12} partners the CarH repressor to downregulate a photoinducible promoter in *Myxococcus xanthus***. *Mol Microbiol* 2008, **67**:804-819.

19. Takano H, Kondo M, Usui N, Usui T, Ohzeki H, Yamazaki R, Washioka M, Nakamura A, Hoshino T, Hakamata W, et al.: **Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in *Thermus thermophilus***. *Journal of Bacteriology* 2011, **193**:2451-2459.

20. Takano H, Mise K, Hagiwara K, Hirata N, Watanabe S, Toriyabe M, Shiratori-Takano H, Ueda K: **Role and function of LitR, an adenosyl B_{12}-bound light-sensitive regulator of *Bacillus megaterium* QM B1551, in regulation of carotenoid production**. *J Bacteriol* 2015, **197**:2301-2315.

21. Cervantes M, Murillo FJ: **Role for vitamin B_{12} in light induction of gene expression in the bacterium *Myxococcus xanthus***. *J Bacteriol* 2002, **184**:2215-2224.

22. Fernandez-Zapata J, Perez-Castano R, Aranda J, Colizzi F, Polanco MC, Orozco M, Padmanabhan S, Elias-Arnanz M: **Plasticity in oligomerization, operator architecture, and DNA binding in the mode of action of a bacterial B_{12}-based photoreceptor**. *J Biol Chem* 2018, **293**:17888-17905.

• • This paper illustrates how CarH homologs can exhibit plasticity in oligomerization as well as in operator architecture and DNA binding, even as they share a conserved mode of action.

23. Kozlowski PM, Garabato BD, Lodowski P, Jaworska M: **Photolytic properties of cobalamins: a theoretical perspective**. *Dalton Trans* 2016, **45**:4457-4470.

24. Rury AS, Wiley TE, Sension RJ: **Energy cascades, excited state dynamics, and photochemistry in cob(III)alamins and ferric porphyrins**. *Acc Chem Res* 2015, **48**:860-867.

25. Schwartz PA, Frey PA: **5'-Peroxyadenosine and 5'-peroxyadenosylcobalamin as intermediates in the aerobic photolysis of adenosylcobalamin**. *Biochemistry* 2007, **46**:7284-7292.

26. Jones AR: **The photochemistry and photobiology of vitamin B_{12}**. *Photochem Photobiol Sci* 2017, **16**:820-834.

27. Banerjee R, Ragsdale SW: **The many faces of vitamin B_{12}: catalysis by cobalamin-dependent enzymes**. *Annu Rev Biochem* 2003, **72**:209-247.

28. Bridwell-Rabb J, Drennan CL: **Vitamin B_{12} in the spotlight again**. *Curr Opin Chem Biol* 2017, **37**:63-70.

29. Dowling DP, Croft AK, Drennan CL: **Radical use of Rossmann and TIM barrel architectures for controlling coenzyme B_{12} chemistry**. *Annu Rev Biophys* 2012, **41**:403-427.

30. Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML: **How a protein binds B_{12}: A 3.0 A X-ray structure of B_{12}-binding domains of methionine synthase**. *Science* 1994, **266**:1669-1674.

31. Giedyk M, Goliszewska K, Gryko D: **Vitamin B_{12} catalysed reactions**. *Chem Soc Rev* 2015, **44**:3391-3404.

32. Gruber K, Puffer B, Krautler B: **Vitamin B_{12}-derivatives-enzyme cofactors and ligands of proteins and nucleic acids**. *Chem Soc Rev* 2011, **40**:4346-4363.

33. Smith AD, Warren MJ, Refsum H: **Vitamin B_{12}**. *Adv Food Nutr Res* 2018, **83**:215-279.

34. Diez AI, Ortiz-Guerrero JM, Ortega A, Elías-Arnanz M, Padmanabhan S, García de la Torre J: **Analytical ultracentrifugation studies of oligomerization and DNA-binding of TtCarH, a *Thermus thermophilus* coenzyme B_{12}-based photosensory regulator**. *Eur Biophys J* 2013, **42**:463-476.

35. López-Rubio JJ, Padmanabhan S, Lázaro JM, Salas M, Murillo FJ, Elías-Arnanz M: **Operator design and mechanism for CarA repressor-mediated down-regulation of the photoinducible *carB* operon in *Myxococcus xanthus***. *J Biol Chem* 2004, **279**:28945-28953.

36. Philips SJ, Canalizo-Hernandez M, Yildirim I, Schatz GC, Mondragon A, O'Halloran TV: **Allosteric transcriptional regulation via changes in the overall topology of the core promoter**. *Science* 2015, **349**:877-881.

37. Kutta RJ, Hardman SJ, Johannissen LO, Bellina B, Messiha HL, Ortiz-Guerrero JM, Elías-Arnanz M, Padmanabhan S, Barran P, Scrutton NS, et al.: **The photochemical mechanism of a B_{12}-dependent photoreceptor protein**. *Nat Commun* 2015, **6**:7907.

•• This paper used ultrafast spectroscopy to infer heterolytic Co-C cleavage in CarH-bound AdoCbl and pointed to an unprecedented photochemistry.

38. Jost M, Simpson JH, Drennan CL: **The transcription factor CarH safeguards use of adenosylcobalamin as a light sensor by altering the photolysis products**. *Biochemistry* 2015, **54**:3231-3234.

•• This paper showed that CarH enables safe use of AdoCbl as a photosensor by altering photolysis products providing evidence for a distinct AdoCbl photochemistry in CarH.

39. Chatelle C, Ochoa-Fernandez R, Engesser R, Schneider N, Beyer HM, Jones AR, Timmer J, Zurbriggen MD, Weber W: **A green-light-responsive system for the**

control of transgene expression in mammalian and plant cells. *ACS Synth. Biol.* 2018, **7**:1349-1358.

•• This paper reports the use of CarH in the first green-light-responsive gene expression system operative in various mammalian cell lines and plant (*A. thaliana*) protoplasts with high induction, low leakiness and good reversibility.

40. Kainrath S, Stadler M, Reichhart E, Distel M, Janovjak H: **Green-light-induced inactivation of receptor signaling using cobalamin-binding domains**. *Angew. Chem. Int. Ed. Engl.* 2017, **56**:4608-4611.

•• This paper describes the first use of the CarH AdoCbl-binding photosensory module as an optogenetics tool for temporal green-light control of receptor protein kinase-regulated cell signaling in cultured human cells and zebrafish embryos.

41. Wang R, Yang Z, Luo J, Hsing IM, Sun F: **B_{12}-dependent photoresponsive protein hydrogels for controlled stem cell/protein release**. *Proc. Natl. Acad. Sci. U.S.A.* 2017, **114**:5912-5917.

•• This paper describes how the CarH AdoCbl-binding photosensory module can be coupled to SpyTag/SpyCatcher technology to produce protein hydrogels that allow AdoCbl/light-controlled encapsulation and release of proteins and cells.

42. Wei J, Wu, W.-H, Wang, R., Yang, Z., Zhang, W.-B.: **B_{12}-dependent protein oligomerization facilitates layer-by-layer growth of photo/thermal responsive nanofilms**. *ACS Macro Lett.* 2018, **7**:514-518.

• In this paper, the CarH-based technology described by Wang et al., (2017) is extended for building photoresponsive layer-by-layer protein nanofilms suitable for AdoCbl/light-controlled release of protein cargos.

43. Shcherbakova DM, Shemetov AA, Kaberniuk AA, Verkhusha VV: **Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools**. *Annu Rev Biochem* 2015, **84**:519-550.

44. O'Banion CP, Lawrence DS: **Optogenetics: a primer for chemists**. *Chembiochem* 2018, **19**:1201-1216.

45. García-Moreno D, Polanco MC, Navarro-Avilés G, Murillo FJ, Padmanabhan S, Elías-Arnanz M: **A vitamin B_{12}-based system for conditional expression reveals *dksA* to be an essential gene in *Myxococcus xanthus***. *J Bacteriol* 2009, **191**:3108-3119.

46. Sun F, Zhang WB, Mahdavi A, Arnold FH, Tirrell DA: **Synthesis of bioactive protein hydrogels by genetically encoded SpyTag-SpyCatcher chemistry**. *Proc Natl Acad Sci U S A* 2014, **111**:11269-11274.

47. Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M: **Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin**. *Proc Natl Acad Sci U S A* 2012, **109**:E690-697.

48. Reddington SC, Howarth M: **Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher**. *Curr Opin Chem Biol* 2015, **29**:94-99.

49. Banerjee A, Howarth M: **Nanoteamwork: covalent protein assembly beyond duets towards protein ensembles and orchestras**. *Curr Opin Biotechnol* 2018, **51**:16-23.

50. Schloss AC, Williams DM, Regan LJ: **Protein-based hydrogels for tissue engineering**. *Adv Exp Med Biol* 2016, **940**:167-177.

Figure legends

Figure 1. The AdoCbl chromophore and CarH modes of action. **(a)** AdoCbl chemical structure showing Co^{3+} (magenta circle) in the centre of the corrin ring, its lower axial DMB (blue) and the upper axial Ado ligand with its photosensitive Co–C5′ bond (red). **(b)** Mechanism of light-dependent gene regulation by CarH photoreceptors. The cartoon depiction of CarH shows the DBD in blue, the CBD in green-yellow, and the linker between them as a black line. ApoCarH (monomer in some homologs, molten globule oligomer in some others) does not bind operator DNA (left). Binding to intact AdoCbl in the dark yields CarH tetramers, which repress transcription by binding to operator DNA (middle). Known targets subject to CarH control are: *carH* itself, genes for carotenogenesis (*crt*), photolyase (*phr*) and global redox regulator (*spx*). CarH operators comprise three tandem 11-bp direct repeats (grey arrows) or have a fourth (broken arrow), and overlap with the -35 or -10 regions of promoters recognized by RNAP with the major σ factor. The DBD with linker as a dotted line reflects plasticity in DNA binding with only three repeats bound (*T. thermophilus* CarH) or the fourth also included (*B. megaterium* CarH). The consensus repeat logo based on the experimentally characterized CarH operators from both *T. thermophilus* and *B. megaterium* [16••,22••]. UV, blue or green light activates transcription by disrupting the AdoCbl Co–C bond and the tetramer to Cbl-bound monomers or dimers (depending on the homolog), and causing loss of DNA binding.

Figure 2. CarH structures and modular architecture. **(a)** CarH protomer structure in the "dark" tetramer (PDB accession code 5C8D), with its N-terminal MerR family winged-helix DBD connected by a flexible linker to the C-terminal AdoCbl-binding photosensory module composed of the four-helix bundle and Rossmann fold subdomains. Typical sizes

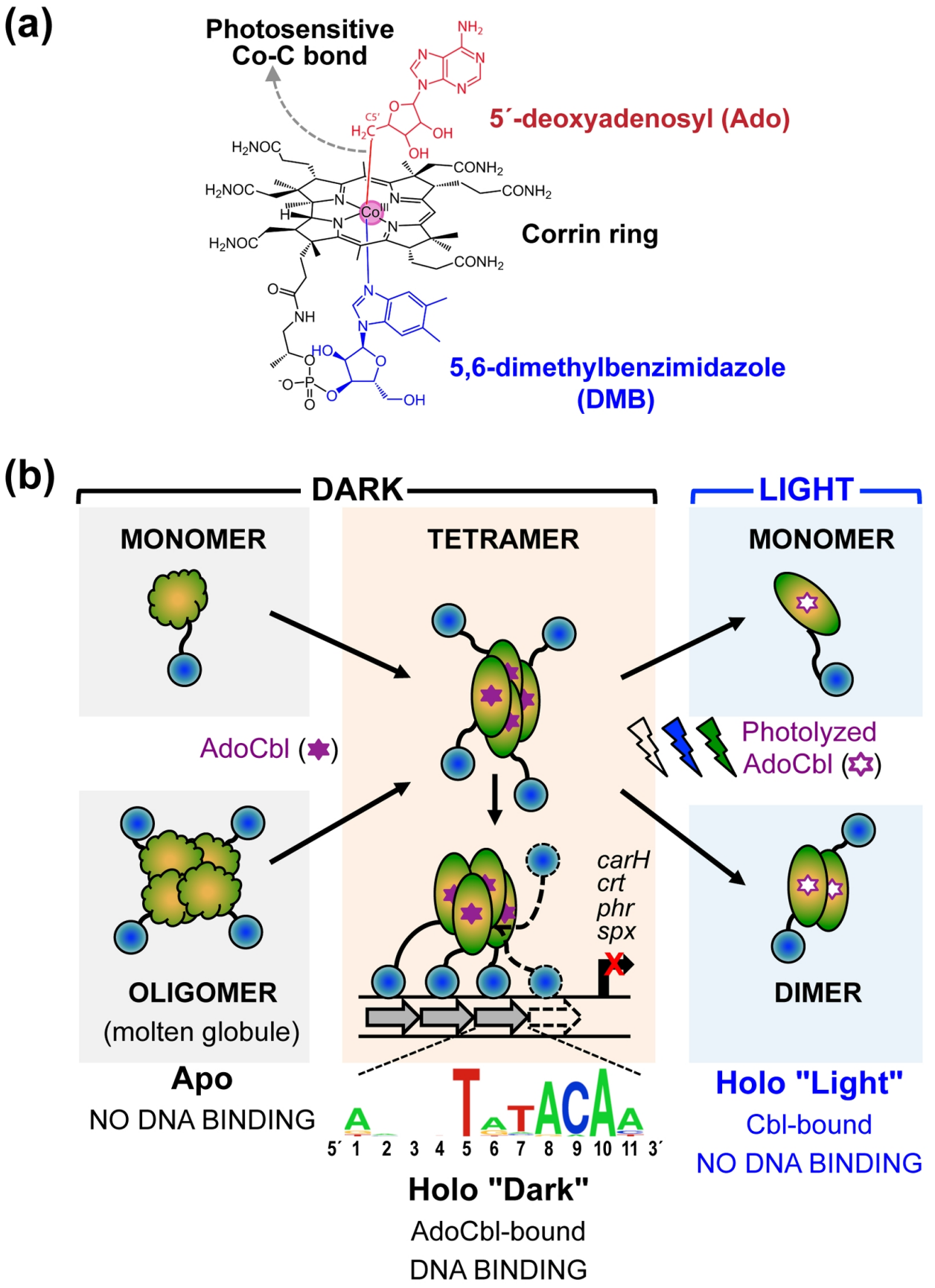
of each module (in parentheses, as number of residues) and their conserved motifs (x=any residue) are indicated [13•]. The same color code is used for a given module in all representations. **(b)** CarH dimer formed by two head-to-tail packed CBDs in the "dark" tetramer structure. **(c)** "Dark" CarH tetramer, a dimer of two AdoCbl-bound dimers, with the DBDs (DNA recognition helix in dark) splayed on the surface. **(d)** "Dark" CarH tetramer bound to DNA (PDB code 5C8E). DNA recognition helices of three of its four DBDs are shown contacting the major groove on the same side of operator DNA. The latter is composed of three DRs (thick arrows with sequences below; conserved bases in bold upper case contact the recognition helix; the -35 promoter element is shaded pink). **(e)** Light-exposed CarH monomer structure (PDB code 5C8F) showing the large shift (>8 Å) in the helix bundle relative to the dark form (transparent).

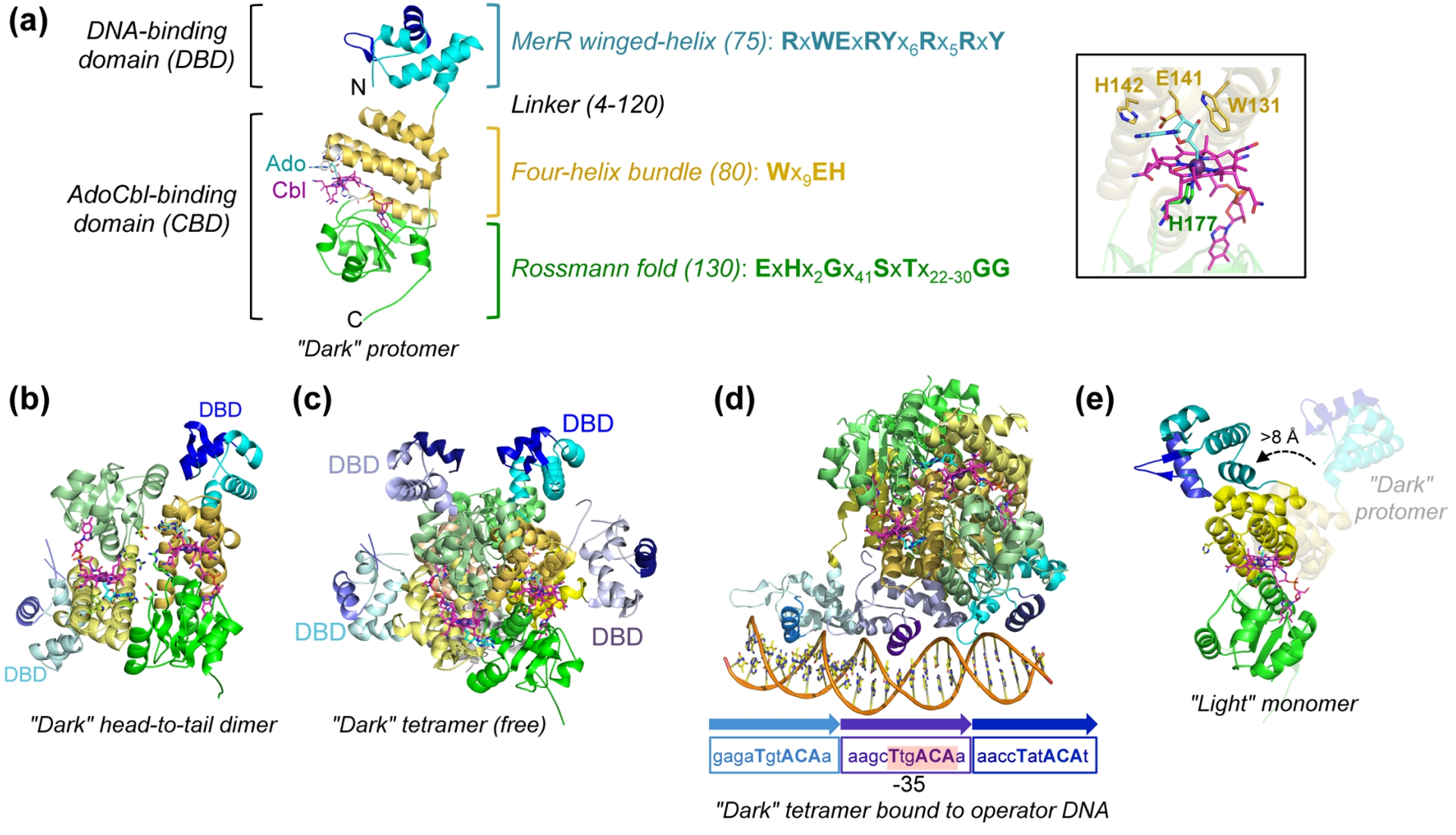
Figure 3. Phylogenetic trees for CarH photoreceptors and standalone CBD proteins. (**a**) Unrooted tree for CarH proteins constructed with 582 CarH homologs [13•] aligned using MUSCLE from which a maximum-likelihood tree (JTT/Jones-Taylor-Thornton model, 95% site coverage cutoff) was calculated using MEGA7 (https://www.megasoftware.net/). The tree was visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). (**b**) Unrooted tree for standalone CBD proteins from the sequences of 159 homologs ([13•]) aligned as in (**a**), refined with Gblocks, and using PhyML (http://www.atgc-montpellier.fr/phyml/) to generate the maximum-likelihood tree that was visualized with FigTree. Colored areas are monophyletic groups corresponding to the indicated bacterial phyla.

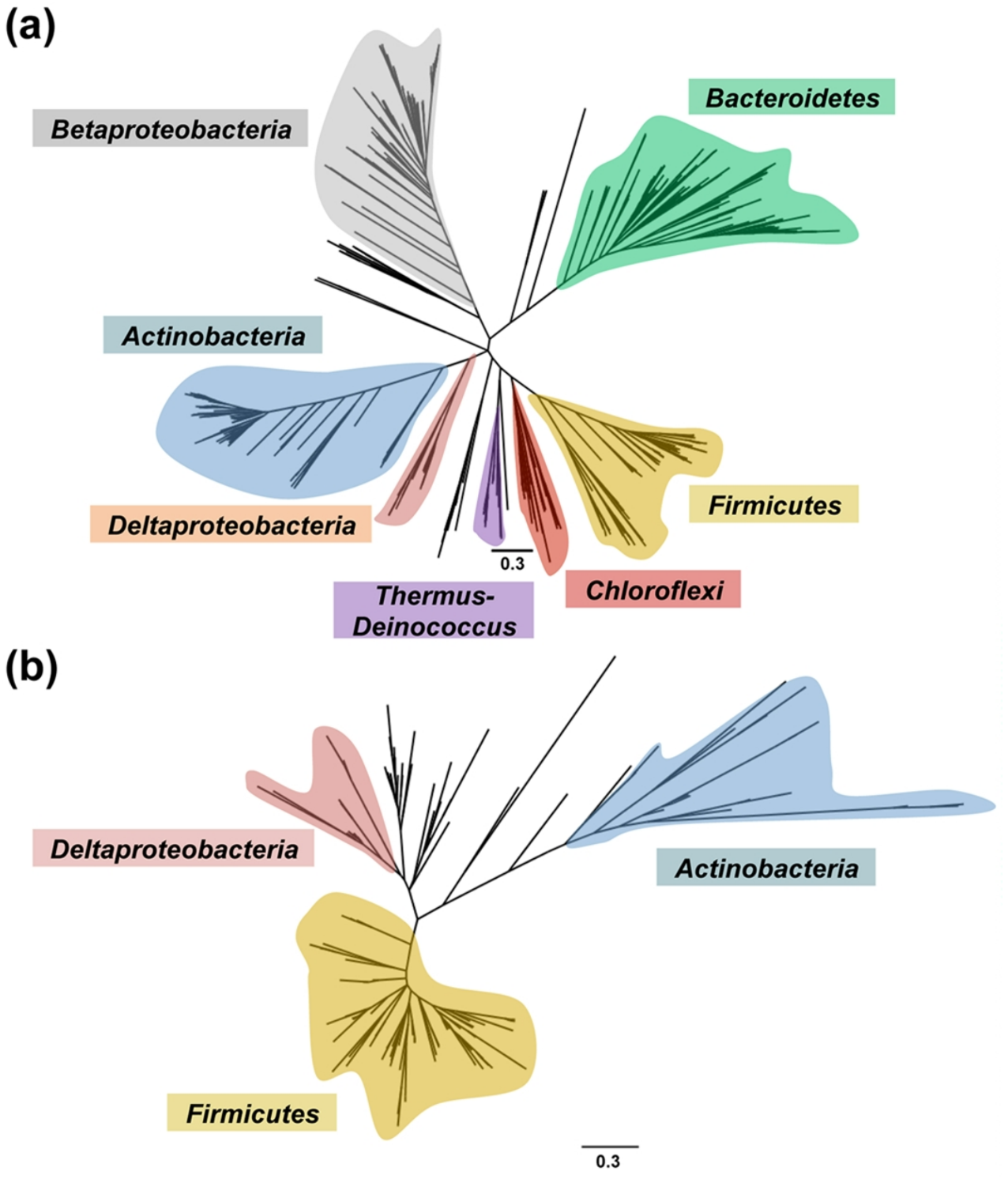
Figure 4. CarH applications in optogenetics and synthetic biology. (**a**) A B_{12}/light-dependent gene expression system for use in bacteria [45]. A plasmid expressing CarH

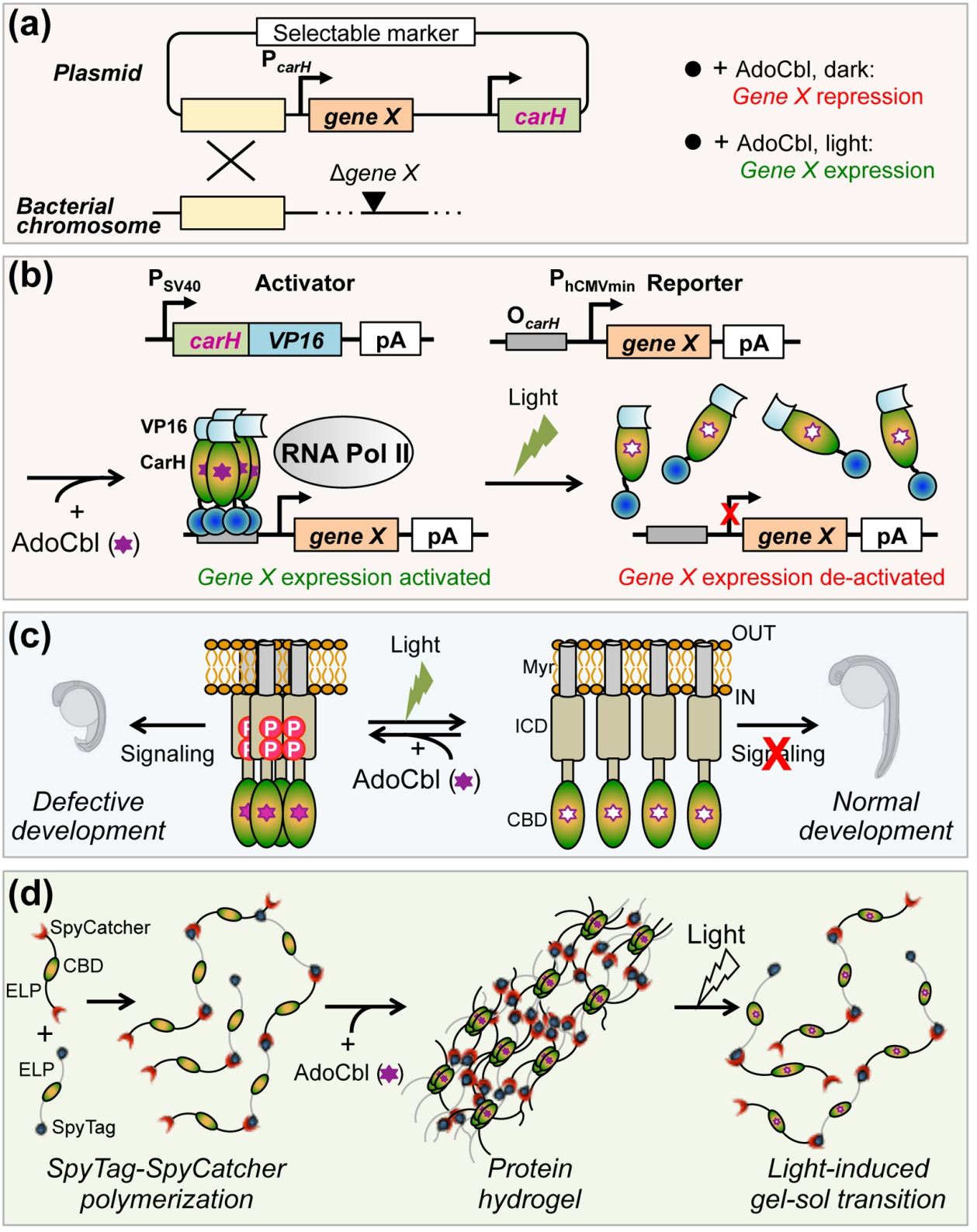
constitutively and the gene of interest (*gene X*) under CarH operator-promoter (P_{carH}) control, a selectable marker, and a DNA segment for chromosomal integration is introduced into the bacterial strain. With AdoCbl present, transformants with the plasmid integrated in the chromosome express *gene X* in the light but not in the dark. (**b**) A CarH AdoCbl/light-responsive gene expression system in mammalian cells and plant (*A. thaliana*) protoplasts [39••]. The "activator" plasmid expresses CarH fused to the *Herpes simplex* transactivation domain VP16 from a constitutive simian virus 40 promoter (P_{SV40}). The "reporter" plasmid with multiple repeats (eight was best) of the CarH operator (O_{carH}) upstream of a human cytomegalovirus-derived minimal promoter ($P_{hCMVmin}$) controls expression of *X*. Both plasmids are supplied with a polyadenylation (pA) segment as shown. CarH-VP16 tetramers formed in the dark with AdoCbl present bind to O_{carH} and recruit the transcription initiation complex for RNAP II-dependent expression of *gene X*. Light abrogates this expression by provoking AdoCbl photolysis and tetramer disruption. (**c**) Green-light control of receptor signaling using CarH CBD [40••]. Expression in the dark, with AdoCbl present, of CarH CBD fused to the intracellular domain (ICD) of murine fibroblast growth factor receptor 1 tyrosine kinase with a myristoylation (Myr) anchor replacing the N-terminal ligand binding and transmembrane domains, results in ICD phosphorylation (P), oligomerization and active signaling. This causes defective development in zebrafish embryos (left). Green light prevents oligomerization and signaling to permit normal development (right). (**d**) Synthesis of AdoCbl/light-dependent protein hydrogels using CarH CBD and SpyTag-SpyCatcher chemistry [41••]. Two telechelic proteins, SpyTag-ELP-CBD-ELP-SpyTag and SpyCatcher-ELP-CBD-ELP-SpyCatcher, where ELP is an elastin-like domain, are separately purified and mixed. Isopeptide linkage between SpyCatcher Lys and a SpyTag Asp sidechains produces linear polymers that, in the dark with AdoCbl present, assemble

as hydrogels via interchain interactions arising from CBD tetramerization. Light exposure disassembles these hydrogels.


(a)
Photosensitive
Co-C bond
5´-deoxyadenosyl (Ado)
Corrin ring
5,6-dimethylbenzimidazole
(DMB)
(b)
DARK
LIGHT
MONOMER
TETRAMER
MONOMER
AdoCbl (✱)
Photolyzed
AdoCbl (✡)
carH
crt
phr
spx
OLIGOMER
(molten globule)
DIMER
Apo
NO DNA BINDING
Holo "Light"
Cbl-bound
NO DNA BINDING
5´ 1 2 3 4 5 6 7 8 9 10 11 3´
Holo "Dark"
AdoCbl-bound
DNA BINDING


(a)
DNA-binding domain (DBD)
AdoCbl-binding domain (CBD)
N
C
Ado
Cbl
MerR winged-helix (75): RxWExRYx6Rx5RxY
Linker (4-120)
Four-helix bundle (80): Wx9EH
Rossmann fold (130): ExHx2Gx41SxTx22-30GG
"Dark" protomer
H142
E141
W131
H177
(b)
DBD
DBD
"Dark" head-to-tail dimer
(c)
DBD
DBD
DBD
DBD
"Dark" tetramer (free)
(d)
gagaTgtACAa
aagcTtgACAa
aaccTatACAt
-35
"Dark" tetramer bound to operator DNA
(e)
>8 Å
"Dark" protomer
"Light" monomer


(a)
Betaproteobacteria
Bacteroidetes
Actinobacteria
Deltaproteobacteria
Firmicutes
0.3
Thermus-
Deinococcus
Chloroflexi
(b)
Deltaproteobacteria
Actinobacteria
Firmicutes
0.3


(a)
Plasmid
Selectable marker
P carH
gene X
carH
Δgene X
Bacterial chromosome
+ AdoCbl, dark:
Gene X repression
+ AdoCbl, light:
Gene X expression
(b)
P SV40
Activator
carH
VP16
pA
P hCMVmin
Reporter
O carH
gene X
pA
VP16
CarH
RNA Pol II
+
AdoCbl
Light
gene X
pA
Gene X expression activated
Gene X expression de-activated
(c)
Signaling
Defective development
Light
+
AdoCbl
Myr
ICD
CBD
OUT
IN
Signaling
Normal development
(d)
SpyCatcher
CBD
ELP
+
ELP
SpyTag
SpyTag-SpyCatcher polymerization
+
AdoCbl
Protein hydrogel
Light
Light-induced gel-sol transition